

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



**“FAKE NEWS Y VERIFICADORES: UN ANÁLISIS AL
IMPACTO DE LA DESINFORMACIÓN Y A LA RESPUESTA
DE LOS MEDIOS PERIODÍSTICOS VERIFICADORES EN
CHILE.”**

SEBASTIÁN ALEJANDRO CAMPOS MUÑOZ

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA**

Profesor Guía: Marcelo Mendoza
Profesor Correferente: Ricardo Ñanculef

Agosto - 2023

DEDICATORIA

Para mi familia, mis amigos y en especial para mi viejo, que quizás ya no puede verme terminar esta etapa de mi vida, pero si por alguna de las vicisitudes del universo puedes leer esto, quiero que sepas que tu hijo lo logró.

AGRADECIMIENTOS

RESUMEN

Resumen— La expansión de medios digitales ha impulsado las *Fake News*, un fenómeno global arraigado en la vida cotidiana y agravado en Chile durante crisis como el estallido social de 2019 y la pandemia Covid-19. En este contexto, este estudio se enfoca en entender la desinformación en Chile y analizar cómo la abordan los medios verificadores del país. Para lograrlo, se construyó un *dataset* que recopiló el trabajo de tres medios de verificación chilenos. Se enriqueció con información contextual y estandarizó las etiquetas de verificación entre los medios. Además, se construyeron y analizaron los árboles de propagación de contenido verificado compartido en Twitter, brindando información sobre respuestas y retweets asociados a la desinformación. Finalmente, se obtuvieron métricas que revelaron datos esenciales, como tiempos de respuesta de las verificaciones, dinámicas de propagación, presencia de bots, emocionalidad en las discusiones y diferencias en la viralidad entre verificación y contenido verificado.

Palabras Clave— Desinformación, Medios verificadores, Fact Check, Fake News Chile

ABSTRACT

Abstract— The expansion of digital media has fueled the proliferation of "Fake News," a global phenomenon ingrained in everyday life and exacerbated in Chile during crises such as the 2019 social outbreak and the Covid-19 pandemic. In this context, this study focuses on comprehending misinformation in Chile and analyzing how the country's fact-checking media address it. To achieve this, a dataset was constructed, compiling the work of three Chilean fact-checking media outlets. It was enriched with contextual information, and verification labels were standardized across the media. Additionally, the propagation trees of verified content shared on Twitter were constructed and analyzed, providing insights into responses and retweets associated with misinformation. Finally, metrics were obtained that revealed crucial data, including verification response times, propagation dynamics, presence of bots, emotional aspects in discussions, and disparities in virality between verification and verified content.

Keywords— Misinformation, Fact-checking media , Fact Check, Fake News Chile

GLOSARIO

RRSS : Redes Sociales, API : Application Programing Interface

ÍNDICE DE CONTENIDOS

RESUMEN	IV
ABSTRACT	IV
GLOSARIO	VI
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABLAS	IX
INTRODUCCIÓN	1
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	3
1.1 Contexto social y situación actual	3
1.2 Importancia de los <i>bots</i>	4
1.3 Contexto e Impacto inicial del estudio	5
1.4 Objetivos	6
1.4.1 Objetivo General	6
1.4.2 Objetivos Específicos	6
CAPÍTULO 2: MARCO CONCEPTUAL	7
2.1 Definición de conceptos	7
2.2 Creación de <i>Datasets</i>	9
2.3 Acerca de las <i>Fake News</i>	9
2.4 Detección de <i>bots</i>	10
2.5 Propagación de la desinformación en redes sociales.	13
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	15
3.1 Creación del dataset.	15
3.2 Creación de árboles de propagaciones de Twitter.	18
3.3 Métricas	21
3.3.1 Métricas del dataset original.	21
3.3.2 Caso de estudio: Métricas de propagación (A partir de los árboles de propagación)	23
CAPÍTULO 4: RESULTADOS Y ANÁLISIS	24
4.1 Resultados	24
4.1.1 Tiempos de respuesta	24
4.1.2 Diferencia de Reacciones	25
4.1.3 Análisis de sentimiento	29
4.1.4 Participación de Bots	30
4.1.5 Resultados métricas del Caso de estudio: Árboles de propagación.	32

4.2	Análisis	35
4.2.1	Análisis resultados Tiempos de respuesta	35
4.2.2	Análisis resultados Diferencia de reacciones.	36
4.2.3	Análisis resultados Análisis de sentimiento.	37
4.2.4	Análisis resultados Bots	37
4.2.5	Análisis métricas Caso de estudio: Arboles de propagación.	38
CAPÍTULO 5: CONCLUSIONES		40
5.1	Trabajo Futuro	41
REFERENCIAS BIBLIOGRÁFICAS		43

ÍNDICE DE FIGURAS

1	Caras de la desinformación.	4
2	Árbol de definición del problema.	6
3	La imagen representa las 1150 características que se puede extraer utilizando el <i>framework</i> planteado.	12
4	La imagen representa un ejemplo de cascada con los valores de todas sus medidas a través del tiempo.	14
5	Evolución de la profundidad según tiempo promedio. La línea de color rojo punteada se encuentra en la profundidad 20	34
6	Evolución del tamaño de los arboles según tiempo promedio.	34
7	Evolución de la anchura máxima de los arboles según tiempo promedio.	35

ÍNDICE DE TABLAS

1	Cantidad de noticias según medio de verificación.	15
2	Traducción de etiquetas de verificación.	18
3	Nueva distribución de verificaciones con respecto a las nuevas categorías.	18
4	Cantidad de días promedio de cada medio verificador. Estas tablas consideran todas las publicaciones de cada medio	24
5	Cantidad de días promedio de cada medio verificador. Para el caso de Fastcheck se consideran las publicaciones con menor o igual a 13 días de retraso en la verificación, para el caso de decodificador son las publicaciones menor o iguales a 8 días y para Factcheck menor o iguales a 128 días.	25
6	Diferencia de <i>Likes</i> entre la desinformación compartida en Twitter y las publicaciones de los medios verificadores en Instagram.	26
7	Diferencia de comentarios entre la desinformación compartida en Twitter y las publicaciones de los medios verificadores en Instagram.	27
8	Diferencia de <i>Likes</i> entre la desinformación compartida en Instagram y las publicaciones de los medios verificadores en Instagram.	27

9	Diferencia de comentarios entre la desinformación compartida en Instagram y las publicaciones de los medios verificadores en Instagram.	28
10	Clasificación de sentimiento de publicaciones verificadas. En este caso, la herramienta utilizada es PySentimiento	29
11	Clasificación de sentimiento de publicaciones verificadas. En este caso, la herramienta utilizada es Spanish Analyser Spacy	30
12	Resumen de los resultados obtenidos utilizando la aplicación Botometer. En <i>header</i> de la primera columna se encuentra la certeza mínima usada para determinar si una cuenta es un bot o no.	31
13	Resumen de los resultados obtenidos utilizando la aplicación BotcheckCL. En <i>header</i> de la primera columna se encuentra la certeza mínima usada para determinar si una cuenta es un bot o no.	32
14	Profundidad en árbol de propagación de respuestas. KS-tests: falso y verdad: $D = 0,168, p \sim 0,08$, falso e impreciso: $D = 0,124, p \sim 0,70$, verdadero e impreciso: $D = 0,275, p \sim 0,01$	33
15	Tamaño en árbol de propagación de respuestas. KS-tests: falso y verdad: $D = 0,206, p \sim 0,01$, falso e impreciso: $D = 0,152, p \sim 0,45$, verdadero e impreciso: $D = 0,195, p \sim 0,17$	33
16	Anchura máxima en árbol de propagación de respuestas. KS-tests: falso y verdad: $D = 0,175, p \sim 0,06$, falso e impreciso: $D = 0,135, p \sim 0,61$, verdadero e impreciso: $D = 0,195, p \sim 0,17$	33

INTRODUCCIÓN

Las redes sociales han ganado cada vez más relevancia a nivel internacional como medios de consumo de información actual, presentándose como alternativas vigentes a los formatos más tradicionales, como la televisión y la radio. Sin embargo, este aumento en su papel en la vida cotidiana de la ciudadanía ha traído consigo un importante efecto secundario, debido a que estos nuevos medios digitales no siempre aplican rigurosidad en la exposición de información al público, lo que resulta en la difusión de noticias o afirmaciones que no siempre son completamente veraces o precisas. Esta falta de veracidad puede originarse de manera no intencionada o deliberada, pero en ambos casos, su impacto en la opinión pública puede ser significativo, ya que este tipo de contenido puede propagarse rápidamente y llevar a conclusiones erróneas entre los consumidores.

Es esencial destacar que se han realizado investigaciones sobre este fenómeno en contextos de alta relevancia social, como las elecciones presidenciales de Estados Unidos en 2016 [Allcott y Gentzkow, 2017], la pandemia de COVID-19 [Rocha *et al.*, 2021] y eventos bélicos como, por ejemplo, el conflicto entre Rusia y Ucrania [BABACAN y TAM, 2022]. Esta problemática no solo surge de la falta de control en la veracidad de las publicaciones en las redes sociales, sino que también se ha utilizado la desinformación como una herramienta para influir en la opinión pública y modificar percepciones en temas relevantes [Bastick, 2021].

Como contrapeso a esta tendencia, surgieron instituciones periodísticas especializadas en la verificación de información viral tanto en medios tradicionales como digitales. Su objetivo es mitigar el impacto de las noticias falsas en la sociedad. Este fenómeno también ha tenido eco en Chile, donde se han vivido eventos de gran relevancia social, como el estallido social en 2019, la pandemia en 2020, la propuesta de una nueva constitución y las elecciones presidenciales en 2021. Durante este período, medios verificadores chilenos como FastCheck, Decodificador y Factchecking UC tuvieron la tarea de enfrentar la avalancha de desinformación que inundó tanto los medios digitales como los tradicionales en el país.

Bajo el contexto anterior, se vio necesario realizar un estudio que utilizara todas las verificaciones realizadas por medios verificadores con el objetivo de entender de mejor manera el comportamiento de las *Fake News* en los medios digitales chilenos. Para lograr lo anterior, esta investigación se centró en la creación de un dataset que contiene verificaciones realizadas por medios nacionales, como Fastcheck, Decodificador y FactChecking UC, entre los años 2019 y 2021. A partir de estas verificaciones, se enriqueció el conjunto de datos con información adicional, como las URL de los contenidos verificados, el medio de publicación y las redes sociales del medio verificador. También se normalizaron las etiquetas de veracidad de cada medio, siguiendo investigaciones previas [Amazeen, 2015] [Lim, 2018] y el asesoramiento de expertos en el campo. El resultado es un *dataset* coherente y homogéneo en términos de etiquetas de veracidad.

Por otro lado, se procedió a extraer datos adicionales de todo el contenido compartido por

la red social Twitter. Estos datos adicionales se pueden resumir en tres componentes: Datos sobre la cuenta y el tweet publicado, todas las respuestas al tweet original y todos los retweets al tweet original. Este proceso se hizo con la finalidad de crear árboles de propagación [Vosoughi *et al.*, 2018] de los cuales se obtendrán las dinámicas de virilización de los tweets dependiendo de su veracidad. Estas dinámicas se obtuvieron en el contexto de un proyecto del IMFD (Millenium Institute Foundational Research on Data) [Mendoza *et al.*, 2023] por lo que en este documento se planteará este desarrollo como un caso de estudio posible gracias al dataset creado con anterioridad.

Finalmente, se obtuvieron cuatro métricas a partir del *dataset* que tienen el objetivo de obtener características del contenido verificado y de los medios verificadores relacionados. Estas métricas son: Tiempos de respuesta entre el medio verificador y la desinformación, Diferencia entre las reacciones y comentarios entre las publicaciones de los medios verificadores en redes sociales y la desinformación, análisis de sentimientos de los comentarios en las publicaciones en Twitter de la desinformación y la distribución de bots de la desinformación publicada en Twitter.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

1.1. Contexto social y situación actual

En la última década ha ocurrido un fenómeno que ha impactado profundamente en la percepción de la población chilena con respecto al mundo que los rodea, resulta que con la creciente disponibilidad de internet dentro de la demografía chilena y la disminución de confianza con respecto a los medios tradicionales, las redes sociales comenzaron a tomar un rol fundamental a la hora de ser medios en los cuales los ciudadanos buscan información respecto a la situación nivel país, transformándose en uno de los medios informativos más relevantes en la actualidad (superando en ciertos periodos de tiempo a la televisión o los diarios nacionales)[Nuñez y Fernández, 2020]. Las razones de este fenómeno son variadas, sobre todo si consideramos que esta situación no es un fenómeno meramente local sino que ha sido transversal en el occidente del mundo, en el caso específico de nuestro país, los principales detonantes son dos: La constante interacción e inmediatez de las plataformas digitales y la campaña social contra los medios tradicionales a partir del 18 de octubre del año 2019.

