



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE COMPUTACIÓN

Aprendizaje automático y procesamiento del lenguaje natural en la comunicación política online y las redes sociales

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires
en el área de Ciencias de la Computación

Federico Albanese

Director: Dr. Esteban Z. Feuerstein

Director Adjunto: Dr. Leandro Lombardi

Consejero de Estudios: Dr. Matías López y Rosenfeld

Lugar de Trabajo: Instituto de Investigación en
Ciencias de la Computación,
Universidad de Buenos Aires - CONICET

Buenos Aires, 2024

APRENDIZAJE AUTOMÁTICO Y PROCESAMIENTO DEL LENGUAJE NATURAL EN LA REDES SOCIALES

El presente trabajo doctoral examina las dinámicas de interacción en redes sociales y el impacto de los medios de comunicación en la formación y difusión de opiniones políticas. Utilizando técnicas de machine learning, procesamiento de lenguaje natural, análisis de grafos y estimación de impacto causal, esta investigación analiza datos de millones de usuarios y publicaciones provenientes de Twitter, Facebook, Instagram y Reddit.

Hemos identificado cómo las estructuras de comunidades virtuales y los patrones de interacción contribuyen a fenómenos como la polarización política, la confirmación de sesgos y la formación de cámaras de eco.

Además, hemos desarrollado una nueva metodología para la detección eficiente de tópicos en textos cortos, típicos de las redes sociales, que supera en rendimiento a las técnicas tradicionales en tareas como la detección no supervisada de tópicos y en sistemas de recomendación.

Nuestros experimentos también muestran una relación causal significativa entre la cobertura mediática de ciertos temas y las variaciones en las encuestas de opinión pública durante procesos electorales. Los resultados indican que el impacto de estos temas en la opinión sobre candidatos políticos depende de su mera mención, independientemente del sesgo o sentimiento con el que se presenta el tema.

Por otro lado, hemos examinado la interacción recíproca entre la agenda de temas utilizada por los medios y la utilizada por los usuarios de redes sociales, destacando cómo ambas se influencian mutuamente en la configuración de un discurso público.

En términos del comportamiento de los usuarios, observamos diferencias significativas en cómo se reciben los mensajes de los políticos dependiendo de la plataforma, y cómo los usuarios cambian de comunidades políticas y de ideología cuando son ignorados dentro de una comunidad online.

La presente tesis también analiza cómo la difusión de noticias con distintos sesgos políticos en comunidades cerradas refuerza la confirmación de sesgos, limita el debate de ideas diferentes y favorece la propagación de información falsa.

Por último, se presenta un análisis adversarial de los sistemas de recomendación en

redes sociales, mostrando la eficacia de algoritmos maliciosos como SAVAGE en escenarios donde los usuarios están distanciados dentro de la red, lo cual es crucial para entender cómo las recomendaciones de contenido pueden ser manipuladas en contextos políticos.

Este estudio no solo busca ofrecer una visión clara de las interacciones en redes sociales y el impacto de los medios de comunicación en el debate político, sino que también sugiere enfoques para un análisis más efectivo y ético en la era digital.

Palabras claves: Procesamiento del Lenguaje Natural, Aprendizaje Automático, Redes Sociales, Ciencias Sociales Computacionales, Comunicación Política.

MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING STRATEGIES FOR SOCIAL NETWORKS AND ONLINE POLITICAL COMMUNICATION.

This doctoral thesis examines the dynamics of user interaction on social media and the impact of mass media on the formation and dissemination of political opinions. Using machine learning techniques, natural language processing, graph analysis, and causal impact estimation, this research analyzes data from millions of users and posts from Twitter, Facebook, Instagram, and Reddit. We have identified how the structures of online communities and interaction patterns contribute to phenomena such as political polarization, bias confirmation, and the formation of echo chambers.

Furthermore, we have developed a new methodology for the detection of topics in short texts, typical in social media, which outperforms traditional techniques in tasks such as unsupervised topic detection and recommendation systems.

Our experiments also show a significant causal relationship between media coverage of certain topics and variations in public opinion polls during electoral processes. The results indicate that the impact of these topics on opinions about political candidates depends solely on their mention, regardless of the bias or sentiment with which the topic is presented.

On the other hand, we have examined the reciprocal interaction between the agenda of topics used by the mass medias and the agenda used by social media users, highlighting how both mutually influence each other, shaping the public discourse.

In terms of user behavior, we observe significant differences in how messages from politicians are received depending on the platform, and how users change political communities and ideologies when ignored within an online community, measured through PageRank.

This thesis also analyzes how the dissemination of news with different political biases in closed communities reinforces bias confirmation, limits the debate of different ideas, and fosters the spread of fake news.

Lastly, an adversarial analysis of recommendation systems in social media is presented, demonstrating the effectiveness of malicious algorithms like SAVAGE in scenarios where users are distanced within the network, which is crucial for understanding how content

recommendations can be manipulated in political contexts.

This study not only seeks to provide a clear view of interactions on social media and the impact of media on the political debate but also suggests approaches for a more effective and ethical analysis in the digital era.

Keywords: Natural Language Processing, Machine Learning, Social Media, Computational Social Science, Political communication.

Índice general

Parte I	Introducción	1
1..	Contexto	3
1.1.	Redes sociales	3
1.2.	Medios masivos de comunicación	4
1.3.	Machine Learning	4
2..	Motivación	7
2.1.	Polarización política y cámaras de eco	7
2.2.	Medios masivos de comunicación en la conversación online	8
2.3.	Limitaciones y consideraciones éticas	8
3..	Aportes de este trabajo	11
Parte II	Detección de tópicos	13
4..	Introducción	17
5..	Métodos previos	19
5.1.	Métodos de reducción dimensional	19
5.2.	Métodos de agrupación para microblogging	20
5.3.	Otros métodos	21
6..	Detección de tópicos en comunidades	23
6.1.	Agrupación por comunidades	23
6.2.	Experimentación	23
6.2.1.	Dataset	23
6.2.2.	Evaluación	24
6.2.3.	Resultados	25
6.3.	Conclusiones	26
Parte III	Medios de comunicación online: Análisis de noticias y sus efectos	27
7..	Introducción	31
8..	La influencia de los Medios de comunicación en la opinión pública	33
8.1.	Dataset	33
8.1.1.	Datos de encuestas	33
8.1.2.	Artículos de noticias	33
8.2.	Métodos	34
8.2.1.	Procesamiento del lenguaje natural	34
8.2.2.	Medidas de correlación y causalidad	35
8.3.	Resultados	36

8.3.1. Número total de menciones	36
8.3.2. Análisis de sentimiento	38
8.3.3. Detección de tópicos	39
8.3.4. Combinación de sentimiento y análisis de temas	41
8.3.5. Causalidad	43
8.4. Conclusiones	44
9.. La influencia entre los Medios de comunicación y las redes sociales	47
9.1. Métodos	47
9.1.1. La agenda de los medios	47
9.1.2. Agenda Pública: Google Trends y la red social Twitter	47
9.2. Resultados	48
9.2.1. Descomposición en tópicos	48
9.2.2. La agenda mediática y la agenda pública	48
9.2.3. Diversidad de Agenda	52
9.2.4. Distancia entre la Agenda de los medios y la Agenda pública	54
9.2.5. Sesgo de agenda en diferentes medios de comunicación	54
9.2.6. Caso de estudio Santiago Maldonado: sesgo de cobertura	56
9.3. Conclusiones	57
Parte IV Redes sociales: Análisis de la comunicación y la dinámica de opiniones de los usuarios	61
10..Introducción	65
11..Comunicación de los políticos	67
11.1. Introducción	67
11.2. Marco teórico sociológico e hipótesis	67
11.3. Metodologías y experimentos	69
11.3.1. Construcción del Dataset	69
11.3.2. H1: Tópicos y Temas en común	70
11.3.3. H2: Sentimiento y Negatividad en Twitter	72
11.3.4. H3: Interpelación	73
11.4. Conclusiones	77
12..Usuarios que cambian de comunidad: Análisis de sus características topológicas y sus tópicos de interés	79
12.1. Introducción	79
12.2. Dataset	80
12.3. Métodos	80
12.3.1. La red de retweets	80
12.3.2. Detección de comunidades	81
12.3.3. Características topológicas	82
12.3.4. Detección de tópicos	82
12.3.5. Análisis de importancia	83
12.4. Resultados	84
12.5. Conclusiones	85

13.. Usuarios que cambian de comunidad: Análisis de sentimiento y opinión política	87
13.1. Introducción	87
13.2. Métodos	87
13.2.1. Detección de comunidades	87
13.2.2. Análisis de Sentimiento	88
13.3. Resultados	88
13.3.1. Comunidades en la red de retweets	88
13.3.2. Sentimiento de las comunidades	89
13.3.3. Usuarios que cambian de comunidad	90
13.4. Conclusiones	92
14.. Efecto causal de compartir noticias	93
14.1. Introducción	93
14.2. Dataset	94
14.2.1. Reddit	94
14.2.2. La dimensión partidista de los subreddits	94
14.2.3. Medios de comunicación	96
14.2.4. Artículos de noticias	96
14.2.5. Características del dataset	96
14.3. RQ1: El engagement y el sesgo político de los medios	97
14.3.1. Configuración experimental	97
14.3.2. Resultados y discusión	98
14.4. RQ2: El engagement de la comunidad y credibilidad de los medios	99
14.4.1. Configuración experimental	99
14.4.2. Resultados	99
14.5. Conclusiones	99
 Parte V Sistemas de recomendación de las redes sociales	101
15.. Introducción	105
16.. Ataques adversariales a los sistemas de recomendación de las redes sociales	107
16.1. Introducción	107
16.1.1. Predicciones de aristas en grafos	107
16.1.2. Ataques adversariales en la predicción de aristas	107
16.2. Datos	108
16.3. Métodos	108
16.3.1. Ataque Adversarial	108
16.3.2. Métricas de evaluación	109
16.4. Resultados	110
16.4.1. Análisis de SAVAGE	110
16.4.2. Comparación con otros métodos	111
16.5. Conclusiones	113

Parte VI Conclusiones	115
17..Conclusiones	117

Parte I

INTRODUCCIÓN

1. CONTEXTO

1.1. Redes sociales

Las redes sociales tecnológicamente mediadas son plataformas digitales que facilitan la interacción y conexión entre individuos o grupos de individuos a través de la creación de perfiles personales. En ellas, los usuarios pueden compartir contenido multimedia (texto, imágenes y/o videos) y responder o reaccionar al mismo. Estas plataformas actúan como espacios virtuales donde los usuarios pueden establecer relaciones con otros usuarios, intercambiar información y expresar sus opiniones.

La historia de las redes sociales tecnológicamente mediadas se remonta a la década de 1990, cuando la noción de conexión digital entre individuos comenzó a tomar forma con plataformas como Six Degrees, GeoCities y Classmates.com. En 1998, Six Degrees implementó el concepto de usuarios que podían crear perfiles y conectarse con los perfiles de familiares o amigos, sentando las bases para las futuras redes sociales. En el 2003 comienza a aumentar la popularidad de las redes sociales con sitios web como Friendster, donde los usuarios podían compartir con sus contactos contenido online. Friendster es la primera red social en tener más de un millón de usuarios. Sin embargo, es con MySpace que se logra un alcance mundial, llegando a tener más de 115 millones de visitas mensuales. En 2006 fue incluso la página web más visitada en los Estados Unidos, superando a Yahoo! y Google.

El lanzamiento de Facebook en 2004, inicialmente dirigido a estudiantes universitarios, transformó radicalmente el panorama de las redes sociales al enfocarse en la identidad real y la autenticidad de las personas. Por otro lado, Twitter (hoy renombrada X) fue fundada en 2006 e introdujo el concepto de microblogging: textos particularmente cortos que permiten a los usuarios compartir pensamientos y actualizaciones de manera concisa y rápida. Con el tiempo, plataformas como Instagram (2010) y Snapchat (2011) redefinieron la interacción a través de la imagen y el contenido efímero, respectivamente. Además, las redes sociales se diversificaron con la aparición de LinkedIn para fines profesionales y plataformas como TikTok, que se destacan por el contenido de videos.

En la actualidad, las redes sociales han alcanzado una inmensa magnitud, convirtiéndose en un componente integral de la vida cotidiana de muchas personas. Con miles de millones de usuarios activos en diversas plataformas, como Facebook, Instagram, Twitter o Reddit, estas redes han logrado establecer conexiones instantáneas y brindar una plataforma para compartir experiencias, noticias y contenido multimedia. La masividad de estas redes sociales no solo se refleja en la cantidad de usuarios, sino también en su impacto cultural, social y político. Su influencia se extiende desde la esfera personal hasta la esfera pública, transformando la manera en que se consumen noticias. Este alcance masivo plantea desafíos y oportunidades significativos, desde abordar la forma de informarse hasta la polarización extrema.

La gran cantidad de contenido generado diariamente en las redes sociales ha superado con creces la capacidad humana para analizarlo de manera eficiente. Ante este desafío, el machine learning surge como una herramienta esencial para extraer patrones, identificar tendencias y procesar información a gran escala [129]. Los algoritmos de aprendizaje automático aplicados al análisis de datos en redes sociales permiten una clasificación au-

tomatizada del contenido compartido y el reconocimiento de patrones de comportamiento de los usuarios. En particular, el uso de modelos de aprendizaje automático en grafos (o Graph Machine Learning) [115] y las técnicas de procesamiento del lenguaje natural [110] posibilitan el caracterizar las propiedades topológicas de las interacciones entre usuarios y analizar el tema, emoción y positividad/negatividad de los textos compartidos, entre otras aplicaciones.

1.2. Medios masivos de comunicación

Los primeros registros históricos escritos de comunicación de noticias datan del año 59 AC, con el denominado Acta diurna del Imperio Romano, donde se comunicaba públicamente lo discutido por el Senado. Sin embargo, el acceso a las actas era muy limitado ya que no existía la posibilidad de reproducción y distribución masivas. A finales del siglo XV, luego de la invención de la imprenta, surgieron los periódicos de papel, permitiendo una mayor llegada de las noticias. Pero fue recién entre los siglos XVIII y XIX que los gobiernos y partidos políticos comenzaron a tener medios propios, con el fin de comunicar noticias transmitiendo sus propios sesgos y opiniones.

Con la llegada de Internet, el alcance de los medios de comunicación se volvió más masivo, incrementando drásticamente su público. Pero también aumentó la cantidad de voces presentes ya que la barrera de entrada para comunicar noticias disminuyó significativamente. En este contexto resulta de interés entender el efecto que tienen los medios sobre una población, sobre todo en contextos políticos. En particular, los medios masivos de comunicación pueden reflejar la realidad desde su propia perspectiva, determinando la importancia relativa otorgada a diferentes temas. Este proceso se conoce como marcar una agenda temática o agenda-setting [148]. De esta forma, los medios pueden llegar a condicionar la percepción de los individuos sobre determinados hechos y afectar sus temas de interés. En la Parte 3 de esta tesis se investigó sobre el rol de los medios en marcar la agenda temática que discute la población, y los efectos que tienen las noticias en contextos electorales. A su vez, en la Parte 4 Sección 4, caracterizamos el comportamiento de los usuarios cuando otros usuarios les comparten noticias con su mismo sesgo político o con el sesgo político opuesto.

Todos estos análisis son posibles por otra característica que tienen los medios online de la actualidad: el acceso al texto de múltiples noticias y el historial del comportamiento de los usuarios en respuesta en las redes sociales. Tal como se mencionó en la sección anterior, dichos análisis se realizan generalmente con modelos de Machine Learning.

1.3. Machine Learning

Alan Turing y Claude Shannon plantaron las semillas conceptuales de la Inteligencia Artificial a principios del siglo XX [196, 217]. Sin embargo, no fue hasta la década de 1950 cuando el aprendizaje automático (o Machine Learning) empezó a ganar prominencia, con trabajos fundamentales como el Perceptrón de Frank Rosenblatt [190].

En las décadas siguientes, surgieron una gran variedad de algoritmos, entre ellos los árboles de decisión [141], los modelos estadísticos predictivos [76] y las redes neuronales [129]. De todas formas, había una falencia de aplicaciones reales debido a la ausencia de grandes conjuntos de datos y la falta de poder de cómputo suficiente para entrenar estos modelos.

De la mano de Internet, aumentaron significativamente los registros y bases de datos: y los servicios de cómputo en la nube incrementaron la posibilidad de entrenar modelos de forma remota. Es por ello que fue recién en este siglo que los modelos de aprendizaje automático incrementaron dramáticamente su uso y variedad.

Entre las técnicas más utilizadas hoy en día, se destacan los embeddings, o vectores que representan a objetos en un espacio multidimensional [151]. En esta tesis se utilizan embeddings para describir desde textos de noticias hasta usuarios de redes sociales.

Por otro lado, las redes neuronales profundas han evolucionado desde que Rosenblatt presento el perceptrón simple. En particular, se han propuesto diferentes arquitecturas que tienen en cuenta las características y simetrías de los datos para mejorar la performance. Por ejemplo las Long short-term memory (LSTM) [93] para textos y series de tiempo, las redes convolucionales (CNN) [128] para imágenes o las redes convolucionales en Grafos (GCN) [115] para redes. Teniendo en cuenta que las redes sociales se pueden describir como grafos donde los nodos son usuarios y las artista son las interacciones, estas últimas cobran relevancia en el contexto de esta tesis y son usadas en la Parte 5.

2. MOTIVACIÓN

La capacidad de analizar grandes cantidades de datos permite a los investigadores identificar y comprender las dinámicas sociales y conductas en las redes sociales. En esta línea, las herramientas de machine learning facilitan el análisis de múltiples problemáticas asociadas a las redes sociales gracias a estudiar patrones de interacción, influencia y comportamiento tanto de los usuarios como de las comunidades virtuales.

2.1. Polarización política y cámaras de eco

La polarización política ha aumentado drásticamente en las últimas décadas [1, 126]. En este contexto, es importante diferenciar entre dos tipos de polarización: “polarización idiológica” (ideological polarization) y “polarización afectiva” (affective polarization) [122]. La primera se define como las diferencias en opiniones e ideologías [59]. Por otro lado, la polarización afectiva se refiere a la medida en que las personas les agradan sus aliados políticos y les desagradan sus oponentes políticos [102]. En otras palabras, es la distancia emocional y social entre miembros de diferentes grupos políticos. Si bien una mayor polarización puede tener efectos positivos, como una mayor participación política [122, 218], también tiene sus consecuencias negativas.

Debido a que los usuarios pueden elegir con quiénes interactuar y, sobre todo, con qué usuarios no interactuar entre los millones de posibles usuarios de las redes sociales, se generan las cámaras de eco, o “echo chambers”: un fenómeno de las redes sociales donde los usuarios tienden a interactuar principalmente con individuos que comparten perspectivas, opiniones y valores similares [53, 180]. Esta auto-selección de contactos y conexiones contribuye a la formación de burbujas informativas, donde la exposición del usuario se ve limitada a puntos de vista afines, dando lugar a una mayor polarización y a la circulación de información falsa o “fake news” [52, 63]. En estas cámaras de eco, la información se refuerza y amplifica dentro de grupos homogéneos, creando un entorno en el que las creencias preexistentes son reforzadas y los desacuerdos son minimizados. Este fenómeno puede tener implicaciones significativas en la forma en que las personas perciben la realidad, ya que la exposición constante a información similar puede distorsionar la objetividad y dificultar la comprensión de opiniones divergentes, llevando en algunos casos a posiciones políticas extremas [29, 31, 101, 211]. De todas formas, aunque varios estudios previos afirman que la exposición a contenido que refuerza nuestras perspectivas lleva a una radicalización política [75, 156, 157], se debe destacar que hay investigaciones que relativizan estos resultados y muestran que no siempre conducen a una mayor polarización [162].

Además es importante remarcar que si bien dichas cámaras de echo se pueden observar en una gran diversidad de temas y países [23, 53, 55, 86, 152, 180], hay algunas situaciones donde observan conversaciones globales entre ambos partidos [29].

La aplicación de técnicas de machine learning y procesamiento del lenguaje natural permite el estudio de todos estos patrones de interacción usando grandes volúmenes de datos de redes sociales. Posibilitando así la implementación de estrategias para fomentar una mayor diversidad de perspectivas y diálogo constructivo en las plataformas de redes sociales. En particular, resulta de interés entender en qué contextos los usuarios “rompen”

sus propias burbujas de sesgo, cambiando el conjunto usuarios con los que interactúan y las consecuencias de dichos cambios. ¿Qué características tienen los usuarios que cambian de comunidad? Los usuarios que cambian con quienes interactúan, ¿cambian su posicionamiento ideológico? En la Parte 4 de esta tesis buscaremos responder dichas preguntas.

2.2. Medios masivos de comunicación en la conversación online

Los medios masivos de noticias desempeñan un papel central al interactuar con múltiples individuos a través de sus contenidos [147]. Es gracias a los mismos que pueden tener la capacidad para marcar la agenda temática, es decir, determinar qué temas son considerados relevantes y deben ser discutidos por la población y qué temas no, teniendo un impacto significativo en la percepción pública [148]. La selección de enfoques y la cantidad de cobertura en las noticias pueden influir directamente en la formación de opiniones y actitudes de la audiencia [34]. La omisión selectiva de ciertos tópicos en favor de otros también puede sesgar la información que llega a los individuos [41]. Cuando los medios abordan positiva o negativamente un tema o candidato político, pueden afectar la percepción pública sobre el mismo, lo que a su vez repercute en procesos electorales.

Es por ello que en la Parte 3 de esta tesis buscamos caracterizar el efecto de los sesgos ideológicos de los medios, específicamente en contextos electorales, y su impacto en la formación de opiniones y en los votos en elecciones presidenciales. Por otro lado, los usuarios pueden elegir compartir noticias dentro de su comunidad, amplificando el discurso de los medios. Sin embargo, la respuesta de la comunidad depende del sesgo de la noticia compartida. No genera el mismo efecto si la noticia compartida comparte el sesgo político con la comunidad reforzando su ideología, que sí tiene un sesgo político opuesto y la desafía. Entender la reacción de los usuarios cuando se les presenta información con diferentes sesgos políticos resulta fundamental para generar ecosistemas de diálogo, sin toxicidad y reduciendo la polarización. En la Parte 4 de esta tesis abordamos esta temática en detalle.

2.3. Limitaciones y consideraciones éticas

Se deben destacar varias limitaciones propias del análisis en redes sociales. En primer lugar, la presencia de bots o cuentas automáticas que pueden distorsionar los efectos reales que se busca observar [15]. Para ello, existen múltiples técnicas y filtros que se pueden usar, como por ejemplo eliminar usuarios que publican contenido en frecuencias anormales (más de una vez por minuto durante 24 horas al día) o los usuarios cuyo contenido publicado es exactamente igual al de muchas otras cuentas, mostrando una estrategia coordinada que busca amplificar automáticamente un mensaje.

En segundo lugar, se debe tener en cuenta el sesgo de los datos al momento de las conclusiones. Si bien los usuarios en redes sociales reflejan las intenciones y comportamientos de los individuos que controlan dichos usuarios, las redes sociales no son un espejo completo de la realidad. Dependiendo de la población, del país y la red social, puede haber sectores de la población que no se vean representados de forma estadísticamente correcta. Por otro lado, las redes sociales también muestran ciertos tópicos con los que los usuarios interactúan, y se omiten otros, que también pueden ser parte de su realidad diaria. Es por ello que toda conclusión extraída de un análisis estadístico de contenido de redes sociales aplica al contexto de las redes sociales y no necesariamente se aplica de forma directa a

otros contextos. De todas formas, dichos análisis siguen resultando de interés teniendo en cuenta la preponderancia, popularidad e influencia de las redes sociales [166].

Por otro lado, se debe tener en cuenta los aspectos éticos [28]. Todos los datos analizados a lo largo de esta tesis son públicos. No se utilizaron datos de cuentas privadas, con el fin de no violar la privacidad de los individuos.

Además, se debe tener en cuenta casos famosos como el de la empresa Cambridge Analytica [92], donde utilizaron minería y análisis de datos de redes sociales para influir directamente en los resultados de las elecciones presidenciales estadounidenses del 2010. Es por ello que cobran relevancia los trabajos académicos que buscan caracterizar y cuantificar la influencia y comportamiento de las redes sociales. En particular, en la Parte 5 de esta tesis abordamos el estudio de un usuario malicioso que busca atacar el algoritmo de recomendación que puede tener una red social para fomentar la difusión de un mensaje o la popularidad de un usuario.

3. APORTES DE ESTE TRABAJO

En esta tesis nos proponemos hacer un aporte a la caracterización de la comunicación política online y el efecto que genera en los usuarios de redes sociales utilizando herramientas de aprendizaje automático y procesamiento del lenguaje natural.

Para ello, en primer lugar se descargaron y confeccionaron múltiples datasets de redes sociales, bajando millones de posts de Facebook, Instagram, Twitter y Reddit. Además se aplicaron modelos para detectar el sentimiento (positividad/negatividad) [143], las emociones expresadas [71], el tópico [20, 35, 226] o características topológicas de la red de interacción de usuarios como el PageRank o el Betweenness centrality [66]. También se usaron desde redes neuronales profundas en grafos [115] hasta modelos de inferencia causal [170] en los diferentes casos de estudio.

De esta forma, se buscó caracterizar en primer lugar el rol de los medios de noticias y su influencia. Se encontró que los medios efectivamente tienen una influencia estadísticamente significativa en los resultados electorales. Es interesante destacar que los temas de las noticias influyen en mayor medida que el sesgo de dichas noticias (sentimiento positivo/negativo). Debido a la importancia de la temática de las noticias, cobra relevancia entender quién le impone la agenda temática a quién: ¿Los usuarios a los medios o los medios a los usuarios? Nuestro caso de estudio reveló que tanto los medios como los usuarios tienen la capacidad de imponer un tema al otro, según el caso. Por ejemplo, se observó que fue primero en redes sociales que el tópico Santiago Maldonado ¹ cobró relevancia, y luego de varios días el mismo fue adoptado por los medios. Por otro lado, cuando los medios hablan consistentemente de un tema, el mismo logra permear en la agenda temática de las redes sociales.

Asimismo, en cuanto al comportamiento de los usuarios en redes sociales, se observó que las cuentas de los políticos varían su mensaje dependiendo de la red social, maximizando la cantidad de likes. Esto lleva a que se realicen posts más tóxicos en Twitter, donde esto se premia por el público, pero menos tóxicos en Facebook e Instagram. Por otro lado, también encontramos que los usuarios que cambiaron de comunidad política a lo largo del tiempo tenían valores estadísticamente más bajos de *PageRank_{CR}*. Una posible interpretación de estos dos últimos resultados es que los usuarios cambian con quién interactúan cuando no tienen fuertes afinidades con otros usuarios, sus mensajes no tienen respuesta y no son “escuchados” por su comunidad. Además, nuestros resultados muestran que los usuarios que cambian de comunidad no solamente cambian con quienes interactúan, sino que también manifiestan un cambio en su opinión política.

Por último, se analizó el comportamiento de SAVAGE, un algoritmo adversarial en grafos, en relación con la topología de la red. Los ataques de SAVAGE y el estudio realizado es particularmente útil en casos del mundo real, como sugerencias de amigos/seguidores en redes sociales donde un usuario malicioso busca influir artificialmente en las recomendaciones de un usuario objetivo.

Los trabajos expuestos en esta tesis caracterizan el rol de los medios masivos de comunicación online y las interacciones en redes sociales desde diversas perspectivas. Estos trabajos son el fruto del trabajo colaborativo con especialistas de distintas áreas y fueron

¹ Santiago Maldonado fue un activista político que desapareció luego de una protesta mapuche en el año 2017. Se lo encontró muerto 77 días después.

presentados y publicados tanto en congresos como en revistas nacionales e internacionales. Los resultados de esta tesis fueron presentados en los siguientes trabajos:

1. Albanese, F., & Feuerstein, E. (2021). Community pooling: LDA topic modeling in Twitter [Oral Presentation]. International Conference on Machine Learning Conference: LatinX in AI (LXAI) Research Workshop 2021.
2. Albanese, F., & Feuerstein, E. (2021). Improved topic modeling in twitter through community pooling. In String Processing and Information Retrieval: 28th International Symposium, SPIRE 2021, Lille, France, October 4–6, 2021, Proceedings 28 (pp. 209-216). Springer International Publishing.
3. Albanese, F., Pinto, S., Semeshenko, V., & Balenzuela, P. (2020). Analyzing mass media influence using natural language processing and time series analysis. *Journal of Physics: Complexity*, 1(2), 025005.
4. Pinto, S., Albanese, F., Dorso, C. O., & Balenzuela, P. (2019). Quantifying time-dependent media agenda and public opinion by topic modeling. *Physica A: Statistical Mechanics and its Applications*, 524, 614-624.
5. Albanese, F., Feuerstein, E., Kessler, G., & Ortiz de Zárate, J. M. (2023). Aprendizaje automático para el análisis crossplataforma de la comunicación política: Gobierno y oposición argentinos en Facebook, Instagram y Twitter. *Cuadernos. info*, (55), 256-280.
6. Albanese, F., Feuerstein, E., Lombardi, L., & Balenzuela, P. (2023). Characterizing community-changing users using text mining and graph machine learning on Twitter. AMW 2023: 15 Alberto Mendelzon International workshop, Chile.
7. Albanese, F., Feuerstein, E., & Balenzuela, P. (2024) Polarization dynamics: A study of individuals shifting between political communities on social media. *Journal of Physics: Complexity*, URL <http://iopscience.iop.org/article/10.1088/2632-072X/ad679d>
8. Albanese, F., Wu, S. & Cheng, L. (2024) A Causal Study of Community Engagement toward Partisan Content and Misinformation on Reddit. [Paper en preparación, versión preliminar presentada en KHIPU Latin America Meeting in Artificial Intelligence 2022]
9. Albanese, F., Trappolini, G., Scarlino L. & Silvestri, F. (2024). The Impact of Source-Target Node Distance on Vicious Adversarial Attacks in Social Network Recommendation Systems. Proceedings of the International Workshop on Graph-Based Approaches in Information Retrieval at the 46th European Conference on Information Retrieval (ECIR 2024).

Parte II

DETECCIÓN DE TÓPICOS

Esta sección contiene resultados que fueron publicados en los siguientes papers:

- Albanese, F., & Feuerstein, E., (2021). Community pooling: LDA topic modeling in Twitter [Oral Presentation]. International Conference on Machine Learning Conference: LatinX in AI (LXAI) Research Workshop 2021.
- Albanese, F., & Feuerstein, E. (2021). Improved topic modeling in twitter through community pooling. In String Processing and Information Retrieval: 28th International Symposium, SPIRE 2021, Lille, France, October 4–6, 2021, Proceedings 28 (pp. 209-216). Springer International Publishing.

4. INTRODUCCIÓN

En las ultimas décadas ha habido un aumento significativo en la cantidad de contenido generado y compartido por las personas que queda registrado online en redes sociales. Por ejemplo, Twitter genera más de 500 millones de posts por día y Facebook genera más de 800 millones de posts y comentarios diariamente¹. Es por ello que surge el interés de tener herramientas automáticas que analicen y clasifiquen dichos textos [100]. En particular, los modelos de procesamiento del lenguaje natural que permitan detectar de qué tema se esta hablando tienen una gran variedad de aplicaciones que van desde la rápida advertencia de desastres naturales [65] hasta la caracterización del debate político [216].

Esta sección esta dedicada al procesamiento del lenguaje natural para la detección de tópicos. Primero describiremos algunas técnicas clásicas que son ampliamente utilizadas en la literatura y, en particular, a lo largo de esta tesis. Luego, proponemos una nueva técnica para la detección de tópicos basadas en la identificación de comunidades en grafos. Trabajos anteriores señalaron que los métodos clásicos tienen problemas con textos cortos (muy comunes en redes sociales) [140]. Frente a este problema, proponemos un método que agrupa en un documento los posts cuyos autores pertenecen a la misma comunidad en la red social, aumentando la longitud de cada documento y reduciendo el número total de los mismos. Luego, comparamos el método propuesto contra otros esquemas de agrupación en términos de calidad de agrupación, recuperación de documentos, tareas de clasificación de aprendizaje automático y tiempo de ejecución, y mostramos empíricamente que este nuevo esquema mejora el rendimiento en comparación con los métodos anteriores en dos conjuntos de datos heterogéneos de Twitter. Los resultados muestran una mejora significativa tanto en términos de tiempo de ejecución como en la calidad de tópicos encontrados.

¹ Fuente: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=2ef31ab160ba>

5. MÉTODOS PREVIOS

5.1. Métodos de reducción dimensional

Un corpus o conjunto de documentos contiene múltiples textos que buscan clasificarse automáticamente según el tema de que tratan. Para ello es necesario primero representar matemáticamente a un documento. Esta tarea puede realizarse de varias formas. Por ejemplo, se puede representar como un vector, donde cada coordenada representa una palabra del vocabulario y su valor la cantidad de veces que aparece dicha palabra. Sin embargo, esta representación no es la mejor para encontrar tópicos ya que existen palabras muy frecuentes con poco significado, como los artículos, conectores y preposiciones, habitualmente llamadas stop words. Otra forma de describir matemáticamente a un documento es mediante la representación *frecuencia del término - frecuencia inversa del documento (tf-idf)*. El valor de cada componente está dado por la frecuencia de cada término en el texto (tf) ponderada por una medida de especificidad (idf) [226]. De esta forma, palabras muy frecuentes en todos los documentos no obtienen valores altos, mientras que palabras muy frecuentes en un documento pero no en los demás, sí. Luego, los vectores se compilan en una matriz documento-término M , de dimensión d (número de documentos) por t (número de términos), que viene dada por la cantidad total de palabras en el corpus después de la eliminación de las stop words [78]. El siguiente paso es encontrar cómo se pueden agrupar los documentos por similitud en grupos llamados tópicos, donde un tópico se define como un grupo de textos similares que tratan aproximadamente sobre el mismo tema. Para detectar los temas principales en el corpus se aplica luego un algoritmo de reducción dimensional sobre la matriz M . De esta forma, un tópico queda definido matemáticamente como una combinación lineal de términos.

Se pueden aplicar distintos algoritmos de reducción dimensional. Por ejemplo PCA (Principal Component Analysis) [210], aunque este método exige que la descomposición en tópicos sea perpendicular entre sí y permite proyecciones en tópicos negativos. Es decir, puede dar como resultado que un documento tenga un componente negativa en un tópico (negativa no en el sentido de análisis de sentimiento, sino de proyección dimensional), lo cual no permite una interpretación intuitiva del resultado. Es por ello, que suelen usarse otros algoritmos, como Non-Negative Matrix Factorization (NMF) [226]. NMF descompone la matriz M como el producto de dos matrices no negativas, H y W (ver ecuación Eq.(5.1)), donde la primera es una matriz documento-tema y nos da la representación de los documentos en el espacio de temas, mientras que el segundo es una matriz tema-término y trae los temas descritos en el espacio de términos, de donde podemos extraer las palabras clave que definen a cada uno. La dimensión interna n en la ecuación (5.1) es el número de temas esperados, que es un parámetro que debe establecerse antes de la descomposición.

$$M^{(d \times t)} \sim H^{(d \times n)} \cdot W^{(n \times t)} \quad (5.1)$$

Esta no-negatividad hace que las matrices resultantes sean más fáciles de interpretar y muy adecuadas para la detección de temas porque proporciona la representación de los artículos en el espacio de temas (matriz H) donde cada columna proporciona el grado de pertenencia de cada artículo a un tema determinado. Además, la matriz W proporciona

la combinación de palabras que describen cada tema.

Otro algoritmo común para la detección de tópicos usando reducción dimensional es Latent Dirichlet Allocation (LDA) [35], un modelo bayesiano generativo. LDA tiene como hipótesis que la cantidad de palabras N en un documento d sigue una distribución de Poisson, la proporción de los tópicos t en los documentos sigue una distribución de Dirichlet (dando nombre al método en cuestión) y cada palabra w pertenece a un tópico siguiendo una distribución Multinomial. Para encontrar los k tópicos presentes en un corpus comienza asignando de forma aleatoria cada palabra a uno de los tópicos. Luego, de forma iterativa para cada documento y para cada palabra, calcula la probabilidad p para reasignar la palabra a un nuevo tópico siguiendo la fórmula:

$$p = p(t|d) * p(w|t) \quad (5.2)$$

A cada paso se asume que todas las palabras están bien ubicadas excepto la palabra en cuestión. Con la intención de aprender las distribuciones de las variables latentes a partir del corpus se utilizan los métodos de Inferencia Variacional o Gibbs Sampling. LDA se ha utilizado ampliamente para modelar temas en diferentes áreas, como la ciencia médica [169], la ciencia política [54], la ciencia computacional social [176] y la ingeniería de software [79].

5.2. Métodos de agrupación para microblogging

En la práctica, el análisis de contenido en servicios de microblogging puede ser especialmente desafiante debido a que los textos son cortos y, a menudo, vagamente coherentes [140, 150]. Dado que Twitter se ha convertido en una plataforma donde se genera, comparte y consume una gran cantidad de contenido, este problema se ha vuelto de interés para la comunidad científica. Sin embargo, utilizar cada post como un documento hace que la matriz documento-término M sea rala. Es decir, no se proporciona datos adecuados de co-ocurrencia de términos dentro de los documentos [150], debido a lo corto de los textos. Esto llevó a la idea de que la agregación de textos de posts similares da lugar a documentos más largos y a una mejor descomposición de temas mediante métodos como NMF o LDA.

En esta sección, describimos cinco métodos propuestos en la literatura y en el próximo capítulo proponemos uno.

- **No agrupado:** El enfoque clásico que trata cada tweet como un documento único. Este es el punto de referencia para la comparación de todos los métodos.
- **Agrupación por autor:** Todos los posts o tweets escritos por un solo usuario se agregan en un solo documento. El número de documentos es igual al número de usuarios. Este método de agrupación supera en performance al esquema no agrupado [96].
- **Agrupación por hashtag:** En este esquema, un documento consiste en todos los tweets que mencionan un hashtag específico. Un tweet que contiene varios hashtags aparece en varios documentos. Los tweets sin hashtags se consideran como documentos individuales. De esta forma se agrupan los textos en base a ejes temáticos. Se ha demostrado que la agregación de tweets de esta manera supera a no agrupar los textos y a agrupar los textos por usuario en algunas métricas y para algunos conjuntos de datos [150].

- **Agrupación por conversación:** Un documento consiste en todos los tweets en un árbol de conversación (es decir, un tweet, todos los tweets escritos en respuesta a él, las respuestas a esas respuestas, y así sucesivamente). Este esquema agrupa tweets de diferentes autores y con varios hashtags que pertenecen a una conversación [17].
- **Agrupación basada en la red:** Los usuarios de Twitter se agrupan si responden o son mencionados en un tweet o en respuestas a un tweet. Cada documento único consiste en todos los tweets de un grupo de usuarios. A diferencia de la agrupación por conversación, solo se consideran las respuestas directas a un tweet original, ya que una conversación puede cambiar su tema con el tiempo. Este esquema de agrupación mostró mejores resultados que los métodos anteriores en la mayoría de las tareas y conjuntos de datos (aunque no en todos los casos) [164].

5.3. Otros métodos

Si bien los métodos de detección de tópicos usando reducción dimensional son ampliamente usados en la literatura, no son los únicos métodos existentes. En esta tesis, en la sección 11, utilizamos también Top2Vec (Angelov, 2020), que estima la cantidad de temas sin necesidad de validar previamente la coherencia de cada posible valor, reduciendo el tiempo de cómputo. Esta técnica, además, no precisa eliminar stop words (artículos, preposiciones, etc.) ni normalizar el texto para su uso y permite identificar de forma determinística de qué tema habló un posteado dado.