Una vez que las redes sociales se instauran en el colectivo como medio para informarse, comienza a ocurrir que la desinformación se convierte rápidamente en un factor relevante en la vida cotidiana del ciudadano chileno, debido a que no todo usuario proveedor de información resulta ser un usuario oficial y/o confiable a la hora de entregar noticias, por lo que esto desemboca a tener grandes cantidades de contenido verificado y no verificado que se propaga por igual dentro de estos nuevos medios digitales. Esta nueva forma de vivir la desinformación puede llegar a ser muy dañina a la hora de tener una visión objetiva sobre la realidad de hechos, situaciones, dichos, noticias, etc, debido a que la información falsa suele propagarse con notable rapidez dentro de sectores específicos, más aún si es que la información refuerza alguna creencia previa del sector en cuestión, generando como consecuencia un potencial para reafirmar prejuicios perjudicando, además, la labor del periodismo para intentar desvelar la verdad de la información en cuestión. Por otra parte, si un grupo está expuesto a hechos falsos que refuerzan sus creencias, este grupo será más escéptico frente a noticias periodísticas verídicas que contradigan estas creencias reforzadas, causando la percepción de que los medios oficiales están siendo injustos y sesgados respecto a los puntos de vista personales del grupo, esto ocurre debido que una vez es 'validado' personalmente un hecho que es acorde a la opinión del grupo, el grupo tiende a desestimar cualquier tipo de información que vaya en contra de las creencias establecidas[Ribeiro *et al.*, 2017] incluso si provienen de fuentes oficiales y/o públicamente confiables, lo cual provoca que estos grupos se vuelvan puntos de propagación de *Fake News* que estén de acuerdo con ellos.

1.2. Importancia de los bots

Por otra parte, uno de los factores más importantes a la hora de propagarse desinformación por las redes sociales es el uso de *bots*. Un *bot* es una pieza de software automatizada que en éste contexto se encarga de generar contenido y establecer interacciones con humanos en redes sociales, los *bot* no necesariamente tienen fines maliciosos ya que puede ser utilizado para difundir noticias de interés publico o coordinar actividades de voluntariado, sin embargo, actualmente se ha presentado un aumento de usos maliciosos de bots para influir en la opinión pública. Dentro de sus usos negativos se puede destacar: la creación de falso apoyo político a entidades en redes sociales, propagar propaganda terrorista, manipular el mercado de acciones, propagar rumores relacionados a teorías conspirativas y propagación de noticias falsas [Varol *et al.*, 2017]. De esta forma, los *bots* se vuelven un factor extremadamente relevante a la hora de analizar la propagación de la desinformación en redes sociales, siendo un punto que amplifica los efectos de las *fake news* y, además, son una pieza fundamental a la hora de modificar la percepción de los usuarios frente a este tipo de desinformación [Küçük y Can, 2020]. Para representar los factores más relevantes que intervienen en la desinformación, se resume pictóricamente en la Figura 1



Figura 1: Caras de la desinformación.

Fuente: [Providel y Mendoza, 2021].

La imagen anterior representa las tres caras de la desinformación. Usuarios maliciosos crean noticias falsas para obtener un beneficio monetario o cambiar la opinión pública, los *bots* amplifican los efectos de propagación de noticias falsas para que, finalmente, los usuarios adopten una postura frente a la desinformación, propagándola voluntaria o involuntariamente.

1.3. Contexto e Impacto inicial del estudio

Para contextualizar, debido al creciente uso de las redes sociales y, por consiguiente, el aumento de desinformación [Nuñez y Fernández, 2019], nació una nueva forma de realizar periodismo llamado '*fast-checking*'. Este nueva rama del periodismo se centra en la verificación de información viral y las comprueba sometiénola a un proceso de verificación de fuentes estricto con el cual, finalmente, se le asigna una etiqueta que representa la veracidad de la información expuesta en el contenido. Estas etiquetas pueden ser: información falsa, información imprecisa, información verdadera, información fantasiosa, etc. Si bien el nacimiento de estos verificadores ha sido de gran ayuda a la hora de dismantelar hechos falsos, en la actualidad estas agrupaciones poseen financiamiento y tecnologías limitadas para realizar su trabajo, lo que dificulta el poseer una calidad de verificación constante, amplia y rápida que esté acorde a la velocidad de generación/propagación de noticias falsas. Bajo este contexto, se ha visto necesario realizar un estudio que logre aportar en el análisis de la problemática de desinformación masificada en Chile y también que logre analizar el impacto que tiene actualmente los medios verificadores a la hora de desmontar información falsa dentro de las redes sociales, Concretamente, algunos medios periodísticos chilenos que se dedican a la verificación de desinformación en redes sociales los cuales se verían beneficiados por un estudio de estas características son: FastCheck, FastChecking UC, Decodificador, Observatorio de datos, Polígrafo del mercurio, entre otros. Estas agrupaciones aportan con contenido verificado de forma diaria que serían primordiales para la investigación a realizar.

Cabe destacar que esta investigación sentará las bases para entender el impacto de la desinformación y de los verificadores en las redes sociales chilenas, aportando al periodismo nacional y, finalmente, a toda la población, contribuyendo una visión objetiva sobre la propagación de la desinformación en las redes sociales chilenas, sus características y el impacto de los medios verificadores a la hora de ser un medio de contención para esta problemática. A modo de resumen, se presenta el árbol de definición del problema en la Figura 2.

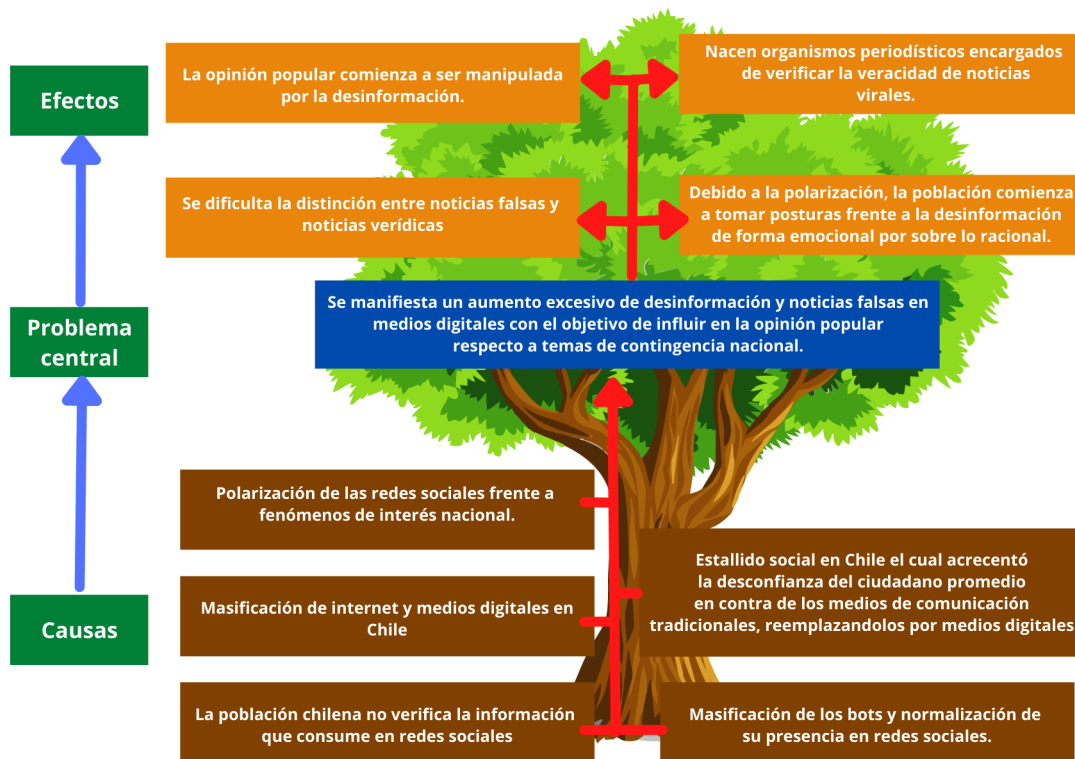


Figura 2: Árbol de definición del problema.
Fuente: Elaboración propia

1.4. Objetivos

1.4.1. Objetivo General

Analizar cuantitativamente la información y desinformación propagada en redes sociales en Chile con la ayuda de verificadores nacionales.

1.4.2. Objetivos Específicos

- Construir un *dataset* de noticias verificadas con la ayuda de medios verificadores nacionales a través de el cruce de sus base de datos.
- Estandarizar verificaciones, obtener fuentes y distinguir indicadores cuantitativos manuales.
- Distinguir indicadores deductivos/automáticos y analizar resultados.

CAPÍTULO 2

MARCO CONCEPTUAL

Con el objetivo de tener un apoyo teórico/técnico a la hora de realizar la investigación, se tendrá en consideración diversa bibliografía preliminar.

2.1. Definición de conceptos

Para comenzar, existen diversos conceptos importantes relacionados a las noticias falsas que es necesario definir a la hora de realizar un estudio serio al respecto. La definición de estos términos será extraído de trabajos relacionados, los cuales son:

- *Misinformation* (Desinformación): Los autores Briony Swire y David Lazer definen el concepto en su estudio de la desinformación respecto a la salud pública como 'información contraria a cualquier tipo de consenso epistémico realizado por la comunidad científica respecto a un fenómeno'[Swire-Thompson y Lazer, 2019]. Esta definición se puede extender a 'cualquier pieza de información contraria a cualquier consenso epistémico alcanzado por la aplicación sistemática de una metodología'[Providel y Mendoza, 2021]. Cabe destacar que el estatus de falso o verdadero de una pieza de información varía según la evidencia disponible en cada instante de tiempo.
- *Disinformation* (Desinformación intencionada): Los autores Briony Swire y David Lazer también definen otro concepto relevante a la hora de analizar la información falsa, el cual es '*Disinformation*' (o desinformación intencionada). Este concepto es una variación del concepto desinformación y es definido como 'un esfuerzo coordinado de intencionadamente poner a circular desinformación en medios informativos con el objetivo de tener una ganancia económica, de poder o de reputación'[Swire-Thompson y Lazer, 2019]. Por su definición, podemos notar que la detección de '*Disinformation*' está fuertemente ligada a la intencionalidad del usuario emisor, por lo que no siempre es fácil poder categorizar desinformación en esta categoría.
- *Fake News* (Noticias falsas): Los autores Xinyhi Zhou y Reza Zafarani abordan el problema de definir lo que es una '*Fake News*', intentando que el término no se solape con otros conceptos como '*Misinformation*' o '*Disinformation*'. Bajo este contexto, se define *Fake News* como 'cualquier artículo, aclamado, declaración, discurso, post u otros tipos de información relacionado a figuras públicas o organizaciones, falso'[Zhou y Zafarani, 2020]. Cabe destacar que esta información no necesariamente debe haber sido redactada por un periodista, y en esta definición no se considera la intencionalidad del emisor, por otra parte, no toda la información del documento

debe ser falsa, dando espacio a matices de lo que puede ser considerado una 'Fake News'.

- **Bot:** Un Bot es un software encargado de realizar acciones repetitivas o responder a comandos enviados por sus usuarios. Bajo el contexto de este estudio, un bot realizará acciones como publicar reiteradas veces un mensaje o repetir mensajes de otros usuarios. Es importante definir este termino ya que el autor Stefano Cresci reconoce a los Bot como uno de los factores principales y más relevantes a la hora de analizar la propagación de desinformación en redes sociales [Cresci, 2020].

Por otra parte, existen algunos conceptos que se usan en informática que se utilizarán en esta investigación, estos conceptos son:

- **API:** El término API es una abreviación de la frase *Application Programming Interface* y básicamente es una interfaz de un programa de computación la cual tiene como objetivo la facilitación de comunicación entre el programa que posee la interfaz y otro programa externo. Esto generalmente se utiliza para que otros programas puedan utilizar procesos/resultados calculados por el programa que posee la API sin necesidad de conocer la manera en la cuál los calcula, de forma que el programa que necesita la información sólo debe enviar un mensaje utilizando la API respetando una estructura previamente definida para luego recibir el resultado requerido como respuesta. Generalmente las API's vienen acompañadas con una documentación que describe la manera de utilizarlas.
- **Árbol (estructura de datos):** El árbol es un tipo de estructura de dato abstracta la cuál se puede definir en dos componentes principales : Nodos y aristas. Un nodo es una estructura simple que alberga algún tipo de información y puede estar relacionado de alguna forma a otro nodo. La información que puede albergar un nodo puede ser cualquier tipo, y generalmente está limitada a al contexto en el cuál fue creado el nodo. Por otra parte, un arista es una conexión entre nodos que generalmente representa un tipo de relación que poseen esos nodos, esta relación puede ser definida de cualquier forma y generalmente esta limitada en el contexto el cuál está representado. Por otra parte, un árbol es una estructura que utiliza estos dos elementos entablando una serie de reglas:
 - Existe un nodo inicial y principal del árbol el cuál se le conoce como nodo raíz.
 - Todos los nodos deben tener un nodo 'padre' con el cuál estarán conectado a través de un arista (exceptuando por el nodo raíz).
 - A su vez, todos los nodos que tienen un padre se le conocerá como nodos hijos.
 - Todos los nodos hijos sólo pueden tener un único padre, pero pueden poseer una cantidad de hijos ilimitada.

Siguiendo las reglas anteriores, se puede construir una estructura que es muy buena representando información jerárquica segmentada por niveles (generaciones), estructura similar a lo que podría ser un árbol genealógico.

2.2. Creación de *Datasets*

Además de la revisión de los trabajos anteriores, se analizará literatura que haya investigado *datasets* similares al planteado en esta investigación. Uno de los trabajos a revisar será el realizado por Isabelle Augenstein et al. el cual logra recaudar un total de 34.918 artículos en inglés provenientes de 26 sitios dedicados al *Fast-Checking*. Todos estos datos provienen de 38 páginas web distintas de las cuales se extrajo 43.837 artículos los cuales fueron filtrados a través de diversos métodos [Augenstein et al., 2019] (por ejemplo, juntando artículo repetidos o eliminando artículos con etiquetas ambiguas). Por otra parte, también se revisará el trabajo de Gautam Kishore y Durgesh Nandini el cual logró recaudar 5182 artículos relacionados al Covid de 92 diferentes páginas de fast-checking, cabe recalcar que los artículos provienen de diversos idiomas siendo el Top-4 de idiomas con más representantes el inglés, español, portugués y francés en ese orden. Para crear el *dataset* se corrigió manualmente todas las URL de las fuentes originales de los artículos de *fast-check*, se eliminaron los artículos de los cuales no se pudo recuperar el título y se eliminaron duplicados. Otro de los pre-procesamientos relevantes fue la corrección ortográfica y la abreviación de los artículos recolectados, utilizando la librería *TextBlob* de *python* para lograr este objetivo [Shahi y Nandini, 2020].

2.3. Acerca de las *Fake News*

Por otra parte, también será pertinente revisar trabajos realizados en el análisis de *Fake News*. Debido a la enorme cantidad de métodos existentes para trabajar con este tipo de datos, se le dará énfasis a revisar preliminarmente trabajos recopilatorios encargados de presentar y resumir distintas formas de abordar la información dentro de las *Fake News*. Uno de estos estudios es el realizado por Francesco Pierri y Stefano Ceri [Pierri y Ceri, 2019], el cual dividió el estado del arte en tres metodologías principales:

- **Análisis basado en contenido:** Son trabajos que se han centrado principalmente en analizar sólo el contenido de las noticias (considerando título, cuerpo, fuente, etc.). Esto no excluye el uso de meta datos, como lo podría ser la etiqueta otorgada por un organismo de *fast-checking*.
- **Análisis basado en contexto:** Son estudios centrados en el análisis basado en las interacciones de los usuarios frente a una pieza de información, estas interacciones podrían ser comentarios, *likes*, *re-tweets*, etc.
- **Análisis basado en contexto y contenido:** Son trabajos que mezclan ambos métodos anteriormente mencionados. En proporción son los menos representados, pero han mostrado buenos resultados a la hora de realizar tareas como la detección de noticias falsas.

Cabe destacar que el estudio también hace énfasis en estudios relacionados a modelos de difusión de noticias falsas y Técnicas de mitigaciones de noticias falsas apoyadas por organizaciones dedicadas al *fast-checking*. Dentro de este contexto, también se analizará el trabajo de Karishma Sharma, Feng Qian y He Jiang [Sharma *et al.*, 2019], el cuál es trata tópicos similares con respecto al trabajo descrito con anterioridad con la diferencia de que es más amplio en los temas que abarca, agregando información sobre los set de datos más relevantes en el estado del arte y centrándose en temas distintos como la extracción de características de los datos y definiendo otros factores relevantes a la hora de trabajar con datos relacionados a las *Fake News*.

Finalmente, se analizarán los trabajos realizados en español. Para entender este sector del estado del arte, se tomará como base el estudio realizado por Eliana Providel y Marcelo Mendoza [Providel y Mendoza, 2021], en su investigación recopilatoria analizaron investigaciones a través de diversos focos de mira los cuales pueden ser, por ejemplo, las representaciones de texto que utilizaron los estudios para realizar variadas tareas o los distintos algoritmos utilizados para diversas tareas de *NLP* como lo son la detección de *fake news*, la detección de postura de usuarios y la detección de *bots*, estos algoritmos consideran tanto trabajos realizados utilizando *Machine Learning* como *Deep Learning*. Por otra parte, también se considera todo el trabajo realizado en términos de creación de recursos en español como lo pueden ser la creación de corpus, *embeddings*, *dataset* centrados en distintas tareas, etc.

2.4. Detección de *bots*

Como mencionamos con anterioridad, los *bots* son un aspecto fundamental a la hora de analizar la propagación de desinformación en línea.

Existen diversos estudios que abordan la problemática de detección de *bots*, uno de ellos es el trabajo de Onur Varol *et al.* [Varol *et al.*, 2017] el cual propone un *framework* de detección de *bots* en Twitter. Este *framework* está centrado principalmente en la detección de características de la cuenta a analizar, dentro de estas características se señalan seis distintos tipos:

- Características basadas en usuario: Corresponden a características extraídas de la *meta data* entregada por la API de Twitter. Estas características pueden incluir número de seguidores, número de tweets producido por el usuario, descripción del perfil del usuario y configuración del perfil del usuario.
- Características basadas en amigos: Considerando que Twitter es una red social donde la clave es la interacción entre usuarios, es necesario extraer características derivadas de la relación usuario-seguidores. Se consideran cuatro tipo de relaciones: retweet, menciones, ser retweeteado y ser mencionado. Para cada grupo se extrae (separadamente) información como lo es el lenguaje utilizado, la hora local, la popularidad, etc.

- Características basadas en redes: A partir de las interacciones entre usuarios, se puede generar redes de comunicación que acarreen información crucial acerca del flujo de información y la relación entre usuarios. Se proponen tres tipos de redes de las cuales puede resultar un análisis relevante: redes de retweets, redes de menciones y redes de co-ocurrencia de *hashtag*. Para el caso de las redes de menciones y retweets, cada usuario corresponde a un nodo, mientras que una arista dirigida une dos usuarios en dirección al flujo de información de la mención o el retweet (en dirección al usuario mencionado o retweeteado). Por otra parte, la red de co-ocurrencia de *hashtag* corresponde a una red no-dirigida donde los nodos son *hashtags*, dos nodos están conectados si es que ocurren en un mismo tweet. Todas las redes mencionadas con anterioridad tienen sus aristas con un peso relacionado a la frecuencia de las interacciones o co-ocurrencias. Para cada red, se extrae información como el *in-out weight degree*, distribuciones, densidades y *clustering* (comunidades).
- Características temporales: Este tipo de características pueden ser relevantes a la hora de obtener información relacionada a la evolución de campañas en línea. Dentro de las características medidas se destaca: actividad del usuario, promedio de tweets cada ciertos periodos de tiempo y distribución de tiempo entre intervalos de eventos.
- Características basadas en contenido y lenguaje: La estructura y características de un mensaje tienen información relevante a la hora de intentar discernir la intencionalidad del mensaje, por ejemplo, un mensaje que busque convencer suele ser informal y de pocas palabras. Las características a extraer en este contexto pueden ser el largo de los tweets y su entropía relacionada, además, la aplicación de técnicas como *part-of-speech (POS) tagging* para detectar componentes del lenguaje natural nos pueden ayudar a obtener la distribución de contenidos como pronombres, verbos, adjetivos, etc.
- Características basadas en sentimientos: El análisis de sentimientos puede recuperar efectivamente y de forma fidedigna las emociones plasmadas en un pedazo de texto e incluso el estado anímico de una conversación completa. Los sentimientos de las redes sociales pueden afectar desde la especulación de mercados financieros hasta la forma de propagar la información en redes sociales. El framework extrae información de sentimientos como lo pueden ser: excitación, aversión, puntaje de dominancia, puntaje de felicidad, polarización y fuerza, puntaje de emoticones.

En resumen, se pueden extraer hasta 1150 características las cuales están plasmadas en la Figura 3.

Table 1: List of 1150 features extracted by our framework.

User meta-data	Screen name length	(***) Happiness scores of aggregated tweets
	Number of digits in screen name	(***) Valence scores of aggregated tweets
	User name length	(***) Arousal scores of aggregated tweets
	Time offset (sec.)	(***) Dominance scores of single tweets
	Default profile (binary)	(*) Happiness score of single tweets
	Default picture (binary)	(*) Valence score of single tweets
	Account age (days)	(*) Arousal score of single tweets
	Number of unique profile descriptions	(*) Dominance score of single tweets
	(*) Profile description lengths	(*) Polarization score of single tweets
	(*) Number of friends distribution	(*) Entropy of polarization scores of single tweets
	(*) Number of followers distribution	(*) Positive emoticons entropy of single tweets
	(*) Number of favorites distribution	(*) Negative emoticons entropy of single tweets
	Number of friends (signal-noise ratio and rel. change)	(*) Emoticons entropy of single tweets
	Number of followers (signal-noise ratio and rel. change)	(*) Positive and negative score ratio of single tweets
	Number of favorites (signal-noise ratio and rel. change)	(*) Number of positive emoticons in single tweets
	Number of tweets (per hour and total)	(*) Number of negative emoticons in single tweets
	Number of retweets (per hour and total)	(*) Total number of emoticons in single tweets
	Number of mentions (per hour and total)	Ratio of tweets that contain emoticons
	Number of replies (per hour and total)	
	Number of retweeted (per hour and total)	
Friends (†)	Number of distinct languages	Number of nodes
	Entropy of language use	Number of edges (also for reciprocal)
	(*) Account age distribution	(*) Strength distribution
	(*) Time offset distribution	(*) In-strength distribution
	(*) Number of friends distribution	(*) Out-strength distribution
	(*) Number of followers distribution	Network density (also for reciprocal)
	(*) Number of tweets distribution	(*) Clustering coeff. (also for reciprocal)
Content	(*) Description length distribution	
	Fraction of users with default profile and default picture	
	(*, **) Frequency of POS tags in a tweet	(*) Timing
	(*, **) Proportion of POS tags in a tweet	(*) Time between two consecutive tweets
Timing	(*) Number of words in a tweet	(*) Time between two consecutive retweets
	(*) Entropy of words in a tweet	(*) Time between two consecutive mentions

† We consider four types of connected users: retweeting, mentioning, retweeted, and mentioned.

‡ We consider three types of network: retweet, mention, and hashtag co-occurrence networks.

* Distribution types. For each distribution, the following eight statistics are computed and used as individual features: min, max, median, mean, std. deviation, skewness, kurtosis, and entropy.

** Part-Of-Speech (POS) tag. There are nine POS tags: verbs, nuns, adjectives, modal auxiliaries, pre-determiners, interjections, adverbs, wh-, and pronouns.