Por otro lado, en la sección 14, aplicamos un método más simple, contando la frecuencia de léxicos (lista de palabras) pre-assignados a tópicos y emociones. En particular, usamos los léxicos de [71].

Finalmente, existen otros métodos de detección de tópicos que no fueron usados en esta tesis. Entre ellos, se encuentran los modelos basados en modelos extensos de lenguaje (o Large Language Models), donde utilizando prompts se extraen los tópicos de un texto [174, 199].

6. DETECCIÓN DE TÓPICOS EN COMUNIDADES

6.1. Agrupación por comunidades

En este esquema novedoso para la detección de tópicos en redes sociales, primero se define un grafo de retweets en términos de $G = (N, E)$, donde los usuarios son los nodos N y los retweets entre ellos son las aristas E [23]. Dado que un usuario puede retwittear múltiples veces los tweets de otros usuarios, las aristas tienen pesos. Una comunidad en una red social es un grupo de usuarios que interactúan principalmente entre sí pero no con otros grupos [180]. Determinamos estas comunidades utilizando el método Louvain para la detección de comunidades [36], que busca maximizar la modularidad mediante un algoritmo de optimización goloso. Por lo tanto, cada comunidad agrupa a los usuarios según sus interacciones. En nuestro método, agrupamos en un solo documento todos los tweets escritos por todos los usuarios en cada comunidad. Por lo tanto, hay tantos documentos como comunidades en la red de retweets. En comparación con la mayoría de los esquemas anteriores, el número de palabras en un documento es mayor y el número de documentos es menor, lo que resulta en una matriz de co-ocurrencia de palabras más densa, lo cual es beneficioso para el algoritmo LDA [17].

6.2. Experimentación

En esta sección se describe el armado experimental y los resultados obtenidos al comparar nuestro método con técnicas de agrupación clásicas.

6.2.1. Dataset

Con el fin de evaluar los esquemas en diferentes escenarios y mostrar la robustez de la metodología, utilizamos dos conjuntos de datos de diferentes características. Nuestros experimentos utilizaron datos de la Streaming API de Twitter ¹. De manera similar a trabajos anteriores, construimos dos conjuntos de datos diversos recopilando posts que contenían diferentes términos o palabras clave y luego se le asignó como etiqueta la palabra clave presente en el mismo[6, 96, 150, 164]. Así mismo, eliminamos todos los tweets que contenían múltiples palabras clave, para preservar la singularidad de las etiquetas de los tweets, lo cual fue importante para nuestro análisis. Además, procesamos los tweets en minúsculas y eliminamos las stop-words. Todos los tweets están en inglés. Los dos conjuntos de datos son:

Conjunto de datos genérico: 115,359 tweets desde el 15 de diciembre hasta el 16 de diciembre de 2020, sobre una amplia gama de temas y recopilados usando las siguientes palabras clave (porcentaje de tweets recuperados por cada una): música (36,78 %), familia (23,94 %), salud (17,21 %), negocios (14,90 %), películas (4,70 %), deportes (2,44 %).

Conjunto de datos de eventos: 328,452 tweets del 20 de enero de 2021. Un conjunto de datos compuesto por tweets pertenecientes a un evento específico: el día de la inauguración del presidente de EE. UU., Biden. Utilizamos las siguientes palabras clave: Biden (69,45 %), joebiden (21,75 %), kamalaharris (4,74 %), inauguration2021 (4,04 %).

¹ <https://developer.twitter.com/en>

6.2.2. Evaluación

Dado que no hay una manera estándar de evaluar modelos de detección de tópicos, trabajos anteriores evaluaron los métodos de agrupación propuestos utilizando diferentes métricas o tareas. Con el fin de presentar un análisis completo y exhaustivo, en este capítulo evaluamos los esquemas mediante las múltiples métricas utilizadas en los diferentes trabajos anteriores: métricas de agrupación de temas (Pureza e Información Mutua Normalizada) [4, 17, 89, 150, 164, 181], una tarea de clasificación supervisada de aprendizaje automático [80, 96], una tarea de recuperación de documentos [6, 17] y el tiempo total de ejecución [17]. Explicamos brevemente cada una de ellas.

Pureza: Definimos cada clúster como un tema y asignamos los tweets al tema correspondiente con la probabilidad más alta (una cantidad estimada con LDA). La pureza de un clúster mide la fracción de tweets en un clúster que tienen la etiqueta de consulta asignada al clúster [194]. Formalmente, sea T_i el conjunto de tweets en el clúster de temas LDA i y Q_j el conjunto de tweets con la etiqueta de consulta j . Sea $T = \{T_1, T_2, \dots, T_{|T|}\}$ el conjunto de tamaño $|T|$ de todos los T_i y sea $Q = \{Q_1, Q_2, \dots, Q_{|Q|}\}$ el conjunto de tamaño $|Q|$ de todos los Q_j . Entonces, la pureza se define de la siguiente manera:

$$\text{Pureza}(T, Q) = \frac{1}{|T|} \sum_{i \in \{1 \dots |T|\}} \max_{j \in \{1 \dots |Q|\}} |T_i \cap Q_j| \quad (6.1)$$

Un puntaje de pureza más alto refleja una mejor representación del clúster y una mejor descomposición LDA.

Información Mutua Normalizada (IMN): La IMN mide la calidad del clúster utilizando la teoría de la información y se define formalmente de la siguiente manera:

$$\text{IMN}(T, Q) = \frac{2I(T, Q)}{H(T) + H(Q)} \quad (6.2)$$

donde $I(\cdot, \cdot)$ es la información mutua, una métrica de la cantidad de información que una variable aleatoria contiene sobre otra variable aleatoria. Luego, se normaliza usando la entropía $H(\cdot)$ de cada una de las variables, ya que esta da una medida de la incertidumbre o aleatoriedad en la misma [194].

Los valores mínimos y máximos de IMN son 0 cuando las etiquetas y los clústeres son conjuntos independientes y 1 cuando los resultados del clúster coinciden exactamente con todas las etiquetas.

Tarea de clasificación supervisada de aprendizaje automático: Para la tarea supervisada de aprendizaje automático, seguimos un esquema básico de evaluación de clasificación de aprendizaje automático [96]. Separamos el conjunto de datos en dos partes (entrenamiento y prueba), entrenamos un clasificador con el primero y evaluamos en el segundo. El primer 80% de los tweets (según el momento en que fueron publicados) se asignaron al conjunto de entrenamiento y el otro 20% al conjunto de prueba. Para esta evaluación, usamos la descomposición en tópicos como las características de un tweet, y la palabra clave con la que fue recolectado como su etiqueta. Entrenamos un clasificador de naive Bayes [155] y reportamos la medida F (F1 score) global en el conjunto de prueba.

Tarea de recuperación de documentos: También evaluamos la descomposición de temas de los diferentes métodos de agrupación en una tarea de recuperación de documentos (information retrieval), utilizando la misma división de entrenamiento-prueba que la tarea de clasificación supervisada. Utilizamos cada tweet en el conjunto de prueba como una

Tab. 6.1: Características de los documentos (doc) para los distintos métodos de agregación de textos.

Método	#doc genérico	evento	Max #palabras/doc genérico	Media #palabras/doc genérico	Media #palabras/doc evento
No agregado	115,359	328,452	783	1,023	137
Autor	36,526	87,883	36,029	11,240	369
Hashtag	34,624	59,388	820,689	3,736,132	8295
Conversación	35,484	67,276	12,480	41,024	141
Red	36,882	88,314	59,195	90,391	385
Comunidades	24,657	31,303	2,077,085	5,284,617	874
					1379

consulta y devolvemos los tweets más similares del conjunto de entrenamiento, según su descomposición de temas LDA. Si el tweet recuperado tiene la misma etiqueta de consulta, lo consideramos relevante. Más concretamente, la metodología es la siguiente: aplicamos LDA utilizando las diferentes técnicas de agrupación en el conjunto de entrenamiento, para cada tweet en el conjunto de prueba calculamos su descomposición de temas, calculamos la similitud coseno entre su descomposición de temas y la descomposición de temas de todos los tweets en el conjunto de entrenamiento y recuperamos los 10 tweets de entrenamiento más similares. Luego, calculamos la medida F1 [187] para saber si las categorías de los tweets recuperados coinciden con la categoría del tweet de prueba. Esta tarea recrea un escenario de recomendación de contenido basado en tweets anteriores.

Tiempo de ejecución: El tiempo medido (en segundos) incluye la agrupación de tweets (agregando los tweets en diferentes documentos) y la modelización de temas LDA, que varía según el número total de documentos de cada método de agrupación.

Todos los experimentos sobre detección de tópicos se realizaron utilizando el mismo hardware: una tarjeta gráfica NVIDIA GTX 1080.

6.2.3. Resultados

Para cada esquema de agrupación de textos, replicamos el flujo de trabajo de entrenamiento utilizado en la literatura y utilizamos un modelo LDA con 10 temas [150, 164]. Como mencionamos anteriormente, los trabajos anteriores mostraron que tener una matriz de co-ocurrencia más densa (menos documentos con más palabras cada uno) es beneficioso para LDA [17]. La Tabla 6.1 informa las características del corpus y muestra cómo nuestro modelo propuesto redujo drásticamente el número de documentos e incrementó el número de palabras por documento.

Los resultados de los experimentos se pueden ver en la Tabla 6.2. Los mejores rendimientos se marcan en negrita. La tabla muestra que nuestro método propuesto, la agrupación por comunidades, tiene el mejor rendimiento de todos los métodos examinados en todas las métricas para el Conjunto de Datos Genérico y en todas las métricas excepto la tarea de recuperación para el Conjunto de Datos de Eventos.

Más concretamente, nuestra metodología obtuvo la mejor calidad de agrupación, con los puntajes más altos de Pureza e IMN. Además, nuestros experimentos mostraron que la agrupación por comunidades superó a los esquemas anteriores en la tarea de clasificación supervisada, indicando que esta descomposición de temas fue un buen descriptor de la etiqueta de consulta. En cuanto a la tarea de recuperación de documentos, esta evaluación considera pequeños cambios en la descomposición de temas de un tweet, ya que utiliza

Tab. 6.2: Resultados de los diferentes métodos para los dataset evento (Even) y genérico (Gen).

Método	Pureza		IMN		Clasificación (F1)		Retrieval (F1)		tiempo	
	Gen	Even	Gen	Even	Gen	Even	Gen	Even	Gen	Even
No agregado	0.664	0.733	0.436	0.110	0.814	0.843	0.837	0.893	137	388
Autor	0.696	0.736	0.374	0.149	0.798	0.859	0.839	0.900	429	926
Hashtag	0.724	0.719	0.383	0.066	0.779	0.762	0.839	0.869	1,737	17,758
Conversación	0.658	0.733	0.436	0.110	0.814	0.843	0.835	0.908	738	1,569
Red	0.695	0.736	0.372	0.149	0.798	0.859	0.840	0.910	1131	2,841
comunidades	0.780	0.779	0.439	0.310	0.827	0.889	0.843	0.868	141	340

la similitud coseno entre esta descomposición en lugar de tener en cuenta solo el tema más probable, como hicimos anteriormente con las métricas de agrupación. Los resultados indican que la agrupación por comunidades tuvo el mejor rendimiento en un conjunto de datos genérico donde los temas de las etiquetas (“familia”, “salud” o “negocios”) se diferencian entre sí. En contraste, encontramos que el método basado en la red tiene una mejor puntuación en esta tarea para el conjunto de datos de eventos, donde las etiquetas están estrechamente relacionadas (el presidente y la vicepresidenta de Estados Unidos: “joebiden” y “kamalaharris”). La agrupación por comunidades tiene un mejor rendimiento en todas las tareas y conjuntos de datos, con la única excepción de la tarea de recuperación en el conjunto de datos de eventos.

Finalmente, la agrupación por comunidades tuvo el mejor rendimiento en tiempo entre todos los métodos de agrupación. Dado que la complejidad temporal de LDA depende del número de documentos [171] y la agrupación por comunidades redujo considerablemente el número de documentos al agrupar en un solo documento todos los tweets publicados por los usuarios de cada comunidad (ver Tabla 6.1), se deduce porque nuestro método fue más rápido que todas las demás técnicas de agregación (menos de la mitad del tiempo de ejecución).

6.3. Conclusiones

Presentamos una nueva forma de agrupar tweets para mejorar la calidad de la modelización de temas LDA en Twitter, sin requerir ninguna modificación en el algoritmo subyacente de LDA. La agrupación por comunidades propuesta utiliza la información de interacción de los usuarios y agrupa en un solo documento todos los tweets de los usuarios que pertenecen a una comunidad en la red de retweets.

Nuestro método fue evaluado y comparado con múltiples técnicas de agrupación en diferentes tareas, incluyendo calidad de agrupación, un problema de clasificación supervisada y tareas de recuperación. Los resultados en dos conjuntos de datos heterogéneos indican que la agrupación basada en comunidades supera a todas las demás estrategias de agrupación en todas las tareas y métricas, con la única excepción de la tarea de recuperación en el conjunto de datos de eventos. Además, el análisis del tiempo de ejecución muestra que la agrupación por comunidades tiene una mejora significativa en el rendimiento temporal en comparación con los métodos de agrupación anteriores, debido a su capacidad para reducir el número total de documentos.

Parte III

MEDIOS DE COMUNICACIÓN ONLINE: ANÁLISIS DE NOTICIAS Y SUS EFECTOS

Esta sección contiene resultados que fueron publicados en los siguientes papers:

- Albanese, F., Pinto, S., Semeshenko, V., & Balenzuela, P. (2020). Analyzing mass media influence using natural language processing and time series analysis. *Journal of Physics: Complexity*, 1(2), 025005.
- Pinto, S., Albanese, F., Dorso, C. O., & Balenzuela, P. (2019). Quantifying time-dependent media agenda and public opinion by topic modeling. *Physica A: Statistical Mechanics and its Applications*, 524, 614-624.

7. INTRODUCCIÓN

Los medios masivos de comunicación (o mass media) desempeñan uno de los roles más importantes en el proceso de formación de la opinión pública. Más allá de informar sobre hechos y eventos, los medios de comunicación ofrecen una interpretación de tales eventos, brindando a las personas una forma de entender su relevancia. A través de su capacidad para reflejar la realidad desde su propia perspectiva, los medios masivos de comunicación determinan la importancia relativa otorgada a diferentes temas, un proceso conocido como agenda-setting. La teoría de agenda-setting se resume comúnmente en la cita “maybe media does not tell you what to think, but what to think about” (“quizás los medios de comunicación no te dicen qué pensar, pero sí sobre qué pensar”) [147, 148]. En otras palabras, los medios masivos de comunicación pueden indicar qué es y qué no es importante, y en qué medida lo es. Por lo tanto, el poder de definir la agenda de los medios masivos de comunicación produce un efecto importante, que adquiere relevancia, por ejemplo, en la formación de opiniones políticas durante períodos electorales [73]. En este contexto, hay una diferencia entre los temas cubiertos por los medios masivos de comunicación (Agenda Mediática) y aquellos que el público considera prioritarios (Agenda Pública).

Además, investigaciones previas han demostrado cómo la percepción pública de un evento político se ve modificada por los medios de comunicación [34]. La influencia suele manifestarse en función de los temas enfatizados y omitidos por dichos medios [41]. Otros estudios han demostrado que leer diferentes medios de manera sostenida lleva a las personas a modificar su ideología política, alineando sus votos con el punto de vista editorial de un periódico en particular [163]. Gerber y Dean [77] estudiaron las elecciones a gobernador de 2005 en el estado de Virginia (Estados Unidos) y observaron que la lectura de periódicos afectó las decisiones de algunos votantes y produjo cambios inducidos en la percepción de los políticos.

En las últimas décadas, dada la disponibilidad de datos y recursos computacionales, el análisis cuantitativo de la influencia de los medios de comunicación se ha abordado desde diferentes perspectivas. Por un lado, la aparición de las redes sociales y la disponibilidad de datos en línea enriquecieron la investigación sobre el impacto de los medios de comunicación masivos. Por ejemplo, King et al. [114] detectaron un aumento en el número de tweets sobre un tema específico después de estar expuestos a noticias relacionadas. Yasseri y Bright [229] mostraron que el número de visitas a la página de Wikipedia de un candidato es un buen predictor de cambios en los resultados electorales. La difusión y el consumo de noticias falsas en las redes sociales se analizaron en [14], donde los autores muestran que las personas tienden a compartir noticias falsas que refuerzan su sesgo ideológico. El papel del sentimiento de las noticias (connotación positiva o negativa) se ha explorado en el marco de cómo la connotación se relaciona con la variación de los indicadores económicos [203], cómo afecta las expectativas públicas sobre la economía [97, 137], o cómo moldea la opinión pública sobre un tema en particular [62]. En la misma línea, [24, 83] abordaron cómo el sentimiento de los tweets de bots o usuarios influyentes induce la connotación de las expresiones de sus seguidores en Twitter. También se emplearon técnicas como el modelado de temas para describir la dinámica de los temas en los medios de comunicación [119, 120, 177] y cómo se relaciona con la respuesta de la audiencia.

Por otro lado, también se han implementado modelos computacionales con el fin de evaluar diferentes mecanismos de interacción entre los medios de comunicación y los individuos, y cómo los medios de comunicación influyen en la formación de opiniones públicas colectivas [81, 175, 198]. Como ejemplos, en [202] se muestra que los periódicos y Twitter tienen una reacción opuesta a los cambios de las tasas de desempleo; en [87], se explora las similitudes y diferencias entre los medios tradicionales y las redes sociales en cómo se comunican noticias sobre tiroteos masivos en Estados Unidos; en [13], los autores muestran cómo las fluctuaciones de la actividad de Twitter en diferentes regiones dependen de la ubicación de los ataques terroristas; y en [191], la compleja interacción entre los medios sociales y los tradicionales se sigue a lo largo del tiempo en un conjunto de cuestiones predefinidas, pero generales.

Es importante señalar que los artículos citados anteriormente trabajan sobre un solo tema o sobre un conjunto de temas predefinidos, generalmente seleccionados por los investigadores. Sin embargo, se puede realizar una selección de temas basada en datos utilizando una herramienta frecuentemente empleada en el análisis de grandes corpus de documentos: la detección automática de tópicos (previamente descrita en el capítulo 4 de esta tesis).

En los siguientes dos capítulos de esta sección se describen dos trabajos que buscan caracterizar la comunicación online de los medios de comunicación:

Primero, estudiar la influencia que tiene la Agenda Mediática en contextos electorales. Los resultados muestran una relación causal significativa entre algunos temas y la diferencia en las encuestas de opinión pública entre Clinton y Trump durante las elecciones presidenciales del 2016. En otras palabras, podemos afirmar que la mención de ciertos temas (como por ejemplo inmigración o escándalos de corrupción) en noticias periodísticas impactan significativamente en la opinión pública durante contextos electorales.

En segundo lugar, también analizamos la relación entre la Agenda Mediática y la Agenda Pública, donde esta última está dada por los contenidos buscados y publicados por usuarios en redes sociales. Resulta de interés saber quién le impone la agenda a quién. ¿Son los medios quienes adaptan su agenda para apelar a los intereses del público? ¿O son los usuarios quienes modifican su agenda temática a partir de la influencia de los medios de comunicación?

8. LA INFLUENCIA DE LOS MEDIOS DE COMUNICACIÓN EN LA OPINIÓN PÚBLICA

En este capítulo, estudiamos la relación entre los medios de comunicación y la opinión pública utilizando una combinación de análisis de sentimiento y detección de temas en artículos de noticias. Como prueba de concepto, aplicamos esta metodología a un caso de estudio particular que es la campaña presidencial estadounidense de 2016, analizando artículos periodísticos donde se menciona a los candidatos involucrados. Consideramos el número de menciones de cada candidato, su contenido de sentimiento y la evolución de la cobertura relativa de un conjunto de temas de los mismos artículos (agenda política de los medios). Más allá del caso particular estudiado, la metodología introducida en este trabajo es muy general y puede extenderse fácilmente a otros casos de interés.

8.1. Dataset

En 2016 se celebraron elecciones presidenciales en Estados Unidos, en las que se enfrentaron la candidata demócrata Hillary Clinton y el candidato republicano Donald Trump, que finalmente ganó las elecciones. Centramos nuestro análisis en dos tipos de datos: las encuestas, como proxy de la opinión pública y las noticias de cuatro de los principales medios de comunicación estadounidenses. Todo el período analizado comprende desde el 28 de julio (última convención del partido, donde se definieron formalmente los candidatos) hasta el 8 de noviembre de 2016 (fecha de elecciones).

8.1.1. Datos de encuestas

Analizamos un total de 263 encuestas nacionales realizadas por diferentes agencias (un promedio de 2,7 encuestas por día), en las que se midió la previsión de votos de cada candidato en un intervalo de unos pocos días (alrededor de 3-5 días). Estos datos, que muestran el resultado de diferentes encuestadores, fueron descargados del sitio web de Real Clear Politics (RCP) [178]. De acuerdo a lo especificado por RCP, estas encuestas nacionales pertenecen a una muestra demográfica equilibrada.

La figura 8.1 muestra la serie temporal del porcentaje potencial de votos para cada candidato (panel superior) y la diferencia entre estas series temporales (el porcentaje de Clinton menos el porcentaje de Trump, panel inferior). Cada punto de la serie temporal representa el promedio de la semana anterior. Es decir, se utilizó un promedio de ventana móvil de 7 días, lo que significa que cada punto de la serie temporal toma en cuenta un promedio de 19 encuestas.

La figura 8.1 muestra que Clinton mantuvo una ventaja sobre Trump durante todo el período. Sin embargo, esta ventaja se vio afectada por algunas variaciones.

8.1.2. Artículos de noticias

Seleccionamos las ediciones en línea de The New York Times, Fox News, CNN y USA Today para realizar nuestro análisis. El criterio de selección es que son los medios de comunicación masiva más populares en términos de búsquedas en línea en todo el territorio de Estados Unidos. The New York Times es un periódico clásico con sede en la ciudad de

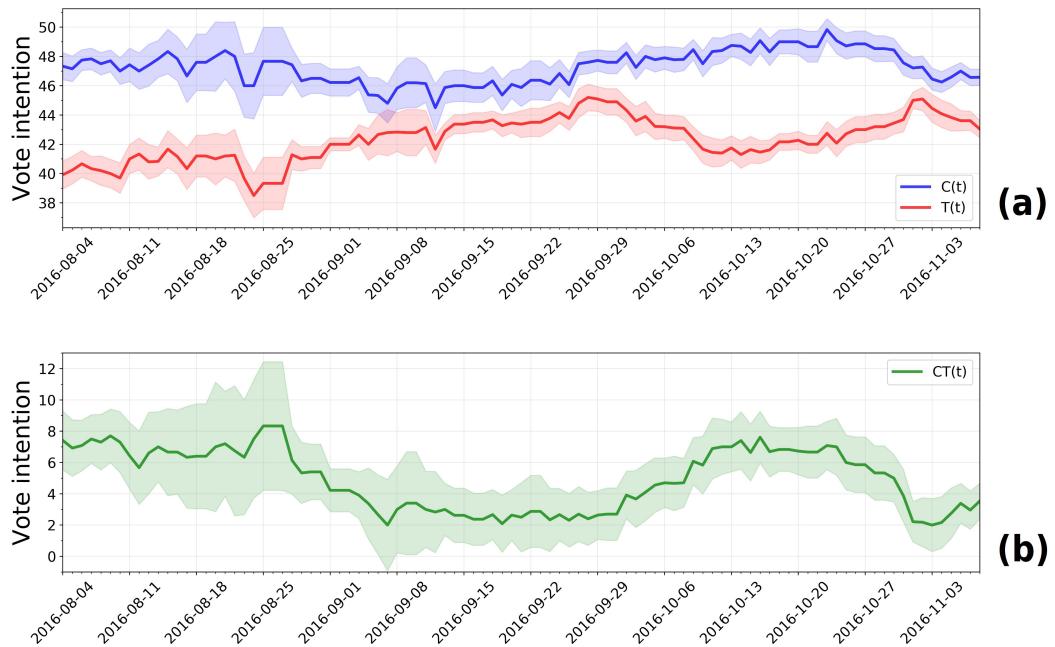


Fig. 8.1: **Encuestas de intención de voto de las elecciones presidenciales de Estados Unidos de 2016.** Arriba: Las curvas de Hillary Clinton (candidata demócrata) y Donald Trump (candidato republicano) para las elecciones presidenciales de Estados Unidos de 2016. Abajo: La diferencia entre estas curvas.

Nueva York, con influencia y lectores en todo el mundo, y Fox News y CNN son canales de televisión que transmiten a muchos países alrededor del mundo. USA Today es un periódico estadounidense de distribución internacional, de mercado medio, que se publica diariamente.

Seleccionamos los artículos que contienen al menos el nombre de uno de los dos principales candidatos: Hillary Clinton (Demócrata) y Donald Trump (Republicano). El corpus analizado está compuesto por un total de 15.175 artículos: 5.672 de The New York Times, 5.750 de Fox News, 2.920 de CNN y 833 de USA Today. Incluimos todos los artículos que mencionan a uno o ambos candidatos, independientemente de la sección a la que pertenecían.

8.2. Métodos

Centramos nuestro análisis en técnicas de minería de textos, de las que extraemos información útil aplicando tanto análisis de sentimiento como detección de tópicos a artículos de noticias.

8.2.1. Procesamiento del lenguaje natural

Análisis de sentimiento

Para medir la frecuencia de menciones positivas y negativas de un candidato determinado, implementamos un algoritmo de análisis de sentimiento. El análisis de sentimiento se

realizó a través de modelos recursivos profundos para la composición semántica aplicados a árboles de sentimiento [201], en particular mediante la implementación de procesamiento de lenguaje natural Stanford CoreNLP [143]. Este algoritmo consiste en armar un árbol a partir de la estructura gramatical y un análisis sintáctico de cada frase. Luego, a cada palabra (nodo) se le asigna un valor de sentimiento, tomado de una base de datos: muy positivo, positivo, neutral, negativo o muy negativo. Además, este algoritmo tiene en cuenta si las palabras son intensificadoras, apaciguadoras, negadoras, etc. El algoritmo asigna un valor de sentimiento a cada nodo comenzando desde los nodos externos. Después de varias iteraciones, termina asignando el valor de sentimiento correspondiente a la frase total.

Existen varios algoritmos para realizar análisis de sentimiento, como aquellos basados en la extracción de características de las oraciones [64] o enfoques basados en léxico para la minería de opiniones [154, 208]. Dado que nuestro corpus de noticias está formado por frases gramaticalmente correctas, el Stanford CoreNLP es el algoritmo adecuado para realizar análisis de sentimiento.

Detección de tópicos

Además del análisis de sentimiento, también realizamos una detección de temas en el corpus de los artículos de noticias utilizando técnicas de aprendizaje no supervisadas como se implementó en [177] y se describe en el capítulo 5 de esta tesis. De esta forma, se representó a los artículos de noticias como vectores numéricos con la representación tf-idf y luego se detectaron los temas principales en dicho corpus aplicando NMF.

Para calcular la cobertura temática de los medios de comunicación, estimamos la cantidad de artículos y su importancia relativa en cada tema. Por este motivo definimos el peso del tema i (T_i) como el producto de la cantidad de artículos de noticias (ponderados por el grado de pertenencia del documento) y la extensión del artículo. La cobertura se puede definir diariamente (distribución temporal) o para todo el período (distribución promedio). La ecuación (8.1) muestra la cobertura para un solo día d ,

$$T_i(d) = \sum_j l(j) \cdot h_{ji} \cdot \delta_{d_j, d}, \quad (8.1)$$

donde $l(j)$ es el número de palabras en el documento j ; h_{ji} (elemento de la matriz H) es el grado de pertenencia del documento j al tema i ; d_j es la fecha del documento j ; y δ es el delta de Kronecker, que vale 1 si $d_j = d$ y vale 0 en caso contrario. Dado que cada vector de documento puede tener varios componentes distintos de cero, un documento puede tratar sobre más de un tema. Para reducir el ruido, finalmente aplicamos un filtro lineal con una ventana deslizante de siete días de ancho y normalizamos los perfiles temporales.

8.2.2. Medidas de correlación y causalidad

Correlación de Spearman

El coeficiente de correlación de Spearman es una medida no paramétrica de correlación de rango. Evalúa qué tan bien se puede describir la relación entre dos variables utilizando una función monótona. Mientras que la correlación de Pearson mide la relación lineal entre las dos variables, la correlación de Spearman evalúa las relaciones monótonas entre ellas. Antes de calcular la correlación de Spearman, eliminamos la tendencia lineal respectiva, si la hubiera, de todas las series temporales para evitar correlaciones espurias.

Causalidad de Granger

El test de causalidad de Granger [84] determina si una serie de tiempo es capaz de pronosticar otra. Dada una serie de tiempo estacionaria, x_t , modelada mediante un modelo de promedio móvil autorregresivo Eq.(8.2), la prueba de causalidad de Granger determina si el modelo de Eq.(8.3) es mejor que el modelo de la ecuación (8.2). En otras palabras, esto significa que la información adicional proporcionada por una segunda serie temporal, $y_{t-\tau}$, mejora la predicción de x_t .

La cantidad de términos que deben incluirse en la ecuación (8.2) y (8.3) se determinan estudiando tanto la autocorrelación como la correlación parcial de x_t , donde los w_t son términos de ruido blanco, y los θ , ϕ y β son solo coeficientes [200]. Si β es significativamente diferente de cero, podemos decir que y_t tiene una relación causal con x_t . Se puede notar que cuando β es cero, se recupera el modelo de la ecuación (8.2).

$$x_t = \sum_i \phi_i x_{t-i} + \sum_j \theta_j w_{t-j} + w_t \quad (8.2)$$

$$x_t = \sum_i \phi_i x_{t-i} + \sum_j \theta_j w_{t-j} + w_t + \beta y_{t-\tau} \quad (8.3)$$

8.3. Resultados

Aquí analizamos los datos de las encuestas y los artículos de noticias (extrayendo el contenido de sentimiento y la descomposición temática) en el período electoral desde el 28 de julio hasta el 8 de noviembre de 2016. Este período comprende desde la última convención de los partidos en 2016, donde se definieron formalmente los candidatos, hasta la fecha de las elecciones.

8.3.1. Número total de menciones

Como primera aproximación, comparamos la serie temporal de las encuestas con el número total de menciones de cada candidato en ambos medios [229]. Estas curvas se muestran en la Figura 8.2. Este primer análisis se realizó independientemente del contexto y la connotación sentimental en la que aparecían las frases.

Calculamos el coeficiente de correlación de Spearman [131] entre el número total de menciones de cada candidato (Figura 8.2) con el diferencial entre los datos de las encuestas (Figura 8.1 (b)). Tenemos en cuenta que los cambios en la cobertura mediática pueden no reflejarse instantáneamente en las encuestas, ya sea porque la escala temporal de cómo los medios pueden ejercer influencia no está clara, o la fecha de publicación de las encuestas es posterior a la recopilación de datos. Por lo tanto, calculamos una correlación con lag o desfasaje entre las series temporales para un rango de rezagos.

Encontramos que el número de menciones de ambos candidatos en el New York Times, CNN y USA Today se correlaciona positivamente con la diferencia entre Clinton y Trump en las encuestas, con un coeficiente de correlación promedio de 0,663 (NYT), 0,426 (CNN) y 0,246 (USA) respectivamente. Esto significa que cuando aumenta el número de menciones en estos medios, aumenta la diferencia Clinton menos Trump, sin importar qué candidato se menciona. Por otro lado, estas correlaciones son negativas para series temporales similares en Fox News, con un coeficiente de correlación promedio de -0,476. En este caso, cuando aumenta el número de menciones a cualquiera de los candidatos en Fox

News, la diferencia Clinton menos Trump disminuye. Estos resultados son estadísticamente significativos ($p < 0,001$) para un desfase entre 7 y 15 días para The New York Times, Fox News y CNN, mientras que para USA Today son significativos para un desfase entre 12 y 15 días. ($p < 0,05$). Se pueden llegar a conclusiones similares al estudiar las menciones de los candidatos por separado.

Los signos de correlación dependen de los medios de comunicación independientemente de los candidatos. Para profundizar en las causas de este comportamiento, aplicamos análisis de sentimiento y modelado de temas en los artículos.

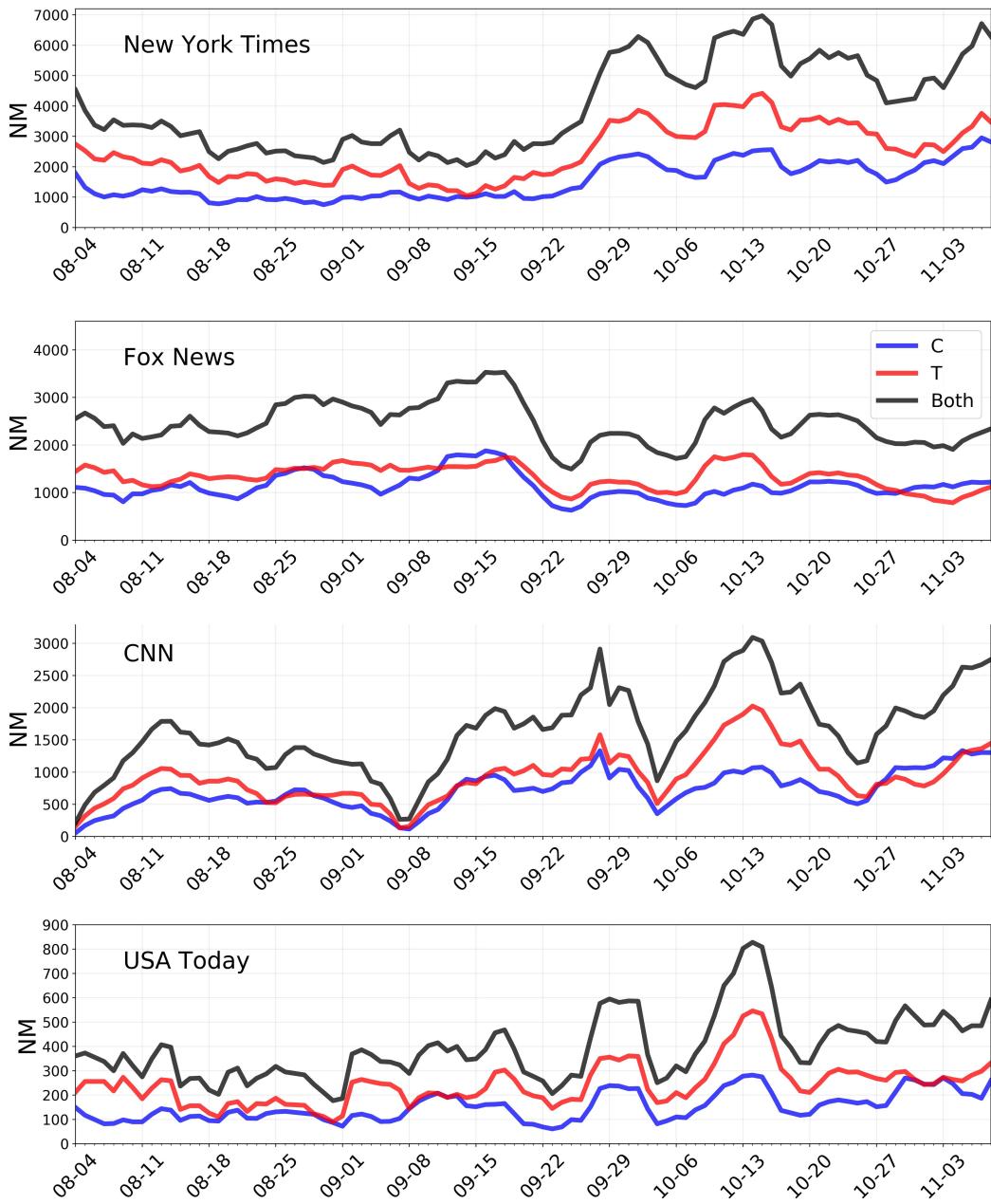


Fig. 8.2: **Series temporales del número total de menciones.** Las curvas muestran el número total de menciones de Hillary Clinton (en azul), Donald Trump (en rojo) y la suma de ambos (en negro) para cada medio de comunicación (de arriba a abajo, New York Times, Fox News, CNN y USA Today).

8.3.2. Análisis de sentimiento

Con el objetivo de estudiar la connotación con la que se menciona a cada candidato, aplicamos el algoritmo clasificador de sentimientos descrito anteriormente. El procedimien-

to es el siguiente:

1. En cada texto, detectamos frases que mencionan términos “Hillary”, “Clinton”, “Donald” o “Trump”. En el caso de que se menciona más de un candidato, separamos las oraciones mediante análisis sintáctico.
2. Aplicamos el análisis de sentimiento para estas oraciones y contamos la cantidad de menciones positivas, negativas y neutrales para cada uno de los candidatos. En este paso, los modelos recursivos profundos para la composición semántica juegan un papel central, ya que el análisis sintáctico de una oración permite comprender cuándo el texto se refiere a un determinado candidato de manera positiva o negativa.

Luego de este procedimiento, registramos para cada día del período estudiado, la cantidad de frases relacionadas con cada candidato, así como su sentimiento. Con base en esta clasificación, definimos una estadística de sesgo de sentimiento SB (Ec. (8.4)), donde $\#C_+$ ($\#C_-$) representa la fracción de menciones positivas (negativas) de Hillary Clinton y $\#T_+$ ($\#T_-$) representa la fracción de menciones positivas (negativas) de Donald Trump, en un medio de comunicación determinado. SB es una medida del sesgo hacia uno de los candidatos: si $SB > 0$, el sesgo es positivo hacia Clinton en comparación con Trump, y por otro lado, si $SB < 0$, el sesgo es positivo hacia Trump.

$$SB = (\#C_+ - \#C_-) - (\#T_+ - \#T_-) \quad (8.4)$$

Calculamos el valor de SB para cada medio. Los resultados de este análisis son $SB_{NYT} = 0,162 \pm 0,004$ para The New York Times, $SB_{USA} = 0,160 \pm 0,010$ para USA Today, $SB_{CNN} = 0,094 \pm 0,006$ para CNN y $SB_{FN} = 0,046 \pm 0,005$ para Fox News. En todos los casos, rechazamos que SB sea un valor negativo ($p < 0,001$), mediante bootstrapping [68, 69]. Aunque en todos los casos SB es un valor positivo, la estadística Sentiment Bias es significativamente baja para Fox News, mientras que no encontramos diferencias significativas entre The New York Times y USA Today.

En resumen, encontramos que $SB_{NYT}, SB_{USA} > SB_{CNN} > SB_{FN}$. Por ejemplo, esto sugiere que The New York Times, USA Today y CNN parecen mencionar a Hillary Clinton de una manera más positiva que Fox News.

El hecho de que SB sea positivo para todos los medios sugiere que este análisis por sí solo no es suficiente para entender el comportamiento del diferencial entre Clinton y Trump en las encuestas. La dependencia temporal de la estadística Sentiment Bias ($SB = SB(t)$) proporcionará una idea útil para comprender las diferencias de comportamiento entre ambos medios. Para avanzar en esta dirección necesitamos analizar primero los temas tratados y su relevancia relativa.

8.3.3. Detección de tópicos

Clasificamos el corpus de noticias en seis temas y denominamos Agenda de Medios a la importancia relativa que cada medio otorga al conjunto de temas, calculada por la Ec. (8.1) y definido en [177].