*** For each feature, we compute mean and std. deviation of the weighted average across words in the lexicon.

Figura 3: La imagen representa las 1150 características que se puede extraer utilizando el *framework* planteado.

Fuente: [Varol et al., 2017].

Por otra parte, también se han realizado herramientas la cuales utilizan los principios mencionados con anterioridad. Una de ellas (la cuál se utilizará más adelante) es *botometer*. Botometer es un modelo de *machine learning* supervisado el cuál fue entrenado en catorce datasets los cuales poseen cuentas de Twitter reconocidas como humanos y otras reconocidas como bots. Este modelo utiliza características de las cuentas para clasificarlas, estas características se dividen en seis clases distintas:

- Perfil de usuario
- Amigos
- Conexiones

- Temporalidad
- Contenido y lenguaje
- Sentimientos

Juntando los elementos de cada categoría, se reúnen más de 1000 características que describen a profundidad una cuenta de Twitter. Por otra parte, el modelo de predicción utilizado es una combinación de *Random Forests* especializados en alguna serie de características de la cuenta, los cuales juntos forman una estructura llamada *Ensemble of Specialized Classifiers* (ESC). Finalmente, cabe destacar que Botometer está entrenado en cuentas de Twitter que se comunican en inglés, por lo que su enfoque principal es en la identificación de bots en este idioma. Sin embargo, también pueden realizar una predicción a cuentas de diversos idiomas, en estos casos el modelo predictivo no utiliza ninguna característica relacionada al lenguaje. [Yang *et al.*, 2022]

2.5. Propagación de la desinformación en redes sociales.

La propagación de desinformación es un factor importante a analizar, ya que nos entrega una perspectiva realista de la importancia que tienen las noticias falsas v/s noticias verdaderas en términos de viralidad en redes sociales. En general, una de las redes sociales más cooperadoras para este tipo de estudios es Twitter, debido a que provee una amplia gama de herramientas para extraer la información necesaria desde su plataforma. Bajo este contexto, uno de los estudios más relevantes dentro de esta área el realizado por Vosoughi *et al.* [Vosoughi *et al.*, 2018]. En esta investigación, se realizó un análisis de aproximadamente 126.000 tweets los cuales interactuaron con aproximadamente 4.5 millones de personas. Los autores consideraron la propagación de información según su veracidad, considerando tres categorías de noticias principales: verdaderas, falsas y mixto (poseen hechos parcialmente verdaderos o parcialmente falsos). Por otra parte, se considera como noticia cualquier afirmación respecto a un hecho, aunque este no sea necesariamente viral o altamente compartido, además, introducen el concepto de cascada. Una cascada en twitter comienza cuando un individuo realiza una afirmación respecto a un tema, esta afirmación puede ser únicamente a través de texto o utilizando métodos multimedia como fotos, videos o fuentes externas, de esta forma, la cascada se comienza a formar cuando esta información comienza a ser compartida (en este caso, utilizando retweets). Un solo rumor puede tener distintas cascadas asociadas, ya que diversos usuarios puede compartir la misma afirmación de forma singular e individual, por lo que dos afirmaciones iguales que compartan la misma fuente corresponden a dos cascadas de propagación distintas. El tamaño mínimo de una cascada es 1 (cuando el tweet con la afirmación no recibió retweets) mientras que el tamaño total se calculará sumando todas las cantidades de retweets que tenga la afirmación original. Además, se definen las siguientes características de una cascada:

- *Depth*: El numero de "saltos.^a a través del tiempo entre tweets y retweets a partir del tweet original. Al navegar el grafo de retweets, cada transición entre retweets corresponde a un salto.
- *Maximun Breadth*: La cantidad de usuarios participando en la cascada en cualquier punto donde se pueda calcular *depth*.
- *Structural Vitality*: Corresponde a una medida que calcula la interpolación entre la propagación del contenido si es que se considera un solo punto de alta propagación o si se considera varios pequeños puntos de propagación.

Para entender de mejor manera estas medidas, se presenta la Figura 4.

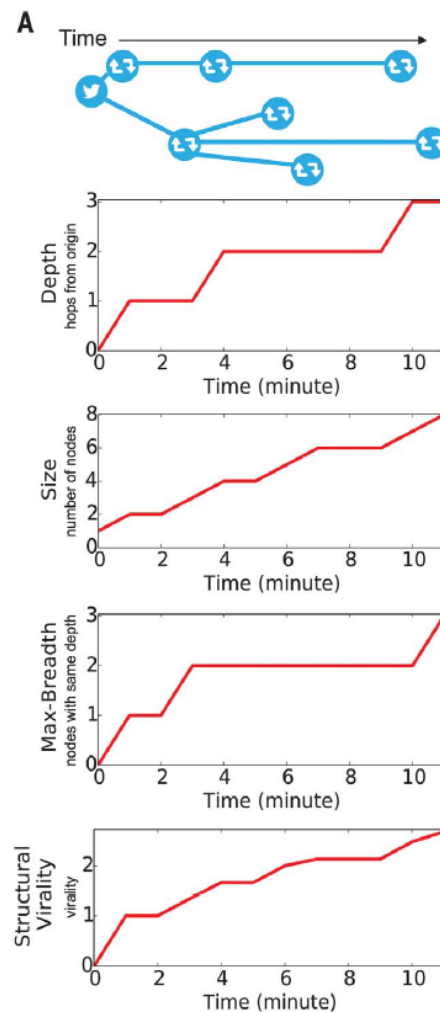


Figura 4: La imagen representa un ejemplo de cascada con los valores de todas sus medidas a través del tiempo.

Fuente: [Vosoughi et al., 2018].

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

En esta sección se procederá a explicar todo el estudio realizado, desde el proceso de creación del dataset inicial, su complementación, la extracción de los árboles de retweets y finalmente el análisis de los datos extraídos.

3.1. Creación del dataset.

El *dataset* de verificaciones se creó a partir de tres medio verificadores chilenos: FastCheck.cl, FactChecking.cl y Decodificador.cl. Cada medio verificador entregó a este estudio cada una de las verificaciones que realizaron entre las fechas 18 de octubre de 2019 hasta el 18 de octubre de 2021, estas fechas fueron seleccionadas debido a que son dos años que cubren eventos como: El estallido social, La pandemia Covid-19, la Nueva Constitución, Elecciones 2021, entre otros. En resumen, el número de noticias entregada por cada medio verificador se describe a continuación:

Medio de verificación	Cantidad de noticias
FastCheck	656
FactChecking.cl	169
Decodificador.cl.	175

Tabla 1: Cantidad de noticias según medio de verificación.
Fuente: Elaboración Propia.

Cada una de las verificaciones corresponde a cada chequeo de noticias realizado y publicado en las páginas web oficiales pertenecientes a cada uno de los medios verificadores. La obtención de los datos fue solicitada directamente a los medios mencionados con anterioridad, con previa autorización de su utilización para un estudio de esta índole. De esta manera, se construyó un *dataset* el cuál estaba conformado por las siguientes columnas:

- Título.
- Contenido.
- Fecha de publicación.
- Metadata acerca del usuario de verificación.
- Categoría de verificación (verdadera,falsa,imprecisa,etc).

Posteriormente, a partir de la información preliminar obtenida en el paso anterior, se procedió a complementar los datos obtenidos realizando una extracción manual de información relacionada a los chequeos. Esta información se extrajo a partir de la información disponible en la publicación original del chequeo, obteniendo de esta forma la siguiente información:

- URL de la noticia ya verificada. Existen diversos lugares donde las verificaciones fueron publicadas, por lo cual los *Link* se dividen en:
 - *Link* de publicación en la página web oficial del verificador.
 - *Link* de publicación en Facebook.
 - *Link* de publicación en Twitter.
 - *Link* de publicación en Instagram.
- Las fechas de creación relacionadas a las publicaciones del punto anterior.
- Tipo de medio por el cuál se propagó la información a verificar, por ejemplo, RRSS, televisión, radio, etc.
- Nombre del medio por el cuál se propagó la información a verificar, en el caso de redes sociales correspondería al nombre de la red social, en caso de ser un medio televisivo corresponde al nombre del programa/canal de televisión por el cual se propagó, etc
- URL de la publicación a verificar (en caso de que esté disponible), con su respectiva fecha de publicación.
- Formato del contenido a verificar (texto , vídeo, imágenes, etc)

Después de haber extraído toda la información anteriormente mencionada, se procedió a homologar las etiquetas de cada uno de los medios verificadores entre sí. Para contextualizar, cada medio verificador participante en esta investigación posee sus propias etiquetas para describir lo que es 'verdadero', 'falso' o algún punto intermedio, por lo que es necesario realizar un proceso de homogeneización de forma de tener una sola etiqueta para cada caso. Para poder realizar este proceso, se tomó en consideración dos modelos propuestos para realizar este proceso:

- El modelo propuesto por [Amazeen, 2015], el cuál fue utilizado en el contexto de buscar y definir una manera de poder medir el porcentaje de 'acuerdo' que existe entre dos o más medios verificadores a la hora de evaluar la veracidad de una afirmación. Este modelo debido a la variabilidad de los métodos de evaluación de veracidad de los distintos medios verificadores, plantea que con fin de simplificar y homogeneizar las clasificaciones de veracidad, es mejor considerar una aproximación binaria. De esta forma, crea dos divisiones para las clasificaciones de los medios verificadores: afirmaciones que muestran una presencia de alguna imprecisión o las afirmaciones que

presentan una ausencia de cualquier tipo de imprecisión. En conclusión, su acercamiento elimina los grados de imprecisión (o grado de falsedad) y se limita a catalogar las etiquetas de verificación como 'precisas' e 'imprecisas'.

- En contraste al modelo anterior, la siguiente propuesta revisada fue la de [Lim, 2018]. En esta ocasión, con el objetivo de querer comparar la fiabilidad entre los dos medios de *fastchecking* más grandes de Estados Unidos (FastCheck y Politifact), se requirió establecer una equivalencia entre las distintas etiquetas de veracidad entre los distintos medios de verificación. Para realizar este proceso, se hizo una especie de 'traducción' entre etiquetas con la ayuda del Editor Jefe del medio verificador FastCheck, en el cual se le asignó a las etiquetas del medio con más etiquetas las etiquetas del medio con menor etiquetas, dejando así una cantidad de etiquetas igual a la cantidad de etiquetas que poseía el medio con menos etiquetas. Cabe destacar que esto se realizó de forma manual y cualitativa con la ayuda de un usuario experto en el tema.

En el caso de esta investigación, se consideró ambos acercamientos para poder llegar a un consenso entre las diversas categorías de los medios estudiados. En primer lugar, se dividió el espectro de veracidad en cuatro grandes categorías: Verdadero (verificaciones que señalan que la información verificada es totalmente verdad), Impreciso (verificaciones que señalan que la información verificada presenta distintos grados de verdad y falsedad), Falso (verificaciones que señalan que la información verificada es totalmente falsa) e Inchequeable (verificaciones que por alguna razón o circunstancia no se puede definir su veracidad). Luego, se hizo una conversión entre todas las etiquetas de todos los medios a algunas de estas cuatro categorías, este proceso se hizo con la ayuda de dos expertos en Fact-Checking: Fabián Padilla (CEO de FastCheck) y Enrique Nuñez (Editor general de Factchecking.cl). Finalmente, también se utilizó como guía las definiciones entregada por cada medio verificador para entender el significado de cada una de sus etiquetas, de esta forma, se obtuvo la siguiente forma de clasificación:

Etiqueta original	Nueva Etiqueta
Ciencia Ficción	Falso
Creible	Verdadero
Creible, pero	Impreciso
Cuestionable	Inchequeable
Engañoso	Impreciso
Fake	Falso
Impreciso	Impreciso
No es creible	Falso
Real	Verdadero
Se puso creativo	Impreciso
Seria creible, pero	Impreciso
Verificado	Verdadero
Falso	Falso
Inchequeable	Inchequeable

Tabla 2: Traducción de etiquetas de verificación.
Fuente: Elaboración Propia.

Posterior al proceso de homogeneización, la distribución de verificaciones en las distintas categorías de veracidad quedó plasmado de la siguiente manera:

Veracity	FastCheck	Decodificador	FactCheckingUC	Total
Verdadero	284	38	53	375
Falso	250	86	37	373
Impreciso	122	42	85	249
Inchequeable	0	3	0	3
Total	656	169	175	1000

Tabla 3: Nueva distribución de verificaciones con respecto a las nuevas categorías.
Fuente: Elaboración Propia.

3.2. Creación de árboles de propagaciones de Twitter.

El siguiente paso relacionado a este proceso tiene que ver con las creación de árboles de propagación (Sección 2.5)[Vosoughi *et al.*, 2018]. Los árboles de propagación se dividen en dos tipos:

- Árboles de conversación: Corresponden a la estructura generada a partir de todas las respuestas realizadas a un tweet. Esta estructura considera tanto las respuestas directas al tweet raíz como las respuestas a respuestas.

- Árboles de retweets: Considera la estructura generada a partir de todos los retweets realizados a un tweet. Esta estructura considera tantos los retweets al tweet raíz como a los retweet de retweets.

Para realizar este proceso, se procedió a seleccionar todas las verificaciones las cuales su fuente de información/desinformación provinieran de Twitter. Cada URL de Twitter posee un número único que corresponde a un ID único que permite extraer la información relacionada al tweet con la ID en cuestión, junto a esto, se utilizó la API de Twitter como herramienta de extracción de datos.

Primeramente se explicará el paso a paso para realizar la reconstrucción de los árboles de conversación:

- Se extrae la ID de cada URL de cada tweet.
- Se realiza una *query* general en la API de twitter donde se rellena el campo *conversation_id*
- Se envía la *request* y se obtienen todas las respuestas relacionadas a ese tweet.
- Cada respuesta obtenida tiene un campo llamado *in_reply_to_user_id* el cuál señala a que usuario está respondiendo la respuesta actual y también posee un *user_id* el cual señala el ID del usuario al cual responde, de esta forma (y considerando la fecha de creación de cada respuesta), se puede asociar cada respuesta al comentario que está respondiendo.
- Una vez asociada cada respuesta a su respectivo comentario, el resultado obtenido corresponde al árbol de conversación del tweet original.

Por otro lado, el proceso para realizar los árboles de retweets es el siguiente:

- Se deben obtener todos los tweets relacionados a un tweet, para eso se realizó una búsqueda de texto utilizando uno de los *end-points* disponible en la API. Este *end-point* permite realizar una búsqueda de texto a lo largo de **todos los retweets** hechos al usuario.
- Utilizando el *end-point* , se procede a buscar el texto del tweet al que se le desea obtener los retweets en todos los retweets hechos al usuario creador del tweet en cuestión, de esta manera, se obtienen todos los retweets del tweet base.
- Los retweets obtenidos por la API no poseen ninguna forma de estructura, por lo que el árbol debe inferirse. Para lograr ese proceso, se hacen los siguientes supuestos:
 - Un usuario suele seguir a las personas que retweetea.

- Si un usuario encuentra más de un retweet a un mismo tweet, se asume que está realizando retweet al retweet más cercano cronológicamente al momento que el usuario actual está realizando el retweet.
- Se considera que es más probable que un usuario realice retweet al tweet original que a un retweet.

Una vez considerados esos supuestos, se puede continuar con el procedimiento.

- Se ordenan los retweets cronológicamente desde el más antiguo al mas nuevo.
- Luego, siguiendo el orden anterior, se recorrió cada retweet aplicando el siguiente procedimiento:
 - Para cada retweet, se comprobó si el usuario de este retweet sigue a la cuenta que publicó el tweet original, en caso de que así sea, se asume que el retweet había sido realizado al retweet original.
 - En caso de no seguir al usuario original, se busca entre los retweets que sucedieron cronológicamente antes del retweet actual y, en el caso de que el usuario actual siga a alguno de los usuarios que retuitearon con anterioridad, se infiere que el retweet se realizó a ese retweet. Si ocurre que un usuario sigue a más de un usuario que haya retuiteado con anterioridad, se asume que es más probable que el usuario actual haya retuiteado al usuario que haya realizado su retweet cronológicamente más cerca al retweet del usuario actual.

El proceso anterior consistentemente logra reconstruir una aproximación del árbol real de propagación de retweets. Lamentablemente, esta metodología presenta un problema, resulta que en el proceso anterior se debe verificar constantemente la pregunta '¿El usuario A sigue al usuario B?', pregunta la cual resulta ser bastante costosa para la API de Twitter. Este fenómeno tiene como consecuencia que la ejecución de ese procedimiento se vuelve muy lento rápidamente, llegando a tardar varios días de procesamiento para los tweets más grandes. Para mitigar este problema en los árboles más grandes, se usaron dos métodos alternativos para agilizar el proceso:

- Utilizar un número reducido de usuarios: Para esta alternativa, se obtuvo todos los usuarios que participaban en cada uno de los árboles (tanto de respuestas y de retweets) y se clasificaron en grupos según en la cantidad de arboles en los cuales participaban. Luego, se procedió a considerar para la creación de arboles sólo los usuarios que participaban en dos o más arboles, esto bajo el supuesto de que los usuarios que participaban en más conversaciones son más relevantes para la estructura del árbol. En caso de que el proceso se volviera a ralentizar debido a la cantidad de retweets, se redujo el universo de usuarios a los que participaron en tres o más arboles, luego en más de cuatro y así sucesivamente.

- Utilizar un grafo parcial de la red de seguimiento chilena en Twitter: Para esta alternativa, se utilizó un grafo de usuarios el cual poseía una muestra de todas las relaciones de seguimiento de cuentas chilenas, alcanzando a recolectar las interacciones de aproximadamente diez millones de personas. Este grafo se construyó a partir de las cuentas de Twitter de las personas que participaron en el proceso constituyente ocurrido en Chile entre los años 2021 y 2022 (más bien conocidos como constituyentes). Teniendo como referencia estas cuentas, se obtuvo todos los seguidores de los constituyentes y también los seguidores de los seguidores, logrando de esta forma obtener una red parcial de seguimiento. Las consultas realizadas a este grafo resultaban ser considerablemente más rápidas en comparación a la API de Twitter, pero no poseía todos los usuarios que participaban en los arboles de interacciones.

Cabe destacar que por cada árbol que resultara muy lento computarlo con el proceso original, se utilizó ambas alternativas para compensarlo, quedando como resultado final el árbol más grande obtenido entre los dos procesos alternativos.

3.3. Métricas

A partir de todo el trabajo realizado en la sección anterior, se procederá a obtener dos tipos de métricas distintas:

- Métricas del dataset original: Son métricas obtenidas a partir de la información obtenida del *dataset* de verificaciones e información adicional obtenida a través de la API de Twitter. Estas métricas **no hacen uso** de los arboles de propagación, y fueron creadas/calculadas con el fin de poder analizar el comportamiento de la desinformación en Chile/Verificadores en Chile a través de diversos focos. Cabe destacar que estas métricas fueron obtenidas únicamente en el marco de esta investigación.
- Métricas de propagación: Son métricas obtenidas a través de la deducción de información utilizando como fuente el comportamiento estructural de los arboles de propagación de Twitter. Cabe destacar que estas métricas fueron obtenidas en el marco de un proyecto/investigación [Mendoza *et al.*, 2023] perteneciente al IMFD (Instituto Milenio Fundamentos de los Datos), por lo que su autoría no pertenece al autor de este documento. No obstante, son expuestos los resultados de estas métricas con el fin de evidenciar las diversas posibilidades que ofrece el *dataset* de desinformación chilena.

3.3.1. Métricas del dataset original.

Respecto a las métricas del dataset original, se dividen en dos tipos:

- Métricas con referencia a los medios verificadores: Estas métricas están enfocadas en analizar alguna característica que competa a los medios verificadores. Las métricas de esta categoría son:
 - Tiempos de respuesta: Esta métrica busca confirmar si existe alguna tendencia entre la velocidad de verificación de una desinformación y su veracidad. Para esto, se calculará la diferencia de días entre la fecha de publicación de la verificación en las redes oficiales del medio verificador y la fecha de publicación de la desinformación. En caso de que algún medio verificador tenga más de una fecha para una única verificación (en que de que, por ejemplo, la publicación se publique un día antes en la página web con respecto a la redes sociales), se considerará la fecha más antigua.
 - Diferencia de reacciones: Esta métrica busca evidenciar si los medios verificadores logran llegar a más personas en relación a la desinformación distribuida en internet. Debido a la naturaleza variada de las redes sociales y las fuentes de desinformación, para esta métrica sólo se considerará la desinformación viralizada a través de Twitter e Instagram (debido a que son las redes con más fácil acceso a la información de las publicaciones), y para el caso del medio verificador sólo se considerará Instagram (esto debido a que para todos los medios verificadores, sólo Instagram muestra ser una red social con interacciones suficientes para ser relevantes en el análisis). Por otro lado, para facilitar el análisis, solo se considerará la diferencia entre likes y comentarios entre las publicaciones para determinar a la cantidad de gente que llegó/ la discusión que causó.
- Métricas con referencia a la desinformación: Estas métricas están enfocadas en analizar algún comportamiento en la desinformación presente en el *dataset*. Las métricas en esta categoría son:
 - Análisis de sentimiento: Esta métrica busca evidenciar si existe una tendencia a los sentimientos positivos/negativos en los comentarios de una desinformación dependiendo de su veracidad. Para esto, se analizarán los comentarios de todas las publicaciones viralizadas en twitter y, como herramientas de análisis de sentimiento, se utilizó PySentimiento y Spanish Analyzer Spacy (con el fin de evitar resultados sesgados por la herramienta). Cada pieza de desinformación se catalogará con un valor entre -1 y 1 dependiendo del promedio de sentimientos de sus comentarios, siendo -1 sentimientos negativos y + 1 sentimientos positivos. Por otra parte, se clasificó cada publicación con tres categorías: Positiva, Neutra o negativa, cada categoría se asigna dividiendo el espacio numérico en tres partes iguales y asignando una categoría a cada una de las partes, de esta forma las publicaciones con un promedio cercano al -1 son negativas, cercano al 0 es neutro y cercano al 1 es positivo.
 - Participación de Bots: Esta métrica busca evidenciar si en los Retweets a la desinformación propagada en twitter, existe una tendencia entre la veracidad de la información y la naturaleza de la cuenta que la propaga realizando retweet (con

naturaleza se refiere a si el usuario es probablemente un bot o no). Para obtener la naturaleza de las cuentas se utilizó dos herramientas: Botometer y botcheckcl. Por otra parte, la métrica buscaba ver la distribución de retweets según veracidad considerando los usuarios que son considerados Bots y los usuarios que no son Bots. Cabe destacar que este análisis se realizó considerando, aproximadamente, los 3500 usuarios más frecuentes dentro de todos los arboles de retweets, esto se realizó de esta manera debido a que el proceso de detección de bots es un proceso costoso, por lo que se redujo el universo de usuarios a un número más manejable y que se pudiera analizar en un tiempo acotado.

3.3.2. Caso de estudio: Métricas de propagación (A partir de los árboles de propagación)

Respecto a las métricas del caso de estudio, se debe aclarar que en el proyecto mencionado sólo se obtuvo estos comportamientos para los árboles de respuestas, teniendo esto en cuenta se obtuvieron tres métricas distintas:

- Profundidad: Corresponde al camino de largo más grande presente en el árbol de propagación comenzando desde la raíz. En este contexto correspondería a la concatenación de respuestas más larga dentro del árbol.
- Tamaño: Corresponde a la cantidad de respuestas presentes en el árbol.
- Anchura Máxima: El concepto 'anchura' por sí solo corresponde a la cantidad de respuestas que posee un nivel dentro del árbol. El nivel de cada respuesta está identificado como la distancia en 'saltos' (Sección 2.5) que se encuentra la respuesta actual de la respuesta raíz. De esta forma, la anchura máxima corresponde a la cantidad de respuestas que tiene el nivel con más respuestas.

Cada una de estas métricas está medida, además, en función de su evolución a lo largo del tiempo debido a que cada respuesta posee un *timestamp* del momento en el cual fue creado.

CAPÍTULO 4

RESULTADOS Y ANÁLISIS

En este capítulo se procederá a recopilar, presentar y analizar cada uno de los resultados obtenidos en la sección 3.3.

4.1. Resultados

4.1.1. Tiempos de respuesta

Para comenzar se procederá a mostrar los tiempos de respuestas sin procesamiento de cada uno de los medios verificadores.