El primer tema es sobre elecciones en general y está representado por palabras como campaña, elección, candidato, etc. Este resultado es consistente con el hecho de que analizamos noticias políticas durante el período de campaña. Elegimos descartarlo dada

su falta de especificidad. Los otros cinco temas revelan los temas discutidos durante la campaña electoral, que etiquetamos y describimos brevemente como:

- “Controversia de emails”: cubre la famosa controversia que enfrentó Hillary Clinton durante las elecciones debido al uso de su servidor privado de correo electrónico para comunicaciones oficiales.
- “Fundación Clinton”: trata sobre las acusaciones de posibles conflictos de intereses debido al hecho de que Clinton era Secretaria de Estado y aceptaron donaciones extranjeras en su fundación.
- “Economía”: trata particularmente sobre impuestos, ingresos, empleos y negocios.
- “Inmigración”: trata sobre la discusión de las políticas migratorias entre México y Estados Unidos, planteada en la campaña de Donald Trump.
- “Relaciones exteriores”: trata sobre la política exterior de los Estados Unidos. En particular, se centra en ISIS y la hipotética interferencia de los rusos en el proceso electoral.

Las palabras clave que definen estos cinco temas están representadas en las nubes de palabras de los paneles superiores de la Figura 8.3 y corresponden a las palabras más significativas que describen la similitud entre los artículos de noticias agrupados en un tema determinado. Vale la pena señalar que este es un método no supervisado y por lo tanto las palabras clave surgen del corpus de noticias analizado y no fueron elegidas arbitrariamente. La figura 8.3 también muestra la evolución temporal de estos temas.

La Agenda comparativa de cada medio se puede visualizar fácilmente en la Figura 8.4. Observamos que el New York Times enfatiza los temas *Economía* y *Relaciones Exteriores*, mientras que Fox News da más cobertura a *Fundación Clinton* e *Inmigración*. Por otro lado, podemos ver que tanto CNN como USA Today cubren *Economía*, compartiendo este interés con el New York Times, pero también prestan atención al tema *Controversia de emails* como lo hace Fox News.

Como hicimos antes con el número de menciones, calculamos la correlación de Spearman para un rango de desfases entre la evolución de la cobertura del tema y la dispersión entre candidatos. Encontramos que casi todos los coeficientes de correlación son significativos para retrasos de alrededor de 10 o 15 días, excepto “*Relaciones exteriores*”, que no muestra una correlación significativa para tres de los cuatro medios de comunicación considerados (ver Tabla 8.1).

Los resultados mostrados en la Tabla 8.1 indican que el contexto en el que se menciona a los candidatos juega un papel clave en la correlación con la intención de voto. Por ejemplo, la serie temporal correspondiente a los temas *Controversia de emails*, *Fundación Clinton* e *Inmigración* se correlaciona negativamente con la dispersión entre candidatos (es decir, la cobertura de estos tres temas empeora la imagen de Clinton), a la inversa de lo que pasa con los temas *Economía* y *Relaciones Exteriores* en el New York Times. Es notorio, como podemos ver en la Figura 8.4, que estos dos temas son más enfatizados por New York Times y CNN, y el primero también por USA Today, mientras que Fox News cubre con mayor intensidad los tres primeros, lo que parece afectar la intención de Clinton. Estos resultados son consistentes con el cálculo de *SB* en la subsección anterior.

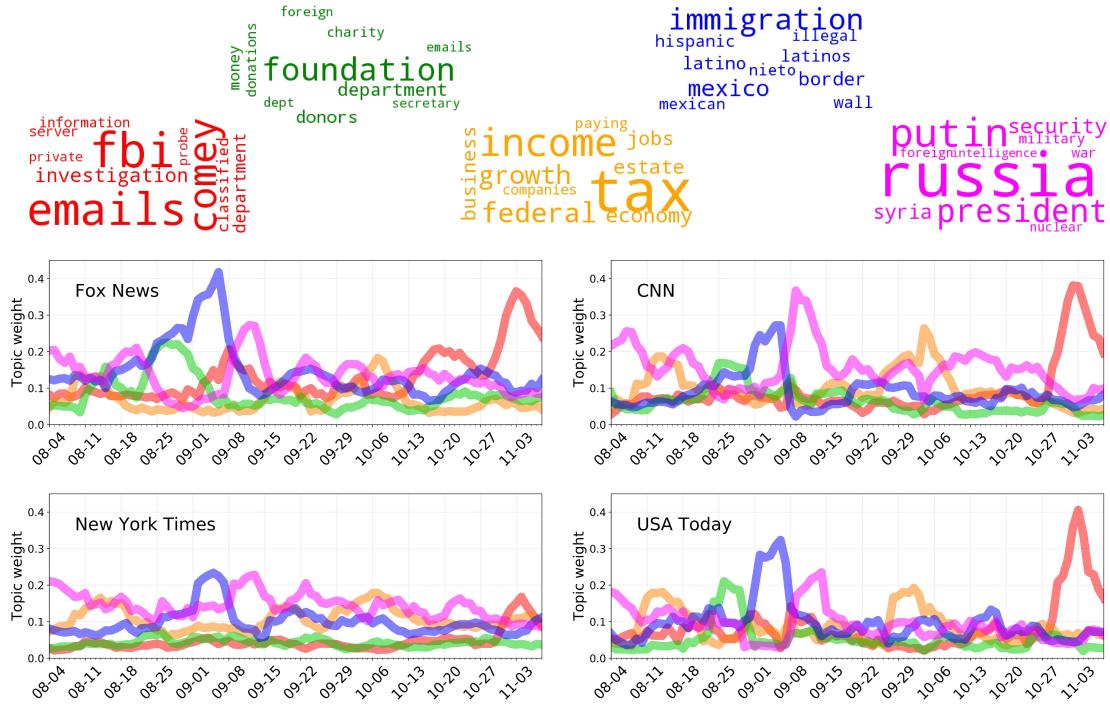


Fig. 8.3: Cobertura de temas para cada medio de comunicación. Mostramos la evolución temporal de los temas durante todo el período, donde también se puede observar la diferencia entre estos medios de comunicación. Cada tema está especificado por las nubes de palabras en la parte superior de la figura (y también se señalan en el texto principal). Tema 1 (rojo): Controversia de emails; Tema 2 (verde): Fundación Clinton; Tema 3 (amarillo): Economía; Tema 4 (azul): Inmigración; Tema 5 (magenta): Relaciones exteriores.

Tab. 8.1: Correlación lineal.

Tópicos	NYT (SRL)	Fox (SRL)	CNN (SRL)	USA (SRL)
Controversia de emails	-0.46 (10-20)	-0.42 (11-20)	-0.45 (13-20)	-0.54 (10-20)
Economía	0.56 (4-20)	0.59 (8-20)	0.48 (5-18)	0.40 (10-15)
Fundación Clinton	-0.53 (3-20)	-0.43 (15-20)	-0.53 (1-20)	-0.40 (5-12)
Inmigración	-0.42 (0-12)	-0.44 (5-20)	-	-0.47 (12-20)
Relaciones Exteriores	0.44 (17-20)	-	-	-

Coeficiente de correlación de Spearman promedio entre la cobertura del tema y la dispersión entre candidatos en un rango de rezagos estadísticamente significativo (SRL) ($p < 0,001$). No se informan valores no significativos.

8.3.4. Combinación de sentimiento y análisis de temas

Aunque la estadística de sesgo de sentimiento (SB) promedio es positiva independientemente del medio de comunicación, existe un período de tiempo en el que, por ejemplo, $SB_{FN}(t) < 0$ o $SB_{CNN} < 0$ (Ver Figura 8.5). Aquí proponemos un análisis combinado para comprender mejor este comportamiento.

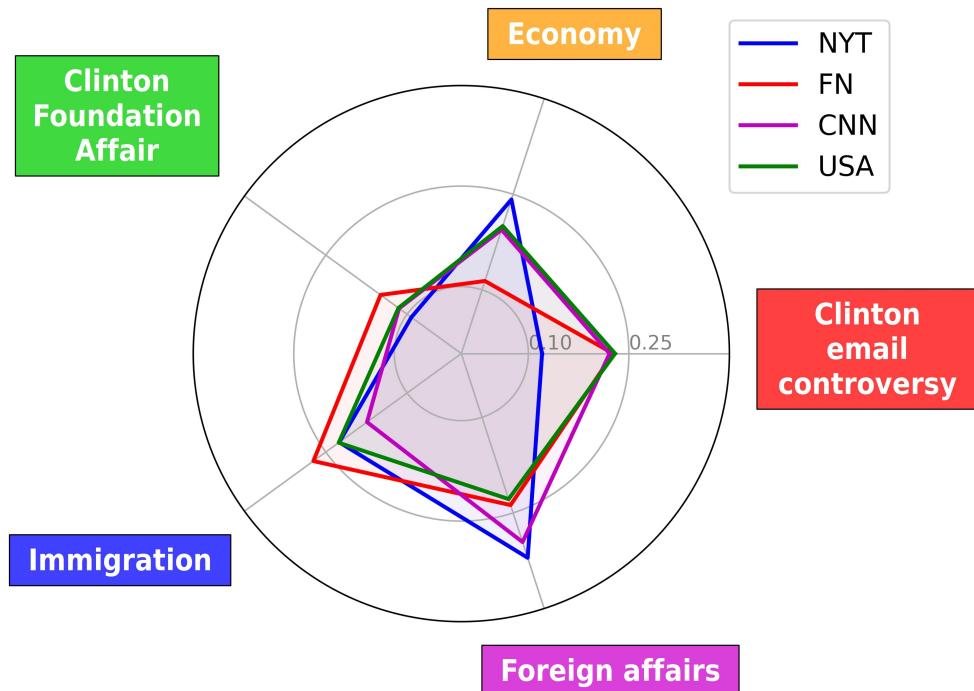


Fig. 8.4: Cobertura de temas para cada medio de comunicación durante todo el período.

El gráfico radar muestra la cobertura acumulativa. Es decir, la agenda de cada medio de comunicación sobre los cinco temas.

En la tabla 8.2, calculamos la estadística Sentiment Bias (SB) para cada tema, discriminado por medio de comunicación. Observamos que en los primeros tres temas el signo de *SB* coincide con el signo de la correlación mostrada en la Tabla 8.1, mientras que el signo del último tema coincide con la única correlación significativa de esa tabla. La única discrepancia está en el tema *Inmigración*. Estos resultados revelan que el análisis de sentimientos es mucho más informativo cuando los artículos de noticias se descomponen en temas que cuando se agrupan todos juntos.

El éxito de este análisis combinado se puede ver nuevamente en la Figura 8.5 donde, además de la evolución temporal de *SB* para cada medio, podemos ver los gráficos radar de las agendas para dos fechas específicas, que pertenecen a períodos donde $SB_{FN}(t), SB_{CNN} < 0$. En esas fechas, podemos ver que la diferencia entre agendas puede explicarse en parte por el tema de la controversia sobre los correos electrónicos de Clinton, que fue más enfatizado por Fox News y CNN que por el New York Times, más notorio al final del período.

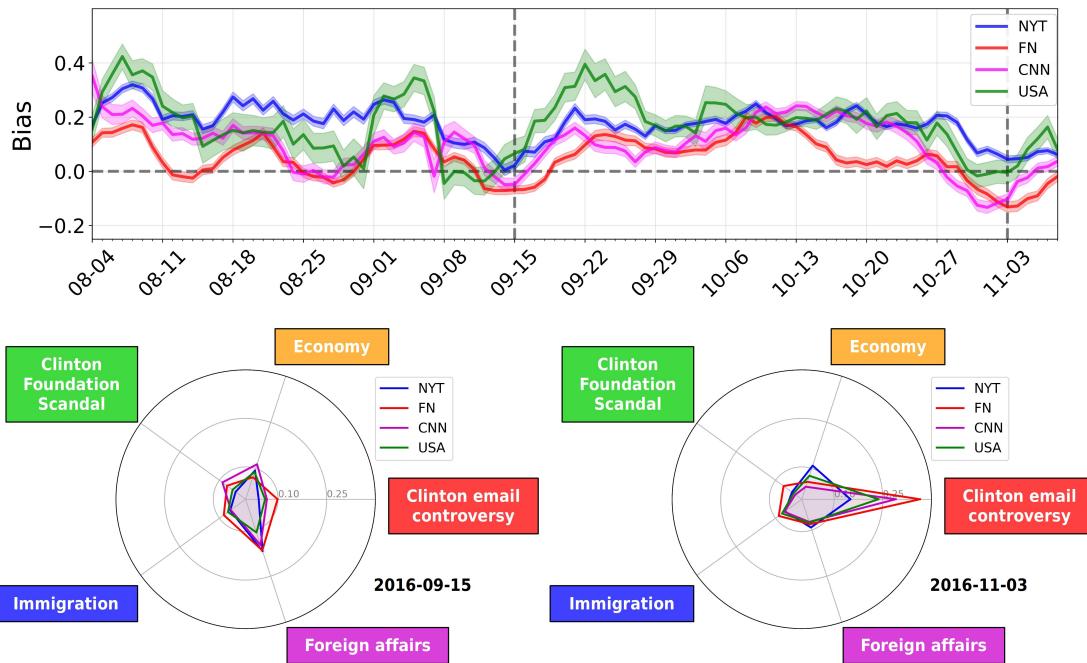


Fig. 8.5: **Serie temporal del sesgo de sentimiento y gráficos radar.** Las curvas en la parte superior muestran la evolución temporal de la estadística de Sesgo de Sentimiento SB para cada periódico, New York Times en azul, Fox News en rojo, CNN en magenta y USA Today en verde. Las figuras en la parte inferior son los gráficos radar de la cobertura sobre los cinco temas para dos días específicos: 2016-09-15 y 2016-11-03 (líneas verticales punteadas en el panel superior), que corresponden a los valores más negativos de SB para Fox News y CNN.

Tab. 8.2: El análisis de sentimiento de cada tópico.

Tópico	SB_{NYT}	SB_{FN}	SB_{CNN}	SB_{USA}
Controversia de emails	-0.475	-0.429	-0.302	-0.315
Economía	0.332	0.070	0.152	0.168
Foundación Clinton	-0.256	-0.257	-0.304	-
Inmigración	0.501	0.347	0.306	0.382
Relaciones Exteriores	0.146	0.053	0.115	0.166

La estadística Sentiment Bias SB se calcula con las noticias de cada tema y de cada diario junto con el signo de la correlación del mismo tema con la diferencia entre Clinton y Trump en las encuestas. En todos los casos rechazamos la hipótesis de que SB tiene signo opuesto con $p < 0,001$. No se informan valores no significativos.

8.3.5. Causalidad

En esta sección, buscamos una relación causal entre la dispersión de las encuestas ($CT(t)$) y las series temporales de los temas aplicando el marco de causalidad de Granger descrito en la sección 8.2.2. Debido a que G no es una serie estacionaria, comenzamos

calculando su primera diferencia ΔCT , la cual es estacionaria (Prueba aumentada de Dickey-Fuller [195], $p < 0,05$) y por lo tanto puede modelarse mediante modelos autorregresivos. Al estudiar la autocorrelación total y parcial de ΔCT , notamos que se describe esencialmente mediante un paseo aleatorio según la ecuación. (8.5), con w_t un valor aleatorio estándar distribuido normalmente. Proponemos dos modelos para el análisis causal: uno descrito por la ecuación. (8.5), y el otro incluyendo la información sobre la cobertura de temas dentro de un cierto rezago τ (Eq. (8.6), donde $T_i(t)$ es el peso del tema i en el tiempo t).

$$\Delta CT(t) = CT(t) - CT(t-1) = w_t \quad (8.5)$$

$$\Delta CT(t+\tau) = \beta \cdot \Delta T_i(t) + w_{t+\tau} \quad (8.6)$$

Decimos que la cobertura de un tema determinado afecta efectivamente la dispersión entre candidatos cuando el parámetro β en la ecuación. (8.6) difiere significativamente de cero. Dado que ambos modelos implican sólo primeras diferencias, la interpretación adecuada es que un valor distinto de cero de β implica que el crecimiento o disminución de un tema determinado predice una variación en el diferencial con un cierto desfasaje.

Los temas con β significativamente diferentes de cero ($p < 0,01$), con un signo y un rezago consistentes con los resultados de las correlaciones lineales calculadas en secciones anteriores, son: *Controversia de emails* ($\beta < 0$ y $\tau = 19$) para Fox News; *Economía* ($\beta > 0$ y τ entre 11 y 16) para Fox News, CNN y USA Today; y *Fundación Clinton* ($\beta < 0$ y $\tau = 19$) e *Inmigración* ($\beta < 0$ y $\tau = 10$), ambos para The New York Times.

Finalmente, es importante resaltar el papel del tema *Controversia emails*. Los datos analizados sugieren que este tema juega un papel clave en el periodo de tiempo cercano a la jornada electoral. Esto se puede observar en la cobertura de Fox News y CNN de la Figura 8.3 y en los gráficos de radar de la Figura 8.5, donde este tema tuvo una mayor cobertura durante la última semana. El análisis sugiere que, cuando este tema se vuelve más importante en los medios de comunicación, hay una reducción notoria en la diferencia entre Clinton y Trump. Esta es la razón por la cual nuestro modelo reporta una relación causal entre este tema y la diferencia de las encuestas. Curiosamente eso ocurre al final del periodo, lo que sugiere que este fue un tema clave en el resultado electoral.

8.4. Conclusiones

La influencia de los medios de comunicación en la opinión pública ha sido estudiada desde diferentes perspectivas y metodologías, que van desde experimentos de campo hasta análisis de datos y modelos computacionales. La gran disponibilidad de datos provenientes de los medios de comunicación masivos y redes sociales hace que las técnicas de análisis basadas en datos sean importantes en la investigación de este tipo de cuestiones.

En este capítulo sugerimos un conjunto de herramientas basadas en técnicas de procesamiento del lenguaje natural para ser aplicadas en el estudio de los efectos que los medios de comunicación pueden producir en la opinión pública. En particular, se abordan qué características de los artículos de noticias están relacionadas con cambios mensurables en la opinión pública, utilizando análisis de sentimiento y detección de temas de artículos de

noticias. Específicamente, es importante el contenido de sentimiento de los artículos de noticias o el tema de las menciones.

Cada método por sí solo, el análisis de sentimientos y la descomposición de temas, se ha aplicado ampliamente para estudiar problemas relacionados. Sin embargo, hasta donde sabemos, el uso de la combinación de los dos métodos para analizar el impacto de los medios de comunicación aún no se ha hecho de manera sistemática.

Al aplicar la metodología desarrollada a la cobertura mediática de las elecciones presidenciales estadounidenses de 2016, podemos comprender aspectos clave en la relación entre los medios de comunicación y la opinión pública. En este ejemplo, analizamos artículos de noticias en los que se mencionaba al menos uno de los dos candidatos involucrados. Realizamos un análisis de sentimiento y detección de temas en ese corpus y los comparamos con medidas de intención de voto como proxy de la opinión pública.

Nuestro enfoque permite extraer información útil de este ejemplo, como se puede ver en la lista a continuación:

- El número total de menciones de ambos candidatos en artículos periodísticos se correlaciona positivamente con la diferencia entre Clinton y Trump en las encuestas del New York Times, CNN y USA Today, pero negativamente en Fox News, independientemente del candidato.
- El análisis de sentimiento promedio de los artículos de noticias donde se menciona a ambos candidatos no es suficiente para explicar el comportamiento anterior.
- El análisis temático, que permite apreciar la diferencia entre la agenda de ambos medios, muestra que la cobertura de determinados temas se correlaciona con la diferencia entre Clinton y Trump en las encuestas.
- El análisis de sentimiento discriminado por tema es consistente con estos últimos resultados (excepto Inmigración), dado que los temas a favor de Clinton muestran sentimiento positivo hacia Clinton y los a favor de Trump son negativos hacia Clinton (o positivos hacia Trump).
- Existe una relación causal ($p < 0,05$) en el sentido de Granger entre cuatro temas y la diferencia entre Clinton y Trump en las encuestas, eso significa que estos temas sirven como buenos predictores de las variaciones en las encuestas.
- El tema relacionado con la controversia sobre los correos electrónicos de Clinton parece ser el más relevante porque:
 1. Se correlaciona negativamente con la diferencia entre Clinton y Trump en las encuestas.
 2. Tiene una relación causal significativa con la diferencia entre Clinton y Trump en las encuestas, lo que, como dijimos antes, determina este tema como un buen predictor de la variación en las encuestas.
 3. Explica los valores negativos de la estadística Sentiment Bias de Fox News ($SB_{FN}(t)$) y CNN ($SB_{CNN}(t)$).

Es importante señalar que la elección de los medios de comunicación no limita el análisis realizado, ya que podemos agregar tantos medios como sea necesario.

A lo largo de este capítulo se analizo el efecto de los medios de comunicación y su agenda temática en la opinión publica. Sin embargo, los individuos de una sociedad también manejan una agenda: los temas de los que hablan y les interesan. De esta forma, resultar interesante analizar el papel de una agenda publica, de las redes sociales, y su relación a la impuesta por los medios masivos de comunicación. En el próximo capítulo ahondaremos en este tema.

9. LA INFLUENCIA ENTRE LOS MEDIOS DE COMUNICACIÓN Y LAS REDES SOCIALES

En este capítulo, proponemos realizar un estudio de los tópicos de artículos periodísticos para analizar la dinámica entre los medios de comunicación y las agendas públicas en redes sociales. Definimos *Agenda de Medios* como la distribución en evolución temporal de la cobertura de un tema por diferentes medios de comunicación y *Agenda Pública* como la distribución del interés público en el mismo espacio temático, observando tanto las búsquedas en Google como la actividad en la red social Twitter. Aplicamos este método para estudiar la dinámica de la agenda de los medios argentinos debido a nuestra familiaridad con los antecedentes políticos, pero como se puede ver a lo largo del trabajo, la metodología implementada es general y puede extenderse fácilmente a otros conjuntos de datos.

9.1. Métodos

9.1.1. La agenda de los medios

Analizamos un período de tres meses de la agenda mediática argentina compuesto por un corpus de artículos periodísticos publicados entre el 31 de julio de 2017 y el 5 de noviembre de 2017. Los artículos provienen de la sección política de las ediciones online de los diarios argentinos *Clarín*, *La Nación*, *Página12* y del portal de noticias, *Infobae*. Los dos primeros lideran la venta de ediciones impresas en la ciudad de Buenos Aires, pero *Clarín* llega aproximadamente al doble de lectores de *La Nación*, y a diez veces a los lectores de *Página 12* [61], quienes fueron elegidos por su orientación política de centro izquierda. Por otro lado, *Infobae* tiene el sitio web más visitado, mucho más que los de *Clarín* y *La Nación* [60]. El corpus analizado está compuesto por 11815 artículos periodísticos: 2908 de *Clarín*, 3565 de *La Nación*, 3324 de *Página 12* y 2018 de *Infobae*. Excepto *Página 12*, todos los artículos fueron tomados de la sección *Política* de los respectivos portales de noticias, mientras que los artículos que pertenecen a la *Página 12* fueron tomados de la sección *El país*.

Al igual que en el capítulo anterior, los artículos se describen como vectores numéricos a través de la representación tf-idf [226]. La dimensión de los vectores ($t = 445993$) está dada por la cantidad total de palabras del corpus después de eliminar las no informativas como preposiciones y conjunciones. Para detectar los temas principales en el corpus, realizamos nuevamente NMF en la matriz de términos del documentos. También de forma análoga al capítulo anterior de la sección 8, definimos la Agenda de Medios como la distribución de los pesos del tema y su evolución en el tiempo.

9.1.2. Agenda Pública: Google Trends y la red social Twitter

Además de la construcción de la Agenda de Medios, es importante tener alguna medida de los intereses públicos y construir lo que llamamos la Agenda Pública. Para lograr este objetivo, tomamos a Google Trends y la red social Twitter como representantes de los intereses públicos online, buscando los mismos temas en el mismo período de tiempo. Aprovechamos las palabras clave de los tópicos para realizar consultas en la herramienta

Google Trends (*GT*) y en la Streaming API de Twitter (*Tw*), y por tanto obtenemos el peso relativo de las búsquedas y tweets en cada plataforma. La distribución de las búsquedas en Google Trends y el número de posts en Twitter es lo que definimos como Agendas Públicas.

En la Tabla 9.1 señalamos las palabras clave involucradas en las consultas. Respecto a Google Trends, obtuvimos directamente la cantidad relativa de búsquedas de las palabras clave. Por otro lado, en Twitter inferimos esa información tomando una muestra de tweets por día. Durante todo el periodo estudiado, recogimos un total de 24.360 posts.

Nótese que la forma en que construimos la Agenda Pública está definida por los temas que se encuentran en la Agenda de Medios y por lo tanto no sería posible encontrar temas de interés público que no hayan sido publicados en los medios de comunicación en el periodo analizado. Sin embargo, la misma limitación intrínseca que proporciona esta metodología nos permite definir ambas agendas en el mismo espacio temático y por tanto realizar medidas de comparación adecuadas, como mostraremos en la siguiente subsección.

9.2. Resultados

9.2.1. Descomposición en tópicos

Inicialmente realizamos una descomposición temática del corpus de noticias del período de tres meses informado anteriormente, centrándonos en los diez temas más importantes. La razón detrás de factorizar el corpus en diez temas se basó en tener una representación dimensional baja del corpus y una interpretación clara de los temas debido a nuestro conocimiento previo del trasfondo político. Descubrimos que esta factorización nos permitió sacar conclusiones útiles. Sin embargo, en futuras investigaciones se pueden tener en cuenta metodologías más sofisticadas para estimar el número de temas en un corpus.

Las palabras clave que definen los diez temas están representadas en las nubes de palabras de Fig 9.1. Dada nuestra interpretación de las palabras clave encontradas en tres de ellos, unimos estos temas como parte de un mismo macrotema al que llamamos *Elecciones*. Lo mismo ocurre con otros dos temas que fueron clasificados como parte de un macrotema denominado *Santiago Maldonado*. Por lo tanto, los diez temas originales quedaron reducidos a siete, los cuales se señalan en la Fig 9.1. El significado de los temas o macrotemas se contextualiza en [176].

Finalmente, siguiendo el procedimiento descrito en la sección anterior, construimos las Agendas como distribuciones de temas cubiertos en medios (Media Agenda) y búsquedas públicas (Agenda Pública) en Google y Twitter.

En un trabajo futuro se podría replicar el análisis utilizando Community Pooling (el método propuesto en el capítulo 6).

9.2.2. La agenda mediática y la agenda pública

En Fig 9.1 mostramos los Medios (MA) y las Agendas Públicas (PA) (discriminadas por Google Trends (GT) y Twitter (Tw)) en una descomposición de siete temas de todo el corpus usando gráficos de radar para representar la distribución promedio. En esta figura también mostramos las nubes de palabras de las palabras clave que definen cada uno de los diez temas originales, donde el tamaño de la palabra refleja su importancia en la definición del tema. En color verde señalamos las palabras involucradas en las consultas de Google

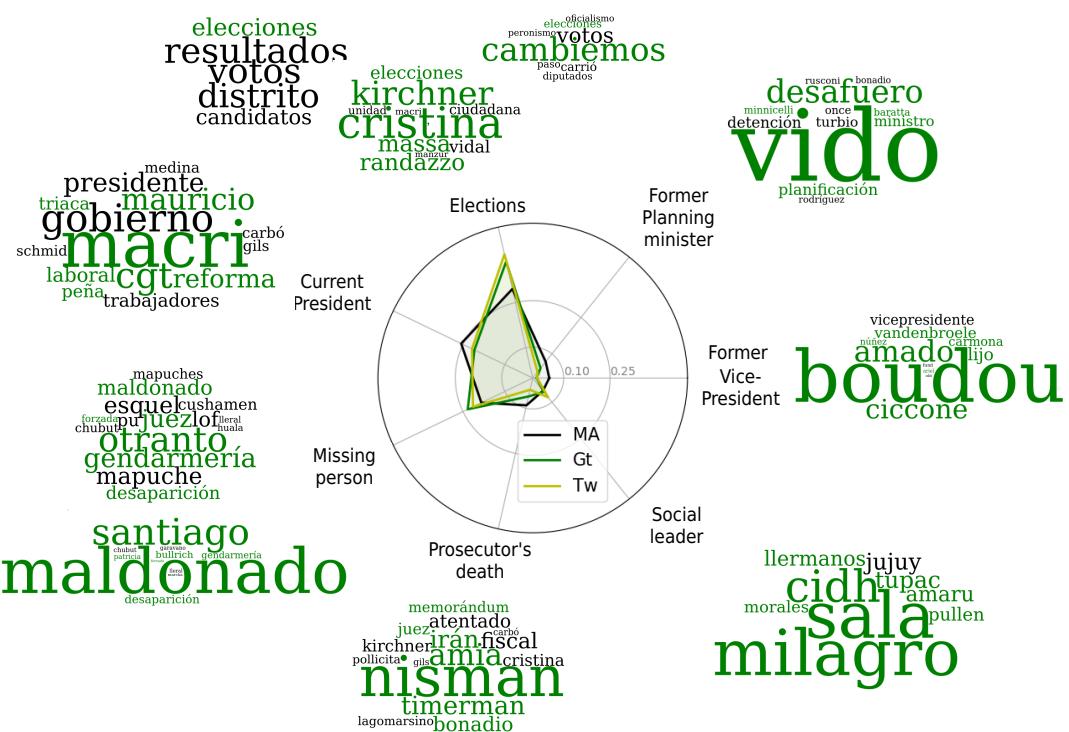


Fig. 9.1: Gráficos de radar de las agendas públicas y de medios representados por la distribución de diez temas (luego reducidos a siete) y sus correspondientes nubes de palabras. La agenda pública está representada tanto por Google Trends (GT) como por Twitter (Tw). Los nombres de los temas se introducen junto con las nubes de palabras que contienen las palabras clave más importantes involucradas en la definición de cada tema. En color verde mostramos las palabras clave utilizadas para definir los temas en las consultas de Google Trends y Twitter (ver Tabla 9.1) y por tanto en nuestra construcción de la Agenda Pública.

Trends y Twitter para construir la Agenda Pública. Las consultas empleadas también se especifican en Table 9.1.

Tab. 9.1: Consultas utilizadas en Google Trends y palabras empleadas en la búsqueda de Twitter (subrayadas) para la construcción de la Agenda Pública.

Tópico	Google Trends + Twitter (subrayado)
Elecciones	<u>elecciones</u> + <u>cambiamos</u> + <u>cristina kirchner</u> + <u>massa</u> + <u>randazzo</u>
Persona desaparecida	<u>santiago maldonado</u> + <u>juez otranto</u> + <u>patricia bullrich</u> + <u>gendarmería</u> + <u>desaparición forzada</u>
Ministro de planificación	<u>de vido</u> + <u>desafuero</u> + <u>ministro de planificación</u> + <u>minnicelli</u> + <u>baratta</u>
Presidente Macri	<u>mauricio macri</u> + <u>cgt</u> + <u>reforma laboral</u> + <u>peña</u> + <u>triaca</u>
Milagro Sala	<u>milagro sala</u> + <u>cidh</u> + <u>tupac amaru</u> + <u>pullen llermanos</u> + <u>morales</u>
Fiscal Nisman	<u>nisman</u> + <u>amia</u> + <u>memorándum</u> con <u>irán</u> + <u>timerman</u> + <u>juez bonadio</u>
Ex-Vicepresidente Boudou	<u>amado boudou</u> + <u>ciccone</u> + <u>ariel lijo</u> + <u>vandenbroele</u> + <u>núñez carmona</u>

Podemos ver que tanto GT como Tw se ven similares en esta representación, pero muestran diferencias específicas con la Agenda de Medios. Por ejemplo, se observa un mayor interés de la audiencia por el tema *Santiago Maldonado* que los medios de comunicación, o viceversa, se produce un menor interés por el tema *Fiscal Nisman*. Sin embargo, esta representación estática no logra mostrar la compleja dinámica de la evolución de las agendas y la importancia de hechos puntuales y específicos que pueden borrar o amplificar sus diferencias.

La evolución temporal de las agendas en el espacio temático se muestra como gráficos de las agendas en la Fig 9.2. El gráfico de relieve proporciona una visualización clara del peso relativo de los temas al mismo tiempo que su clasificación. En Fig 9.2, también destacamos algunos eventos importantes relacionados con la dinámica de los temas. Es posible apreciar cómo el tema principal cambia con el tiempo y observar las diferencias cualitativas entre las agendas. En particular, se pueden ver algunas diferencias entre las Agendas Públicas que no se observaron en Fig 9.1, como por ejemplo, la persistencia de los temas principales es mayor en Twitter que en Google Trends. Esto es más evidente al final del periodo analizado, donde los temas tratados en Google Trends cambian con mayor frecuencia que en Twitter.

También se calcularon las correlaciones lineales entre los mismos temas en Agenda Medios y Agenda Pública. En todos los casos, encontramos que las correlaciones son positivas y estadísticamente significativas, como se muestra en la Tabla 9.2. Interpretamos esto como una validación de los temas encontrados en el corpus y las palabras clave que lo describen. Aunque estamos particularmente interesados en aquellos períodos en los que las agendas difieren, se espera que los medios y los intereses públicos sigan en general un patrón similar, impulsado principalmente por acontecimientos externos.

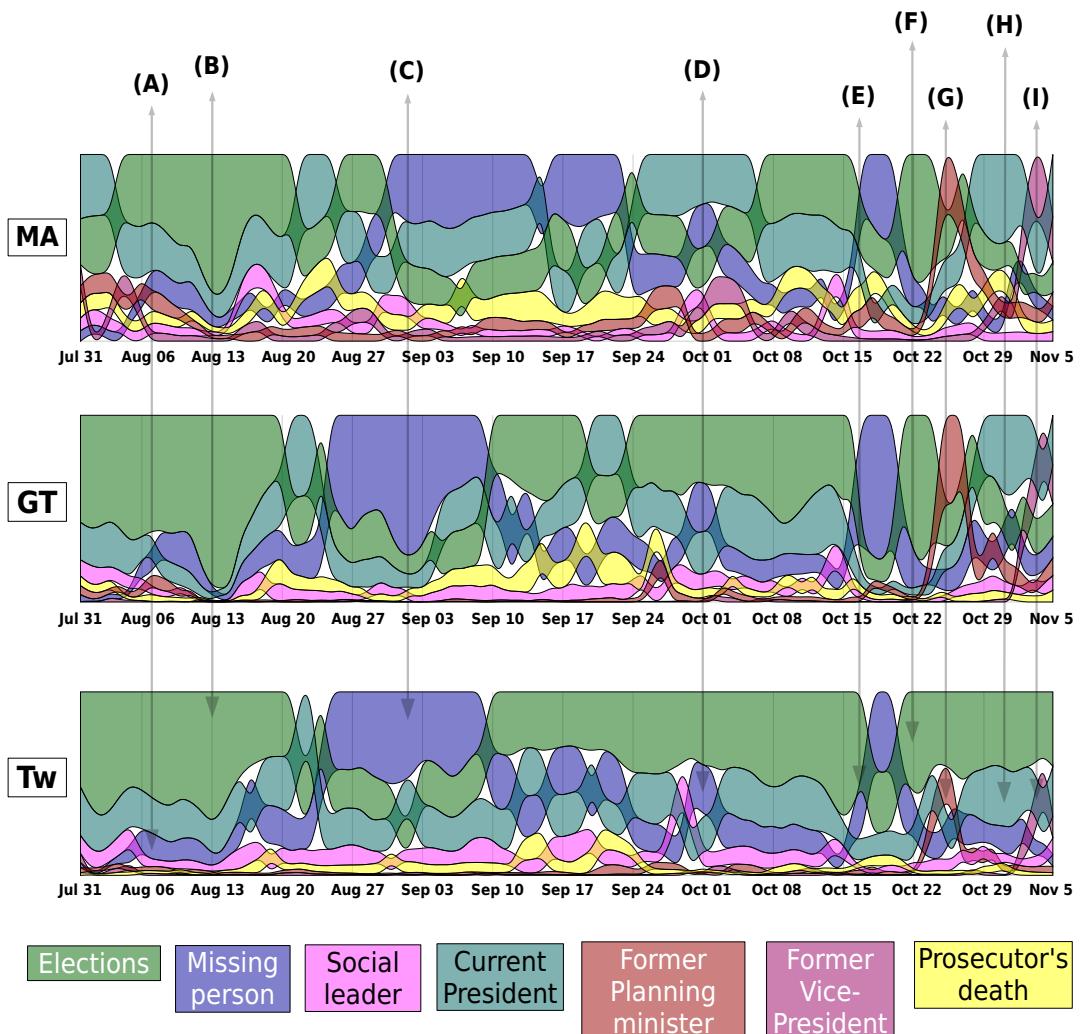


Fig. 9.2: Gráfico de la Agenda de los Medios (MA) y las Agendas Públicas dependientes del tiempo extraídos de Google Trends (GT) y Twitter (Tw). Los anchos y clasificaciones de las curvas codifican el peso relativo del tema. Además, se señalan algunos eventos importantes relacionados con los temas: **A**: Primeras noticias de la desaparición de Santiago Maldonado; **B**: Elecciones primarias; **C**: Marcha por la desaparición de Santiago Maldonado; **D**: Marcha por la desaparición de Santiago Maldonado; **E**: Aparición del cuerpo de Santiago Maldonado; **F**: Elecciones generales; **G**: Detención de Julio De Vido; **H**: Debate de la reforma laboral; **I**: Detención de Amado Boudou

Tab. 9.2: Correlación entre los perfiles temporales de los topicos de la agenda publica y de la agenda mediatica. Todos los valores son estadisticamente significativos ($p < 10^{-9}$), excepto (*), que es significativo con $p < 0,05$.

Tópico	Correlación MA y GT	MA y Tw	GT y Tw
Elecciones	0.81	0.59	0.75
Persona desaparecida	0.68	0.76	0.89
Ministro de planificación	0.92	0.82	0.87
Presidente Macri	0.77	0.75	0.63
Milagro Sala	0.49	0.25*	0.57
Fiscal Nisman	0.56	0.59	0.75
Ex-Vicepresidente Boudou	0.90	0.92	0.97

9.2.3. Diversidad de Agenda

¿Qué tan dominante es el tema principal? ¿El grado de predominio de un tema determinado en la Agenda de Medios se refleja en la Agenda Pública? La diversidad es una variable clave cuando se abordan múltiples temas [40], debido a que nos dice cómo se distribuye la atención entre los diferentes temas de discusión. Como se propuso en [40], utilizamos la entropía de Shannon normalizada H para cuantificar la diversidad dentro de nuestro marco.

$$H = - \sum p(x_i) * \ln p(x_i) \quad (9.1)$$

donde x_i representa un tema y $p(x_i)$ representa la proporción de atención que recibe ese tema.

En la Fig 9.3 podemos ver el valor de H en función del tiempo para las tres agendas. Es importante prestar atención a aquellos períodos de tiempo en los que la diversidad es menor de lo habitual. Este efecto es notoriamente más pronunciado en la Agenda Pública dada por GT, y en particular en cuatro días concretos en los que se pueden detectar cuatro mínimos locales de la entropía de Shannon. Tres de ellos son valores atípicos como se define anteriormente, dos de ellos de GT y uno de Tw. El otro no ha sido identificado como un valor atípico pero es un mínimo pronunciado y por lo tanto un punto de interés en nuestra descripción.