Fastcheck	Total	Verdad	Impreciso	Falso
Total noticias	567	248	102	217
Cantidad días	9857	5405	758	3667
Media días por noticia	17.38	21.79	7.43	16.89

Decodificador	Total	Verdad	Impreciso	Falso
Total noticias	140	34	29	77
Cantidad días	1978	99	170	1709
Media días por noticia	14.12	2.91	5.86	22.19

Factcheck UC	Total	Verdad	Impreciso	Falso
Total noticias	136	29	80	27
Cantidad días	7131	1160	4478	1493
Media días por noticia	52.43	40	55.97	55.29

Tabla 4: Cantidad de días promedio de cada medio verificador. Estas tablas consideran todas las publicaciones de cada medio

Fuente: Elaboración propia

Por otra parte, con el objetivo de hacer un análisis más robusto respecto a las velocidades de cada medio verificador, se procedió a eliminar los *outliers* de los conjuntos de cada medio verificador por separado. Para esto, se realizó un *boxplot* genérico con coeficiente 1.5, y se

procedió a eliminar cada uno de los datos que quedara fuera del rango intercuartílico multiplicado con el coeficiente. De esta manera, los resultados quedan de la siguiente manera:

Fastcheck	Total	Verdad	Impreciso	Falso
Total noticias	502	230	90	182
Cantidad días	1781	758	403	620
Media días por noticia	3.54	3.29	4.47	3.40

Decodificador	Total	Verdad	Impreciso	Falso
Total noticias	119	32	22	65
Cantidad días	279	81	70	128
Media días por noticia	2.34	2.53	3.18	1.96

Factcheck UC	Total	Verdad	Impreciso	Falso
Total noticias	127	26	68	23
Cantidad días	842	735	2328	715
Media días por noticia	31.18	28.29	34.23	31.08

Tabla 5: Cantidad de días promedio de cada medio verificador. Para el caso de Fastcheck se consideran las publicaciones con menor o igual a 13 días de retraso en la verificación, para el caso de decodificador son las publicaciones menor o iguales a 8 días y para Factcheck menor o iguales a 128 días.

Fuente: Elaboración propia

4.1.2. Diferencia de Reacciones

A continuación, se procede a recopilar las diferencias de *likes*/comentarios de cada una de las desinformaciones publicadas en Twitter o Instagram. Una diferencia es positiva cuando la publicación en Instagram del medio verificador **supera** a la desinformación original, mientras que una diferencia negativa es el caso contrario. De esta forma, se obtuvieron las siguientes relaciones para el caso Twitter V/S Instagram:

Todos los medios	Verdad	Impreciso	Falso
Negativos	33	16	38
Positivos	85	27	65

FastCheck	Verdad	Impreciso	Falso
Negativos	16	8	15
Positivos	84	21	65

Decodificador	Verdad	Impreciso	Falso
Negativos	16	5	20
Positivos	1	2	7

FactChecking	Verdad	Impreciso	Falso
Negativos	1	3	3
Positivos	0	4	0

Tabla 6: Diferencia de *Likes* entre la desinformación compartida en Twitter y las publicaciones de los medios verificadores en Instagram.

Fuente: Elaboración propia

Por otro lado, siguiendo la misma línea para el caso de los comentarios:

Todos los medios	Verdad	Impreciso	Falso
Negativos	61	28	69
Neutral	0	1	3
Positivos	57	14	31

FastCheck	Verdad	Impreciso	Falso
Negativos	43	15	43
Neutral	0	0	0
Positivos	57	14	30

Decodificador	Verdad	Impreciso	Falso
Negativos	17	7	23
Neutral	0	0	3
Positivos	0	0	1

FactChecking	Verdad	Impreciso	Falso
Negativos	1	6	3
Neutral	0	1	0
Positivos	0	0	0

Tabla 7: Diferencia de comentarios entre la desinformación compartida en Twitter y las publicaciones de los medios verificadores en Instagram.

Fuente: Elaboración propia

De esta misma forma, se sigue la misma estructura para las comparaciones Instagram con Instagram:

Todos los medios	Verdad	Impreciso	Falso
Negativos	7	9	8
Positivos	22	11	16

FastCheck	Verdad	Impreciso	Falso
Negativos	6	9	7
Positivos	22	10	16

Decodificador	Verdad	Impreciso	Falso
Negativos	1	0	1
Positivos	0	0	0

FactChecking	Verdad	Impreciso	Falso
Negativos	0	0	0
Positivos	0	1	0

Tabla 8: Diferencia de *Likes* entre la desinformación compartida en Instagram y las publicaciones de los medios verificadores en Instagram.

Fuente: Elaboración propia

Por otro lado, siguiendo la misma línea para el caso de los comentarios:

Todos los medios	Verdad	Impreciso	Falso
Negativos	13	9	11
Neutral	0	2	0
Positivos	16	9	13

FastCheck	Verdad	Impreciso	Falso
Negativos	12	9	10
Neutral	0	1	0
Positivos	16	9	13

Decodificador	Verdad	Impreciso	Falso
Negativos	1	0	1
Neutral	0	0	0
Positivos	0	0	0

FactChecking	Verdad	Impreciso	Falso
Negativos	0	0	0
Neutral	0	1	0
Positivos	0	0	0

Tabla 9: Diferencia de comentarios entre la desinformación compartida en Instagram y las publicaciones de los medios verificadores en Instagram.

Fuente: Elaboración propia

4.1.3. Análisis de sentimiento

Como recordatorio, el proceso de análisis de sentimiento se realizó con dos librerías distintas: PySentimiento y Spanish Analyzer Spacy. Esto se realizó de esta manera para evitar obtener conclusiones influidas por el sesgo que podría tener la herramienta a alguna clasificación. Teniendo esto en cuenta, los resultados obtenidos son:

Todos los medios	Verdad	Impreciso	Falso
Positivos	1	0	0
Neutral	14	7	15
Negativos	97	33	79

FastCheck	Verdad	Impreciso	Falso
Positivos	0	0	0
Neutral	11	2	12
Negativos	81	21	54

Decodificador	Verdad	Impreciso	Falso
Positivos	1	0	1
Neutral	2	1	2
Negativos	14	8	21

FactChecking	Verdad	Impreciso	Falso
Positivos	0	0	0
Neutral	1	4	1
Negativos	2	4	4

Tabla 10: Clasificación de sentimiento de publicaciones verificadas. En este caso, la herramienta utilizada es PySentimiento

Fuente: Elaboración propia

Por otro lado, para el caso de Spanish Analyzer Spacy:

Todos los medios	Verdad	Impreciso	Falso
Positivos	2	1	4
Neutral	15	8	14
Negativos	95	31	76

FastCheck	Verdad	Impreciso	Falso
Positivos	2	1	2
Neutral	11	6	10
Negativos	79	16	54

Decodificador	Verdad	Impreciso	Falso
Positivos	0	0	2
Neutral	4	1	3
Negativos	13	8	18

FactChecking	Verdad	Impreciso	Falso
Positivos	0	0	0
Neutral	0	1	1
Negativos	3	7	4

Tabla 11: Clasificación de sentimiento de publicaciones verificadas. En este caso, la herramienta utilizada es Spanish Analyser Spacy

Fuente: Elaboración propia

4.1.4. Participación de Bots

Como recordatorio, se menciona que la detección de bots se realizó utilizando dos librerías distintas: Botometer y BotcheckCL. Cabe destacar que cada librería devuelve un porcentaje de certeza con respecto a la cuenta entregada a analizar, este porcentaje representa la seguridad que tiene la herramienta con respecto a la calidad de la cuenta (bot o no). Con esto en mente, a la hora de clasificar si una cuenta es bot o no hay que tomar una decisión respecto a desde que porcentaje de certeza de la herramienta se debería considerar una cuenta como bot, con el fin de obtener resultados representativos de la realidad. Para determinar este porcentaje, se probó ambas herramientas con cuentas del Gobierno de Chile (presidente y ministros) ya que son cuentas que poseen bastante tráfico, poseen bastante seguidores y tiene una frecuencia de uso bastante altas. Por otra parte, también se probaron las herramientas con cuentas que era públicamente conocidas como bots, ya que de esta forma se podría estimar algún punto de partida para la clasificación de bot o no bot.

Para el caso de Botometer las cuentas del gobierno obtuvieron una certeza entre 0.75 y 0.8, mientras que las cuentas bots obtuvieron una certeza mayor al 0.8, de esta forma se determinó una certeza mínima para la clasificación de bot como 0.8. Por otro lado, para el caso de BotcheckCL las cuentas del gobierno presentaban valores inferiores y cercanos al 0.5, mientras que los bots eran mayor al valor 0.5, por lo que se determinó una certeza mínima de 0.5.

De todas maneras, para evidenciar el comportamiento de los resultados y las conclusiones según la certeza mínima elegida, se procedió a probar cuatro certezas mínimas para cada una de las herramientas comenzando a partir de las certezas mínimas ya mencionadas. Con todo esto en mente, se obtuvieron los siguientes resultados:

>0.80	Retweets	Usuarios	Promedio por usuario	Verdadero	Impreciso	Falso
Bot	16301	1046	15.58	46.61 %	14.05 %	39.34 %
No Bot	32656	2270	14.38	55.14 %	13.29 %	31.57 %

>= 0.81	Retweets	Usuarios	Promedio por usuario	Verdadero	Impreciso	Falso
Bot	4649	309	15.04	33.56 %	16.05 %	50.40 %
No Bot	44308	3007	14.73	54.27 %	13.28 %	32.45 %

>= 0.83	Retweets	Usuarios	Promedio por usuario	Verdadero	Impreciso	Falso
Bot	2013	141	14.27	27.12 %	15.95 %	56.93 %
No Bot	46944	3175	14.78	53.38 %	13.44 %	33.18 %

>= 0.85	Retweets	Usuarios	Promedio por usuario	Verdadero	Impreciso	Falso
Bot	1212	89	13.61	23.18 %	13.54 %	60.81 %
No Bot	47745	3227	14.79	53.04 %	13.48 %	33.48 %

Tabla 12: Resumen de los resultados obtenidos utilizando la aplicación Botometer. En en *header* de la primera columna se encuentra la certeza mínima usada para determinar si una cuenta es un bot o no.

Por otra parte, para el caso de BotCheckCL:

>= 0.5	Retweets	Usuarios	Promedio por usuario	Verdadero	Impreciso	Falso
Bot	15190	934	16.26	51.81 %	13.54 %	34.65 %
No Bot	31314	2197	14.25	52.70 %	13.52 %	33.78 %

>= 0.51	Retweets	Usuarios	Promedio por usuario	Verdadero	Impreciso	Falso
Bot	12130	727	16.69	52.58 %	13.38 %	34.04 %
No Bot	34374	2404	14.29	52.35 %	13.58 %	34.07 %

>= 0.53	Retweets	Usuarios	Promedio por usuario	Verdadero	Impreciso	Falso
Bot	7825	463	16.90	52.64 %	13.58 %	33.78 %
No Bot	38678	2668	14.49	52.36 %	13.51 %	34.12 %

>= 0.55	Retweets	Usuarios	Promedio por usuario	Verdadero	Impreciso	Falso
Bot	4616	266	17.35	53.29 %	12.98 %	33.73 %
No Bot	41888	2865	14.62	52.31 %	13.59 %	34.10 %

Tabla 13: Resumen de los resultados obtenidos utilizando la aplicación BotcheckCL. En en *header* de la primera columna se encuentra la certeza mínima usada para determinar si una cuenta es un bot o no.

4.1.5. Resultados métricas del Caso de estudio: Árboles de propagación.

Para comenzar, se muestran la media, la desviación estándar, el valor mínimo y el valor máximo de las métricas de propagación planteadas en el capítulo anterior.

Profundidad	Media	Desviación estándar	Min	Max
Verdadero	↓ 6.92	8.54	1	78
Falso	9.13	11.54	1	88
Impreciso	9.92	10.10	1	60

Tabla 14: Profundidad en árbol de propagación de respuestas. KS-tests: falso y verdad: $D = 0,168, p \sim 0,08$, falso e impreciso: $D = 0,124, p \sim 0,70$, verdadero e impreciso: $D = 0,275, p \sim 0,01$.