Un valor menor en la diversidad de agenda se debe a que el tema más importante atrae prácticamente toda la atención del público y de los medios, colapsando la agenda a uno de los temas involucrados. En los gráficos de radar incluidos en la Fig 9.3 podemos ver cómo dos de estos valores atípicos (**a** y **d**) pertenecen al tema *Elecciones*. Están relacionados con las elecciones legislativas primarias y generales que tuvieron lugar el 13 de agosto y el 22 de octubre respectivamente. En todas las agendas estos puntos se detectaron como atípicos excepto el punto (**d**) en la Agenda de Twitter. ¿Por qué es eso? El gráfico de radar de la agenda de Twitter para este día muestra una asociación entre el tema *Elecciones* y el *Presidente Macri*, disminuyendo la importancia de este tema. Las discusiones en Twitter sobre elecciones aparecen también en el punto (**c**), cuando las otras agendas parecen ser más diversas. Por otro lado, y a pesar de no catalogarse como outlier, también nos centramos en el punto (**b**) porque la Entropía de Shannon en la Agenda de Google muestra un mínimo (agenda colapsante) que no se corresponde ni en los Medios ni en las Agendas de Twitter. Rastreando el contexto vemos que pertenece al tema *Santiago Maldonado* y esta

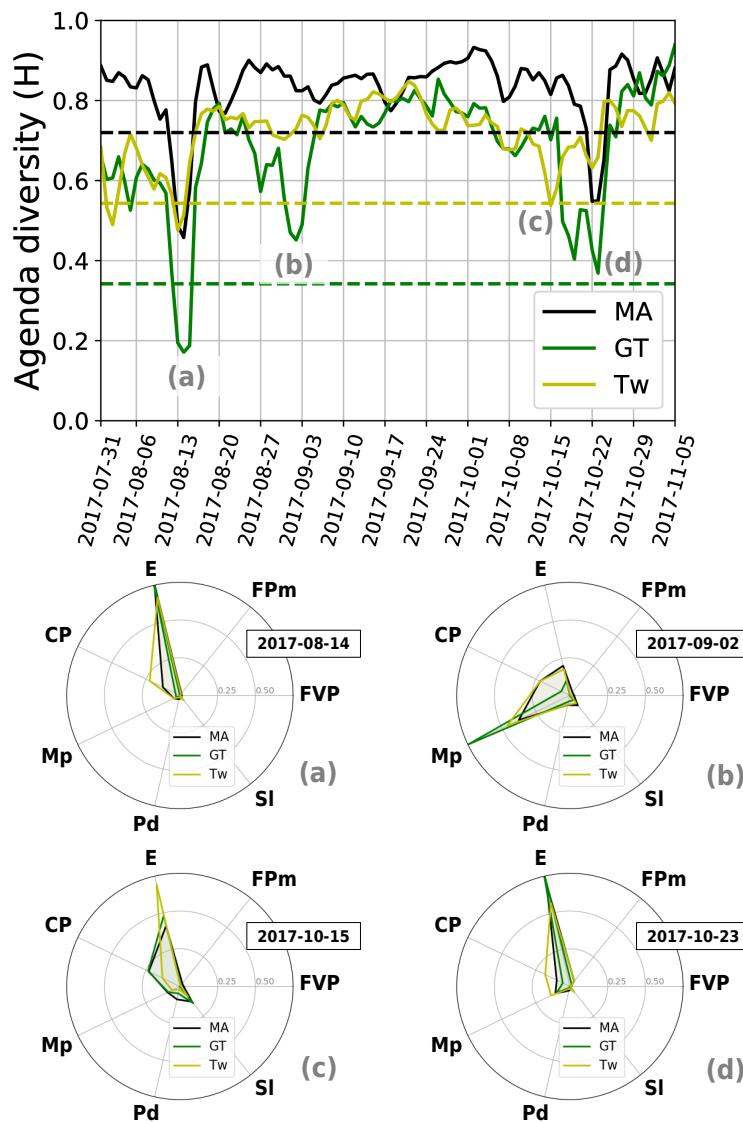


Fig. 9.3: entropía de Shannon (H) como métrica de diversidad de agenda. La Agenda Pública muestra un comportamiento menos diverso que la Agenda de Medios como se puede observar en el gráfico superior. Las líneas horizontales corresponden a las vallas interiores inferiores de cada señal para identificar valores atípicos. Los gráficos de radar relacionados muestran la agenda en los días seleccionados donde las series de tiempo presentan abandonos (puntos a-d), lo que indica que el tema más importante capta la mayor parte de la atención del público. **E:** Elecciones; **FPm:** Ministro de Planificación; **FVP:** Ex vicepresidente; **SI:** Líder social Milagro Sala; **Pd:** Muerte del Fiscal Nisman; **Mp:** Santiago Maldonado; **CP:** Presidente Macri.

fecha corresponde a la manifestación que se realizó un mes después de la desaparición de Santiago Maldonado.

A partir de la medida de H también hemos observado que la mediana de la diversidad de la Agenda Pública es estadísticamente significativamente menor que la mediana de la Agenda de Medios. Específicamente, $H_{GT} = 0,73$ y $H_{Tw} = 0,74$ son estadísticamente significativamente más bajos que $H_{MA} = 0,85$ con $p < 10^{-18}$, mientras que no hay una diferencia significativa entre los primeros dos. Sin embargo, en la Fig 9.3 podemos ver que GT muestra abandonos más abruptos en la diversidad en respuesta a eventos específicos. De todo este análisis podemos concluir que, dado un conjunto finito de temas, la Agenda Pública es menos diversa que la Agenda de Medios, porque el público parece centrarse en los temas más importantes de lo que los medios pueden hacerlo, tal vez debido a decisiones editoriales.

9.2.4. Distancia entre la Agenda de los medios y la Agenda pública

Si bien la diversidad es una propiedad de cada distribución, una pregunta natural al comparar diferentes distribuciones es qué tan similares son. Dadas nuestras descripciones de las agendas como distribuciones que evolucionan en el tiempo, podemos compararlas calculando la distancia de Jensen-Shannon. En este contexto, los valores atípicos en fechas seleccionadas corresponden a divergencias entre los Medios y la Agenda Pública: Eventos específicos cuando los intereses públicos no coinciden con la oferta mediática. En la Fig 9.4 mostramos la distancia Jensen-Shannon entre Medios y Agendas Públicas en función del tiempo. Nos centramos en tres puntos que parecen bastante relevantes. En todos los casos, las distribuciones temáticas de estos días mostradas muestran que el incremento en la distancia entre agendas se debe a un mayor interés de la opinión pública por el tema *Santiago Maldonado*.

Los puntos **(c)** y **(d)** de la Fig 9.4 muestran que tanto el público como los medios resaltan este tema, pero los medios no ignoran otros temas, por lo que la distancia correspondiente entre ellos puede interpretarse como falta de diversidad en la Agenda Pública como se discutió en la última sección. Por otro lado, los puntos **(a)** (tomamos este punto por ser un máximo local a pesar de no ser un valor atípico) y **(b)** muestran un mayor interés del público en el tema *Santiago Maldonado* que no se refleja en los Medios. En la Fig 9.2 podemos ver que este tema pasa a ser el más importante en interés público (tanto en GT como en Tw) días antes de que suceda en la Agenda de Medios. Este hecho se puede asociar a una campaña en redes sociales (como Facebook y Twitter) a favor de la aparición de Santiago Maldonado que se llevó a cabo el 26 de agosto. Esta campaña fue masiva e inicialmente subestimada por los principales medios de comunicación argentinos.

Finalmente, es importante decir que la distancia Jensen-Shannon, junto con la medición de la diversidad de agenda dada por la entropía de Shannon, da una idea del comportamiento independiente de la audiencia y los medios en determinadas fechas particulares. Su identificación puede ser un punto de partida para estudiar la reacción de los medios ante un cambio en los intereses de la audiencia.

9.2.5. Sesgo de agenda en diferentes medios de comunicación

En este apartado dejamos de lado la Agenda Pública y estudiamos la composición de la Agenda de Medios en cada medio de comunicación. En la Fig 9.5 mostramos las agendas de los medios representadas como diagramas de radar de forma análoga a la que se muestra

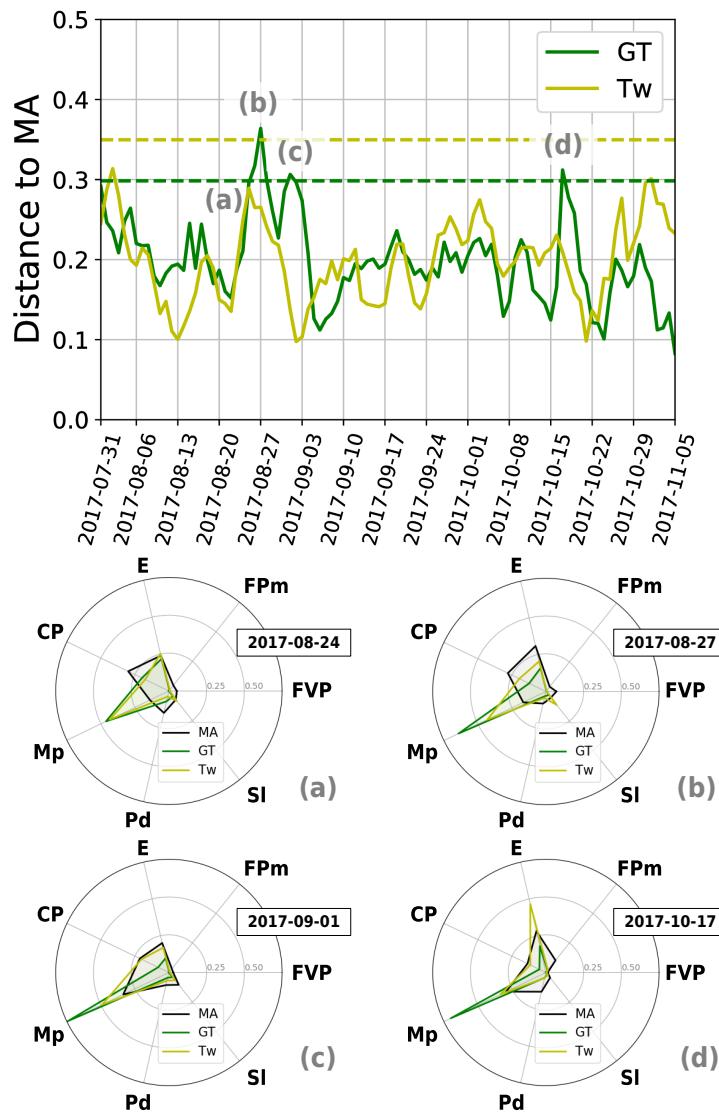


Fig. 9.4: Distancia de Jensen-Shannon entre los medios y las agendas públicas en función del tiempo Las distancias mayores se deben a un mayor interés de la audiencia en el tema *Santiago Maldonado* lo que disminuye el interés en otros temas. Por otro lado, la Agenda de Medios aún mantiene cierto grado de diversidad. **E**: Elecciones; **FPm**: Ministro de Planificación; **FVP**: Ex vicepresidente; **SI**: Líder social Milagro Sala; **Pd**: Muerte del Fiscal Nisman; **Mp**: Santiago Maldonado; **CP**: Presidente Macri.

en la Fig 9.1. Es importante recordar que los temas son los mismos que se introdujeron en las nubes de palabras de la Fig 9.1, pero al momento de calcular los pesos de los temas, los artículos fueron discriminados por medio.

Fig 9.5 muestra cualitativamente las diferencias entre las agendas periodísticas. Por ejemplo, podemos ver como el diario Página 12 le da más importancia a los temas *Santiago Maldonado* y *Milagro Sala*, mientras reduce al mínimo la cobertura del tema *Ministro de Planificación* como lo hicieron los demás. La diferencia en la cobertura se conoce como *sesgo de cobertura* [58].

El sesgo observado es consistente con lo esperado de los medios analizados y refleja el clima político altamente polarizado observado en la sociedad argentina. Durante la gestión de Cristina Fernández de Kirchner (2007-2015), el gobierno se enfrentó a varios medios de comunicación. Esto llevó a medios como Clarín, La Nación y el portal de noticias Infobae a ser muy críticos con el gobierno de Fernández, enfatizando las denuncias de corrupción relacionadas con el mismo, como se puede ver en la importancia que se le da a los temas *Exministro de Planificación* y *Exvicepresidente*. Por otro lado, Página 12 tiene un sesgo ideológico opuesto [235, 236], apoyando al gobierno anterior de Cristina Kirchner y por ende siendo muy crítico con el gobierno de Mauricio Macri, haciendo especial énfasis en temas relacionados con derechos humanos, como se puede observar nuevamente en la cobertura dada a los temas *Líder social Milagro Sala* y *Santiago Maldonado*.

Dada la importancia del tema *Santiago Maldonado*, dedicaremos el siguiente apartado al análisis del mismo.

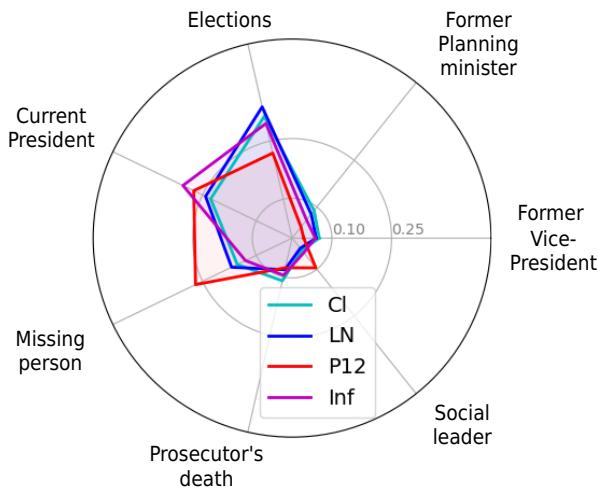


Fig. 9.5: Gráfico de radar de las distribuciones La figura muestra de forma cualitativa el sesgo en las diferentes agendas periodísticas. Por ejemplo, el mayor interés de Página 12 (P12) por el tema *Santiago Maldonado* y su menor cobertura en el *Ex Ministro de Planificación* respecto a los demás diarios.

9.2.6. Caso de estudio Santiago Maldonado: sesgo de cobertura

En la subsección anterior hemos visto que un solo tema fue responsable de la reducción de la diversidad de agenda, de la disimilitud temporal entre la agenda pública y mediática

y del sesgo de cobertura entre diferentes medios: El tema *Santiago Maldonado*. Además, este es el más adecuado para discutir porque:

- Causó un gran impacto tanto en los medios como en la audiencia;
- su cobertura se despliega plenamente a lo largo del lapso de tiempo analizado.

En la Fig 9.6 panel (a), mostramos el peso relativo del tema *Santiago Maldonado* tanto para la Agenda Pública como para los Medios. Luego de la cobertura inicial, las agendas parecen diferenciarse alrededor del 15 de agosto, cuando el tema comienza a cobrar más importancia en las Agendas Públicas que en las de Medios. Alrededor del 24 de agosto, el tema aumenta abruptamente en el interés de la audiencia mientras que la reacción en los medios es más lenta. Esta fecha está muy cerca del 26 de agosto, cuando se realizó una campaña en las redes sociales. Luego de ese evento, los medios aumentan su cobertura sobre el tema. Este comportamiento nos llevó a la siguiente pregunta: ¿Es este un caso donde las redes sociales imponen la agenda mediática?

Cuando miramos más de cerca la cobertura mediática podemos ver que no todos los medios tuvieron el mismo comportamiento. La Fig 9.6 panel (b) muestra el perfil temporal del tema *Santiago Maldonado* para cada medio. Notoriamente, Página 12 fue el medio que rápidamente prestó atención a este tema desde sus inicios, señalado con un círculo discontinuo en el panel (b). El rango de fechas donde se produce esta mayor cobertura anticipan el interés del público mostrado en el panel (a).

Para integrar la información del panel (a) y (b) de la Fig 9.6, calculamos la cobertura acumulada del tema *Santiago Maldonado*, que se muestra en la Fig Panel 9.6 (c). Definimos la cobertura acumulada del tema *Santiago Maldonado* como la integración numérica entre la fecha inicial y la fecha actual (normalizada por la cobertura total) del perfil temporal del tema en la agenda de Medios, Google y Twitter, además de la agenda de la Página 12. Esta cantidad nos muestra cómo se ha ido acumulando la atención mediática y pública desde los primeros acontecimientos.

En el panel (c) de la Fig 9.6 se sugiere una compleja dinámica de establecimiento de agenda dentro del tema *Santiago Maldonado*: El periódico menos vendido (Página 12) desencadenó un debate público a través de una mecanismo de retroalimentación reforzado por búsquedas reiterativas en Google y discusiones en redes sociales. Debido a este creciente interés público, el resto de medios se vieron obligados a prestar atención a este tema y finalmente, el tema pasa a ocupar también un lugar destacado en la Agenda de Medios.

Además del análisis realizado anteriormente, hay dos hechos importantes que deben mencionarse sobre el tema *Santiago Maldonado*: Primero, la desaparición de una persona es un tema muy sensible en la sociedad argentina (debido a los recuerdos de los casos de personas desaparecidas durante la última dictadura cívico-militar), lo que puede explicar por qué este tema en particular despertó el interés de la audiencia; y segundo, como se mencionó anteriormente, Página 12 estuvo particularmente interesada en cubrir este tema desde el principio (debido a su oposición a la administración presidencial), mientras que el resto de los medios no lo hicieron hasta que el tema fue destacado por la audiencia (Agenda Pública).

9.3. Conclusiones

Los medios de comunicación juegan un papel fundamental en la formación de opinión y por lo tanto, es de vital importancia contar con una descripción cuantitativa precisa de

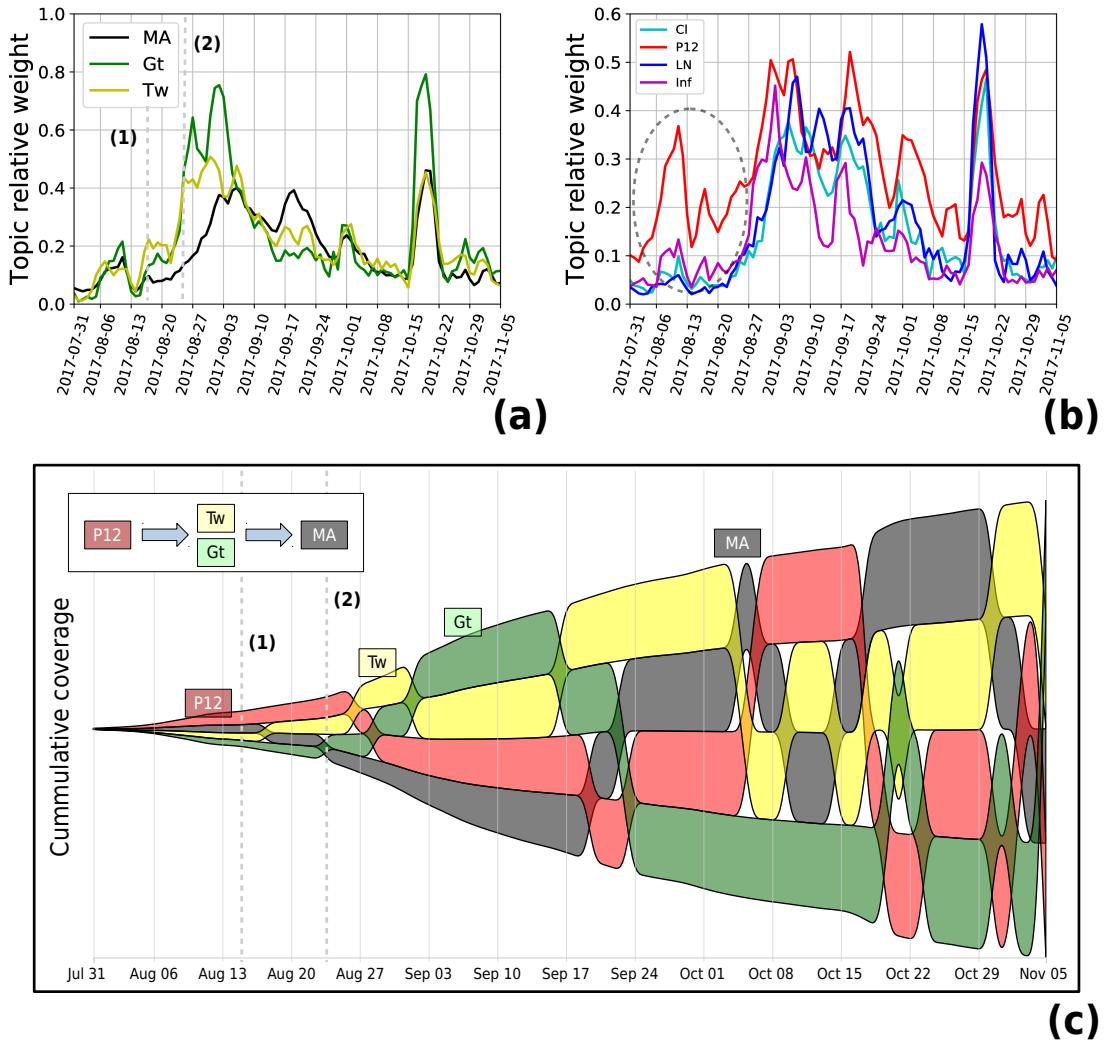


Fig. 9.6: **Interacción de agenda en el tema Santiago Maldonado.** Los perfiles temporales del panel (a) muestra que las Agendas Pública y de Medios parecen diferenciarse alrededor del 15th de agosto (línea gris vertical (1)) y la audiencia aumenta abruptamente su interés en el tema alrededor del 24th de agosto (línea gris (2)). Sin embargo, cuando vemos al medio de comunicación por separado en el panel (b), observamos que Página 12 muestra una mayor cobertura al inicio de este tema (señalado con un círculo discontinuo), antes de aumentar la atención del público. Al calcular la cobertura acumulada, representada como un gráfico de relieve en el panel (c), sugerimos que el tema fue establecido primero por la Página 12 y luego el interés de la audiencia parece causar la cobertura del resto de los medios.

los Medios y la Agenda Pública y su relación en el marco de la teoría del establecimiento de agenda. En este trabajo desarrollamos un marco, mediante la implementación de un algoritmo de detección de temas, para describir las Agendas Públicas y de Medios como una distribución de pesos (que mide la importancia de cada tema) en un espacio temático común, que emerge intrínsecamente del corpus de textos y que evoluciona en el tiempo. En este marco, es posible desarrollar métricas adecuadas para medir propiedades como la diversidad o la distancia entre agendas.

En concreto, hemos encontrado que un tema muy atractivo centra la atracción del público más de lo observado en los medios, lo que mantiene un cierto grado de diversidad y una gama más amplia de intereses temáticos. Curiosamente, también mostramos que las distancias entre agendas pueden emplearse para detectar rápidamente aquellos períodos en los que el público puede tener un comportamiento independiente respecto de los medios.

La misma metodología nos permite comparar las agendas de los medios analizados y cuantificar el sesgo de cobertura en los diferentes temas. En particular, detectamos diferencias importantes en el tema *Santiago Maldonado*, donde el respectivo sesgo de los medios en contra y a favor de la administración del gobierno parece ser relevante para explicarlas.

Además, el análisis del tema *Santiago Maldonado* presentado en Fig 9.6 muestra evidencia concreta de un caso de dinámica compleja de agenda-setting, en la que no todos los medios juegan el mismo papel y se produce una dinámica de retroalimentación entre los medios y las redes sociales que debe ser analizada detenidamente en el marco de la teoría del agenda-setting.

Parte IV

REDES SOCIALES: ANÁLISIS DE LA COMUNICACIÓN Y LA DINÁMICA DE OPINIONES DE LOS USUARIOS

Esta sección contiene resultados que fueron presentados en los siguientes trabajos:

- Albanese, F., Feuerstein, E., Kessler, G., & Ortiz de Zárate, J. M. (2023). Aprendizaje automático para el análisis crossplataforma de la comunicación política: Gobierno y oposición argentinos en Facebook, Instagram y Twitter. *Cuadernos. info*, (55), 256-280.
- Albanese, F., Feuerstein, E., Lombardi, L., & Balenzuela, P. (2023). Characterizing community-changing users using text mining and graph machine learning on Twitter. *AMW 2023: 15 Alberto Mendelzon International workshop*, Chile.
- Albanese, F., Feuerstein, E., & Balenzuela, P. (2024) Polarization dynamics: A study of individuals shifting between political communities on social media. *Journal of Physics: Complexity*, URL <http://iopscience.iop.org/article/10.1088/2632-072X/ad679d>
- Albanese, F., Wu, S. & Cheng, L. (2023) A Causal Study of Community Engagement toward Partisan Content and Misinformation on Reddit. [Paper en preparación, versión preliminar presentada en KHIPU Latin America Meeting in Artificial Intelligence 2022]

10. INTRODUCCIÓN

La proliferación de las redes sociales ha marcado un hito significativo en la comunicación online al democratizar el acceso y la participación en la generación de contenido. Décadas atrás, eran solo los grandes medios (diarios, radios y canales televisión) quienes podían difundir de forma masiva información y opiniones. Pero actualmente un usuario puede publicar en redes sociales y llegar a una audiencia masiva e internacional. Este cambio de paradigma ha otorgado a los usuarios un papel activo en la construcción y difusión de narrativas.

En la sección anterior se caracterizó la influencia y rol que tienen los grandes medios de comunicación online y su influencia sobre los usuarios. En esta sección nos centraremos en el comportamiento que tienen los usuarios particulares. Primero nos enfocaremos en el rol de las cuentas oficiales de los políticos, donde podemos ver diferencias claras en su comportamiento y en la respuesta de sus seguidores dependiendo de la red social en cuestión, ya sea Twitter, Instagram o Facebook. Después estudiamos a los usuarios en general y las comunidades políticas que forman al interactuar más con ciertos usuarios que con otros. En particular, resultan de interés los usuarios que cambian de comunidad política, quienes en un primer momento interactúan con usuarios con un determinado sesgo político y luego con usuarios con un sesgo opuesto. Caracterizamos sus tópicos de interés, su sentimiento con respecto a políticos y sus características topológicas en la red de interacciones. De esta forma pudimos detectar una relación estadísticamente significativa entre cambios de las características topológicas y el sentimiento de los textos. Es decir, los usuarios que cambian de comunidad no solo cambian con quiénes interactúan, sino también cambian su sesgo político.

Finalmente caracterizamos el efecto causal que hay cuando un usuario comparte noticias políticas con su comunidad en Reddit, ya sean noticias de medios de comunicación confiables o no. Los resultados del análisis muestran un efecto causal significativo del sesgo político de un artículo y el engagement que recibe. En particular, se observa que los artículos con sesgo opuesto a la comunidad tienden a tener un número neto menor de likes en comparación con los artículos con mismo signo político, incentivando el sesgo de confirmación y desincentivando debates con ideas opuestas. También mostramos que existe una relación causal negativa entre la veracidad de los medios y el engagement de un artículo, incentivando la difusión de las fake news.

11. COMUNICACIÓN DE LOS POLÍTICOS

11.1. Introducción

En secciones anteriores caracterizamos el rol online de los medios de comunicación masivos y su interacción con los individuos. En esta sección nos centraremos en analizar y estudiar el rol de los usuarios individuales. En particular, las cuentas oficiales de políticos y cómo son sus estrategias de comunicación en las diferentes plataformas. Resulta de interés entender qué diferencias y similitudes muestran sus publicaciones dependiendo de la plataforma en cuestión.

El análisis cross-platform toma como objeto de estudio y unidad de análisis a cada usuario y las distintas redes sociales con las que interactúa en forma frecuente [189]. Existe un incipiente campo de estudios de comunicación política cross-plataforma de campañas electorales de los países centrales como Estados Unidos [38], Alemania [206], Noruega [161] y Suecia [125]. Estudios más conceptuales se preguntan cómo la lógica política influye en la arquitectura de los diferentes medios interactivos [44, 167] y otros comparan medios tradicionales con Twitter [111].

En términos metodológicos, estos trabajos han recurrido al análisis de metadatos (likes, retuits, etc.), en menor medida a técnicas cualitativas [204] y al análisis de discursos con aproximaciones novedosas [206]. Ellos revelan que los políticos y los partidos utilizan estrategias diferentes en cada plataforma y que todavía hay una predilección por Facebook a pesar del lugar central que tiene Twitter en los debates y polémicas. Asimismo, observan un creciente uso de Instagram (y, hasta hace pocos años, de Snapchat) aunque todavía hay pocos trabajos sobre Tik Tok. Subrayan la necesidad de innovaciones metodológicas para incrementar el alcance y la rigurosidad de los estudios [90, 132, 161, 209]. En efecto, por su parte, Stier et al. [206] señalan tres limitaciones de la mayoría de los trabajos sobre comunicación política en medios digitales. La primera es que en general se basan en períodos de campañas, pero pocos dan cuenta de la comunicación en tiempos ordinarios; la segunda es que suelen centrarse en una sola plataforma; y la tercera que analizan más metadatos que en los textos.

Es por ello que en esta sección buscamos superar estas limitaciones y analizar las publicaciones en Facebook (FB), Twitter (TW) e Instagram (IG) de 50 políticos argentinos del oficialismo y la oposición durante el año 2020. Las 50 figuras políticas argentinas fueron seleccionadas en términos de cargos, responsabilidades o relevancia dentro del Gobierno a nivel nacional (Frente de Todos), que llamaremos Oficialismo y de la Oposición (Juntos por el Cambio) que impera en algunas provincias y las principales ciudades. Se debe destacar que el año 2020 fue un año sin elecciones nacionales e incluye la pandemia de COVID-19. Nuestro marco teórico articula teorías de agenda setting desarrolladas en secciones anteriores y sociología de problemas públicos.

11.2. Marco teórico sociológico e hipótesis

Los análisis propuestos en este capítulo se fundamentan en distintas teorías de la comunicación política como los estudios de agenda setting [22] y la teoría del framing o encuadre [192, 193]. Teniendo en cuenta las diferencias entre las plataformas de redes

sociales, presuponemos que los mensajes pueden diferenciarse en tres aspectos:

1. En relación a la agenda, esto es hablar sobre tópicos distintos en cada plataforma.
2. En virtud del encuadre, es decir, hablar de los mismos temas, pero enmarcados de manera distinta según la red.
3. En su dimensión interpelativa o vocativa, esto es, respecto al receptor al que irían dirigidos. Las opciones 1 y 2 serían mutuamente excluyentes, en cambio la dimensión 3 puede combinarse con la 1 o con la 2 (p.ej. puede mantenerse el tema y el encuadre, pero variar en una plataforma y en otro a quién estaría dirigido).

En cuanto a las diferencias entre Oficialismo y Oposición, Kelly et al. propone la teoría de la propiedad, donde afirma que los políticos priorizan hablar principalmente de temas que los benefician [113]. Por ejemplo, en Estados Unidos a los Demócratas les convenía hablar de integración racial y de bienestar, y a los Republicanos de crimen y de seguridad nacional. Los análisis y resultados del capítulo 8 muestran justamente como los medios de comunicación siguen dicho comportamiento. Por otro lado, Ansolabehere y Iyengar argumentan que la propiedad no es una estrategia convincente para las audiencias y que puede ser necesario “montar la ola” (“to ride the wave” [21], centrándose en los temas del momento, bajo el riesgo de ser considerado cínico o no sintonizar con las preocupaciones del público [21]).

Respecto a las diferencias entre las distintas redes sociales, Bossetta define como “estructura de la red” a las normas técnicas que regulan la relación entre usuarios en cada plataforma [38]. Por lo tanto, presuponemos que Twitter impulsa una conversación interpelativa de tipo “de uno al otro” (to-each-other) puesto que favorece la polémica entre usuarios con ideas diferentes ya que por defecto no se seleccionan a los seguidores; mientras en las otras dos, los seguidores suelen ser personas más afines y no conforman espacios habituales de controversias; son más propicias a una comunicación de tipo “cada uno sin escuchar al otro” (past-each-other): el emisor elige sobre qué temas publicar y puede orientar la agenda con menor injerencia de contrincantes. En este contexto y basándonos en Kaplan, Park & Ridout [109] suponemos que Oficialismo y Oposición tienen más probabilidad de hablar de los mismos temas (baja propiedad) en Twitter y de temas distintos (alta propiedad) en Facebook e Instagram. En otras palabras, conjeturamos que Oficialismo y Oposición eligen (o no les queda más opción) una red para debatir y la(s) otra(s) para promoverse en los temas que se consideran más fuertes. Pero también supusimos que en tiempos ordinarios como el que estudiamos (no de campaña electoral), el espacio político no es solo de confrontación con el contrincante y de celebración de las propias acciones, sino que habría mensajes comunes tanto para el oficialismo como de la oposición en los que sea menos plausible la controversia.

En virtud de lo anterior, proponemos dos hipótesis al respecto:

- H1a Tópicos: Cada espacio elige Facebook e Instagram para hablar sobre los temas en los que tiene propiedad, mientras que Twitter se convierte en la plataforma en las que se debaten los temas sin propiedad exclusiva de uno u otro grupo.
- H1b Temas en común: Los temas en común entre Oficialismo y Oposición no solo incluyen confrontaciones sino también coincidencias o temas de baja conflictividad potencial.

Si bien Oficialismo y Oposición pueden hablar de temas distintos, otra opción es que hablan de los mismos temas con un encuadre distinto. Es decir, una valoración diferente, y a menudo opuesta, sobre dichos temas. En este sentido, la Teoría de la Valoración dentro de los estudios de encuadre se centra en analizar los recursos lingüísticos por medio de los cuales las personas llegan a expresarse [144]. Esta teoría está atenta a la valoración y la emoción de los discursos que denotan una diferente posición. Por lo tanto, conjeturamos que una diferencia entre Oficialismo y Oposición será la valoración sobre los principales temas de agenda. Así, un mismo tópico tendrá una connotación positiva para unos y negativa para los otros. En particular, presuponemos que la negatividad estará sobre todo en Twitter, puesto que es la red de la polémica. Es por ello que nuestras hipótesis son las siguientes:

- H2a Sentimientos: Oficialismo y Oposición suelen enunciar mensajes con sentimiento (positividad/negatividad) distinto dependiendo de la red por la que se expresan.
- H2b Negatividad en Twitter: Twitter es la plataforma donde hay mayor proporción de mensajes que expresan sentimientos negativos debido a la mayor frecuencia de interacciones confrontativas.

Nuestra última hipótesis se vincula al hecho que en Facebook e Instagram los contenidos se muestran principalmente a las cuentas que el usuario sigue mientras que en Twitter es en base a tópicos de interés. Esto promueve un mayor debate entre los usuarios, no solo en círculos caracterizados por la homofilia [149] sino también por gente con otros puntos de vista. A partir de esto, nuestra hipótesis es:

- H3 Interpelación: Los políticos tienden a interpelar entre sí más en Twitter que en Instagram y Facebook.

11.3. Metodologías y experimentos

En esta sección detallaremos las técnicas y métodos que aplicamos para armar los datasets pertinentes y testear nuestras hipótesis.

11.3.1. Construcción del Dataset

Alberto Fernández fue el presidente de la Argentina desde 2019 al 2023. Él resultó elegido en los comicios junto a Cristina Fernández de Kirchner como Vicepresidente encabezando el Frente de Todos, una alianza entre distintas corrientes del peronismo que venció al ex-Presidente Mauricio Macri, quien buscaba su reelección con la coalición Juntos por el Cambio. Esta alianza se conformaba por Propuesta Republicana (PRO), Unión Cívica Radical (UCR), Coalición Cívica ARI y Peronismo Republicano; lo que en este trabajo llamamos Oposición mientras que a los primeros Oficialismo. Para construir nuestro corpus seleccionamos 50 figuras políticas 25 del oficialismo y 25 de la oposición, de características lo más homogéneas posibles en ambos grupos en cuanto a cargos, responsabilidades o notoriedad, asegurándonos de que todos tuvieran cuentas oficiales en Facebook, Twitter e Instagram (ver Tabla 1 y 2 del Anexo del paper [10] para un mayor detalle). Del oficialismo elegimos 12 personalidades que tuvieron cargos en el Poder ejecutivo (principales Ministros y primera línea del Poder Ejecutivo Nacional), y 13 Senadores y Diputados de

distintas provincias y con alta exposición pública. De la oposición seleccionamos 11 políticos que tuvieron cargos ejecutivos de los cuales 7 eran actuales al momento del estudio (Intendentes de las principales urbes y Gobernadores) y 4 anteriores (Ex Presidente, ex Gobernadora de la Provincia de Buenos Aires, Presidente del Pro y ex Ministra de Seguridad, ex Gobernador de la Provincia de Mendoza y Presidente de la UCR) y 14 Diputados y Senadores relevantes.

Luego, con las APIs de Twitter y de CrowdTangle [32], descargamos todos los posteos que publicaron durante 2020 en las 3 plataformas, totalizando 150 cuentas (3 por cada figura política) y 84.435 posteos, de los cuales 56.622 son de Twitter, 16.133 de Facebook y 11.680 de Instagram. A pesar de que las imágenes son una componente importante del modo de comunicación en Instagram [30, 72], para este trabajo nos hemos limitado a analizar el texto de los posteos. Un primer hallazgo es que los políticos realizan más del doble de publicaciones en Twitter que en Facebook e Instagram juntos.

11.3.2. H1: Tópicos y Temas en común

Método

Para probar la H1 precisábamos identificar temas propios y temas comunes de Oficialismo y Oposición en cada plataforma. Algoritmos tradicionales para la detección y modelado de tópicos, como LDA [35], descriptos en secciones anteriores, requieren que se les provea a priori la cantidad de tópicos en los que se quiere dividir el corpus, y por ello la coherencia de la división resultante depende de que dicho parámetro coincida con el real, lo que exige realizar pruebas con distintos parámetros hasta encontrar el valor correcto. Como este dataset era voluminoso y diverso, se hubiese necesitado realizar numerosos intentos hasta llegar a dicho valor [188]. Por ese motivo recurrimos a una técnica más reciente, Top2Vec [20], reduciendo el tiempo de cómputo. Y, en efecto, según Top2Vec fueron 1028 los tópicos discutidos. Esta técnica, además, no precisa eliminar stopwords (artículos, preposiciones, etc.) ni normalizar el texto para su uso y permite identificar de forma determinística de qué tema habló un posteo dado. Luego de identificar los temas, indagamos cuáles pertenecen a cada sector político. La categorización entre temas propios y comunes la definimos así:

- Tema propio: Tópico en el cual el 95 % o más de los posteos provienen del mismo sector político.
- Tema común: Tópico en el cual cada grupo produjo entre el 45 % y 55 % de los posteos.

Resultados H1a (Tópicos)

Para verificar H1a medimos las proporciones de propios y comunes en cada red. Fue necesario normalizar la cantidad de posteos por político y por red social, ya que como dijimos los políticos realizan por lo general más posteos diarios en Twitter que en Instagram y Facebook. Los siguientes dos gráficos muestran dichas proporciones:

Twitter es la red más utilizada para los temas compartidos y Facebook e Instagram para los propios, y podemos confirmar que hay una agenda diferente en las plataformas, por lo cual se confirma la H1a.

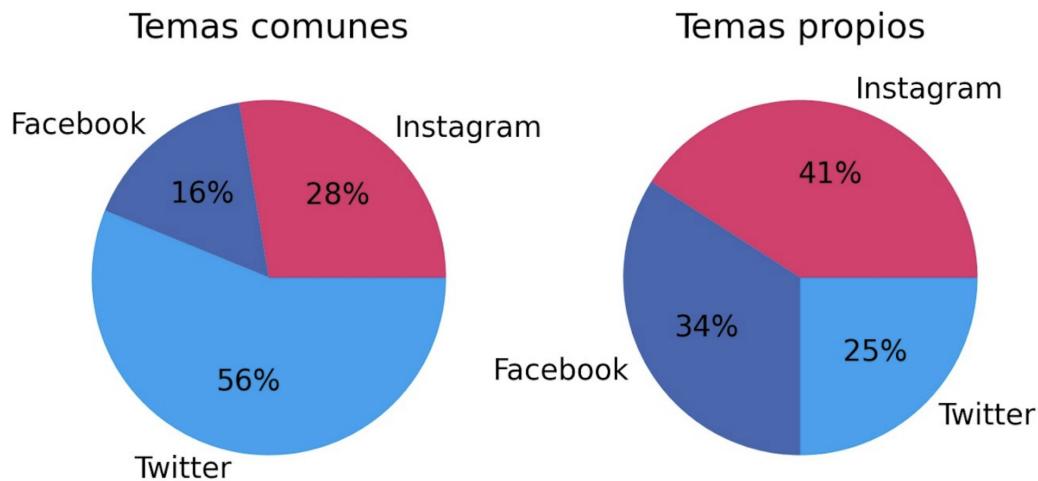


Fig. 11.1: Proporción de posteos por red social de tópicos que pertenecen a un único sector político (temas propios) y compartidos por ambos (temas comunes).

Resultados H1b (Temas en común)

¿Qué características tienen los temas propios?

Para elucidar este interrogante analizamos las principales palabras y posteos de cada tópico por grupo:

- Oficialismo

1. Defensa del río#DefendamosElRío: CampañaEslogan del oficialismo nacional en la Ciudad Autónoma de Buenos Aires (CABA), gobernada por la oposición y referida a la discusión sobre qué hacer con terrenos fiscales adyacentes al Río de la Plata.
2. Derechos de las mujeres: Campañas del oficialismo.
3. #Levantarnos: Campaña del oficialismo tendiente a salir de la crisis económica y la pandemia.
4. Revolución de las viejas#LaRevoluciónDeLasViejas: Campaña por los derechos de las mujeres mayores
5. Campañas anti discriminación: Campañas para luchar contra la xenofobia, machismo, homofobia, clasismo, entre otros.

- Oposición

1. Cifras COVID San Isidro: Informes sobre casos COVID-19 en San Isidro, distrito gobernado por la oposición.
2. Informes Covid CABA: Informes sobre casos COVID-19 en Capital Federal, distrito gobernado por la oposición.
3. Voluntariado para cuidar a los mayores: Programa del gobierno de la CABA para asistir a la gente mayor durante la cuarentena.
4. Encuentros virtuales con vecinos: Actividades virtuales con vecinos de los distritos gobernados por la oposición

5. Críticas al kirchnerismo: Críticas al sector del oficialismo representado por Cristina Fernández de Kirchner

Los temas son coincidentes con las agendas de cada sector en los medios de comunicación. En efecto, el oficialismo hace eje sobre campañas de gestión y cuestiones sociales o de derechos mientras que la oposición se refiere a la gestión en sus distritos y critica al oficialismo (focalizándose en el sector representado por la Vicepresidenta Cristina Fernández de Kirchner).

¿Cuáles son los temas en común?

Para verificar la H1b, nos propusimos ver qué tópicos tuvieron una participación similar por partido, lo que definimos como tópicos comunes. A su vez, dividimos a los mismos en controversiales y no controversiales dependiendo de si los políticos tienen o no opiniones opuestas que discuten o debaten. Detectamos los siguientes temas, entre otros:

- Temas comunes no controversiales

1. Saludo y reconocimiento a trabajadores: Saludos a los bomberos, trabajadores de la salud y otros trabajadores en su día.
2. Condolencias por fallecimientos: en ocasión de la muerte de figuras del campo político (p.ej un Juez federal, ex senador nacional o ex Gobernador de una provincia).
3. Aniversario Guerra de Malvinas: en ocasión del aniversario de la Guerra de Malvinas contra el Reino Unido en 1982.
4. Cuidado jubilados: Mensajes de la importancia de cuidar a los jubilados en pandemia.
5. Aniversarios patrios: Mensajes por los aniversarios patrios como el Día de la Independencia.

- Temas comunes controversiales

1. Vacuna Sputnik: Discusiones sobre dicha vacuna de origen ruso: el gobierno nacional posteaba sobre su compra y la oposición denunciaba que era de baja efectividad.
2. Menciones a Ginés: Menciones Ginés González García, ex-ministro nacional de salud. Mientras que el oficialismo anunciaría actividades con él, la oposición lo criticaba por su gestión.

Confirmamos así que la hipótesis H1b se cumple, ya que la mayoría de los tópicos en común son no controversiales, con la excepción de la vacuna Sputnik y las menciones al Ministro de Salud.

11.3.3. H2: Sentimiento y Negatividad en Twitter

Método H2a (Sentimiento) ¿Hay diferencias en el encuadre?

Para responder a este interrogante y testear la hipótesis H2a buscamos caracterizar la positividad y negatividad de los mensajes [3] mediante la aplicación sobre los posts de

una red neuronal convolucional [127] para el análisis de sentimiento. Luego, con un test estadístico intentamos detectar si había diferencias significativas entre la proporción de mensajes positivos y negativos de cada una de las redes.

Resultado

No hallamos disparidades importantes. Por lo tanto, no se verificó con este método nuestra hipótesis de que los políticos se expresan con positividad o negatividad distinta dependiendo de la red social.

Método H2b (Negatividad en Twitter)

Realizamos nuevos testeos con los novedosos desarrollos en torno a la toxicidad: un mensaje se considera tóxico si por su tenor rudo e irrespetuoso puede generar que el interlocutor abandone una conversación (Fortuna, 2020). Para medir la toxicidad usamos la API de Perspective [225] que utiliza redes neuronales profundas para el procesamiento del lenguaje natural pre entrenadas para dicha tarea. Este algoritmo le asigna a cada texto un valor entre 0 y 1, que representa la probabilidad de que el mensaje sea tóxico. Siguiendo la metodología utilizada por otros autores [98], definimos un valor de corte por encima del cual consideramos a un mensaje como tóxico.

Resultado

Al cuantificar la cantidad de mensajes tóxicos en cada red social, se observó que la proporción, si bien era chica en las tres redes sociales, en Twitter era considerablemente mayor siendo de un 7,6 % contra un 1,2 % en Instagram y un 0,4 % en Facebook. Ahora bien, ¿por qué los políticos tienen incentivos para publicar mensajes con mayor toxicidad en una red respecto de las otras dos? Descubrimos que en Twitter la mayor toxicidad se corresponde con una mucho mayor cantidad de likes, pero no sucede lo mismo en las otras redes. En la siguiente figura se presenta el resultado en cada red social:

Al realizar un test de Spearman [118], en Twitter observamos una correlación estadísticamente significativa (p valor menor a 0.05) y positiva entre la cantidad de likes y la toxicidad de un post con un coeficiente de correlación de 0.15, Por otro lado, en Facebook dicho coeficiente es menor a la mitad, de 0.08. En Instagram ni siquiera hay una correlación significativa entre toxicidad y likes. Concluimos que, tal como afirma la H2b, los políticos tendrían incentivos para ser tóxicos en Twitter, pero no tanto en Facebook y en Instagram, ya que en la primera los mensajes tóxicos reciben una cantidad mucho mayor de likes.

11.3.4. H3: Interpelación

Método 1

Para testear nuestra tercera hipótesis hubo que realizar distintos pasos. Primero, comparamos la forma discursiva de los mensajes políticos para observar diferencias entre las plataformas. Apelamos a técnicas de procesamiento del lenguaje natural que miden la similaridad de los textos según la cantidad y significatividad de las palabras compartidas: dos textos se consideran más parecidos si comparten muchas palabras que no son muy comunes en el resto del corpus. Para ello vectorizamos el texto mediante la técnica de

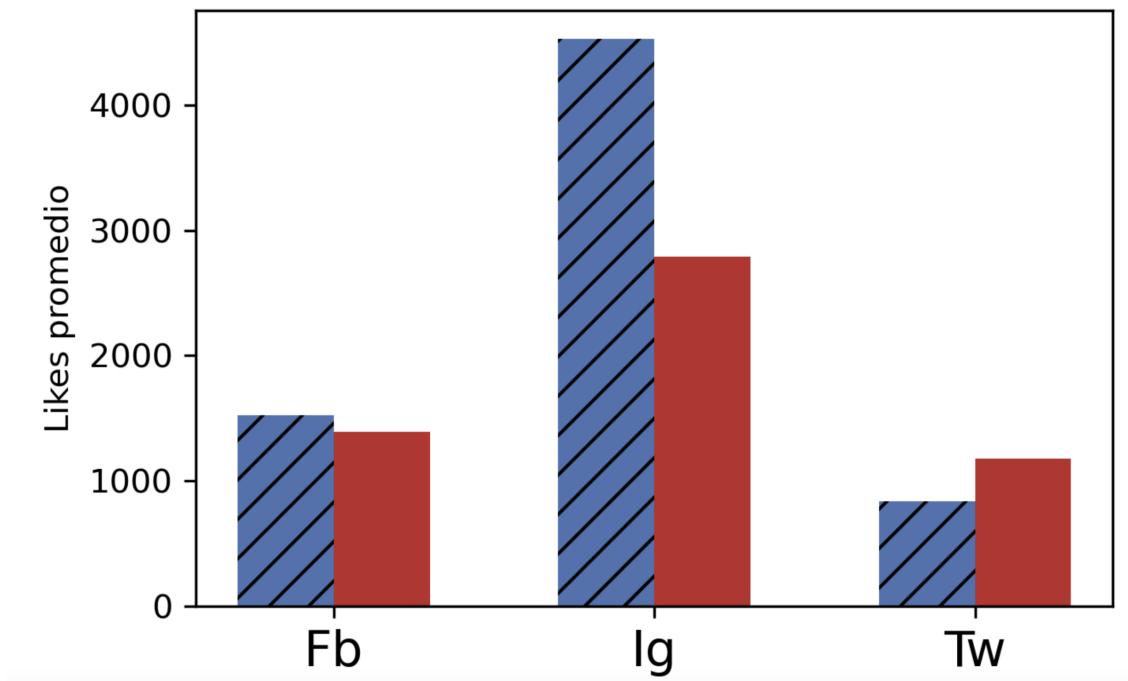


Fig. 11.2: Cantidad de likes promedio en mensajes mensajes tóxicos (rojo) y no tóxicos. (azul) en cada plataforma.

conteo de frecuencia de palabras: Term frequency – Inverse document frequency (Tf-idf) y medimos la similaridades de los mismos a través de la similaridad coseno [194]. Así encontramos que las cuentas del oficialismo y oposición en Twitter tenían, globalmente, un gran parecido entre sí, mucho más que respecto a las otras dos redes. Para captar la particularidad de Twitter analizamos los patrones de distribución de las palabras a fin de descubrir cuáles se encontraban juntas con mayor frecuencia. A través de la técnica de descomposición en valores singulares (SVD) hallamos los principales grupos de palabras (dimensiones) y luego, entrenamos un árbol de decisión [210] para predecir a qué plataforma pertenecía cada usuario: un árbol de decisión entrenado para clasificar los textos según su pertenencia a Twitter, Facebook o Instagram, puede detectar si hay un grupo de palabras que se usa principalmente en una red social y no en las demás. El árbol fue entrenado sobre el 75 % de las cuentas seleccionadas al azar, dejando el 25 % restante para calcular su performance (test set). Cada cuenta fue representada por la concatenación de todos sus posteos.

Resultados 1

Respecto a la eficacia del modelo, de las 26 instancias de la clase 1 (FB/IG) se predijeron correctamente 24 y de las 12 de la clase 2 (Twitter), 10. De esta forma, la exactitud (accuracy) del modelo predictivo es de un 89,4 % y el área bajo la curva ROC [155] es 0.919. Luego nos enfocamos en la dimensión 50, aquella que más significativamente separaba y clasificaba los textos según la red social. Esta dimensión la llamamos Interpelativa debido a que las principales palabras más utilizadas son: usted, renuncia, saludos, buen día y espalda. Si bien se ven algunas palabras relacionadas a cuestiones o consignas co-

yunturales como “renuncia” o “espalda” nos resultó llamativo que la principal palabra sea “usted” ya que ésta puede denotar un diálogo de interpelación directa con otro usuario y presumiblemente otro político. A continuación se enumeran los posteos más importantes dentro de la dimensión Interpelativa:

1. “Usted, sí. <https://t.co/zjRiDBvEvE>” (FerIglesias)
2. “@SolciPlata Usted, en cambio, sí.” (FerIglesias)
3. “@clarigv1 A usted” (WolffWaldo)
4. “@Damian_Deglauve @WorldGrace saludos!” (gabicerru)
5. “Soy yo la que lo quiere a usted, @caramellocumpa!!! <https://t.co/1qZjlEvADi>” (fvallejoss)
6. “@shetpwk94 Que tengas un buen día Delfi!!! No salgas de tu casa !!! Cuídate mucho” (alferdez)
7. “Si usted insistía en adjudicar esta compra con sobreprecios, hubiéramos realizado la denuncia al PAMI. Pero entendemos que ha procedido como corresponde.” (gracielaocana)

En la lista se observa que todos los posteos son de diálogos directos, es decir quien los genera está interpelando a otro usuario, no necesariamente en un tono negativo (lo que concuerda con lo visto al intentar distinguir las redes mediante el análisis de sentimiento).

La siguiente figura es un histograma de los usuarios diferenciados por red social según cuánto usaban las palabras de dicha dimensión interpelativa (usando el score obtenido con SVD).

Se observa que la mayoría de las cuentas de Twitter están sobre la derecha del eje X, lo que significa que tienen una componente significativa en esta dimensión (usaron frecuentemente las palabras asociadas a la misma). Por otro lado, casi todas las barras azules y naranjas (cuentas de Facebook e Instagram respectivamente) están sobre la izquierda del eje X. Es interesante notar que, en algunos casos, cuentas en distintas redes pertenecientes a una misma persona se encuentran en lugares opuestos respecto al valor 0 del eje X. Esto nos indica que esa persona tuvo una forma de comunicar distinta en Instagram y Twitter respecto a las palabras asociadas a dicha dimensión. Un caso destacado es el del Diputado de la oposición Fernando Iglesias, quien era un vocero importante contra el gobierno, cuya cuenta de Twitter tuvo un score de 0.25 sobre el eje X mientras que en su cuenta de Instagram el score fue de casi -0.3, ocupando extremos opuestos del gráfico.

Método 2

A continuación, medimos la frecuencia relativa de cada término en cada red social. Para esto graficamos las palabras según la importancia que tienen en Twitter y en Facebook+Instagram (Ver figura 11.4), posicionando en el eje X el score en Twitter y en el eje Y el score en las otras dos. Agregamos una línea roja indicando equivalencia, es decir que aquellos términos posicionados en o cerca de ella tienen frecuencias parecidas en ambos casos.

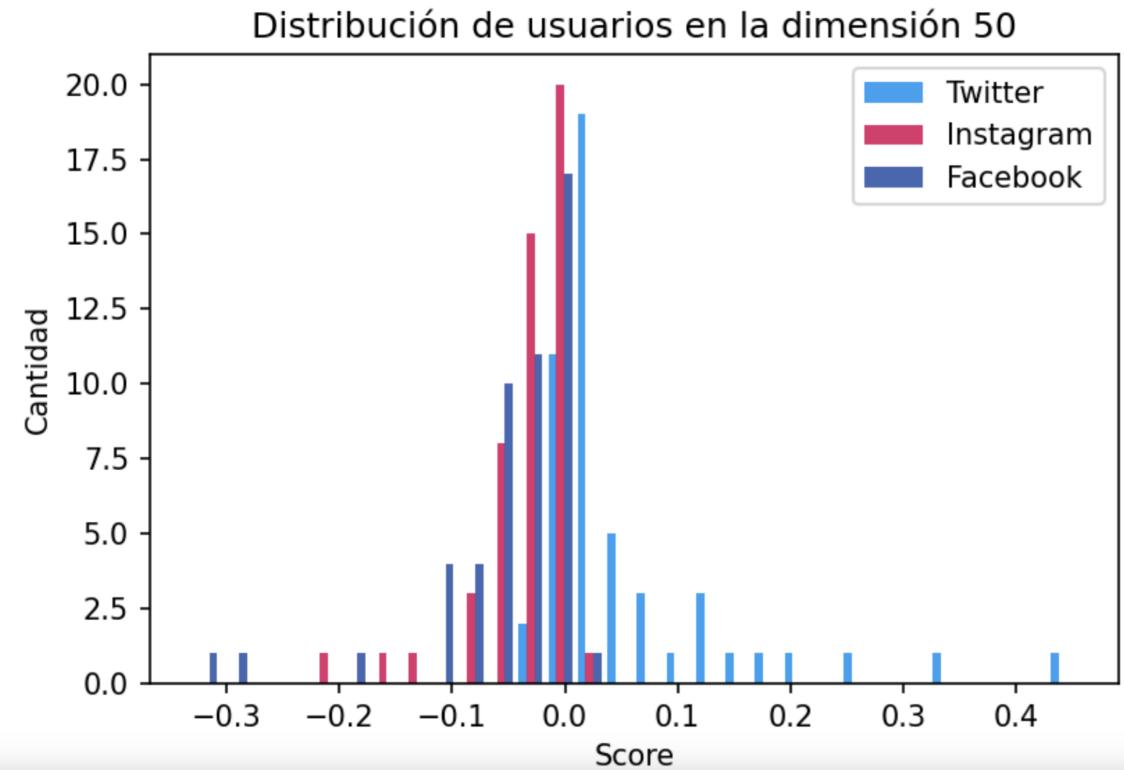


Fig. 11.3: Scores obtenidos sobre la dimensión Interpelativa por las cuentas. En el eje X se encuentran los scores y en el Y la cantidad de cuentas con dicho score.

Resultados 2

Se verifica que “usted” y “vos” tienen una importancia muchísimo mayor en Twitter, con un score mayor a 0.4 en esa red y menor a 0.2 en Facebook e Instagram. El término más importante en Twitter, es “alferdez”, que es nombre de la cuenta del ex-Presidente Alberto Fernández. Vemos también que los términos referidos a gestión como “obras” y “vecinos” (muy usado por los gobiernos municipales) son importantes en FB e IG y no en Twitter, lo que sugeriría que esas plataformas son más elegidas para comunicar la gestión pública.

Método 3

Para seguir intentando validar la H3 (Twitter es el terreno para las interacciones) testeamos el uso de los distintos pronombres en las plataformas; en concreto la frecuencia normalizada de ciertas palabras interactivas (la cantidad de veces que aparece dicha palabra dividida por la cantidad total de palabras usadas en esa red).

Resultados 3

En el gráfico observamos que los pronombres en segunda persona como “vos” y “usted” son más usados en Twitter que en Facebook o Instagram y lo mismo sucede con expresiones dirigidas a otro interlocutor, como “buen día”, “hola” o “saludos”. En contraposición, el pronombre en primera persona “yo” aparece en mayor medida en Facebook que en

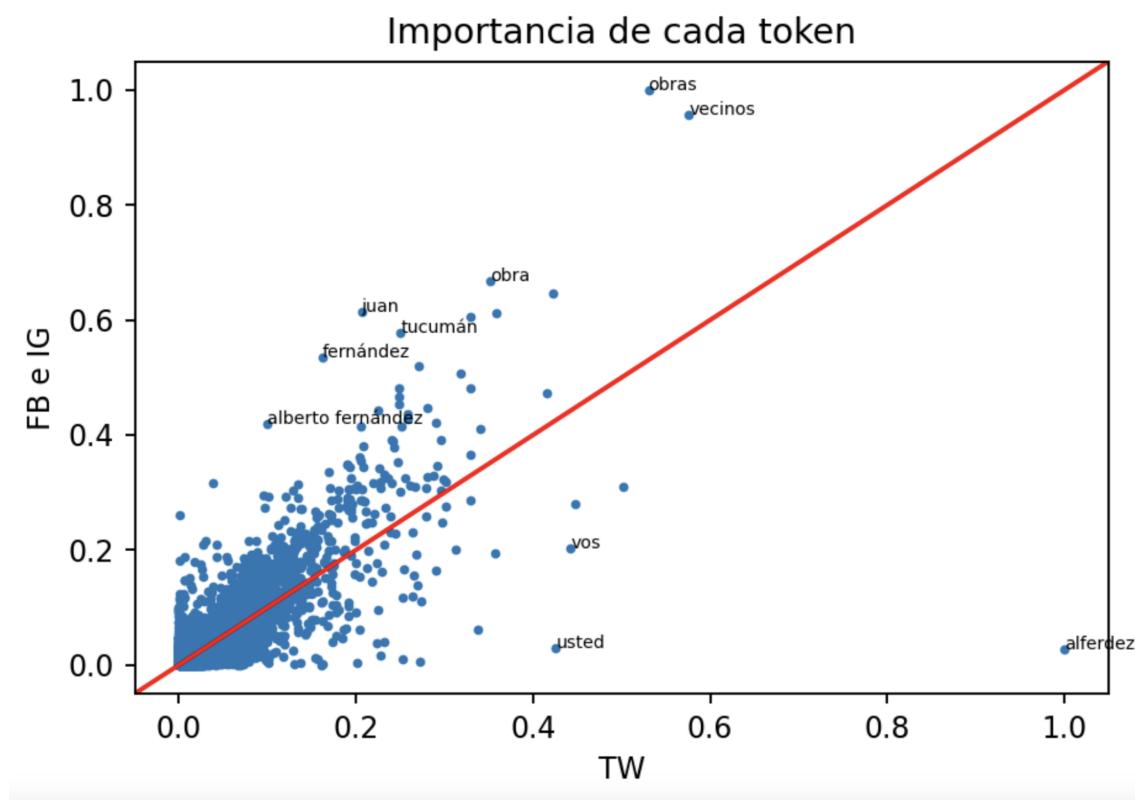


Fig. 11.4: Frecuencia de términos por red social.

Instagram o Twitter. También los pronombres de tercera persona son más usados en Facebook (él) e Instagram (ella) que en Twitter.

Estos resultados refuerzan la hipótesis 3 antes confirmada de que en Twitter hay un mayor diálogo entre usuarios, mientras que Facebook e Instagram son redes menos interpelativas y se usan más los pronombres personales de primera o tercera persona.

11.4. Conclusiones

En esta sección nos propusimos elucidar similitudes y diferencias en la comunicación de 50 usuarios, correspondientes a políticos argentinos del oficialismo y la oposición, analizando sus posteos durante 2020 en Facebook, Instagram y Twitter. Logramos caracterizar el comportamiento de dichos usuarios, viendo que hablan más de los temas sobre los que tienen propiedad en Facebook y en Instagram que en Twitter. La arquitectura de esta última posibilita un campo de controversias entre seguidores afines y no afines; a diferencia de las dos primeras, donde predomina la afinidad política. Más aún, el rasgo principal es la interpelación en general a otros actores políticos y el carácter controversial de esta plataforma se verifica también en que hay una mayor proporción de mensajes con toxicidad, que son los que más adhesión generan. En otras palabras, en Twitter la toxicidad es recompensada. Es por ello que centrar la mirada y la investigación solo en Twitter refuerza una imagen conflictiva de la política y con poca referencia a las acciones políticas concretas, una razón de peso para seguir a los mismos emisores en las otras plataformas. Observamos también temas en común de baja conflictividad tanto rituales y celebraciones

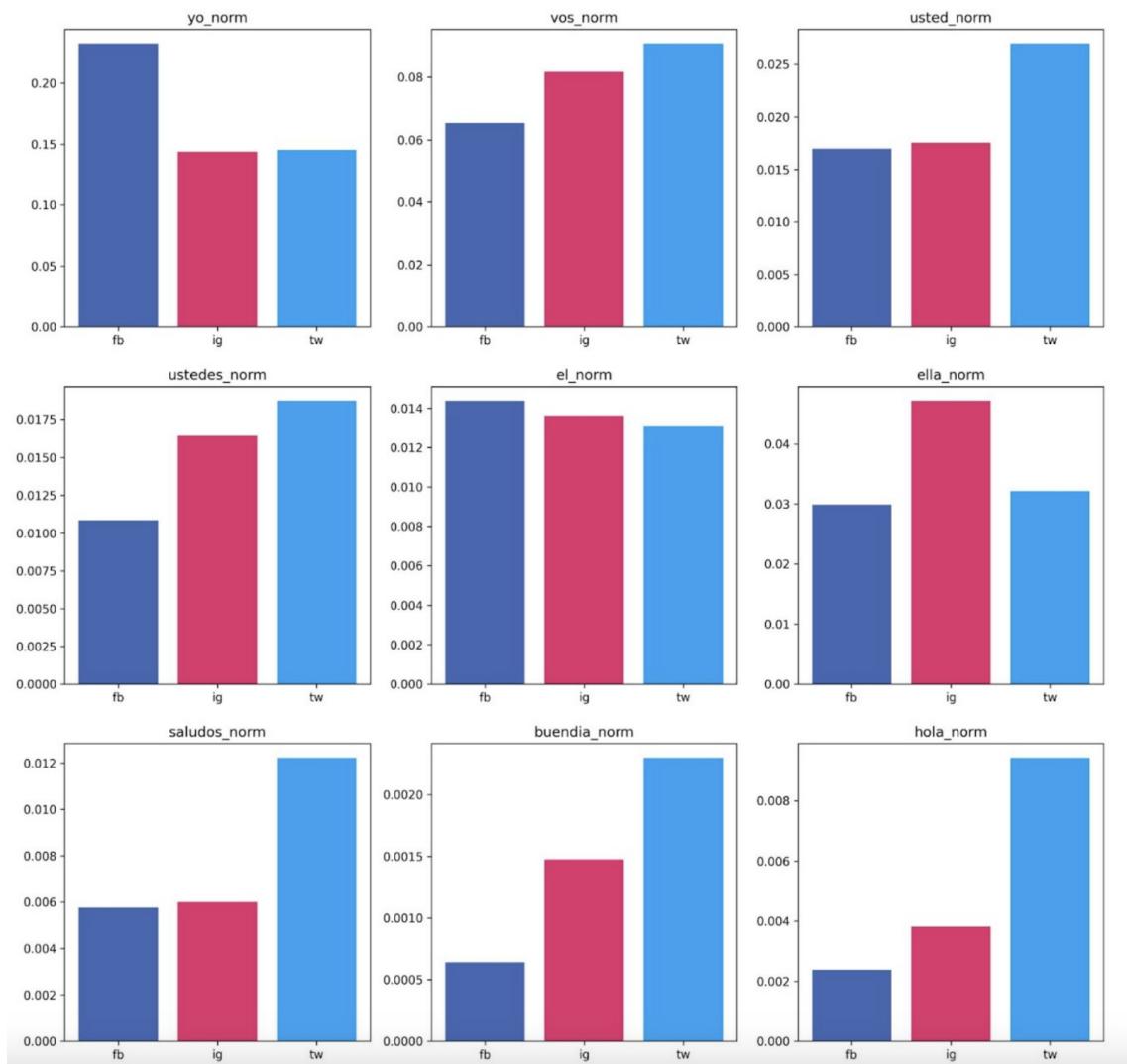


Fig. 11.5: Frecuencia normalizada de cada término en las tres redes sociales: Facebook (fb), Instagram (ig) y Twitter (tw).

comunes, pero sobre todo, referencias a las políticas implementadas.

En cuanto a sus aportes para la comunicación política, este trabajo muestra un enfoque cross-plataforma para captar mejor la complejidad y matices de la comunicación política actual de usuarios políticos, tanto en la producción de mensajes, como en la interacción con otros usuarios.

12. USUARIOS QUE CAMBIAN DE COMUNIDAD: ANÁLISIS DE SUS CARACTERÍSTICAS TOPOLOGICAS Y SUS TÓPICOS DE INTERÉS

12.1. Introducción

En las redes sociales interactúan personas con diferentes opiniones políticas y diversos orígenes. Sin embargo, esta diversidad no se traduce en debates enriquecedores entre usuarios con diferentes perfiles porque tienden a agruparse según sus creencias, constituyendo comunidades homogéneas conocidas como cámaras de eco [105]. Aruguete et al. se centró en la interacción entre usuarios en contextos políticos y describió cómo los usuarios de Twitter enmarcan eventos políticos compartiendo contenido exclusivamente con usuarios de ideas afines formando dos comunidades bien definidas [23]. Una estructura partidista segregada con una conexión extremadamente limitada entre comunidades de usuarios con diferentes orientaciones políticas en las redes de retweets se puede encontrar en múltiples períodos, en diferentes contextos y países como, por ejemplo, las elecciones intermedias del Congreso de Estados Unidos de 2010 [55], las elecciones federales canadienses de 2011 [86] o tweets sobre la muerte del presidente venezolano Hugo Chávez [152]. También se pueden encontrar comunidades bien definidas en diferentes plataformas [53, 180].

Trabajos anteriores mostraron las consecuencias dramáticas y los efectos negativos de las comunidades cerradas y las cámaras de eco, que incluyen el aumento del discurso negativo, el discurso de odio y el extremismo político [135], el sesgo de confirmación (es decir, la tendencia de los usuarios a buscar y recibir información que fortalezca su narrativa preferida) [180] y la difusión de rumores infundados y noticias falsas [52, 63].

En este capítulo, proponemos un marco de aprendizaje automático para caracterizar a los usuarios que rompen esta lógica y cambian con quién interactúan: los usuarios que cambian la comunidad (es decir, los usuarios de Twitter que primero pertenecían a una comunidad bien definida y luego comience a interactuar principalmente con diferentes usuarios que cambian a otra comunidad). Analizar a los usuarios que cambian de comunidad política puede ofrecer información valiosa sobre la compleja dinámica de la política electoral, ya que pueden ser el factor decisivo en el resultado de una elección.

Se construyeron y utilizaron tres conjuntos de datos para demostrar que la metodología se puede generalizar fácilmente a diferentes escenarios. Es decir, examinamos tres conjuntos de datos de la red Twitter construidos con tweets de: elecciones parlamentarias de Argentina de 2017, elecciones presidenciales de Argentina de 2019 y tweets de 2020 sobre Donald Trump. Para cada conjunto de datos, analizamos dos períodos de tiempo diferentes e identificamos las comunidades más grandes correspondientes a las principales fuerzas políticas. Utilizando información topológica del grafo y detectando temas de discusión de la primera red, construimos y entrenamos un modelo que clasifica si un individuo cambiará su comunidad. Además, analizando el modelo se puede encontrar las características relevantes de los usuarios que cambian de comunidad.

12.2. Dataset

Tal como se comentó en el capítulo 6, Twitter tiene varias API disponibles para desarrolladores, entre las que se encuentra la Streaming API que permite al desarrollador descargar en tiempo real una muestra de los tweets que se suben a la red social filtrándolos por idioma, términos, hashtags, etc. [153]. Los datos se componen del id del tweet, el texto, la fecha y hora del tweet, el id de usuario y el nombre de usuario, entre otras características. En caso de ser un retweet, también contiene la información de la cuenta de usuario del tweet original.

Para esta investigación, recopilamos tres conjuntos de datos en dos períodos de tiempo diferentes: elecciones parlamentarias de Argentina de 2017 (2017ARG), elecciones presidenciales de Argentina de 2019 (2019ARG) y tweets de Donald Trump en Estados Unidos de 2020 (2020US). Para el conjunto de datos argentino, la API de Streaming se utilizó durante la semana anterior a las elecciones primarias (del 7 de agosto al 13 de agosto de 2017 y del 5 de agosto al 12 de agosto del 2019 respectivamente) y la semana anterior a las elecciones generales (del 15 octubre al 20 octubre 2017 y del 20 octubre al 27 octubre del 2019 respectivamente). Las palabras clave se eligieron según los cuatro principales partidos políticos presentes en las elecciones. Para el conjunto de datos de Estados Unidos de 2020, utilizamos “realDonaldTrump” (la cuenta oficial del presidente Donald Trump) como palabra clave y las semanas desde el 9 de mayo hasta el 16 de mayo y desde el 10 de junio al 16 de junio de 2020 como primer y segundo período respectivamente. Los detalles se pueden encontrar en el paper [11]. Hemos analizado más de 9 millones de tweets y más de 1,5 millones de personas en total.

12.3. Métodos

En esta sección presentaremos la metodología empleada para caracterizar a los usuarios. Describimos cómo calculamos cada característica e implementamos un modelo supervisado que clasifica a los usuarios que cambiaron su comunidad con el tiempo. Estos modelos nos permiten resaltar qué rasgos son características relevantes de los usuarios.

12.3.1. La red de retweets

Representamos la interacción entre individuos en términos de un grafo $G = (N, E)$, donde los usuarios son nodos (N) y los retweets entre ellos son las aristas (E). Teniendo en cuenta que un usuario puede ser retuiteado varias veces por otro usuario, usamos un grafo dirigido y ponderado. Sin embargo, cuando un usuario n_1 retuitea un tweet escrito por otro usuario n_2 , ¿la dirección de la arista debe ser de n_1 a n_2 o de n_2 a n_1 ? Esta definición tiene implicaciones importantes. En el primer escenario, las aristas representan indicadores de los “influencers” y generadores de contenido importantes. En el segundo escenario, las aristas representan el flujo de información a través de la red, desde la fuente hasta el usuario que difunde el mensaje. De hecho, no existe un consenso claro en la literatura científica sobre qué dirección se debe dar a las aristas: mientras algunos autores [18, 117, 227] utilizan la primera, otros [55, 152, 185] prefieren la segunda. A priori, no podemos decir qué dirección es mejor para nuestro propósito, por lo que decidimos calcular las características topológicas en ambos escenarios. Llamamos a las direcciones de las aristas RC (de Retweeter a Content Creator) y CR (de Content Creator a Retweeter).

En la Fig. 12.1, podemos visualizar la red de retweets para cada período de tiempo y conjunto de datos. En el caso del conjunto de datos de Estados Unidos, la mayoría de los usuarios se concentran en dos grupos, lo que refleja la polarización política en ese país. Por otro lado, en el conjunto de datos argentino podemos identificar dos grandes grupos y también algunos más pequeños. Las visualizaciones de los grafos se producen con el diseño Force Atlas 2 utilizando el software Gephi [104].

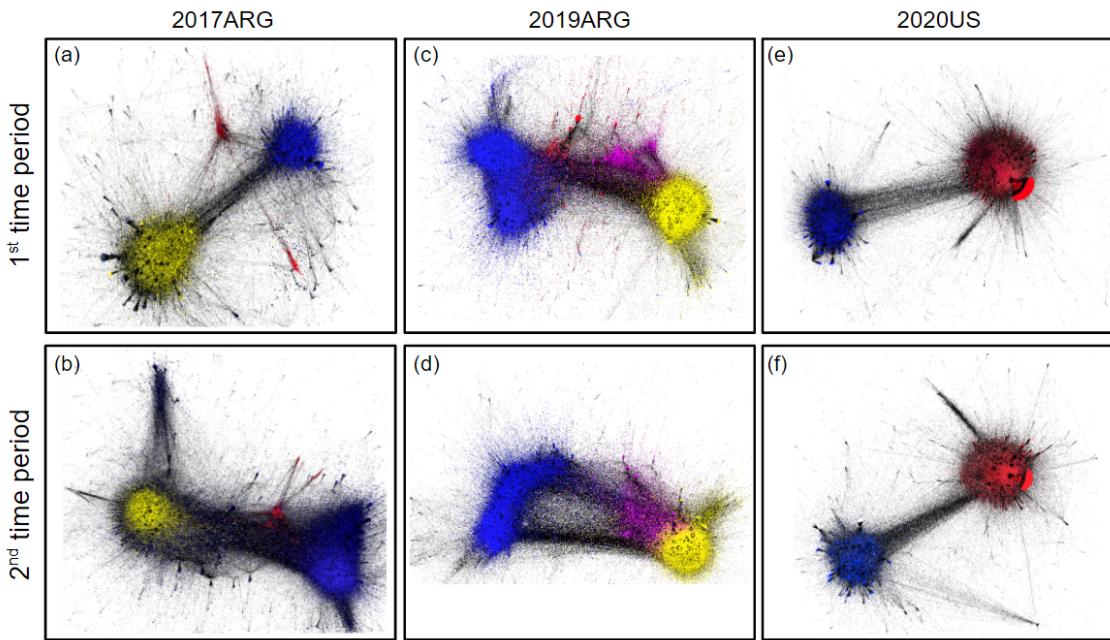


Fig. 12.1: Red de retweets: (a) elecciones primarias de Argentina del 2017 ; (b) elecciones generales de Argentina del 2017; (c) elecciones primarias de Argentina del 2019; (d) elecciones generales de Argentina del 2019; (e) primer periodo del 2020 en Estados Unidos y (f) segundo periodo. Cada nodo es un usuario (coloreado dependiendo de su comunidad) y cada arista es un retweet entre dos usuarios.

12.3.2. Detección de comunidades

En un grafo dado, una comunidad es un conjunto de nodos fuertemente conectados entre sí y con poca o ninguna conexión con nodos de otras comunidades [228]. Detectamos las comunidades en la red de retweets para cada conjunto de datos utilizando el método de Louvain [36]. Dada su estocasticidad, seguimos la solución propuesta por Lancichinetti et al. [124] que ejecuta el método varias veces (100 en nuestro caso). Luego, en este trabajo sólo se consideraron los nodos que siempre fueron asignados consistentemente a la misma comunidad en todas las iteraciones, para minimizar la posibilidad de un etiquetado incorrecto. También consideramos únicamente a los usuarios que recibieron o realizaron más de 5 retweets en cada período de tiempo.

A pesar de que el algoritmo encontró varias comunidades, solo consideramos las 4 más grandes para los conjuntos de datos argentinos y las 2 más grandes para los conjuntos de datos estadounidenses, ya que contienen más del 90 % de los usuarios. Examinamos el texto de los tweets y los usuarios con mayor grado de cada una de las comunidades más grandes y encontramos que cada una tenía una clara orientación política correspondiente a los cuatro

partidos políticos más importantes en la elección (siendo “Cambiemos”, “Unidad Ciudadana”, “Partido Justicialista” y “1 País” para 2017ARG y “Frente de Todos”, “Juntos por el Cambio”, “Consenso Federal” y “Frente de Izquierda-Unidad” para 2019ARG). En cuanto al conjunto de datos de Estados Unidos de 2020, las dos comunidades más grandes correspondían a cuentas de republicanos y demócratas. Estados Unidos tiene un sistema político bipartidista como se puede ver en la Fig. 12.1, donde sólo dos grandes clusters concentran casi todos los usuarios e interacciones. En contraste, los conjuntos de datos argentinos tienen dos comunidades principales y también algunas comunidades menores. Esta topología de red con clusters altamente conectados y polarizados se había reportado en trabajos anteriores [23, 55, 205].

12.3.3. Características topológicas

Dado que los conjuntos de datos analizados comprenden dos momentos de la red de retweets separadas en el tiempo, necesitamos caracterizar completamente a los usuarios en las primeras redes para identificar adecuadamente a aquellos usuarios que cambian su comunidad. Con este objetivo, calculamos las siguientes métricas para cada usuario de la red:

1. Grado: El número total de conexiones de un nodo.
2. Grado de entrada: El número de conexiones que recibe un nodo.
3. Grado de salida: El número de conexiones que emite un nodo.
4. PageRank: La importancia de un nodo dentro de la red.
5. Betweenness centrality: La influencia de un nodo sobre el flujo de información en la red.
6. Coeficiente de clustering: La medida de la tendencia de los nodos a formar comunidades o clusters.
7. Afiliación: La comunidad detectada.

Como mencionamos anteriormente, es importante tener en cuenta que la dirección de las aristas de la red afecta drásticamente el valor de estas métricas. En consecuencia, los calculamos con ambas interpretaciones. Todas estas métricas se utilizaron como características en la tarea de clasificación de aprendizaje automático y en el análisis de importancia de las características.

12.3.4. Detección de tópicos

Las características descritas anteriormente se basan en la interacción del usuario y surgen de la topología de la red de retweets. También caracterizamos los temas de discusión durante el primer período de cada conjunto de datos analizando los textos de los tweets.