Fuente: [Mendoza *et al.*, 2023]

Tamaño	Media	Desviación Estándar	Min	Max
Verdadero	485	1701	3	14237
Falso	457	669	2	3570
Impreciso	↑ 679	1101	4	5321

Tabla 15: Tamaño en árbol de propagación de respuestas. KS-tests: falso y verdad: $D = 0,206, p \sim 0,01$, falso e impreciso: $D = 0,152, p \sim 0,45$, verdadero e impreciso: $D = 0,195, p \sim 0,17$.

Fuente: [Mendoza *et al.*, 2023]

Anchura Máxima	Media	Desviación Estándar	Min	Max
Verdadero	332	1235	1	10813
Falso	299	469	1	2833
Impreciso	↑ 380	543	1	1883

Tabla 16: Anchura máxima en árbol de propagación de respuestas. KS-tests: falso y verdad: $D = 0,175, p \sim 0,06$, falso e impreciso: $D = 0,135, p \sim 0,61$, verdadero e impreciso: $D = 0,195, p \sim 0,17$.

Fuente: [Mendoza *et al.*, 2023]

Además, se recopila el la información dinámica de cada una de las medidas de tamaño presentadas con anterioridad. Cabe destacar que los siguientes gráficos evidencias los comportamientos promedio de crecimiento a través del tiempo de cada una de las medidas distinguidos según su veracidad:

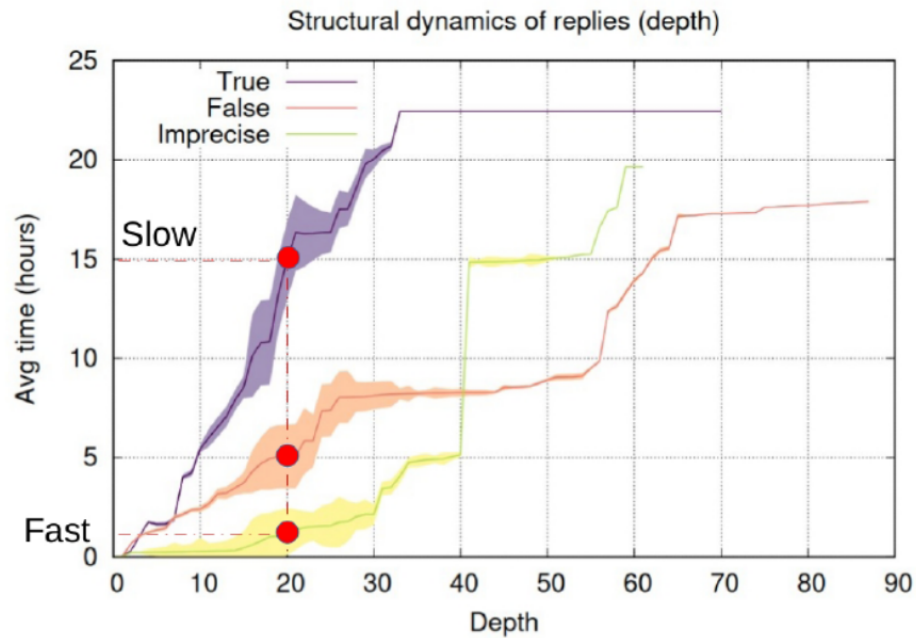


Figura 5: Evolución de la profundidad según tiempo promedio. La línea de color rojo punteada se encuentra en la profundidad 20

Fuente: [Mendoza *et al.*, 2023].

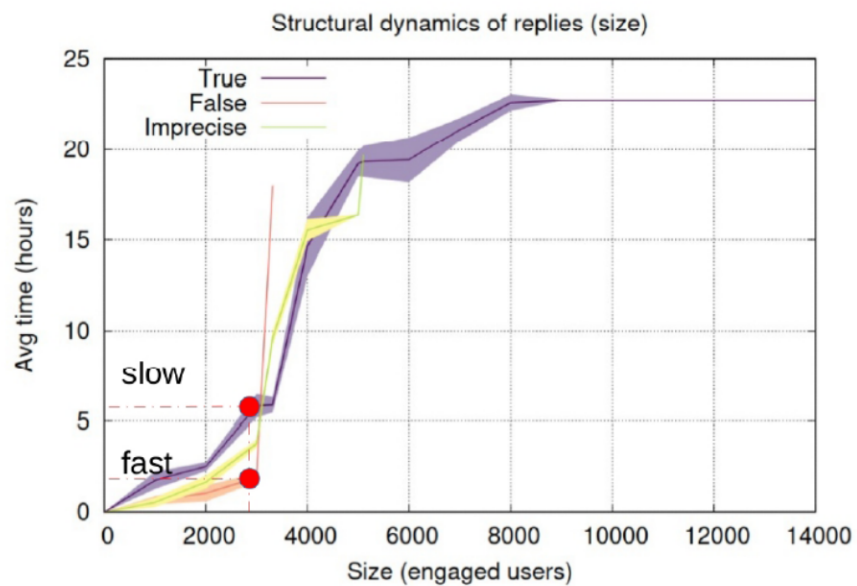


Figura 6: Evolución del tamaño de los arboles según tiempo promedio.

Fuente: [Mendoza *et al.*, 2023].

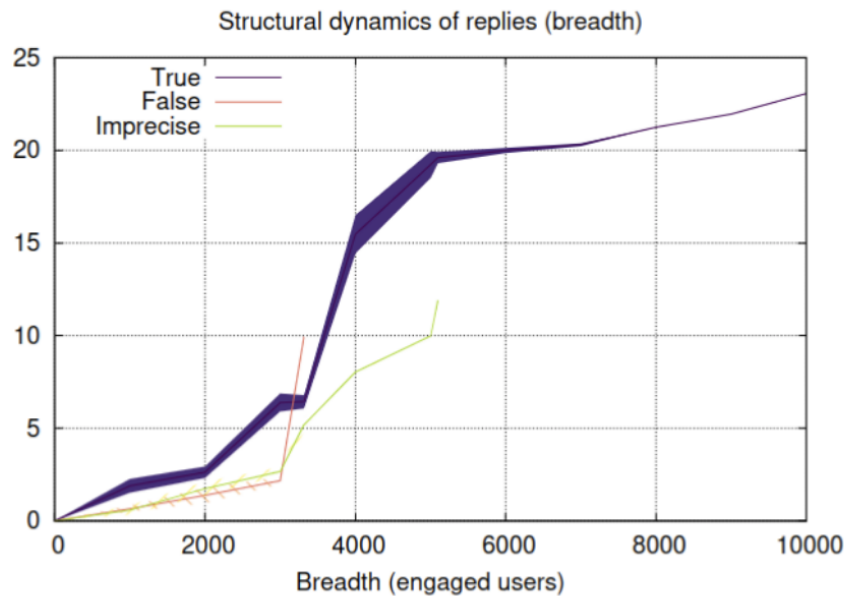


Figura 7: Evolución de la anchura máxima de los árboles según tiempo promedio.
Fuente: [Mendoza *et al.*, 2023].

4.2. Análisis

4.2.1. Análisis resultados Tiempos de respuesta

Para comenzar, podemos evidenciar que la evolución en los tiempos de respuesta entre la Tabla 4 y la Tabla 5 al quitar los *outliers*, reduciendo los tiempos promedios totales de Fastcheck a aproximadamente un 20 % con respecto a los tiempos originales, a aproximadamente un 16 % en el caso de decodificador y a aproximadamente a un 59 % en el caso de FactCheck UC. Por otra parte, podemos además notar la cantidad de *outlier* que se tuvieron son: Para el caso FastCheck correspondieron a un 11.4 % de las noticias, para el caso de Decodificador fueron un 15 % de las noticias y para el caso de FactCheck UC corresponde a un 6.6 % de las noticias. Con esto en consideración, podemos notar que un pequeño porcentaje de verificaciones en el *dataset* original provoca un aumento considerable en los promedios de los tiempos de respuesta de los medios verificadores, esto evidencia de cierta manera que los medios verificadores no siempre logran otorgar públicamente una verificación en cortos periodos de tiempo (entiéndase como corto período menos de una semana). De todas formas, esto no necesariamente debe significar algo negativo, ya que sea a estudiado en el pasado el como retienen las personas en su memoria la veracidad de noticias verificadas, y se ha demostrado que existe una tendencia a recordar con mayor precisión a largo plazo la veracidad de noticias verificadas si es que la persona hizo un juicio respecto a la veracidad de la noticia por ella misma, provocando que la verificación profesional sirva para reafirmar o desmentir creen-

cias previas de la persona con respecto a la información obtenida [Brashier *et al.*, 2021]. Por otra parte, hay que recordar que sobre todo en las redes sociales la importancia de una noticia no suele durar mucho considerando que el flujo de información por estos medios es frenético y dinámico, por lo que tampoco es una buena idea entregar una verificación a una noticia cuando ya dejó de importarle al público general.

Por otro lado, podemos notar que FactCheck UC muestra promedios muy por encima con respecto a los otros medios verificadores, esto ocurre debido a que el medio FactCheck publica sus verificaciones por lotes una vez por semestre ya que estas verificaciones se hacen en un marco de un curso en la Universidad Católica, por lo que no es del todo justo hacer una comparación con respecto a velocidades. Por otra parte, para el caso de Factcheck y Decodificador (considerando los tiempos sin *outliers*) podemos notar una tendencia en que las verificaciones imprecisas suelen ser más lentas de verificar en general, esto de cierta manera reafirma la intuición de que las medias verdades (o medias falsedades) son un espectro complejo de analizar, ya que una media verdad puede implicar que los dichos a verificar se pueden descomponer en una serie de hechos los cuales no todos son ciertos/totalmente ciertos y que incluso pudo haber una manipulación (intencional o no) de la realidad de una situación con el fin de llevar al lector oyente a conclusiones erróneas sobre la información original ¹.

Finalmente, con respecto a FastCheck y Decodificador, podemos notar que entre estos dos medios el medio Decodificador resulta ser más rápido en promedio que FastCheck, esto no necesariamente debe significar que un verificador es más cercano al 'óptimo' o es mejor, debido a que la velocidad de una verificación muchas veces es proporcional a la naturaleza de la noticia verificada. Por ejemplo, regularmente se hacen verificaciones a la Cuenta Pública expuesta por el Presidente de la República todo los años, estas verificaciones suelen ser más rápidas y concisas ya que se analizan afirmaciones puntuales de los dichos del presidente, por lo que es erróneo hacer comparaciones de eficiencia con sólo esta información.

4.2.2. Análisis resultados Diferencia de reacciones.

Con respecto a la diferencia de reacciones de la tabla 6 , podemos notar que existen un 44 % más de noticias con diferencia de reacción positiva en información que resulta ser verdad, un 25 % más de noticias con diferencia de reacción positiva en información que resulta ser imprecisa y un 26 % más de noticias con diferencia de reacción positiva en información que resulta ser falsa. Esto es una noticia positiva desde la perspectiva de los medios verificadores ya que, si consideramos los *likes* como una medida de aprobación publica y alcance, los medios verificadores en conjunto suelen superar a las fuentes de información original. Por otra parte, podemos notar en la tabla 8 también muestra una tendencia positiva para los medios verificadores, con un 51 % más de noticias con diferencia de reacción positiva en información que resulta ser verdad, un 51 % más de noticias con diferencia de reacción positiva en información que resulta ser verdad, un 51 % más de noticias con diferencia de reacción positiva en información que resulta ser verdad, un 10 % más de noticias con diferencia de reacción

¹<https://www.fastcheck.cl/sistema-de-calificacion/>

positiva en información que resulta ser imprecisas y un 33 % más de noticias con diferencia de reacción positiva en información que resulta ser falsas. Notar que esta diferencia positiva ocurre en dos redes sociales fundamentalmente distintas como lo son Twitter e Instagram, por lo que en general se logra tener más alcance/más aprobación. No obstante, cabe destacar que si observamos los datos desagregados de la tabla 6 y 8, la mayor parte del peso de estos resultados se lo lleva el medio verificador FastCheck ya que es el medio más popular y con más verificaciones del país, mientras que el resto de medios presentan una tendencia a tener diferencia de reacciones negativas.

Por otro lado, si observamos las tabla 7 y la tabla 9, podemos notar que los resultados son distintos, en esta ocasión los medios verificadores no siempre poseen diferencias positivas con respecto a la desinformación. Esto no necesariamente es negativo, ya que los comentarios reflejan de cierta manera la discusión generada por la publicación en redes sociales, y si consideramos que las informaciones verificadas suelen ser polémicas (por lo tanto, se prestan para discusiones largas y extensas).