De manera similar a trabajos [9, 176] descriptos en las secciones 8 y 9, primero los tweets se describieron como vectores a través de la representación tf-idf [184] y usamos 3-gramas y un diccionario de stop-words que no sólo contenía artículos, preposiciones, pronombres y algunos verbos sino también los nombres de los políticos, partidos y palabras

Tab. 12.1: Características de los datasets

	2017ARG	2019ARG	2020US
<i>#Individuos</i>	21134	26118	116854
<i>#Comunidades</i>	4	4	2
<i>#Featuresdeltexto</i>	9	7	6
<i>#Featuresdelgrafo</i>	10	10	10
<i>Trainingset</i>	14159	17499	78292
<i>Testset</i>	6975	8619	38562

como “elección”. Luego, realizamos NMF [226] para agrupar nuestro corpus de textos en temas. De esta forma, los usuarios fueron caracterizados por un vector donde cada celda corresponde a uno de los temas y su valor al porcentaje de tweets que el usuario publicó con ese tema.

12.3.5. Análisis de importancia

Dado que nuestro objetivo era caracterizar a los usuarios que cambian de comunidad y comienzan a interactuar con usuarios de otros clusters, implementamos un modelo de aprendizaje automático que clasifica a los usuarios y luego realizamos un análisis de importancia de las características (“Feature Importance analysis”). Las instancias del modelo fueron los usuarios de Twitter que estuvieron activos durante ambos períodos [43] y pertenecían a una de las comunidades más grandes en las redes de ambos períodos. En consecuencia, se redujo el número de usuarios considerados en esta etapa. Los individuos se caracterizaron mediante un vector de características con componentes correspondientes a las métricas topológicas listadas previamente y otros correspondientes al porcentaje de tweets en cada uno de los temas. La información utilizada para construir estos vectores de características se recopiló únicamente del primer período de tiempo, para evitar la fuga de datos al realizar la predicción. El objetivo era un vector binario que toma el valor 1 si el usuario cambia de comunidad entre el primer y segundo período de tiempo y 0 en caso contrario. El resumen del dataset se muestra en la Tabla 12.1.

Aplicamos la técnica de aumento de gradiente XGBoost [47], que utiliza un ensamble de modelos predictivos y ha demostrado ser eficiente en una amplia variedad de escenarios supervisados superando a otros modelos anteriores [160]. Usamos una división aleatoria de 67/33 entre el conjunto de datos de entrenamiento y evaluación. Para realizar un ajuste de hiperparámetros de los modelos XGBoost, utilizamos el método de búsqueda aleatoria [33] sobre el conjunto de datos de entrenamiento con validación cruzada triple.

Finalmente, realizamos una permutación aleatoria de los valores de las features entre los usuarios para comprender cuáles de ellas son los más importantes en el rendimiento de nuestro modelo (usando el llamado algoritmo de Permutación de Importancia de las Características [16]). De esta manera, podríamos identificar las características más importantes que separan a los usuarios que sí cambian su comunidad de aquellos que no cambian con quiénes interactúan.

Tab. 12.2: Performance de los modelos de XGBoost (ROC AUC)

	2017ARG	2019ARG	2020US
XGB (text)	0.7339	0.6683	0.6839
XGB (graph)	0.7664	0.7995	0.7425
XGB (text + graph)	0.7925	0.8019	0.7614

12.4. Resultados

Entrenamos tres modelos diferentes de XGBoost para cada conjunto de datos: el primero se entrenó solo con las características obtenidas mediante minería de texto (cuántos tweets de los tópicos seleccionados habla el usuario); un segundo modelo fue entrenado solo con características obtenidas a través de análisis del grafo (grado, PageRank, centralidad, coeficiente de clustering y afiliación del cluster); y el último modelo fue entrenado con todos los datos. De esta manera, podríamos comparar la importancia del procesamiento del lenguaje natural y el análisis del grafo para la tarea de clasificar usuarios que cambian de comunidad.

En la Tabla 12.2 podemos ver el área bajo la curva ROC [186] de los diferentes modelos para cada conjunto de datos. El mejor rendimiento lo obtiene en todos los casos el modelo de aprendizaje automático construido con todas las características de los usuarios, que es capaz de clasificar de manera más eficiente a los usuarios que cambiaron de comunidad. Este resultado es esperado, ya que un conjunto de modelos logra tener suficiente profundidad y robustez para comprender la información de la red, los temas de los tweets y las características del grafo de los usuarios. Además, el modelo entrenado con características del grafo superó al modelo con solo características de texto en los tres casos.

Realizamos una permutación aleatoria de los valores de las características entre los usuarios para el modelo entrenado con todas las características (texto + grafo). Encontramos que la característica más importante en todos los casos corresponde a la importancia de la conectividad del nodo: $PageRank_{CR}$, donde las aristas apuntan desde la fuente del tweet (el creador del contenido) hasta el usuario que retuiteó. Los coeficientes de importancia de las características del $PageRank_{CR}$ son 1635 (2017ARG), 2836 (2019ARG) y 843 (US2020). Todas las demás características muestran coeficientes aún más bajos. En particular, el otro $PageRank_{RC}$ (correspondiente a la otra dirección de las aristas) tenía coeficientes de características importantes de 717, 1202 y 527 para cada conjunto de datos respectivamente (una reducción mayor que 40 %). Esto significa que existe una clara dirección privilegiada de aristas para la tarea de detectar a los usuarios que cambiaron de comunidad.

Al comparar los promedios de $PageRank_{CR}$ (PR) de estos usuarios con los usuarios que no cambiaron de comunidad, observamos que estos últimos tuvieron valores más altos en todos los casos (Tabla 12.3). Aplicamos el test de Kolmogorov-Smirnov [94] a las distribuciones de PR de cada conjunto y encontramos que estas diferencias fueron estadísticamente significativas en todos los casos ($p < 0,001$). El $PageRank$ mide qué tan relevante o importante es un usuario en la red de retweets en función de los retweets de sus mensajes y la importancia de los usuarios que retweetearon. La dirección de $PageRank_{CR}$ representa el flujo de información en una red, comenzando desde el creador del tweet y luego difundiéndose a través de la red. El hecho de que los usuarios que cambiaron de comunidad tuvieran valores de $PageRank_{CR}$ estadísticamente más bajos significa que estos

Tab. 12.3: $PageRank_{CR}$ promedio de los usuarios que cambiaron de comunidad y de los usuarios que no cambiaron de comunidad.

	2017ARG	2019ARG	2020US
Cambia	1.32e-5	3.81e-6	3.16e-6
No cambia	1.55e-5	4.43e-6	3.47e-6

usuarios eran menos relevantes para la conversación y sus mensajes no se difundieron en su comunidad original. Una posible interpretación de estos resultados es que un usuario cambia de comunidad cuando no tiene fuertes afinidades con su comunidad y sus mensajes no tienen respuesta en su entorno.

El hecho de que $PageRank_{CR}$ sea la característica más importante también es consistente con que el modelo entrenado con características del grafo obtiene un mejor AUC que el modelo entrenado con los textos de los tweets en los tres conjuntos de datos.

12.5. Conclusiones

En este análisis presentamos un enfoque de marco de aprendizaje automático para identificar y caracterizar a los usuarios que cambiaron su comunidad por otra. El marco incluye técnicas de procesamiento del lenguaje natural para detectar sus temas de interés y aprendizaje automático en grafos para describir cómo un individuo interactúa con otros usuarios. El marco se aplicó a tres conjuntos de datos diferentes con resultados similares, lo que muestra que la metodología se puede generalizar fácilmente.

Descubrimos que los usuarios que cambiaron de comunidad tenían valores estadísticamente más bajos de $PageRank_{CR}$. Esta característica topológica también fue el indicador más importante de la tarea de clasificación en los tres conjuntos de datos según el análisis de importancia de la característica. En particular, nuestros resultados también muestran que existe una dirección claramente privilegiada en la red para esta tarea, con las aristas que van del creador de contenido al retweeteador. Una posible interpretación de estos dos últimos resultados es que los usuarios cambian con quién interactúan cuando no tienen fuertes afinidades con otros usuarios, sus mensajes no tienen respuesta y no son “escuchados” por su comunidad.

13. USUARIOS QUE CAMBIAN DE COMUNIDAD: ANÁLISIS DE SENTIMIENTO Y OPINIÓN POLÍTICA

13.1. Introducción

En el capítulo anterior se propuso un método de detección de usuarios que cambian de comunidad en redes sociales basado en el método de Louvain para la detección de comunidades, algoritmos de machine learning supervisado como XGBoost y métodos de detección de tópicos. Tanto en ese capítulo como en Zolezzi et al. [233] se estudia el flujo de usuarios entre comunidades políticas. Sin embargo, estos trabajos no incorporan análisis de sentimiento para determinar si estos cambios reflejan solo variaciones en los patrones de interacción o también significan cambios en las perspectivas políticas.

Por lo tanto, en este capítulo estudiamos un conjunto de datos de Twitter con 2 millones de tweets sobre el presidente de los Estados Unidos, Donald Trump, durante el año de las elecciones presidenciales de 2020 en Estados Unidos (el mismo que en el capítulo anterior) [12]. Luego, examinamos los textos publicados utilizando un método de análisis de sentimiento. Finalmente, encontramos correlaciones estadísticamente significativas entre el cambio de comunidad y los cambios en el sentimiento hacia Donald Trump. De esta forma, los resultados muestran que los usuarios no cambian solamente de comunidad, sino que por el contrario, los usuarios que cambian con quienes interactúan también manifiestan un cambio en su opinión política.

13.2. Métodos

13.2.1. Detección de comunidades

Siguiendo el mismo procedimiento que en los capítulos anteriores, se uso la red de retweets, donde donde los nodos (N) corresponden a los usuarios y las aristas (E) representan los retweets entre ellos. Teniendo en cuenta los resultados del capítulo anterior, las aristas apuntan desde el creador de contenido hacia el retweeter, representando el flujo de información a través de la red, que se origina en el creador de contenido y se propaga a los usuarios que amplifican el mensaje.

Los modelos de detección de comunidades que maximizan la modularidad (como el método de Louvain previamente utilizado en esta tesis) tienen un carácter descriptivo [159]. Es decir, al no considerar desviaciones sobre un modelo nulo, no permiten inferir un modelo subyacente sino meramente catalogar y asignar labels a los nodos en base a sus conexiones [173]. En contraposición, el Modelo de Bloques Estocásticos (SBM, por sus siglas en inglés) detecta las comunidades latentes incorporando priors jerárquicos y técnicas de inferencia bayesiana [95]. Por lo tanto, SBM tiene la capacidad de respaldar conclusiones inferenciales [173]. Como en este caso buscamos inferir la perspectiva política de los usuarios analizando su relación con la opinión política mostrada en los textos de los tweets, se utilizó SBM. En particular, usamos la implementación de la biblioteca de Python graph-tool para discernir comunidades dentro de redes dirigidas y ponderadas [172].

13.2.2. Análisis de Sentimiento

El análisis de sentimientos es una técnica de procesamiento de lenguaje natural utilizada para determinar el sentimiento o tono emocional expresado en un texto. Existen múltiples enfoques para clasificar el sentimiento de un texto, como los algoritmos basados en reglas [179], modelos de aprendizaje automático como la Máquina de Soporte Vectorial (SVM) o Naive Bayes (NB) [25] y redes de aprendizaje profundo [230]. Considerando las características de nuestro conjunto de datos (tweets con texto muy corto e informal), utilizamos VADER (Valence Aware Dictionary and sEntiment Reasoner), que es una herramienta de análisis de sentimientos basada en léxico y reglas, diseñada específicamente para contenido de redes sociales [99]. Para cada tweet, VADER proporciona una puntuación de sentimiento. Según la documentación de la herramienta, la puntuación toma valores de 1 a -1, donde una puntuación mayor a 0,05 se considera positiva, una puntuación entre 0,05 y -0,05 es neutral, y una puntuación menor a -0,05 es negativa.

La puntuación de sentimiento de un usuario se define como el promedio de sentimiento de sus tweets.

13.3. Resultados

13.3.1. Comunidades en la red de retweets

Usamos SBM para identificar las comunidades de usuarios de las redes de retweets. Teniendo en cuenta la naturaleza bipartidista de la política estadounidense, caracterizada por la presencia de dos partidos predominantes (Demócratas y Republicanos) [9, 116], configuramos los parámetros de SBM para identificar dos comunidades dentro de las redes.

En la Fig. 13.1, visualizamos la red de retweets y los resultados del algoritmo de detección de comunidades para cada período de tiempo. Tal como se describió en el capítulo anterior, los usuarios se concentran en dos comunidades, mostrando la polarización política en ese país. Las visualizaciones de los grafos fueron producidas con el diseño Force Atlas 2 usando el software Gephi [104].

Asignaríamos las etiquetas “comunidad republicana” a la comunidad que tiene a los usuarios “@realDonaldTrump” y “@Mike_Pence” (las cuentas oficiales de Twitter de los candidatos presidenciales y vicepresidenciales republicanos de 2020) y “comunidad demócrata” a la comunidad que tiene a los usuarios “@JoeBiden” y “@KamalaHarris” (las cuentas oficiales de los candidatos presidenciales y vicepresidenciales demócratas de 2020). Los resultados del análisis de sentimientos discutidos en la siguiente sección apoyan esta asignación de etiquetas.

En la tabla 13.1 mostramos las características principales del análisis de comunidades: el tamaño y porcentaje de ambos grupos en ambos momentos y cuán similares son usando el Índice de Jaccard. Primero, debemos notar que las comunidades republicanas son consistentemente más grandes que las demócratas. Este desequilibrio puede atribuirse al hecho de que los datos fueron recopilados usando la palabra clave “realDonaldTrump”, el presidente republicano de los Estados Unidos en ese momento. En segundo lugar, podemos ver que la mayoría de los usuarios permanecen en ambos grupos cuando se comparan en estos dos momentos: 95,5 % para los republicanos y 92,7 % para los demócratas. Aquellos que no lo hacen son el foco de nuestra investigación en este trabajo.

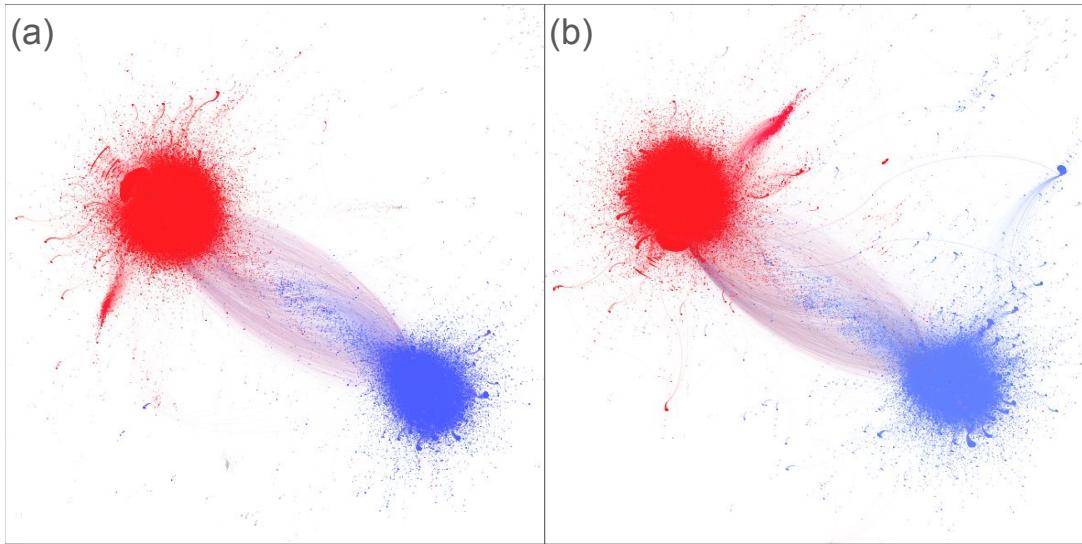


Fig. 13.1: Red de retweets del primer período de tiempo del 9 de mayo al 16 de mayo, t_1 , (a) y del segundo período de tiempo del 10 de junio al 16 de junio, t_2 (b), durante las elecciones presidenciales de Estados Unidos de 2020. Cada nodo es un usuario de Twitter y cada arista (dirigida y pesada) representa los retweets entre dos usuarios. Los nodos están coloreados según su comunidad: los usuarios de la “comunidad republicana” están en rojo y los de la “comunidad demócrata” en azul.

Tab. 13.1: El número total de usuarios de cada comunidad para ambos períodos de tiempo (t_1 y t_2).

	Comunidad republicana	Comunidad demócrata
#usuarios en t_1	12569 (64,1 %)	7044 (35,9 %)
#usuarios en t_2	11945 (63,8 %)	6769 (36,1 %)
Índice de Jaccard	0,955	0,927

13.3.2. Sentimiento de las comunidades

Hipotetizamos que cuando los usuarios de diferentes comunidades políticas mencionan a Donald Trump, sus sentimientos expresados mostrarán discrepancias notables. Esperaríamos que la comunidad republicana tenga un sentimiento más positivo al mencionar a Donald Trump y la comunidad demócrata tenga un sentimiento más negativo al mencionarlo.

En la Fig. 13.2, mostramos la distribución estadística de los valores de sentimiento de los usuarios al mencionar al presidente Donald Trump en cada comunidad y período de tiempo. Dada la naturaleza compleja de los datos, usamos el muestreo bootstrap, un método en el que repetidamente tomamos muestras de datos con reemplazo de la distribución de datos para estimar valores medios e intervalos de confianza, lo que proporciona una representación simple de los datos [91]. La Fig. 13.2 muestra tanto la distribución original de los valores de sentimiento para t_1 (a) y t_2 (b) como la muestra bootstrap de valores medios, también para t_1 (c) y t_2 (d). Los resultados muestran que los usuarios de la comunidad republicana tienen puntuaciones de sentimiento promedio positivas, mientras que los usuarios de la comunidad demócrata tienen puntuaciones de sentimiento promedio ne-

gativas, confirmando nuestra hipótesis. Este resultado es consistente para ambos períodos de tiempo. También encontramos que hay diferencias estadísticamente significativas entre las puntuaciones de sentimiento de las comunidades (p -value $< 0,01$) usando la prueba no paramétrica U de Mann-Whitney [142] y la prueba H de Kruskal-Wallis [121] en t_1 y en t_2 . Además, es importante notar que las diferencias entre las comunidades son estables a lo largo del tiempo.

Si bien nuestro análisis ha revelado puntuaciones de sentimiento distintas para cada comunidad al mencionar a Trump, no todos los miembros de la comunidad se alinean con estos sentimientos. En particular, 71,6 % de los usuarios en la comunidad demócrata expresan negatividad, y 67,7 % de los usuarios en la comunidad republicana tienen una puntuación de sentimiento positiva. Aunque estos porcentajes representan una alineación mayoritaria de sentimiento, se debe notar que lógicamente existen usuarios dentro de estas comunidades cuyos sentimientos se desvían de la tendencia predominante.

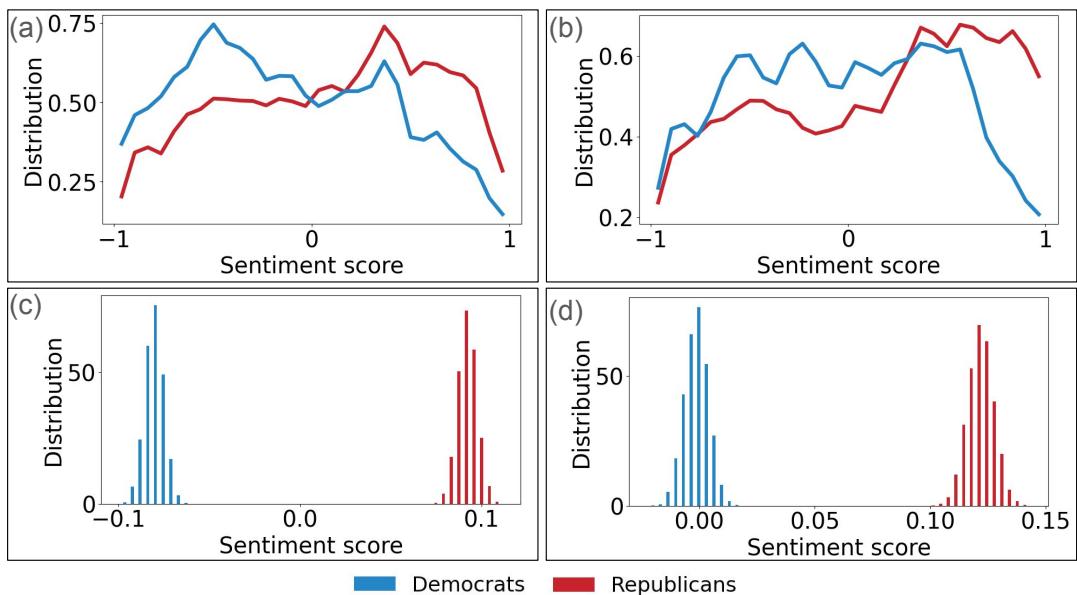


Fig. 13.2: Puntuación de sentimiento de tweets pertenecientes a la comunidad republicana y la comunidad demócrata para el primer período de tiempo del 9 de mayo al 16 de mayo, t_1 , (a) y el segundo período de tiempo del 10 de junio al 16 de junio, t_2 , durante las elecciones presidenciales de EE.UU. de 2020. (c) y (d) muestran la muestra Bootstrap de valores medios en los mismos momentos.

13.3.3. Usuarios que cambian de comunidad

En esta subsección, buscamos explorar las diferencias entre aquellos usuarios que cambian de comunidades cuando se comparan en dos instantes de tiempo diferentes, t_1 y t_2 . Comprender las diferencias entre ellos puede proporcionar información sobre cómo los individuos interactúan con otros usuarios de diversas comunidades ideológicas.

Formalmente, definimos a un usuario como un usuario que cambia de comunidad (u_c) cuando su comunidad en el tiempo t_1 y t_2 son diferentes. Por otro lado, definimos a un usuario como un usuario que no cambia de comunidad (u_{nc}) cuando su comunidad en el tiempo t_1 y t_2 es la misma.

Tab. 13.2: Delta de la puntuación media de sentimiento entre los usuarios de la comunidad republicana y demócrata en t_1 y t_2 ($sentiment_{t_2} - sentiment_{t_1}$). Nota: Las diferencias son estadísticamente significativas (prueba U de Mann-Whitney [142] y la prueba H de Kruskal-Wallis [121]).

	Demócrata en t_1	Republicano en t_1
Demócrata en t_2	$0,075 \pm 0,085$	$-0,066 \pm 0,074$
Republicano en t_2	$-0,036 \pm 0,088$	$0,035 \pm 0,082$

Puntuaciones de sentimiento

Primero analizamos las puntuaciones medias de sentimiento hacia Donald Trump para los usuarios de las comunidades Republicana y Demócrata que mantuvieron (u_{nc}) o cambiaron (u_c) sus afiliaciones comunitarias en t_1 .

Calculamos los valores medios y los intervalos de confianza utilizando el enfoque de submuestreo. Para los usuarios republicanos, la puntuación media de sentimiento en t_1 para aquellos que permanecieron dentro de su comunidad fue $s_{nc} = 0,081 \pm 0,433$, mientras que para aquellos que cambiaron de comunidades, fue $s_c = 0,080 \pm 0,401$. De manera similar, los usuarios demócratas que mantuvieron su afiliación comunitaria exhibieron una puntuación media de sentimiento de $s_{nc} = -0,102 \pm 0,408$, mientras que aquellos que experimentaron un cambio de comunidad mostraron una puntuación media de $s_c = -0,087 \pm 0,456$.

Al realizar pruebas estadísticas no paramétricas U de Mann-Whitney [142] y la prueba H de Kruskal-Wallis [121], determinamos que las diferencias observadas en las puntuaciones de sentimiento entre los usuarios que cambiaron y los que no lo hicieron no eran estadísticamente significativas, lo cual es consistente con el hecho de que los intervalos de confianza exceden los valores medios en un orden de magnitud. Por lo tanto, hay una capacidad predictiva limitada basada en las puntuaciones de sentimiento, en contraste con las características topológicas, descriptas en el capítulo anterior.

Por otro lado, también caracterizamos la variación de la puntuación media de sentimiento entre el primer (t_1) y el segundo (t_2) períodos de tiempo de los usuarios al mencionar al presidente Donald Trump en cada comunidad (la puntuación de sentimiento en t_2 menos la puntuación de sentimiento en t_1).

La tabla 13.2 muestra la variación en las puntuaciones medias de sentimiento entre t_1 y t_2 para los usuarios de la comunidad republicana y demócrata. Los resultados muestran que los usuarios republicanos que mantienen su comunidad (aquellos que estaban en la comunidad republicana en t_1 y permanecen allí en t_2) cambian su puntuación de sentimiento en promedio $0,035 \pm 0,082$. Pero los usuarios que estaban en la comunidad republicana en t_1 y cambian su comunidad a la demócrata en t_2 muestran una variación en su puntuación media de sentimiento de $-0,066 \pm 0,074$. Un efecto equivalente se midió para los usuarios demócratas. Por lo tanto, observamos que los usuarios que cambian sus comunidades también cambian sus puntuaciones de sentimiento de manera diferente a los usuarios que mantienen su comunidad.

Estas variaciones en la puntuación media de sentimiento de los usuarios que cambian su comunidad son estadísticamente significativamente diferentes ($p\text{-value} < 0,01$) a las de los usuarios que no cambian su comunidad, según la prueba no paramétrica U de Mann-Whitney [142] y la prueba H de Kruskal-Wallis [121]. Los resultados muestran que la clasificación de los usuarios que experimentan cambios de comunidad se alinea con cambios

significativos en el sentimiento al mencionar a Donald Trump entre el primer y segundo períodos de tiempo. A pesar de la capacidad predictiva limitada de las puntuaciones de sentimiento, los resultados demuestran una relación entre cambiar de comunidad y cambiar su sentimiento hacia Donald Trump.

13.4. Conclusiones

En conclusión, nuestra investigación ha proporcionado información sobre las dinámicas de los cambios de comunidad en redes sociales y sus implicaciones para el comportamiento de los usuarios en contextos políticos. A través de la introducción de un marco de trabajo de machine learning, hemos identificado y caracterizado a los usuarios que experimentan cambios de comunidad.

Analizar los cambios en el sentimiento expresado por los usuarios durante los cambios de comunidad puede proporcionar información valiosa sobre cómo estas transiciones podrían afectar las opiniones políticas de los individuos. Es importante destacar que hemos encontrado correlaciones estadísticamente significativas entre los cambios de comunidad y las alteraciones en el sentimiento hacia Donald Trump, arrojando luz sobre la interacción entre los cambios ideológicos y las interacciones en línea.

Este capítulo presenta un análisis del sentimiento en tweets y las interacciones de los usuarios. Los resultados contribuyen a avanzar en nuestra comprensión de las dinámicas de las redes sociales y proporcionan herramientas computacionales para identificar y estudiar efectivamente a los usuarios que experimentan cambios de comunidad.

Trabajo futuro podría replicar los análisis en diversos conjuntos de datos en varios contextos, países y redes sociales. Además, examinar un conjunto de datos extendido que abarque varios años, empleando diferentes ventanas de tiempo, podría ofrecer información valiosa.

14. EFECTO CAUSAL DE COMPARTIR NOTICIAS

14.1. Introducción

Como hemos visto en capítulos anteriores, la literatura sobre polarización afectiva sugiere que las personas tienden a interactuar más con aquellos cuyas creencias son similares a las suyas [103]. Esto da como resultado comunidades cerradas donde los miembros interactúan principalmente con personas internas y poco con personas externas, un fenómeno comúnmente conocido como efecto cámara de eco [168] y que se ha observado en múltiples contextos políticos en diferentes países [37, 74] y en plataformas de redes sociales [8, 23].

Muchos estudios han investigado el comportamiento y las interacciones entre usuarios de distintas comunidades políticas [19, 223]. En este contexto, es de particular interés cuando se comparten noticias, ya que estas pueden moldear los puntos de vista políticos de las personas e influir en elecciones o políticas públicas [26, 222]. Por ejemplo, Faris et al. [70] midió la frecuencia con la que los usuarios comparten artículos de medios de comunicación en Twitter y encontró que hay personas en ambos partidos comparten artículos con el sesgo político opuesto al propio. Además, la difusión de información errónea es una gran amenaza para la sociedad moderna. Sorprendentemente, los investigadores han encontrado una mayor interacción con fuentes poco creíbles. Por ejemplo, [67] estudió las reacciones de los usuarios a la información errónea en Facebook e informó que los artículos provenientes de medios que suelen publicar información errónea recibieron seis veces más “Me gusta” (Likes).

Sin embargo, la mayoría de las investigaciones existentes todavía utilizan métodos de análisis de correlación. Se ha utilizado poco el análisis causal [170] que busca descubrir la relación causa-efecto teniendo en cuenta variables externas que median, causan o intervienen en dicha relación. Consideramos que el problema de comparar comportamientos cross-partidistas y co-partidistas puede encajar naturalmente en el marco del análisis causal porque nos permite controlar fácilmente los factores de confusión ocultos. Por ejemplo, al estudiar el engagement de los usuarios en artículos partidistas, se deben controlar los temas de los artículos. Y sólo así podemos obtener una estimación imparcial del efecto causal del partidismo. Para llenar estos huecos en la literatura, planteamos una perspectiva causal para abordar dos preguntas de investigación relacionadas con el engagement de la comunidad en artículos de medios partidistas en la plataforma Reddit:

- RQ1: ¿Cómo afecta al engagement de un artículo el sesgo político del medio de noticias, dependiendo del sesgo de la comunidad?
- RQ2: ¿Cómo afecta al engagement de un artículo la credibilidad del medio de noticias, dependiendo del sesgo de la comunidad? ¿Cuán diferentes son los efectos si se controla por el sesgo político de los medios?

Para responder a estas preguntas, seleccionamos un nuevo conjunto de datos de artículos publicados por medios partidistas estadounidenses y publicaciones en Reddit de esos artículos compartidos con comunidades partidistas. Estimamos el sesgo político de los subreddits usando el método propuesto por Waller y Anderson [219] y extrajimos información sobre temas y emociones de los artículos usando léxicos [71] (tal como se describe

en el capítulo 5 sobre los métodos de detección de tópicos). Luego utilizamos modelos causales, en particular los modelos gráficos probabilísticos, para responder las dos preguntas de investigación. Los modelos causales miden cómo un cambio en un tratamiento T impacta una variable de resultado Y , dado un conjunto de variables de confusión Z [88]. Esta relación de causa y efecto puede cuantificarse mediante el efecto promedio del tratamiento (ATE) [170]. El objetivo del análisis causal es descubrir el proceso subyacente de generación de datos y, por lo tanto, ayudarnos a comprender completamente la relación entre las variables. En este trabajo, buscamos comprender cómo el sesgo político y la credibilidad de un medio alteran causalmente el engagement de los usuarios cuando un artículo partidista se comparte en un subreddit.

Las principales aportes de este capítulo son:

- Mostramos que compartir artículos con un sesgo político opuesto al de la comunidad tienden a causar una disminución significativa en el número neto de upvotes, en comparación con lo que habría pasado si el artículo fuera del mismo sesgo político. Este efecto causal es considerablemente mayor en los subreddits de derecha que en los de izquierda.
- Descubrimos que los usuarios tienden a dar una mayor cantidad de upvotes a fuentes poco confiables, en vez de priorizar los medios de noticias veraces.
- Confeccionamos un conjunto de datos que contiene 174068 publicaciones en Reddit donde se comparten artículos periodísticos de 551 medios de Estados Unidos, junto con las inclinaciones políticas, la popularidad y los puntajes de credibilidad de los medios. Además, replicamos el método de Neural Community Embeddings descrito en [219], permitiéndonos catalogar la orientación política de los subreddits donde se compartieron las noticias.

Hasta donde sabemos, este trabajo es pionero en caracterizar la relación causal entre el sesgo político y la desinformación de las fuentes de noticias sobre el engagement de la comunidad en las redes sociales.

14.2. Dataset

14.2.1. Reddit

Recopilamos un conjunto de 371 millones de publicaciones realizadas en Reddit durante 2021. El conjunto de datos está disponible públicamente a través del servicio Pushshift¹. Para cada publicación, extrajimos la identificación del autor, el nombre del subreddit y la identificación del lugar donde se publicó, la fecha, el texto de la publicación y la dirección URL si se compartió algún enlace.

14.2.2. La dimensión partidista de los subreddits

Para nuestro análisis, utilizamos una muestra de datos que incluía todas las publicaciones que contenían un artículo de noticias en subreddits partidistas de izquierda o derecha. Para identificar subreddits partidistas, replicamos los métodos descritos en [219] en un conjunto de datos más completo y más reciente (siendo el paper original hasta el 2018 y el

¹ <https://files.pushshift.io/reddit/submissions/>

nuestro incluyendo datos del 2021). Primero generamos los embeddings [145] de todos los subreddits usando skip-gram word2vec con muestreo negativo. A continuación, calculamos la dimensión partidista (una puntuación que representa el sesgo político en cada subreddit) generando primero el vector de dimensión social utilizando los subreddits r/democrats y r/Conservatives como semillas y luego midiendo la proyección de cada subreddit a este vector [165].

Validamos las incorporaciones y las puntuaciones partidistas con los datos publicados por Waller and Anderson [219]. Los resultados se presentan a continuación:

- Encontramos una correlación significativa ($p < 0,01$) entre nuestras puntuaciones partidistas y las de [219] utilizando la prueba de correlación de rangos de Spearman.
- Medimos las analogías de la comunidad: los subreddits universitarios deberían estar más cerca de los subreddits de sus ciudades correspondientes que de los subreddits de otras ciudades; de igual forma, los equipos deportivos a sus correspondientes ciudades y a su correspondiente deporte. Waller and Anderson [219] resolvió perfectamente el 72 % de las analogías y en el 96 % de los casos la respuesta correcta estuvo en las 5 comunidades principales. En nuestro caso, los resultados son 72,22 % y 86,75 %, respectivamente.
- Utilizando un conjunto de 50 subreddits que mencionan explícitamente su afiliación partidista, encontramos que nuestras puntuaciones partidistas están fuertemente relacionadas con esta etiqueta (Cohen' $d = 2,26$).

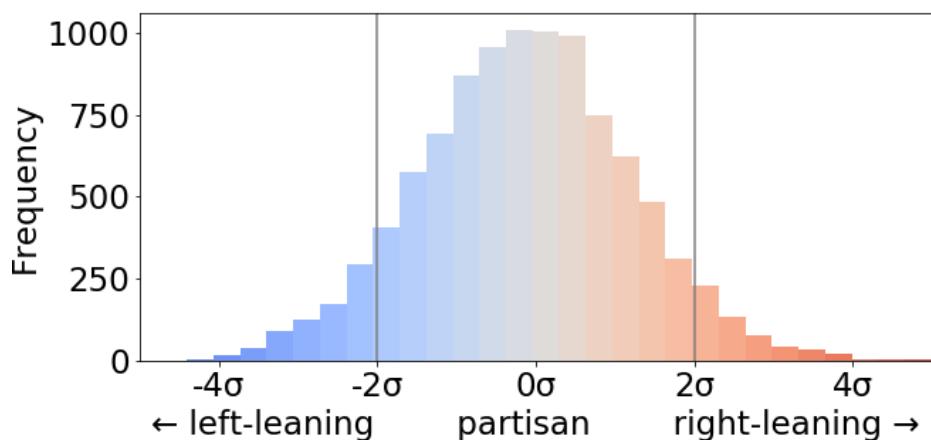


Fig. 14.1: La distribución de las puntuaciones partidistas para las comunidades de Reddit más populares de 10,000. El eje x muestra el número de desviaciones estándar de la puntuación partidista media (es decir, puntuación z). Las comunidades varían de izquierda a derecha y están coloreadas por puntuación z .

La figura 14.1 muestra el histograma de las puntuaciones partidistas de los subreddits. El eje x es el número de desviaciones estándar de la puntuación partidista media (es decir, puntuación z), que van desde la izquierda hasta la derecha. En este trabajo, seleccionamos subreddits partidistas con valores mayores que 2σ en el histograma.

14.2.3. Medios de comunicación

Se buscó caracterizar a los medios de comunicación en cuanto a su sesgo político, la veracidad de sus noticias y su popularidad. Para ello se utilizó la información y las etiquetas provistas por el sitio web de verificación Media Bias/Fact Check (MBFC)², que es un recurso de uso común [39, 223].

Sesgo político: Usamos las etiquetas de sesgo político del sitio MBFC. Estas etiquetas clasifican a cada medio en extrema izquierda, izquierda, centro izquierda, centro, centro derecha, derecha o extrema derecha analizando las posiciones editoriales respecto a temas como el aborto, la política económica, el derecho a las armas o la atención sanitaria. Teniendo en cuenta que el objetivo de este trabajo es analizar publicaciones interpartidistas, colapsamos los medios de extrema izquierda, izquierda y centro izquierda y los medios de extrema derecha, derecha y centro derecha en izquierda y derecha respectivamente. Descartamos los medios centrales. Terminamos con un total de 551 medios (293 de izquierda y 258 de derecha). Combinando el sesgo político del medio que publica el artículo y el sesgo político del subreddit donde se compartió el artículo, definimos el intercambio copartidista cuando tienen el mismo sesgo político y el intercambio interpartidista cuando tienen el sesgo opuesto.

Desinformación: MBFC presenta una etiqueta de desinformación (misinformation) al examinar si difunden noticias falsas, teorías conspirativas o información errónea. Las etiquetas son: muy alto, alto, mayoritariamente veraz, mixto, bajo y muy bajo. Hemos dividido estas seis categorías en tres: fuente veraz (muy alta, alta y mayoritariamente veraz), fuente mixta (mixta) y fuente de información errónea (baja y muy baja).

Popularidad: MBFC también enumera los datos de visitas a la página para determinar la cantidad de tráfico que recibe una fuente. Considera el número de suscriptores de medios impresos y el tamaño del mercado de medios de TV/Radio. Luego, etiqueta los medios en tres categorías: tráfico mínimo, tráfico medio y tráfico alto.

14.2.4. Artículos de noticias

Usando la biblioteca de Python `newspaper3k`³, bajamos el texto de 294787 artículos de noticias de los 551 medios partidistas. En 30,85 % de los casos, las solicitudes de scraping no tuvieron éxito porque el artículo fue eliminado. Por lo tanto, solo consideramos el 69,15 % de artículos accesibles restantes.

14.2.5. Características del dataset

14.1 resume nuestro conjunto de datos de publicaciones de Reddit que compartieron artículos de noticias en un subreddit de izquierda o derecha durante 2021. Se debe notar que este es un conjunto de datos desequilibrado, debido al hecho de que los usuarios tienden a compartir más artículos con el mismo sesgo político y artículos de izquierda. Por lo tanto, realizar un análisis causal controlando por los medios y el subreddit es vital.

² <https://mediabiasfactcheck.com>

³ <https://newspaper.readthedocs.io/>

Tab. 14.1: Descripción de las características de nuestro dataset. Upvotes netos representa el promedio de upvotes - downvotes.

	subreddit	izquierda	derecha
#subreddits	1,514	1,602	
#posts	139,588	64,480	
#upvotes netos	9.51	9.26	
#posts co-partidista	111,872	35,375	
#posts cross-partidista	27,716	29,105	

14.3. RQ1: El engagement y el sesgo político de los medios

14.3.1. Configuración experimental

Nuestro objetivo es medir el efecto causal que tiene el partidismo de un medio sobre el engagement de los usuarios. Los artículos de medios con el mismo sesgo político que el subreddit donde se compartió son publicaciones co-partidistas y los artículos de medios con sesgo político opuesto son publicaciones cross-partidistas. Usamos la puntuación o likes de Reddit, el número neto de upvotes (votos a favor - votos en contra), para cuantificar el engagement de los usuarios en una publicación.

Luego, usamos la biblioteca de Python DoWhy [197] para medir cómo el sesgo del medio afecta el número neto de upvotes. Usamos si un artículo es co-partidistas o cross-partidistas como tratamiento y el número neto de upvotes como resultado. La popularidad de los medios, el sesgo político de los medios, la veracidad de los medios y los temas del artículo son factores de confusión. La figura 14.2 (1) muestra el grafo causal utilizado en los experimentos, donde los nodos representan nuestras variables y todos los factores de confusión pueden tener un efecto causal en el resultado y el tratamiento. Es importante considerar todas estas interacciones en el modelo porque los factores de confusión también pueden tener un efecto en el resultado (ciertos temas pueden causar más upvotes y un tráfico mínimo en los medios menos upvotes) y queremos aislar el efecto del tratamiento.

Utilizamos dos métodos para probar la solidez de los resultados: causa común aleatoria (random common cause) y tratamiento con placebo (placebo treatment). Cuando se utiliza el método de causa común aleatoria, el análisis se repite, pero agregando una variable aleatoria independiente como causa común al conjunto de datos, lo que no debería cambiar significativamente los resultados estimados si el análisis causal es robusto. Para este artículo, utilizamos un umbral de tolerancia de varianza de 10 % del valor estimado original [108]. En el método de tratamiento con placebo, el análisis se repite pero reemplaza la verdadera variable de tratamiento con una variable aleatoria independiente. En este caso, el efecto causal estimado debería ser cero.

Medimos el efecto medio del tratamiento (o ATE, por sus siglas en inglés) mediante regresión lineal y presentamos el efecto de percepción: el ATE sobre el número neto de upvotes en subreddits de derecha/izquierda dividido por el número medio de upvotes en los subreddits de derecha/izquierda. Calculamos los intervalos de confianza ejecutando 100 simulaciones de arranque y reportamos los resultados con un nivel de confianza de 0,95. Finalmente, validamos la solidez de los resultados mediante las dos pruebas antes dichas: causa común aleatoria y tratamiento con placebo.

Repetimos este experimento desagregando según el sesgo político del subreddit donde se publicó el artículo. Por lo tanto, podemos comparar las diferencias entre el efecto causal

Tab. 14.2: Se observa la media del número neto de upvotes en artículos co-partidistas y cross-partidistas en subreddits de derecha y de izquierda. También se reporta el efecto causal (ATE) de realizar una publicación cross-partidista en el número neto de upvotes, con respecto a una publicación co-partidista.

	subreddit	izquierda	derecha
upvotes co-partidistas		9,75	15,54
upvotes cross-partidistas		3,67	4,33
efecto causal	$-71\% \pm 8\%$	$-112\% \pm 12\%$	

en subreddits de derecha y subreddits de izquierda.

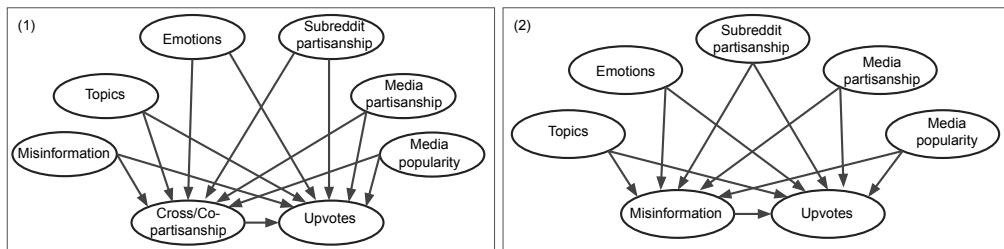


Fig. 14.2: Grafos causales de RQ1 (1) y RQ2 (2).

14.3.2. Resultados y discusión

Una comunidad política online puede interactuar de manera diferente cuando un usuario comparte un artículo político dependiendo del sesgo político de sus medios. La tabla 14.1 muestra que los usuarios tienden a compartir artículos publicados por un medio con la misma inclinación política. Es interesante observar cómo los usuarios interactúan de manera diferente cuando un artículo con una tendencia política opuesta se comparte dentro de su comunidad política. Las publicaciones cross-partidistas tienen en promedio un número neto de upvotes más bajo en comparación con las publicaciones co-partidistas, tanto en subreddits de derecha como de izquierda (Tabla 14.2). Los resultados muestran que los artículos cross-partidistas tienden a provocar una disminución en el número neto de upvotes en comparación con los que habrían tenido si el artículo fuera co-partidista. En particular, los usuarios de subreddits de izquierda tienden a votar un 71 % menos si un artículo proviene de un medio de tendencia opuesta. Por otro lado, hay una interacción significativamente más negativa con los artículos cross-partidistas en subreddits de derecha, donde los artículos cross-partidistas provocan una disminución del 112 % en el número neto de upvotes en comparación con lo que habrían tenido si el artículo tuviera la misma tendencia política. Este resultado es consistente con investigaciones anteriores sobre asimetrías ideológicas [130].

Trabajos anteriores sobre cámaras de eco también encontraron esta correlación y sugieren que la gente prefiere contenidos con su sesgo político [53]. En este análisis causal, consideramos otras variables que podrían tener un efecto en el número de upvotes (como el tema del artículo, las emociones en el artículo, la popularidad de los medios y el sesgo político de los medios) para aislar el efecto del periodismo de los medios, y mostramos que también existe una relación causal. Todos los resultados son estadísticamente

Tab. 14.3: Estimación promedio del efecto del tratamiento de las fuentes de información errónea sobre el número neto de upvotes.

subreddit	izquierda	derecha
media		
izquierda	$-18,7\% \pm 4,1\%$	$16,1\% \pm 7,6\%$
derecha	$-46,8\% \pm 8,7\%$	$-38,1\% \pm 12,6\%$

significativos según la herramienta DoWhy y pasan las pruebas de robustez.

14.4. RQ2: El engagement de la comunidad y credibilidad de los medios

14.4.1. Configuración experimental

Usamos una configuración experimental similar descrita para la RQ1. La principal diferencia es que consideramos la etiqueta de desinformación del artículo por parte de los medios como el tratamiento y el número neto de upvotes como el resultado. Los factores de confusión incluyen la popularidad de los medios de comunicación, el sesgo político de los medios, el sesgo político del subreddit y los temas del artículo. La figura 14.2 (2) muestra el grafo causal. Repetimos el experimento para subreddits de derecha e izquierda y para medios de derecha e izquierda por separado. Como tal, podemos observar cómo los usuarios interactúan de manera diferente según el sesgo político de los subreddits y la fuente del medio.

14.4.2. Resultados

La tabla 14.2 muestra el efecto causal que tienen las fuentes de información errónea sobre el número neto de upvotes. De manera similar a RQ1, utilizamos causa común aleatoria y tratamiento con placebo como métodos de refutación para garantizar la solidez del análisis y que todos los resultados pasen estas pruebas de validación. Un efecto causal negativo indica que las fuentes fácticas tienden a provocar una disminución en el número neto de upvotes y las fuentes de información errónea aumentan el número neto de upvotes.

Los resultados muestran que los usuarios tienden a favorecer significativamente las fuentes de información errónea, en la mayoría de los casos. En particular, los usuarios de subreddits de izquierda tienden a votar 46,8 % menos si un artículo de un medio de derecha proviene de una fuente objetiva. Los efectos causales son significativamente mayores en los medios de derecha que en los de izquierda. Trabajos anteriores habían demostrado la correlación entre la desinformación y los likes en las redes sociales [67]. Este análisis amplía estudios anteriores al proporcionar también una perspectiva causal, aislando el efecto de las fuentes de información errónea y controlando otras variables. El único efecto positivo corresponde a los subreddits de derecha con medios de izquierda. Implica que los usuarios del subreddit de derecha tienden a penalizar las fuentes de desinformación y favorecen las fuentes fácticas si los medios son de izquierda.

14.5. Conclusiones

Trabajos anteriores mostraron una correlación entre las métricas de engagement del usuario y las fuentes de información errónea y las interacciones entre partidos. Este estudio

presenta un análisis causal de estas relaciones y muestra cómo el sesgo político de los medios de comunicación y las fuentes de desinformación impactan el engagement de los usuarios. Los resultados muestran un efecto causal significativo del sesgo político de un artículo y el engagement que recibe. Observamos que los artículos cross-partidistas tienden a tener un menor número neto de upvotes en comparación con los artículos co-partidistas. Este efecto causal negativo es significativamente mayor en los subreddits de derecha que en los de izquierda. También mostramos que existe una relación causal negativa entre la veracidad de los medios y el engagement de un artículo, al controlar múltiples variables de confusión como el tema del artículo, las emociones que retrata, la popularidad de los medios y el sesgo de los medios de comunicación. Los usuarios tienden a dar una mayor cantidad de upvotes a fuentes de información errónea, en vez de priorizar a las fuentes de información objetivas. La única excepción son los usuarios del subreddit de derecha con medios de izquierda, que penalizan a los medios de información errónea.

Parte V

SISTEMAS DE RECOMENDACIÓN DE LAS REDES SOCIALES

Esta sección contiene resultados que fueron publicados en el siguiente paper:

- Albanese, F., Trappolini, G., Scarlino L. & Silvestri, F. (2024). The Impact of Source-Target Node Distance on Vicious Adversarial Attacks in Social Network Recommendation Systems. Proceedings of the International Workshop on Graph-Based Approaches in Information Retrieval at the 46th European Conference on Information Retrieval (ECIR 2024).

15. INTRODUCCIÓN

En capítulos anteriores se analizaron diversos aspectos de las redes sociales, haciendo principal énfasis en el comportamiento de los usuarios, sus interacciones y el contenido que comparten. Sin embargo, en las redes sociales, los usuarios y el contenido que publican no son los únicos actores de interés. Los algoritmos de recomendación propios de las redes sociales también juegan un rol fundamental en qué contenido de qué usuario se muestra a qué otro usuario.

En este contexto, las redes neuronales profundas y los métodos de predicción de aristas permiten la sugerencia personalizada de nuevos contactos dentro de una red social [123]. Su funcionamiento es relevante ya que el contenido que se le muestra al usuario depende (entre otros factores) de los contactos de dicho usuario.

Estos algoritmos de aprendizaje automático muestran probabilidades estimadas para cada usuario, donde la lista de aristas predicha representa un conjunto de nuevos contactos potenciales. Estos algoritmos resultan especialmente cruciales en el contexto de los influencers de las redes sociales y la dinámica de interacciones. Ser seguido por múltiples usuarios o por ciertas comunidades puede interpretarse como un respaldo valioso, que genera diversos beneficios para el usuario, como mayores ingresos por campañas publicitarias o mayor llegada de su mensaje hacia cierto público.

Sin embargo, estos sistemas de recomendación son susceptibles a ataques adversariales, especialmente por parte de usuarios malintencionados que buscan mejorar artificialmente su reputación en las plataformas de redes sociales [46, 136, 234]. Por ejemplo, un político puede buscar que su usuario le aparezca sugerido a todos los usuarios que componen una comunidad política cuyo sesgo sea similar al propio, para aumentar su popularidad. O, por el otro lado, buscar ser sugerido en una comunidad con sesgo opuesto, distante en la red de interacciones.

SAVAGE (SpArse Vicious Attacks on Graph nEworks) es un modelo de ataque que utiliza nodos viciosos para explotar las debilidades en los sistemas de predicción de aristas basados en redes neuronales profundas [214]. Trabajos anteriores se centraron en caracterizar el rendimiento del modelo con diversos conjuntos de datos utilizando pares aleatorios de nodos origen-destino. Sin embargo, la tasa de éxito de un ataque adversarial depende de la distancia en la red entre el nodo de origen (quien realiza el ataque) y el nodo de destino (usuario quien es el objetivo del ataque). En el contexto del ejemplo anterior, el nodo origen sería la cuenta del político y el nodo destino sería el usuario al que quiere que le aparezca su contenido.

Orquestar un ataque exitoso puede resultar más complicado cuando el nodo de origen está lejos del nodo de destino en la red de interacciones de la red social, debido a las características de los algoritmos de aprendizaje automático.

En este capítulo, presentamos un análisis exhaustivo del rendimiento de SAVAGE y sus ventajas sobre los métodos alternativos de ataque adversarial de predicción de aristas en términos de éxito del ataque y optimización de recursos. Utilizando un conjunto de datos de redes sociales compuesto por tweets en contextos políticos, estudiamos el impacto de la distancia origen-destino en el rendimiento general del ataque, que tiene aplicaciones del mundo real, como sugerencias de amigos/seguidores de redes sociales donde el usuario origen puede o no tener amigos o amigos de amigos en común con el usuario objetivo. Al

explorar diferentes escenarios donde los nodos de origen y de destino están ubicados a una distancia variable en el grafo de interacciones, el estudio proporciona una comprensión más profunda de cuán significativamente este factor puede influir en el éxito del ataque.

16. ATAQUES ADVERSARIALES A LOS SISTEMAS DE RECOMENDACIÓN DE LAS REDES SOCIALES

16.1. Introducción

16.1.1. Predicciones de aristas en grafos

El problema de predecir aristas en grafos es un área muy explorada en el análisis de redes [123, 146, 224]. Un grupo fundamental pero eficaz de métodos para la predicción de aristas son las heurísticas. Estos métodos plantean la hipótesis de homofilia. Es decir, una hipótesis de que la similitud, cualquiera sea su definición, influye en la presencia de un vínculo entre dos nodos. En la práctica, estas técnicas emplean heurísticas predeterminadas para estimar la similitud entre nodos, que luego predicen la probabilidad de una arista [134, 138]. Estos enfoques heurísticos se clasifican según la extensión de los vecinos considerados para calcular la similitud de los pares de nodos. Por ejemplo, metodologías como Common Neighbors [158], Jaccard [134] y Preferential Attachment [27] caen bajo heurísticas de primer orden ya que solo toman en cuenta los vecinos inmediatos de los nodos. Por el contrario, Adamic-Adar [2] y las estrategias de asignación de recursos (resource allocation strategies) [232] son heurísticas de segundo orden que involucran una vecindad de los nodos. Otras heurísticas más complejas, conocidas como heurísticas de alto orden, requieren conocimiento de toda la red, con ejemplos que incluyen el índice de Katz [112], rooted PageRank [42] y SimRank [106]. A pesar de su eficacia práctica, los métodos heurísticos a menudo se ven limitados por sus supuestos inherentes sobre la existencia de vínculos. Para abordar estas limitaciones, la predicción de aristas se ha replanteado como un problema típico de clasificación binaria, abordado utilizando enfoques de aprendizaje supervisado [5]. Varios estudios recientes han introducido métodos eficaces basados en redes neuronales profundas en grafos (GNN) para la predicción de aristas [115, 133, 231]. En términos generales, las GNN facilitan el aprendizaje de representaciones de nodos (embeddings) agregando datos del entorno local de cada nodo. Estos embeddings luego se utilizan para la predicción de aristas posterior. Este estudio examina específicamente un ataque adversarial dirigido a metodologías de predicción de aristas basadas en GNN.

16.1.2. Ataques adversariales en la predicción de aristas

A pesar de su eficacia, las investigaciones indican que los modelos de aprendizaje automático son susceptibles a lo que se conoce como ataques adversariales. Estos ataques están diseñados intencionalmente para engañar las capacidades predictivas de los modelos. Para ello, realizan cambios pequeños, pero cuidadosamente elegidos, al input del modelo.

Se ha demostrado que estas tácticas son efectivas en varias áreas importantes, como el análisis de imágenes [82] y la identificación de malware [85]. Además, las técnicas utilizadas en perturbaciones adversariales comparten características con el subcampo de ejemplos contrafactuals [48, 49, 50, 51, 139, 212, 213]. Sin embargo, sólo un número limitado de estudios ha investigado la efectividad de estos ataques adversariales en los modelos de predicción de aristas, particularmente aquellos que utilizan Graph Neural Networks (GNN), que requieren alterar el grafo de entrada. Entre estos estudios cabe

destacar[45, 46, 136, 234], y se puede encontrar un compendio detallado en [56]. Los métodos mencionados generalmente suponen que el atacante tiene control sobre una parte de los nodos existentes. Sin embargo, esta suposición suele resultar poco práctica debido a los costes excesivos que implica. Para contrarrestar esto,[221] introdujo un método que permite a los atacantes crear nodos nuevos (es decir, falsificados o maliciosos) para ataques más eficientes. Además,[220] desarrolló una aproximación lineal de este enfoque para una mejor escalabilidad, y [57] diseñó una perturbación universal capaz de apuntar a múltiples nodos sin dejar de ser efectiva. Tanto en escenarios tradicionales como maliciosos, todos estos métodos restringen el alcance de la modificación del grafo. Este límite generalmente se aplica de dos maneras: ya sea asignando al atacante un presupuesto específico [46, 207] o estableciendo reglas para limitar la variación entre el grafo original y el modificado, como se ve en [136]. En cualquier caso, esto sirve como límite a la cantidad de “recursos maliciosos” que pueden utilizarse para el ataque. Por esta razón se propuso SAVAGE [214], que define al ataque adversarial como una tarea de optimización, logrando el equilibrio entre la efectividad del ataque y la eficiencia en el manejo de recursos. Sin embargo, hasta donde sabemos, faltan trabajos que estudien el impacto de la distancia entre el nodo atacante y el nodo objetivo del ataque. En este capítulo, nos propusimos abordar este problema.

16.2. Datos

Para realizar este análisis se usó un conjunto de tweets políticos. Dicho conjunto es el mismo que fue utilizado en la sección 6 para evaluar distintos algoritmos de detección de tópicos [7].

16.3. Métodos

16.3.1. Ataque Adversarial

Para realizar el ataque, se replicó el procedimiento experimental utilizado en el paper original de SAVAGE [214]. En específico, se entrenó un sistema de predicción de aristas basado en GNN en un entorno transductivo[215], utilizando el 90 % de las aristas existentes como conjunto de entrenamiento y el 10 % restante como conjunto de prueba, logrando una precisión de 80 %. La arquitectura de modelo consta de dos *capas convolucionales* apiladas de tamaño 128 y 64 respectivamente, seguidas por una función de activación ReLU. A estas capas convolucionales les sigue un MLP[182] de dos capas con una función de activación ReLU. Finalmente, el proceso de entrenamiento utiliza la función de pérdida de entropía cruzada binaria y el optimizador Adam con una tasa de aprendizaje de 0,01. Este es el modelo que se utilizará posteriormente para la tarea de predicción de aristas.

El experimento implica probar SAVAGE en el grafo descrito en la sección anterior y emplear el modelo de predicción de aristas GNN para la tarea de predicción de aristas. Esta tarea se realiza en 6 conjuntos diferentes de pares, cada uno de ellos compuesto por 50 pares de nodos *origen – destino* desconectados, para descubrir si se predecirán las aristas dirigidas entre ellos.

Los pares en cada conjunto tienen la misma distancia entre el nodo origen y destino. El rango de distancias usado es [3,8]. Esto mostrará cómo la distancia puede afectar a las predicciones de los modelos.

Para este experimento y de manera similar a trabajos anteriores [214], el número máximo de nodos viciosos adicionales que se pueden introducir en el grafo se ha establecido

en 50. El modelo GNN se itera 150 veces para cada par, ajustando los pesos en cada iteración.

El modelo de predicción de aristas predecirá las aristas en el grafo original y en el perturbado, mostrando la mejora del rendimiento.

16.3.2. Métricas de evaluación

Al evaluar el rendimiento de SAVAGE, se utilizan varias métricas y se comparan con la tarea de predicción de aristas en el grafo imperturbable:

- **Kullback-Leibler Divergence:** Mide la diferencia entre distribuciones de probabilidad [107]. En este contexto, cuantifica el cambio de distribución de grados del nodo antes y después del ataque adversarial. Esta métrica proporciona información sobre cómo la perturbación generada afecta la estructura general del grafo. Dado que este estudio de caso involucra grafos dirigidos, la divergencia KL se calcula por separado tanto para la distribución de grado interno como para la distribución de grado externo del nodo. Un valor más alto de la divergencia KL corresponde a un impacto más fuerte en la topología de la red, lo que sugiere una perturbación más significativa causada por el ataque adversarial. En consecuencia, es deseable una divergencia KL más baja ya que indica menos cambios en la estructura del grafo.
- **Mean Added Nodes:** Esta métrica evalúa la utilización de recursos y la eficiencia del ataque midiendo la cantidad de nodos que se agregan al grafo perturbado. Un valor más bajo indica una mejor optimización de los recursos y una menor complejidad del ataque, ya que implica que se necesitaron menos nodos adicionales para lograr la perturbación deseada.
- **Initial and Final Prediction:** Estas métricas miden las puntuaciones de predicción de aristas para una arista de destino, respectivamente, antes y después del ataque. A cada par origen-destino se le asigna una puntuación de predicción en el rango [0,1], que representa la probabilidad de existencia del vínculo. Dado el modelo de predicción de aristas, la predicción media inicial indica las puntuaciones de predicción reales en el grafo original, mientras que la predicción media final captura los valores predichos en el grafo perturbado. Comparar estas métricas nos permite evaluar la efectividad del ataque.
- **Attack Rate:** Dada la puntuación de predicción calculada de una arista, se utiliza un umbral para determinar si se predice que la arista existirá o no. Si el valor previsto es superior al umbral, el ataque se considera exitoso. La tasa de ataque para un conjunto de pares se calcula como la proporción de ataques exitosos con respecto al número total de intentos de ataque:

$$AttackRate = \frac{\#successful\ attacks}{\#attacks}. \quad (16.1)$$

Es importante reconocer que, a pesar de estar estrechamente relacionada con la métrica de predicción final, la tasa de ataque puede fluctuar dependiendo de la elección del umbral: un umbral más alto conduce a predicciones más precisas, pero puede

resultar en una tasa de ataque más baja. En este caso de estudio, el umbral se establece en 0,6, lo que significa que un ataque se considera exitoso si la probabilidad prevista de la arista objetivo es igual o mayor que 60 %.

16.4. Resultados

16.4.1. Análisis de SAVAGE

En primer lugar, se realizó un ataque adversarial de SAVAGE en el modelo de predicción de aristas a través del grafo de retweets de usuarios de Twitter y se probaron conjuntos de muestras basados en la distancia mencionada anteriormente.

Considerando un resultado agregado en los 300 pares a varias distancias en el rango [3,8], el ataque SAVAGE logró una tasa de ataque de 61,3 %. La predicción final promedio fue de 0,60, en comparación con los 0,48 de la predicción inicial promedio. Eso significa un 25 % de mejora a costa de un promedio de sólo 6,77 de nodos viciosos agregados, lo que implica una sobrecarga de 0,06 % en términos de nodos en la red.

La divergencia KL se calculó para las distribuciones de grado de entrada y salida. La divergencia KL promedio para la distribución de grados de entrada (KL IN) fue de $1,1 \times 10^{-21}$, mientras que para la distribución de grados de salida (KL OUT) fue de $4,1 \times 10^{-4}$. Aunque el valor ideal sería 0, estos valores siguen siendo muy bajos, lo que indica un impacto mínimo en la estructura de la red.

En las Figuras 16.1a y 16.1b se observa que el ataque exhibe un mejor rendimiento en distancias más cortas. aunque la tendencia no sigue un patrón lineal. Específicamente, a una distancia de 3, hay una puntuación de predicción final máxima de 0,69, lo que resulta en una tasa de ataque de 70 %. Esto representa una mejora en la precisión de la predicción de casi 40 % en comparación con el grafo imperturbable, donde la puntuación de predicción era de 0,50.

Por otro lado, el análisis de *divergencia KL* y *media de nodos agregados* no reveló ninguna tendencia o correlación significativa con la distancia origen-destino.

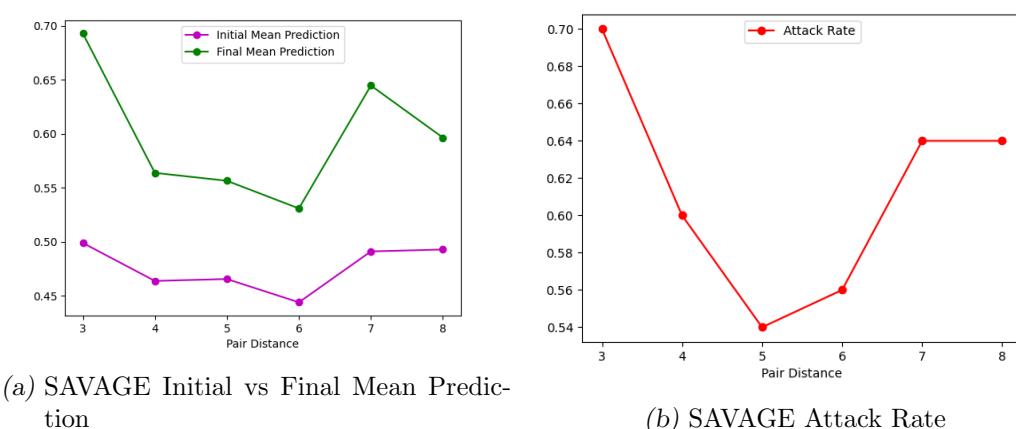


Fig. 16.1: SAVAGE Results on Distance-based Sample Sets

Para proporcionar solidez a los resultados, los intervalos de confianza se calcularon con un nivel de confianza de 95 %, como se muestra en la Tabla 16.1. Estos intervalos se

calcularon utilizando la técnica bootstrap, sampleando con reposición [183].

Tab. 16.1: SAVAGE 95 % confidence intervals for final predictions

Distancia	Predicción
3	$0,684 \pm 0,101$
4	$0,560 \pm 0,101$
5	$0,554 \pm 0,106$
6	$0,531 \pm 0,104$
7	$0,640 \pm 0,100$
8	$0,590 \pm 0,096$

16.4.2. Comparación con otros métodos

Se llevaron a cabo diferentes ataques adversariales en los mismos conjuntos de muestras de pares de nodos, para comparar el rendimiento de SAVAGE con otros modelos. Inicialmente, se consideró una versión adaptada del método de ataque de gradiente iterativo o **AIGA** (por sus siglas en inglés) [46]. Esta versión adaptada fue diseñada específicamente para encajar en el marco de este estudio. El método de ataque de gradiente iterativo (IGA) ha sido reconocido como uno de los ataques adversariales de mayor rendimiento en la predicción de aristas. Por lo tanto, al adaptarlo y probarlo en este contexto, se proporciona una comprensión más completa de la efectividad de SAVAGE en relación con los métodos de ataque de última generación.

Además, en la evaluación se incluyó una línea de base aleatoria simple (RAND). Este enfoque permite agregar/eliminar conexiones en el grafo y activar/desactivar nodos víscos con una probabilidad p , que determina la fuerza del atacante. En este estudio de caso, se consideraron dos líneas de base aleatorias: **RAND-L**, con una probabilidad baja $p_L = 0,25$, y **RAND-H**, con una probabilidad alta $p_H = 0,75$.

Las figuras en 16.2 proporcionan una representación visual de las métricas de evaluación para todos los métodos probados. Los resultados obtenidos confirman la superioridad de SAVAGE sobre otros marcos de ataque en términos de predicción final (*AP*) y tasa de ataque (*AR*) en todas las distancias. Vale la pena señalar que a una distancia de 8, SAVAGE logra una puntuación de predicción media que es más del doble en comparación con AIGA. Esto demuestra que SAVAGE no sólo funciona mejor en distancias más cortas sino que también supera a los métodos en distancias mayores. SAVAGE también mostró su excepcional utilización de recursos. AIGA empleó el número máximo de nodos adicionales disponibles (*AN*) para cada conjunto de muestra, lo que indica una baja optimización de recursos. Por el contrario, SAVAGE demostró un uso mínimo de recursos maliciosos adicionales, superando incluso a RAND-L (que agrega nodos maliciosos con una probabilidad baja).

Se observa una pequeña debilidad de SAVAGE en la evaluación en relación con KL-Divergencia (*KI* y *KO*): a pesar de tener un buen desempeño y exhibir pequeños cambios en la distribución de grados en el grafo perturbado, AIGA logró un mejor desempeño en distancias pequeñas.

Tab. 16.2: Resultados de los distintos métodos. Los modelos fueron evaluados utilizando como métrica la predicción final (*AP*), el attack rate (*AR*), mean added nodes (*AN*) y la KL-Divergence de la distribución de indegree de los nodos (*KI*) y la distribución de outdegree de los nodos (*KO*).

Método	Distancia 3					Distancia 4				
	AP ↑	AR ↑	AN ↓	KI ↓	KO ↓	AP ↑	AR ↑	AN ↓	KI ↓	KO ↓
SAVAGE	0.69	0.70	8.62	$3 \cdot 10^{-21}$	$13 \cdot 10^{-4}$	0.56	0.60	3.98	0.00	0.00
AIGA	0.59	0.60	50	0.00	0.00	0.45	0.44	50	0.00	0.00
RAND-L	0.12	0.1	12.9	$2 \cdot 10^{17}$	$5 \cdot 10^{-3}$	0.15	0.06	12.9	$2 \cdot 10^{17}$	$5 \cdot 10^3$
RAND-H	0.10	0.08	37.54	10^{-17}	$5 \cdot 10^3$	0.12	0.06	37.54	10^{-17}	$5 \cdot 10^{-3}$
Distancia 5										
Método	AP ↑	AR ↑	AN ↓	KI ↓	KO ↓	AP ↑	AR ↑	AN ↓	KI ↓	KO ↓
SAVAGE	0.56	0.54	9.72	$3 \cdot 10^{-21}$	$12 \cdot 10^{-4}$	0.53	0.56	7.28	0.00	0.00
AIGA	0.53	0.52	50	0.00	0.00	0.47	0.48	50	0.00	0.00
RAND-L	0.07	0.02	12.9	$2 \cdot 10^{17}$	$5 \cdot 10^{-3}$	0.13	0.08	12.9	$2 \cdot 10^{17}$	$5 \cdot 10^{-3}$
RAND-H	0.06	0.02	37.54	10^{-17}	$5 \cdot 10^{-3}$	0.10	0.04	37.54	10^{-17}	$5 \cdot 10^{-3}$
Distancia 7										
Método	AP ↑	AR ↑	AN ↓	KI ↓	KO ↓	AP ↑	AR ↑	AN ↓	KI ↓	KO ↓
SAVAGE	0.64	0.64	7.38	0.00	0.00	0.60	0.64	3.68	0.00	0.00
AIGA	0.51	0.52	50	0.00	0.00	0.29	0.26	50	0.00	0.00
RAND-L	0.05	0.00	12.9	$2 \cdot 10^{17}$	$5 \cdot 10^{-3}$	0.16	0.16	12.9	$2 \cdot 10^{17}$	$5 \cdot 10^{-3}$
RAND-H	0.04	0.00	37.54	10^{-17}	$5 \cdot 10^{-3}$	0.13	0.12	37.54	10^{-17}	$5 \cdot 10^{-3}$

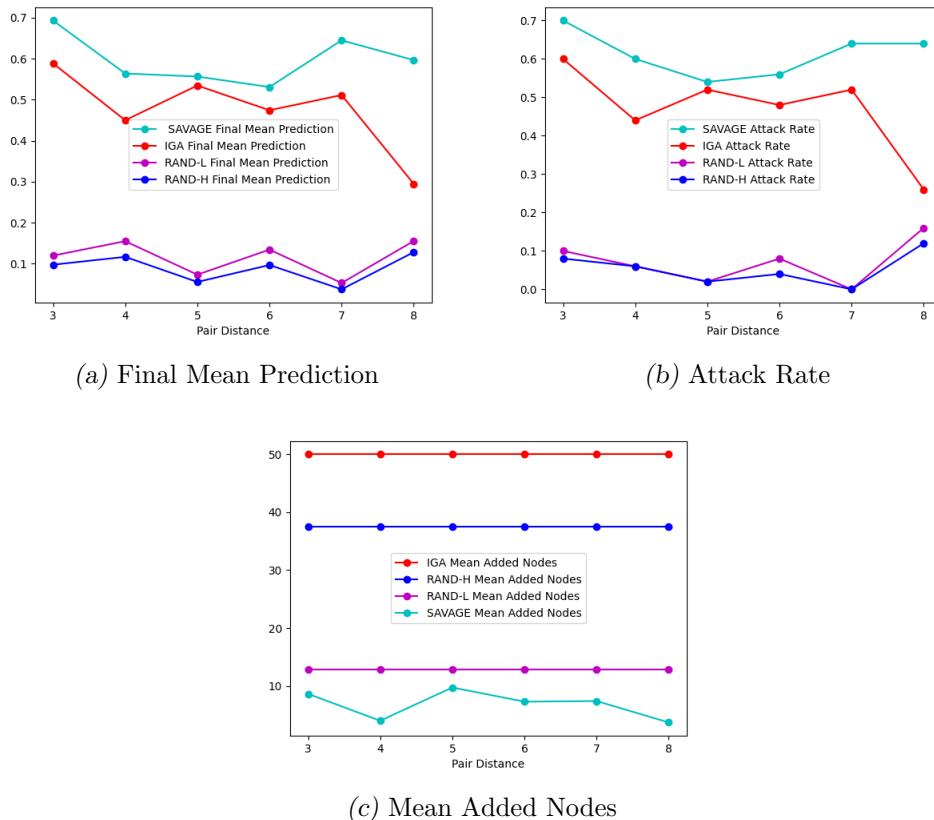


Fig. 16.2: SAVAGE vs IGA vs RAND-L vs RAND-H Comparison

16.5. Conclusiones

Con todos los avances y la popularización de los algoritmos de aprendizaje automático resulta fundamental entender las limitaciones y los problemas de los sistemas de recomendación de usuarios de las redes sociales.

En este trabajo, se analizó el comportamiento de SAVAGE en relación con la topología de la red y la distancia entre los nodos de origen y de destino. Los resultados mostraron el rendimiento del SAVAGE en comparación con los otros métodos. En particular, confirma nuestra hipótesis inicial de que SAVAGE es particularmente bueno en escenarios de grandes distancias entre los nodos, lo cual es particularmente útil en casos del mundo real, como sugerencias de amigos/seguidores en redes sociales donde el usuario origen (el atacante) no está cerca del usuario objetivo.

El trabajo futuro podría centrarse en validar los resultados en escenarios más diversos con otros conjuntos de datos.

Parte VI

CONCLUSIONES

17. CONCLUSIONES

En esta tesis, hemos explorado las complejidades y dinámicas inherentes a las redes sociales, enfocándonos en cómo los patrones de interacción y las estructuras de las comunidades virtuales influyen en la formación y difusión de opiniones, especialmente en contextos políticos. Hemos diseñado y aplicado técnicas avanzadas de machine learning y procesamiento de lenguaje natural para caracterizar los fenómenos de polarización política y cámaras de eco, así como el impacto de los medios de comunicación en la conversación online. A través de nuestros análisis, hemos identificado patrones clave que subrayan tanto la perpetuación de las cámaras de eco como los momentos y mecanismos mediante los cuales los usuarios pueden romper con estas burbujas informativas. Además, hemos considerado las implicaciones éticas y las limitaciones metodológicas asociadas con el estudio de las redes sociales. A continuación, resumiremos los hallazgos principales y discutiremos sus implicaciones.

En la Parte 2 presentamos una nueva forma de agrupar tweets para mejorar la calidad de la modelización de temas en Twitter, sin requerir ninguna modificación en el algoritmo subyacente. En el desarrollo de nuestra investigación, implementamos y comparamos nuestro enfoque utilizando distintas métricas que incluyeron la evaluación de la calidad de agrupación, clasificación supervisada y tareas de recuperación de información. Dichos experimentos buscan replicar aplicaciones reales como son la detección no supervisada de tópicos, la caracterización del contenido online y sistemas de recomendación de contenido para usuarios, respectivamente. Los experimentos, llevados a cabo en dos bases de datos distintas, demostraron que nuestro método de agrupación basado en comunidades superó a las técnicas alternativas en casi todas las tareas y métricas evaluadas. Adicionalmente, nuestros análisis revelaron que nuestro enfoque ofrece ventajas significativas en términos de eficiencia temporal, al reducir de manera efectiva el volumen de documentos a procesar, comparado con métodos de agrupación tradicionales. Estos resultados subrayan la efectividad de la agrupación basada en comunidades, destacando su potencial para optimizar el análisis de redes sociales en diversos contextos aplicativos.

En la Parte 3 abordamos dos aspectos fundamentales relacionados con la influencia de los medios de comunicación en contextos políticos y electorales. En particular, medimos cómo los medios, al marcar una agenda mediática, favorecen o perjudican a ciertos candidatos en contextos electorales. Primero, identificamos una relación causal significativa entre la cobertura de ciertos temas por los medios de comunicación, como la inmigración y los escándalos de corrupción, y las fluctuaciones en las encuestas de opinión pública durante las elecciones presidenciales de 2016 en Estados Unidos. Este hallazgo confirma que la mención de temas específicos en las noticias tiene un impacto notable en la opinión pública en tiempos electorales. Se debe notar que es el tópico quien logra este efecto causal en las encuestas, y no el sentimiento o el sesgo del medio. Eso muestra que no importó el sesgo con el que se habla de un tema (si a favor o en contra), sino que la mera mención de un tema ya es intrínsecamente positivo o negativo para un candidato.

En segundo lugar, en la Parte 3 también exploramos la interacción entre la agenda mediática y la agenda pública, esta última reflejada en los contenidos que los usuarios buscan y comparten en redes sociales. Investigamos la dinámica de influencia mutua: si son los medios los que ajustan su contenido para resonar con los intereses del público, o si son

los usuarios quienes modifican sus temas de interés en respuesta a la cobertura mediática. Estos análisis son cruciales para entender la dirección del flujo de influencia entre los medios y el público en la configuración del discurso público. Nuestros resultados muestran una relación reciproca, donde dependiendo el caso ambos se afectan y retroalimentan mutuamente.

En la Parte 4 centramos nuestro foco en el comportamiento de los usuarios de redes sociales. Primero, analizamos el papel de las cuentas oficiales de políticos en plataformas como Twitter, Instagram y Facebook, observando variaciones significativas en el comportamiento de estos y las respuestas de sus seguidores en cada red social. Los resultados muestran que la toxicidad expresada por los políticos es recompensada con likes en Twitter, debido a su carácter controversial, pero castigada con menor cantidad de interacciones en Facebook e Instagram, llevando a que los políticos adapten sus mensajes dependiendo de la plataforma.

Además, pusimos el foco en usuarios en general y las comunidades políticas que forman. Nuestros experimentos muestran que los usuarios que cambian de comunidad política, usuarios que empezaron a interactuar con otros usuarios cuyas ideas políticas eran opuestas, tenían valores estadísticamente más bajos de *PageRank*. Una posible interpretación de este resultado es que los usuarios cambian con quién interactúan cuando sus mensajes no tienen respuesta y no son “escuchados” por su comunidad. También mostramos que cambios en la comunidad política están significativamente relacionados con cambios en el sesgo de los usuarios (no solo son cambios con quién se interactua, sino también cambian significativamente las ideas que expresan).

Por otro lado, también evaluamos el impacto causal de compartir noticias políticas en Reddit. Los resultados revelan que el sesgo político de un artículo influye significativamente en el nivel de interacción que recibe; los artículos con un sesgo contrario al de la comunidad tienden a recibir menos aprobación, lo que refuerza el sesgo de confirmación y desalienta el debate de ideas contrarias. Además, identificamos una relación causal negativa entre la veracidad de los medios y el engagement de los artículos, promoviendo así la difusión de noticias falsas. Estos hallazgos subrayan la necesidad de estrategias que fomenten un diálogo más equilibrado y veraz en las plataformas sociales.

Finalmente, en la Parte 5 hemos destacado la importancia de comprender los desafíos de los sistemas de recomendación en redes sociales ante el avance y la popularización de los algoritmos de aprendizaje automático maliciosos. Nos centramos en analizar el comportamiento del método SAVAGE en relación con la topología de la red y la distancia entre nodos. Los resultados confirman que SAVAGE es particularmente eficaz en escenarios donde existe una gran distancia entre los nodos, lo cual es relevante en aplicaciones prácticas como las sugerencias de amigos o seguidores en redes sociales. Este aspecto es crucial en contextos políticos, cuando el usuario origen no tiene conexiones directas o indirectas con el usuario objetivo, al cual se le quiere maliciosamente modificar las recomendaciones de usuarios y de contenido que le sugiere una red social.

En conclusión, esta tesis ha buscado contribuir a la comprensión de las dinámicas de interacción en redes sociales y el impacto de los medios de comunicación masivos en contextos políticos. A través del uso de técnicas avanzadas de machine learning y análisis de grafos, hemos explorado en profundidad las características, reacciones y dinámicas de los usuarios y sus perspectivas ideológicas.

Los resultados aquí expuestos refuerzan la necesidad de abordar de manera crítica las interacciones online, ofreciendo una perspectiva sobre cómo los medios de comunicación,

las redes sociales y los algoritmos de recomendación de contenido influyen en las opiniones públicas.

Bibliografía

- [1] Alan I Abramowitz and Kyle L Saunders. Is polarization a myth? *The Journal of Politics*, 70(2):542–555, 2008.
- [2] Lada A. Adamic and Eytan Adar. Friends and neighbors on the web. *Social Networks*, 25(3):211–230, 2003. ISSN 0378-8733. doi: [https://doi.org/10.1016/S0378-8733\(03\)00009-1](https://doi.org/10.1016/S0378-8733(03)00009-1). URL <https://www.sciencedirect.com/science/article/pii/S0378873303000091>.
- [3] Qurat Tul Ain, Mubashir Ali, Amna Riaz, Amna Noureen, Muhammad Kamran, Babar Hayat, and A Rehman. Sentiment analysis using deep learning techniques: a review. *International Journal of Advanced Computer Science and Applications*, 8 (6), 2017.
- [4] Nadeem Akhtar and MM Beg. User graph topic model. *Journal of Intelligent & Fuzzy Systems*, 36(3):2229–2240, 2019.
- [5] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki. Link Prediction Using Supervised Learning. In *Proc. of SDM '06: Workshop on Link Analysis, Counter-Terrorism and Security*, volume 30, pages 798–805, 2006.
- [6] Ghaidaa A Al-Sultany and Hiba J Aleqabie. Events tagging in twitter using twitter latent dirichlet allocation. *International Journal of Engineering & Technology*, 8 (1.5):503–508, 2019.
- [7] Federico Albanese and Esteban Feuerstein. Improved topic modeling in twitter through community pooling. In *International Symposium on String Processing and Information Retrieval*, pages 209–216. Springer, 2021.
- [8] Federico Albanese, Leandro Lombardi, Esteban Feuerstein, and Pablo Balenzuela. Predicting shifting individuals using text mining and graph machine learning on twitter. *arXiv preprint arXiv:2008.10749*, 2020.
- [9] Federico Albanese, Sebastián Pinto, Viktoriya Semeshenko, and Pablo Balenzuela. Analyzing mass media influence using natural language processing and time series analysis. *Journal of Physics: Complexity*, 1(2):025005, 2020.
- [10] Federico Albanese, Esteban Feuerstein, Gabriel Kessler, and Juan Manuel Ortiz de Zárate. Aprendizaje automático para el análisis crossplataforma de la comunicación política: Gobierno y oposición argentinos en facebook, instagram y twitter. *Cuadernos. info*, (55):256–280, 2023.
- [11] Federico Albanese, Esteban Feuerstein, Leandro Lombardi, and Pablo Balenzuela. Characterizing community changing users using text mining and graph machine learning on twitter. In *AMW*, 2023.

- [12] Federico Albanese, Esteban Feuerstein, and Pablo Balenzuela. Polarization dynamics: A study of individuals shifting between political communities on social media. *Journal of Physics: Complexity*, 2024. URL <http://iopscience.iop.org/article/10.1088/2632-072X/ad679d>.
- [13] Abdallah El Ali, Tim C Stratmann, Souncil Park, Johannes Schöning, Wilko Heuten, and Susanne CJ Boll. Measuring, understanding, and classifying news media sympathy on twitter after crisis events. *arXiv preprint arXiv:1801.05802*, 2018.
- [14] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–236, 2017.
- [15] Eiman Alothali, Nazar Zaki, Elfadil A Mohamed, and Hany Alashwal. Detecting social bots on twitter: a literature review. In *2018 International conference on innovations in information technology (IIT)*, pages 175–180. IEEE, 2018.
- [16] André Altmann, Laura Tološi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 2010.
- [17] David Alvarez-Melis and Martin Saveski. Topic modeling in twitter: Aggregating tweets by conversations. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, 2016.
- [18] Giambattista Amati, Simone Angelini, Francesca Capri, Giorgio Gambosi, Gianluca Rossi, and Paola Vocca. On the retweet decay of the evolutionary retweet graph. In *International Conference on Smart Objects and Technologies for Social Good*, pages 243–253. Springer, 2016.
- [19] Jisun An, Haewoon Kwak, Oliver Posegga, and Andreas Jungherr. Political discussions in homogeneous and cross-cutting communication spaces. In *AAAI ICWSM*, 2019.
- [20] Dimo Angelov. Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*, 2020.
- [21] Stephen Ansolabehere and Shanto Iyengar. Riding the wave and claiming ownership over issues: The joint effects of advertising and news coverage in campaigns. *Public Opinion Quarterly*, 58(3):335–357, 1994.
- [22] N Aruguete. El poder de la agenda. política, medios y público. argentina: Biblos. *Cuadernos de Comunicación*, 2015.
- [23] Natalia Aruguete and Ernesto Calvo. Time to# protest: Selective exposure, cascading activation, and framing in social media. *Journal of communication*, 68(3):480–502, 2018.
- [24] Younggue Bae and Hongchul Lee. Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. *Journal of the American Society for Information Science and Technology*, 63(12):2521–2535, 2012.
- [25] Xue Bai. Predicting consumer sentiments from online text. *Decision Support Systems*, 50(4):732–742, 2011.

-
- [26] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 2015.
 - [27] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *ArXiv preprint*, abs/10.1126, 2010. URL <https://arxiv.org/abs/10.1126>.
 - [28] Hanna Barakat and Elissa M Redmiles. Community under surveillance: Impacts of marginalization on an online labor forum. In *AAAI ICWSM*, 2022.
 - [29] Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26(10):1531–1542, 2015.
 - [30] Jennifer Bast. Managing the image. the visual communication strategy of european right-wing populist politicians on instagram. *Journal of Political Marketing*, pages 1–30, 2021.
 - [31] Fabian Baumann, Philipp Lorenz-Spreen, Igor M Sokolov, and Michele Starnini. Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, 124(4):048301, 2020.
 - [32] Lion Behrens and Pirmin Stöckle. Crowdtangle api.
 - [33] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.
 - [34] Timothy Besley and Robin Burgess. The political economy of government responsiveness: Theory and evidence from india. *The Quarterly Journal of Economics*, 117(4):1415–1451, 2002.
 - [35] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
 - [36] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
 - [37] Javier Borge-Holthoefer, Walid Magdy, Kareem Darwish, and Ingmar Weber. Content and network dynamics behind egyptian political polarization on twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 700–711, 2015.
 - [38] Michael Bossetta. The digital architectures of social media: Comparing political campaigning on facebook, twitter, instagram, and snapchat in the 2016 us election. *Journalism & mass communication quarterly*, 95(2):471–496, 2018.
 - [39] Alexandre Bovet and Hernán A Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, 2019.
 - [40] Amber E Boydston, Shaun Bevan, and Herschel F Thomas. The importance of attention diversity and how to measure it. *Policy Studies Journal*, 42(2):173–196, 2014.

- [41] Craig Leonard Brians and Martin P Wattenberg. Campaign issue knowledge and salience: Comparing reception from tv commercials, tv news and newspapers. *American Journal of Political Science*, pages 172–193, 1996.
- [42] Sergey Brin and Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1):107–117, 1998. ISSN 0169-7552. doi: [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X). URL <https://www.sciencedirect.com/science/article/pii/S016975529800110X>.
- [43] Remy Cazabet and Giulio Rossetti. Challenges in community discovery on temporal networks. In *Temporal Network Theory*, pages 181–197. Springer, 2019.
- [44] Andrew Chadwick, James Dennis, and Amy P Smith. Politics in the age of hybrid media: Power, systems, and media logics. In *The Routledge companion to social media and politics*, pages 7–22. Routledge, 2015.
- [45] Jinyin Chen, Yangyang Wu, Xuanheng Xu, Yixian Chen, Haibin Zheng, and Qi Xuan. Fast Gradient Attack on Network Embedding. *arXiv preprint arXiv:1809.02797*, 2018.
- [46] Jinyin Chen, Xiang Lin, Ziqiang Shi, and Yi Liu. Link Prediction Adversarial Attack via Iterative Gradient Attack. *IEEE Transactions on Computational Social Systems*, 7(4):1081–1094, 2020.
- [47] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [48] Ziheng Chen, Fabrizio Silvestri, Gabriele Tolomei, Jia Wang, He Zhu, and Hongshik Ahn. Explain the Explainer: Interpreting Model-Agnostic Counterfactual Explanations of a Deep Reinforcement Learning Agent. *IEEE Transactions on Artificial Intelligence*, pages 1–15, 2022. doi: 10.1109/TAI.2022.3223892.
- [49] Ziheng Chen, Fabrizio Silvestri, Jia Wang, Yongfeng Zhang, Zhenhua Huang, Hongshik Ahn, and Gabriele Tolomei. GREASE: Generate Factual and Counterfactual Explanations for GNN-based Recommendations. *ArXiv preprint*, abs/2208.04222, 2022. URL <https://arxiv.org/abs/2208.04222>.
- [50] Ziheng Chen, Fabrizio Silvestri, Jia Wang, He Zhu, Hongshik Ahn, and Gabriele Tolomei. ReLAX: Reinforcement Learning Agent Explainer for Arbitrary Predictive Models. In *Proc. of CIKM '22*, pages 252–261. ACM, 2022.
- [51] Ziheng Chen, Fabrizio Silvestri, Jia Wang, Yongfeng Zhang, and Gabriele Tolomei. The Dark Side of Explanations: Poisoning Recommender Systems with Counterfactual Examples. In *Proc. of SIGIR '23*, pages 2426–2430. ACM, 2023.
- [52] Daejin Choi, Selin Chun, Hyunchul Oh, Jinyoung Han, et al. Rumor propagation is amplified by echo chambers in social media. *Scientific reports*, 10(1):1–10, 2020.
- [53] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), 2021.

- [54] Raviv Cohen and Derek Ruths. Classifying political orientation on twitter: It's not easy! In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, 2013.
- [55] Michael D Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. In *Fifth international AAAI conference on weblogs and social media*, 2011.
- [56] Enyan Dai, Tianxiang Zhao, Huaisheng Zhu, Junjie Xu, Zhimeng Guo, Hui Liu, Jiliang Tang, and Suhang Wang. A Comprehensive Survey on Trustworthy Graph Neural Networks: Privacy, Robustness, Fairness, and Explainability. *ArXiv preprint*, abs/2204.08570, 2022. URL <https://arxiv.org/abs/2204.08570>.
- [57] Jiazhu Dai, Weifeng Zhu, and Xiangfeng Luo. A Targeted Universal Attack on Graph Convolutional Network by Using Fake Nodes. *Neural Processing Letters*, pages 1–17, 2022.
- [58] Alexander Dallmann, Florian Lemmerich, Daniel Zoller, and Andreas Hotho. Media bias in german online newspapers. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 133–137. ACM, 2015.
- [59] Russell J Dalton. Generational change in elite political beliefs: The growth of ideological polarization. *The Journal of Politics*, 49(4):976–997, 1987.
- [60] DataReportal. ¿cuáles fueron los 10 sitios web más visitados de argentina en 2022? *Ambito Financiero*. URL <https://www.ambito.com/lifestyle/cuales-fueron-los-10-sitios-web-mas-visitados-argentina-2022-n5661750>.
- [61] Instituto Verificador de Circulaciones. <http://www.ivc.org.ar>.
- [62] Claes H De Vreese and Hajo G Boomgaarden. Media effects on public opinion about the enlargement of the european union. *JCMS: Journal of Common Market Studies*, 44(2):419–436, 2006.
- [63] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.
- [64] Kiran S Doddi, Mrs YV Haribhakta, and Parag Kulkarni. Sentiment classification of news article. *Diss. College of Engineering Pune*, 2014.
- [65] Paul S Earle, Daniel Bowden, and Michelle Guy. Twitter earthquake detection: earthquake monitoring in a social world. *Annals of geophysics*, 54(6):708–715, 2011.
- [66] David Easley, Jon Kleinberg, et al. *Networks, crowds, and markets: Reasoning about a highly connected world*, volume 1. Cambridge university press Cambridge, 2010.
- [67] Laura Edelson, Minh-Kha Nguyen, Ian Goldstein, Oana Goga, Damon McCoy, and Tobias Lauinger. Understanding engagement with us (mis) information news sources on facebook. In *ACM IMC*, 2021.

- [68] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.
- [69] Bradley Efron et al. Second thoughts on the bootstrap. *Statistical Science*, 18(2):135–140, 2003.
- [70] Robert M Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. Partisanship, propaganda, and disinformation: Online media and the 2016 us presidential election. *berkman klein center for internet & society*, 2017.
- [71] Ethan Fast, Binbin Chen, and Michael S Bernstein. Empath: Understanding topic signals in large-scale text. In *ACM CHI*, 2016.
- [72] Juan C Figuereo-Benítez, Fidel González-Quiñones, and Juan D Machin-Mastromatteo. Instagram como objeto de estudio en investigaciones recientes. una revisión de literatura con enfoque en revistas científicas. *Ámbitos. Revista Internacional De Comunicación*, (53):9–23, 2021.
- [73] John A Fortunato and Shannon E Martin. The intersection of agenda-setting, the media environment, and election campaign laws. *Journal of Information Policy*, 6:129–153, 2016.
- [74] R Kelly Garrett. Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication*, 2009.
- [75] R Kelly Garrett, Shira Dvir Gvirsman, Benjamin K Johnson, Yariv Tsfati, Rachel Neo, and Aysenur Dal. Implications of pro-and counterattitudinal information exposure for affective polarization. *Human communication research*, 40(3):309–332, 2014.
- [76] Seymour Geisser. *Predictive inference*. Chapman and Hall/CRC, 2017.
- [77] Alan S Gerber, Dean Karlan, and Daniel Bergan. Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52, 2009.
- [78] Martin Gerlach, Hanyu Shi, and Luís A Nunes Amaral. A universal information theoretic approach to the identification of stopwords. *Nature Machine Intelligence*, 1(12):606–612, 2019.
- [79] Malcom Gethers and Denys Poshyvanyk. Using relational topic models to capture coupling among classes in object-oriented software systems. In *2010 IEEE International Conference on Software Maintenance*, pages 1–10. IEEE, 2010.
- [80] Salvatore Giorgi, Daniel Preotiuc-Pietro, Anneke Buffone, Daniel Rieman, Lyle H Ungar, and H Andrew Schwartz. The remarkable benefit of user-level aggregation for lexical-based population-level predictions. *arXiv preprint arXiv:1808.09600*, 2018.
- [81] Juan Carlos González-Avella, Victor M Eguíluz, Mario G Cosenza, Konstantin Klemm, Jose L Herrera, and Maxi San Miguel. Local versus global interactions in nonequilibrium transitions: A model of social dynamics. *Physical Review E*, 73(4):046119, 2006.

-
- [82] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and Harnessing Adversarial Examples. In *Proc. of ICLR '15*, 2015. URL <http://arxiv.org/abs/1412.6572>.
 - [83] Yuriy Gorodnichenko, Tho Pham, and Oleksandr Talavera. Social media, sentiment and public opinions: Evidence from # brexit and # uselection. Technical report, National Bureau of Economic Research, 2018.
 - [84] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
 - [85] Kathrin Grosse, Nicolas Papernot, Praveen Manoharan, Michael Backes, and Patrick McDaniel. Adversarial Examples for Malware Detection. In *Proc. of ESORICS '17*, pages 62–79. Springer International Publishing, 2017. ISBN 978-3-319-66399-9.
 - [86] Anatoliy Gruzd and Jeffrey Roy. Investigating political polarization on twitter: A canadian perspective. *Policy & internet*, 6(1):28–45, 2014.
 - [87] Lauren Guggenheim, S Mo Jang, Soo Young Bae, and W Russell Neuman. The dynamics of issue frame competition in traditional and social media. *The ANNALS of the American Academy of Political and Social Science*, 659(1):207–224, 2015.
 - [88] Ruocheng Guo, Lu Cheng, Jundong Li, P Richard Hahn, and Huan Liu. A survey of learning causality with data: Problems and methods. *ACM Computing Surveys*, 2020.
 - [89] Malek Hajjem and Chiraz Latiri. Combining ir and lda topic modeling for filtering microblogs. *Procedia Computer Science*, 112:761–770, 2017.
 - [90] Uwe Hasebrink and Andreas Hepp. How to research cross-media practices? investigating media repertoires and media ensembles. *Convergence*, 23(4):362–377, 2017.
 - [91] Tim Hesterberg. Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(6):497–526, 2011.
 - [92] Joanne Hinds, Emma J Williams, and Adam N Joinson. “it wouldn’t happen to me”: Privacy concerns and perspectives following the cambridge analytica scandal. *International Journal of Human-Computer Studies*, 143:102498, 2020.
 - [93] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
 - [94] John L Hodges. The significance probability of the smirnov two-sample test. *Arkiv för Matematik*, 3(5):469–486, 1958.
 - [95] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.
 - [96] Liangjie Hong and Brian D Davison. Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics*, pages 80–88, 2010.

- [97] Daniel J Hopkins, Eunji Kim, and Soojong Kim. Does newspaper coverage influence or reflect public perceptions of the economy? *Research & Politics*, 4(4):2053168017737900, 2017.
- [98] Yiqing Hua, Thomas Ristenpart, and Mor Naaman. Towards measuring adversarial twitter interactions against candidates in the us midterm elections. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 272–282, 2020.
- [99] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.
- [100] Rania Ibrahim, Ahmed Elbagoury, Mohamed S Kamel, and Fakhri Karray. Tools and approaches for topic detection from twitter streams: survey. *Knowledge and Information Systems*, 54:511–539, 2018.
- [101] Shanto Iyengar and Kyu S Hahn. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of communication*, 59(1):19–39, 2009.
- [102] Shanto Iyengar, Gaurav Sood, and Yphtach Lelkes. Affect, not ideology: A social identity perspective on polarization. *Public opinion quarterly*, 76(3):405–431, 2012.
- [103] Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood. The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, 2019.
- [104] Mathieu Jacomy, Tommaso Venturini, Sébastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9(6):e98679, 2014.
- [105] Kathleen Hall Jamieson and Joseph N Cappella. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press, 2008.
- [106] Glen Jeh and Jennifer Widom. SimRank: A Measure of Structural-Context Similarity. In *Proc. of KDD '02*, pages 538–543. ACM, 2002. ISBN 158113567X. doi: 10.1145/775047.775126. URL <https://doi.org/10.1145/775047.775126>.
- [107] J. Joyce. *Kullback-leibler divergence*, pages 720–722. 2011.
- [108] Qiao Kang, Xing Song, Xiaying Xin, Bing Chen, Yuanzhu Chen, Xudong Ye, and Baiyu Zhang. Machine learning-aided causal inference framework for environmental data analysis: a covid-19 case study. *Environmental Science & Technology*, 2021.
- [109] Noah Kaplan, David K Park, and Travis N Ridout. Dialogue in american political campaigns? an examination of issue convergence in candidate television advertising. *American Journal of Political Science*, 50(3):724–736, 2006.
- [110] Amir Karami, Morgan Lundy, Frank Webb, and Yogesh K Dwivedi. Twitter and research: A systematic literature review through text mining. *IEEE access*, 8:67698–67717, 2020.

-
- [111] Rune Karlsen and Bernard Enjolras. Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with twitter data. *The International Journal of Press/Politics*, 21(3):338–357, 2016.
 - [112] Leo Katz. A New Status Index Derived from Sociometric Analysis. *Psychometrika*, 18(1):39–43, 1953.
 - [113] Stanley Kelley and Thad W Mirer. The simple act of voting. *American Political Science Review*, 68(2):572–591, 1974.
 - [114] Gary King, Benjamin Schneer, and Ariel White. How the news media activate public expression and influence national agendas. *Science*, 358(6364):776–780, 2017.
 - [115] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *Proc. of ICLR '17*. OpenReview.net, 2017. URL <https://openreview.net/forum?id=SJU4ayYg1>.
 - [116] Ezra Klein. *Why we're polarized*. Simon and Schuster, 2020.
 - [117] Marina Kogan, Leysia Palen, and Kenneth M Anderson. Think local, retweet global: Retweeting by the geographically-vulnerable during hurricane sandy. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 981–993, 2015.
 - [118] Stephen Kokoska and Daniel Zwillinger. *CRC standard probability and statistics tables and formulae*. Crc Press, 2000.
 - [119] Olessia Koltsova and Sergei Koltcov. Mapping the public agenda with topic modeling: The case of the russian livejournal. *Policy & Internet*, 5(2):207–227, 2013.
 - [120] Damir Korenčić, Strahil Ristov, and Jan Šnajder. Getting the agenda right: measuring media agenda using topic models. In *Proceedings of the 2015 Workshop on Topic Models: Post-Processing and Applications*, pages 61–66, 2015.
 - [121] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.
 - [122] Emily Kubin and Christian Von Sikorski. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association*, 45(3):188–206, 2021.
 - [123] Ajay Kumar, Shashank Sheshar Singh, Kuldeep Singh, and Bhaskar Biswas. Link Prediction Techniques, Applications, and Performance: A Survey. *Physica A: Statistical Mechanics and its Applications*, 553:124289, 2020.
 - [124] Andrea Lancichinetti and Santo Fortunato. Consensus clustering in complex networks. *Scientific reports*, 2(1):1–7, 2012.
 - [125] Anders Olof Larsson. Green light for interaction: Party use of social media during the 2014 swedish election year. *First Monday*, 2015.

- [126] Richard R Lau, David J Andersen, Tessa M Ditonto, Mona S Kleinberg, and David P Redlawsk. Effect of media environment diversity and advertising tone on information search, selective exposure, and affective polarization. *Political Behavior*, 39:231–255, 2017.
- [127] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
- [128] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.
- [129] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [130] JooYoung Lee, Siqi Wu, Ali Mert Ertugrul, Yu-Ru Lin, and Lexing Xie. Whose advantage? measuring attention dynamics across youtube and twitter on controversial topics. In *AAAI ICWSM*, 2022.
- [131] Ann Lehman. *JMP for basic univariate and multivariate statistics: a step-by-step guide*. SAS Institute, 2005.
- [132] Seth C Lewis and Oscar Westlund. Actors, actants, audiences, and activities in cross-media news work: A matrix and a research agenda. *Digital journalism*, 3(1):19–37, 2015.
- [133] Boning Li, Yingce Xia, Shufang Xie, Lijun Wu, and Tao Qin. Distance-Enhanced Graph Neural Network for Link Prediction. In *Proc. of ICML '21: Workshop on Computational Biology*, 2021.
- [134] David Liben-Nowell and Jon Kleinberg. The Link Prediction Problem for Social Networks. In *Proc. of CIKM '03*, pages 556–559. ACM, 2003. ISBN 1581137230. doi: 10.1145/956863.956972. URL <https://doi.org/10.1145/956863.956972>.
- [135] Lucas Lima, Julio CS Reis, Philipe Melo, Fabricio Murai, Leandro Araujo, Pantelis Vikatos, and Fabricio Benevenuto. Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 515–522. IEEE, 2018.
- [136] W. Lin, S. Ji, and B. Li. Adversarial attacks on link prediction algorithms based on graph neural networks. In *Proceedings Of The 15th ACM Asia Conference On Computer And Communications Security*, pages 370–380, 2020.
- [137] Juliane A Lischka. What follows what? relations between economic indicators, economic expectations of the public, and news on the general economy and unemployment in germany, 2002-2011. *Journalism & Mass Communication Quarterly*, 92(2):374–398, 2015.
- [138] Linyuan Lü and Tao Zhou. Link prediction in Complex Networks: A Survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150–1170, 2011.

-
- [139] Ana Lucic, Maartje A Ter Hoeve, Gabriele Tolomei, Maarten De Rijke, and Fabrizio Silvestri. CF-GNNExplainer: Counterfactual Explanations for Graph Neural Networks. In *Proc. of AISTATS '22*, pages 4499–4511. PMLR, 2022.
 - [140] Tinghuai Ma, Jing Li, Xinnian Liang, Yuan Tian, Abdullah Al-Dhelaan, and Mohammed Al-Dhelaan. A time-series based aggregation scheme for topic detection in weibo short texts. *Physica A: Statistical Mechanics and its Applications*, 536:120972, 2019.
 - [141] Oded Z Maimon and Lior Rokach. *Data mining with decision trees: theory and applications*, volume 81. World scientific, 2014.
 - [142] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.
 - [143] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.
 - [144] James R Martin and Peter R White. *The language of evaluation*, volume 2. Springer, 2003.
 - [145] Trevor Martin. community2vec: Vector representations of online communities encode semantic relationships. In *NLP+CSS at ACL*, 2017.
 - [146] Víctor Martínez, Fernando Berzal, and Juan-Carlos Cubero. A Survey of Link Prediction in Complex Networks. *ACM Comput. Surv.*, 49(4):1–33, 2016. ISSN 0360-0300. doi: 10.1145/3012704. URL <https://doi.org/10.1145/3012704>.
 - [147] Maxwell McCombs. A look at agenda-setting: Past, present and future. *Journalism studies*, 6(4):543–557, 2005.
 - [148] Maxwell E McCombs and Donald L Shaw. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972.
 - [149] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.
 - [150] Rishabh Mehrotra, Scott Sanner, Wray Buntine, and Lexing Xie. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 889–892, 2013.
 - [151] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
 - [152] Alfredo Jose Morales, Javier Borondo, Juan Carlos Losada, and Rosa M Benito. Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):033114, 2015.

- [153] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M Carley. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In *Seventh international AAAI conference on weblogs and social media*, 2013.
- [154] Aminu Muhammad, Nirmalie Wiratunga, and Robert Lothian. Contextual sentiment analysis for social media genres. *Knowledge-based systems*, 108:92–101, 2016.
- [155] Andreas C Müller and Sarah Guido. *Introduction to machine learning with Python: a guide for data scientists*. .”O'Reilly Media, Inc.”, 2016.
- [156] Diana C Mutz. The consequences of cross-cutting networks for political participation. *American journal of political science*, pages 838–855, 2002.
- [157] Diana C Mutz. *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press, 2006.
- [158] Mark EJ Newman. Clustering and Preferential Attachment in Growing Networks. *Physical review E*, 64(2):025102, 2001.
- [159] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
- [160] Didrik Nielsen. Tree boosting with xgboost—why does xgboost win “every” machine learning competition? Master's thesis, NTNU, 2016.
- [161] Rasmus Kleis Nielsen and Kim Christian Schrøder. The relative importance of social media for accessing, finding, and engaging with news: An eight-country cross-media comparison. *Digital journalism*, 2(4):472–489, 2014.
- [162] Brendan Nyhan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, et al. Like-minded sources on facebook are prevalent but not polarizing. *Nature*, 620(7972):137–144, 2023.
- [163] Felix Oberholzer-Gee and Joel Waldfogel. Media markets and localism: Does local news en espanol boost hispanic voter turnout? *The American economic review*, 99(5):2120–2128, 2009.
- [164] Anaïs Ollagnier and Hywel Williams. Network-based pooling for topic modeling on microblog content. In *International Symposium on String Processing and Information Retrieval*, pages 80–87. Springer, 2019.
- [165] Abi Oppenheim, Federico Albanese, and Esteban Feuerstein. Investigating the relationship between user specialization and toxicity on reddit: A sentiment analysis approach. *arXiv preprint arXiv:2406.03443*, 2024.
- [166] Brian L Ott. The age of twitter: Donald j. trump and the politics of debasement. *Critical studies in media communication*, 34(1):59–68, 2017.
- [167] Diana Owen. New media and political campaigns. 2014.
- [168] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

-
- [169] Michael Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 2011.
 - [170] Judea Pearl. *Causality*. Cambridge university press, 2009.
 - [171] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
 - [172] Tiago P Peixoto. Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4(1):011047, 2014.
 - [173] Tiago P Peixoto. Bayesian stochastic blockmodeling. *Advances in network clustering and blockmodeling*, pages 289–332, 2019.
 - [174] Chau Minh Pham, Alexander Hoyle, Simeng Sun, Philip Resnik, and Mohit Iyyer. Topiccpt: A prompt-based topic modeling framework. *arXiv preprint arXiv:2311.01449*, 2023.
 - [175] Sebastián Pinto, Pablo Balenzuela, and Claudio O Dorso. Setting the agenda: Different strategies of a mass media in a model of cultural dissemination. *Physica A: Statistical Mechanics and its Applications*, 458:378–390, 2016.
 - [176] Sebastián Pinto, Federico Albanese, Claudio O Dorso, and Pablo Balenzuela. Quantifying time-dependent media agenda and public opinion by topic modeling. *Physica A: Statistical Mechanics and its Applications*, 524:614–624, 2019.
 - [177] Sebastián Pinto, Federico Albanese, Claudio O. Dorso, and Pablo Balenzuela. Quantifying time-dependent media agenda and public opinion by topic modeling. *Physica A: Statistical Mechanics and its Applications*, 524:614 – 624, 2019. ISSN 0378-4371. doi: <https://doi.org/10.1016/j.physa.2019.04.108>. URL <http://www.sciencedirect.com/science/article/pii/S0378437119304844>.
 - [178] Real Clear Politics. <https://realclearpolitics.com/>. URL <https://realclearpolitics.com/>.
 - [179] Guang Qiu, Xiaofei He, Feng Zhang, Yuan Shi, Jiajun Bu, and Chun Chen. Dasa: dissatisfaction-oriented advertising based on sentiment analysis. *Expert Systems with Applications*, 37(9):6182–6191, 2010.
 - [180] Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. Echo chambers on facebook. *Available at SSRN 2795110*, 2016.
 - [181] Mauricio Quezada and Barbara Poblete. A lightweight representation of news events on social media. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1049–1052, 2019.
 - [182] P. Raj and P. David. *The digital twin paradigm for smarter systems and environments: The industry use cases*. Academic Press, 2020.

- [183] K. Ramachandran and C. Tsokos. *Mathematical statistics with applications in R*. Academic Press, 2020.
- [184] Juan Ramos et al. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, volume 242, pages 133–142, 2003.
- [185] Bhavtosh Rath, Wei Gao, Jing Ma, and Jaideep Srivastava. From retweet to believability: Utilizing trust to identify rumor spreaders on twitter. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 179–186, 2017.
- [186] Marnie E Rice and Grant T Harris. Comparing effect sizes in follow-up studies: Roc area, cohen's d, and r. *Law and human behavior*, 29(5):615–620, 2005.
- [187] CJ van Rijsbergen. *Information retrieval*. Butterworth-Heinemann, 1979.
- [188] Michael Röder, Andreas Both, and Alexander Hinneburg. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408, 2015.
- [189] Richard Rogers et al. Digital methods for cross-platform analysis. *The SAGE handbook of social media*, pages 91–110, 2017.
- [190] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [191] W Russell Neuman, Lauren Guggenheim, S Mo Jang, and Soo Young Bae. The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64(2):193–214, 2014.
- [192] Dietram A Scheufele. Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass communication & society*, 3(2-3):297–316, 2000.
- [193] Dietram A Scheufele and Shanto Iyengar. The state of framing research: A call for new directions. 2014.
- [194] Hinrich Schütze, Christopher D Manning, and Prabhakar Raghavan. *Introduction to information retrieval*, volume 39. Cambridge University Press Cambridge, 2008.
- [195] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
- [196] Claude E Shannon. Xxi. programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 41(314):256–275, 1950.
- [197] Amit Sharma, Vasilis Syrgkanis, Cheng Zhang, and Emre Kiciman. Dowhy: Addressing challenges in expressing and validating causal assumptions. *arXiv preprint arXiv:2108.13518*, 2021.

- [198] Yasufumi Shibanai, Satoko Yasuno, and Itaru Ishiguro. Effects of global information feedback on diversity: extensions to axelrod's adaptive culture model. *Journal of Conflict Resolution*, 45(1):80–96, 2001.
- [199] Kabir Manandhar Shrestha, Katie Wood, David Goodman, and Meladel Mistica. Do we need subject matter experts? a case study of measuring up gpt-4 against scholars in topic evaluation. In *NL4AI@ AI* IA*, 2023.
- [200] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 3. Springer, 2000.
- [201] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.
- [202] Stuart Soroka, Mark Daku, Dan Hiaeshutter-Rice, Lauren Guggenheim, and Josh Pasek. Negativity and positivity biases in economic news coverage: Traditional versus social media. *Communication Research*, page 0093650217725870, 2017.
- [203] Stuart N Soroka, Dominik A Stecula, and Christopher Wlezien. It's (change in) the (future) economy, stupid: economic indicators, the media, and public opinion. *American Journal of Political Science*, 59(2):457–474, 2015.
- [204] Niels Spierings and Kristof Jacobs. Political parties and social media campaigning: A qualitative comparative analysis of parties' professional facebook and twitter use in the 2010 and 2012 dutch elections. *Acta Politica*, 54:145–173, 2019.
- [205] Leo G Stewart, Ahmer Arif, and Kate Starbird. Examining trolls and polarization with a retweet network. In *Proc. ACM WSDM, workshop on misinformation and misbehavior mining on the web*, volume 70, 2018.
- [206] Sebastian Stier, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. Election campaigning on social media: Politicians, audiences, and the mediation of political communication on facebook and twitter. In *Studying Politics Across Media*, pages 50–74. Routledge, 2020.
- [207] Yiwei Sun, Suhang Wang, Xianfeng Tang, Tsung-Yu Hsieh, and Vasant Honavar. Adversarial Attacks on Graph Neural Networks via Node Injections: A Hierarchical Reinforcement Learning Approach. In *Proc. of TheWebConf '20*, pages 673–683, 2020.
- [208] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.
- [209] Anne Mette Thorhauge and Stine Lomborg. Cross-media communication in context: A mixed-methods approach. *MedieKultur: Journal of media and communication research*, 32(60):16–p, 2016.
- [210] Michael E Tipping and Christopher M Bishop. Mixtures of probabilistic principal component analyzers. *Neural computation*, 11(2):443–482, 1999.

- [211] Christopher K Tokita, Andrew M Guess, and Corina E Tarnita. Polarized information ecosystems can reorganize social networks via information cascades. *Proceedings of the National Academy of Sciences*, 118(50):e2102147118, 2021.
- [212] Gabriele Tolomei and Fabrizio Silvestri. Generating Actionable Interpretations from Ensembles of Decision Trees. *IEEE Transactions on Knowledge Data Engineering*, 33(4):1540–1553, 2021.
- [213] Gabriele Tolomei, Fabrizio Silvestri, Andrew Haines, and Mounia Lalmas. Interpretable Predictions of Tree-based Ensembles via Actionable Feature Tweaking. In *Proc. of KDD '17*, pages 465–474. ACM, 2017.
- [214] G. Trappolini, V. Maiorca, S. Severino, E. Rodola, F. Silvestri, and G. Tolomei. Sparse vicious attacks on graph neural networks. *IEEE Transactions on Artificial Intelligence*, 2023.
- [215] R. Tripodi and M. Pelillo. *Transductive Learning Games for Word Sense Disambiguation*, pages 109–128. 2017.
- [216] Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the international AAAI conference on web and social media*, volume 4, pages 178–185, 2010.
- [217] Alan M Turing. *Computing machinery and intelligence*. Springer, 2009.
- [218] Markus Wagner. Affective polarization in multiparty systems. *Electoral Studies*, 69: 102199, 2021.
- [219] Isaac Waller and Ashton Anderson. Quantifying social organization and political polarization in online platforms. *Nature*, 2021.
- [220] Jihong Wang, Minnan Luo, Fnu Suya, Jundong Li, Zijiang Yang, and Qinghua Zheng. Scalable Attack on Graph Data by Injecting Vicious Nodes. *Data Mining and Knowledge Discovery*, 34(5):1363–1389, 2020.
- [221] Xiaoyun Wang, Minhao Cheng, Joe Eaton, Cho-Jui Hsieh, and Felix Wu. Attack Graph Convolutional Networks by Adding Fake Nodes. *ArXiv preprint*, abs/1810.10751, 2018. URL <https://arxiv.org/abs/1810.10751>.
- [222] Galen Weld, Maria Glenski, and Tim Althoff. Political bias and factualness in news sharing across more than 100,000 online communities. In *AAAI ICWSM*, 2021.
- [223] Siqi Wu and Paul Resnick. Cross-partisan discussions on youtube: Conservatives talk to liberals but liberals don't talk to conservatives. In *AAAI ICWSM*, 2021.
- [224] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2020.
- [225] Ellery Wulczyn, Nithum Thain, and Lucas Dixon. Ex machina: Personal attacks seen at scale. In *Proceedings of the 26th international conference on world wide web*, pages 1391–1399, 2017.

-
- [226] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 267–273. ACM, 2003.
 - [227] Min-Chul Yang, Jung-Tae Lee, Seung-Wook Lee, and Hae-Chang Rim. Finding interesting posts in twitter based on retweet graph analysis. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 1073–1074, 2012.
 - [228] Zhao Yang, René Algesheimer, and Claudio J Tessone. A comparative analysis of community detection algorithms on artificial networks. *Scientific reports*, 6(1):1–18, 2016.
 - [229] Taha Yasseri and Jonathan Bright. Wikipedia traffic data and electoral prediction: towards theoretically informed models. *EPJ Data Science*, 5(1):1–15, 2016.
 - [230] Lei Zhang, Shuai Wang, and Bing Liu. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253, 2018.
 - [231] Muhan Zhang and Yixin Chen. Link Prediction Based on Graph Neural Networks. In *Proc. of NeurIPS '18*, pages 5171–5181, 2018. URL <https://proceedings.neurips.cc/paper/2018/hash/53f0d7c537d99b3824f0f99d62ea2428-Abstract.html>.
 - [232] Tao Zhou, Linyuan Lü, and Yi-Cheng Zhang. Predicting Missing Links via Local Information. *The European Physical Journal B*, 71(4):623–630, 2009.
 - [233] María Victoria Zolezzi, Federico Albanese, and Esteban Feuerstein. Characterizing community structures on social media over time: a graph learning approach. In *Simposio Argentino de Ciencia de Datos y GRANdes DAtos (AGRANDA 2023)-JAIIo 52 (Universidad Nacional de Tres de Febrero, 4 al 8 de septiembre de 2023)*, 2023.
 - [234] Daniel Zügner, Amir Akbarnejad, and Stephan Günnemann. Adversarial Attacks on Neural Networks for Graph Data. In *Proc. of KDD '18*, pages 2847–2856. ACM, 2018. doi: 10.1145/3219819.3220078. URL <https://doi.org/10.1145/3219819.3220078>.
 - [235] Esteban Zunino. The assessment of political news in the media agenda: a methodological proposal for more extensive content analysis. *Communication & Society*, 29(4):235–253, 2016.
 - [236] Esteban Zunino and Natalia Aruguete. La cobertura mediática del conflicto campo-gobierno. un estudio de caso. *Global Media Journal*, 7(14), 2010.