Finalmente, esta tendencia a las diferencias positivas es algo beneficioso para los medios verificadores, debido a que las personas expuestas a verificaciones tienen una tendencia a reducir la probabilidad de creer en *misinformation* [Porter y Wood, 2021].

4.2.3. Análisis resultados Análisis de sentimiento.

Respecto a esta métrica, a partir de la información entregada en la tabla 10 y la tabla 11 podemos determinar que existe una tendencia clara a que la emocionalidad promedio de los comentarios presentes en noticias que fueron verificadas provenientes de Twitter es negativa sin importar la veracidad de la noticia en sí. En el pasado, se comprobó la utilidad de resolver la tarea de *Sentymment Analisis* para poder detectar noticias falsas [Alonso *et al.*, 2021], donde en general se busca identificar los sentimientos expresados en la escritura de la desinformación. En este caso se intentó ver alguna correlación entre las reacciones de los usuarios a cierta información y su veracidad, pero dado los resultados no se puede afirmar que esta correlación exista. Esto puede deberse a la naturaleza de la red social analizada, ya que Twitter es una plataforma la cual en general una persona puede expresarse libremente inclusive si esto implica insultar, usar lenguaje soez o improprios. Esta libertad ha generado que la red social sea bastante tóxica sobre todo en tweets controversiales, ya que las limitaciones a la expresión son casi nulas. Bajo este contexto, es esperable que la reacción de los comentarios a noticias de veracidad dudosa transversalmente evoque sentimientos negativos, ya que de cierta forma responde a la cultura engendrada en la propia red social.

4.2.4. Análisis resultados Bots

Para comenzar, podemos notar en la tabla 12 que la sección con mínima certeza 0.80 los usuarios considerados como Bots realizan en promedio 15.58 retweets los cuales una

46.61 % corresponden a retweets a información verdadera, un 14.05 % corresponde a retweets a información imprecisa y 39.34 % corresponde a retweets a información falsa. Luego, podemos notar que a medida que se aumenta la mínima certeza, existe una clara tendencia a que los retweets promedios por usuario comienzan a disminuir mientras que el porcentaje de noticias falsas aumentó considerablemente. Para el caso de tweets promedio disminuyó desde 15.58 hasta 13.61, mientras que los porcentajes de retweets a noticias falsas aumentó desde 39.34 % a un 60.81 %. Por otro lado, si analizamos la información entregada en la tabla 13 podemos notar una tendencia completamente distinta, en esta ocasión los tweets promedios por usuarios comienzan en 16.26 (mayor promedio que la tabla anterior en su inicio) y termina en un promedio de 17.35 tweets promedios por usuario, mientras que la distribución de veracidad sobre la información retweeteada se mantiene casi igual, siendo alrededor de un 52 % de noticias verdaderas, alrededor de un 13 % de noticias imprecisas y alrededor de un 33-34 % para noticias falsas.

Si bien se ha estudiado en el pasado el uso de bots para la propagación de desinformación, hay que recordar que fundamentalmente los bots en plataformas como Twitter suelen ser cuentas hiper-sociales la cuales tiene una actividad mucho más frecuente y frenética en comparación a una cuenta manejada por un humano, sobre todo a la hora de generar retweets en la plataforma [Himelein-Wachowiak *et al.*, 2021], por lo que resulta contradictorio para el caso de Botometer que a la hora de aumentar la mínima certeza la cantidad de retweets promedio disminuya, ya que si bien el porcentaje de proporción de noticias falsas aumentó (lo cual puede indicar que esos bots fueron utilizados para propagar noticias falsas), el hecho que una de las características principales de los bots que es, por ejemplo, su participación en términos de frecuencia de retweets haya disminuido, lo cual genera ruido a la hora de intentar obtener una conclusión al respecto. Por otra parte, los resultados de BotcheckCL mantiene un promedio de de retweets promedio por usuario creciente, por lo que se mantiene consistente con las funcionalidades básicas de un bot, por otro lado, podemos notar que los usuarios bot mantienen una distribución similar a medida que se aumenta el mínimo de certeza posible, esto se puede explicar en el hecho de que los bots principalmente son utilizados para la propagación de información y viralización, lo cual no necesariamente deben ser siempre de desinformación. Considerando lo explicado anteriormente, existe un tipo de usuario que busca a hacer cyber propaganda con el fin de obtener un beneficio propio propagando desinformación. Estos usuarios suelen ser cuentas falsas manejadas por personas que manipulan fotos e información intencionadamente y buscan llegar a la mayor cantidad de personas posibles, con el fin de desprestigiar, dañar o manipular la opinión pública respecto a un tema en específico [Goswami, 2018]. Con esta nueva variable en mente, existen una alta posibilidad que Botometer esté detectando como bot cuentas que realicen cyber propaganda mientras que BotCheckCL está rescatando efectivamente lo que es un Bot.

4.2.5. Análisis métricas Caso de estudio: Árboles de propagación.

Con respecto a las métricas de propagación, podemos notar que en la tabla 14 los arboles de verdad tienden a ser más superficiales que el resto de árboles, mientras que los arboles de

imprecisión y falsedad muestran una profundidad mucho más alta (y similares entre sí). Esto representa de cierta manera que las discusiones en información falsa e imprecisa era más larga en promedio que a la información verdadera, esto refleja en el fondo que los usuarios discutían más entre ellos en un hilo de conversación con más frecuencia cuando la información tenía presente dudosa veracidad con respecto a su contenido.

Por otro lado, en la tabla 6 podemos notar que el contenido impreciso en promedio llega a una mayor cantidad de personas que otro tipo de contenido. Si consideramos estudios pasados en el tema donde se ha utilizado una clasificación binaria (notando sólo la existencia de la etiqueta *False* para contenido falso e impreciso), podemos notar que este resultado es consistente con estudios previos respecto al tema [Vosoughi *et al.*, 2018], los cuales indican que el contenido falso alcanza a más personas que otro tipo de contenido. Esto puede ocurrir debido a que el contenido falso suele ser polémico y sensacionalista, por lo que la cantidad de comentarios que generen al respecto suele ser mayor al buscar el interés en el usuario que vea la información.

Finalmente, con respecto a las tablas iniciales podemos notar en la tabla 16 que el contenido impreciso en promedio es más ancho que el resto de contenidos.

Con respecto a los gráficos de propagación, podemos notar que si tomamos como punto de referencia la profundidad 20 en el gráfico 5, la velocidad promedio depende bastante dependiendo del tipo de contenido el cuál se esta compartiendo. Por ejemplo, el contenido impreciso toma tan sólo 2 horas en llegar a la profundidad de referencia, mientras que el contenido verdadero toma 15 horas en llegar a esa misma profundidad. De esta manera, podemos además notar que el contenido falso e impreciso llega en promedio a gran profundidad en mucho menos tiempo en comparación al verdadero. Por otra parte, podemos notar en el gráfico 6 que la cantidad de usuarios alcanzados por hora también depende de la veracidad del contenido, ya que por ejemplo para alcanzar 3000 usuarios, el contenido verdadero en promedio necesita 6 horas mientras que el contenido falso necesita tan sólo dos horas. Por otro lado, también podemos notar que la cantidad de usuarios a los que llega el contenido falso e impreciso es mayor al contenido verdadero en sus primeras horas, pero a la larga el contenido verdadero logra alcanzar a mucha más gente que el resto de contenidos. Finalmente, la figura 7 muestra comportamientos similares a los mencionados con anterioridad, donde los contenidos verdaderos suelen ser más lento y llegar a más personas que los contenidos falsos e imprecisos.

Todo este análisis es parte de los resultados obtenidos dentro del proyecto en el cuál se desarrolló esta memoria, y corresponden a una sección dentro de un análisis más grande realizado a partir del dataset creado [Mendoza *et al.*, 2023].

CAPÍTULO 5

CONCLUSIONES

La desinformación es un problema más que vigente en la actualidad, y si bien es un fenómeno que está presente todo el tiempo en redes sociales, está claro que sus momentos de auge principal es en periodos de incertidumbre, polémica, polarización y catástrofe. Este problema se ha mostrado de forma internacional como nacional, y si bien globalmente este fenómeno ha sido profundamente estudiado en múltiples contextos, en Chile la desinformación y los medios verificadores han sido popularizado recientemente en los últimos años. Teniendo en cuenta esta situación, se vio necesario estudiar éste fenómeno en el contexto chileno debido a que como país hemos sufrido diversos eventos que han marcado al país en lapsos pequeños de tiempo, como lo son la pandemia mundial, el estallido social, la nueva constitución, elecciones presidenciales, etc. Y durante estos periodos, la desinformación en los medios digitales proliferó enormemente dentro de los círculos sociales del país.

A lo largo de este estudio, en primer lugar se construyó un dataset el cuál recopiló 1000 verificaciones de distintos medios nacionales las cuales fueron complementadas y homogeneizada, a las cuales se le anexo información de sus respectivos árboles de propagación en el caso de que el contenido verificado se hubiera compartido en la red social Twitter. Este dataset demostró ser lo suficientemente poblado en información como para poder realizar análisis que nos muestren de distintas maneras características de la desinformación propagada a través de las redes sociales chilenas. Por otra parte, respecto a los análisis realizados se pueden obtener las siguientes conclusiones:

- Los medios FastCheck y Decodificador tienen un promedio de menos de una semana para verificar contenido independiente de su veracidad, mientras que Factchecking UC se demora bastante más (esto es debido a la naturaleza del medio periodístico). Este tiempo resulta ser ideal si consideramos que la viralidad y el impacto de los tweets en general ocurren durante los primeros día de existencia, mientras que se ha estudiado en el pasado el beneficio de corregir verificaciones en las personas sobre contenido polémico una vez la persona ya ha emitido un juicio al respecto. De esta manera, se puede decir que el factchecking en FastCheck y Decodificador funcionan de buena manera en términos de tiempos de respuesta.
- Respecto a la diferencia de reacciones, podemos ver con claridad que FastCheck lograr generar consistentemente igual o más impacto en término de likes en sus publicaciones mientras que sus compañeros Decodificador y Factchecking UC no lo logran regularmente debido a su menor popularidad. Por otro lado, en términos de comentarios podemos notar que las diferencias no son tan grandes con respecto al contenido verificado, mostrando que es mucho más difícil causar más discusión en una verificación que en un contenido polémico.
- Respecto a los sentimientos en los contenidos polémicos, se encontró que no existe una relación entre la veracidad de la información y los sentimientos mostrados por los

usuarios en contenido verificado. Esto se explica considerando la naturaleza de Twitter (naturaleza la cuál estimula la libertad de expresión, la cual tiene como consecuencia la proliferación de comentarios con emocionalidad negativa).

- Respecto a los bots dentro del contenido verificado, se pudo notar una disparidad en los resultados dependiendo de la herramienta para detectar bots utilizada. Por un lado, Botometer nos muestra una presencia de bots con tendencia a compartir noticias negativas, la cual sube al aumentar el mínimo de certeza para clasificar al usuarios como bot, esto teniendo efecto que también disminuya la cantidad de retweets promedio por usuario. Por otro lado, botcheckcl al aumentar el mínimo de certeza para clasificar bots, la proporción de retweets por veracidad no cambia manteniéndose constante en el tiempo, por otra parte, la cantidad de retweets promedio aumentó al subir el mínimo de certeza. Considerando el comportamiento esperado de bots y teniendo en consideración la definición de un activista digital, podemos decir que lo que está logrando detectar Botometer es que en los tweets hubo presencia de activistas digitales que buscaron activamente propagar noticias falsas, mientras que botcheckcl mostró la presencia de bots dentro de las cuentas recolectadas.
- Respecto a las métricas de propagación, podemos concluir que existe una tendencia en que el contenidos verdadero suelen ser más lento durante las primeras horas en comparación al contenido falso e impreciso, por otro lado, el contenido verdadero a la larga alcanza mayores tamaños tanto en anchura, profundidad y tamaño. Esta dinámica ya se ha encontrado en el pasado, alineándose con los resultados de los estudios citados en este documento. Finalmente, cabe señalar que los impreciso mostraron comportamientos distintos con respecto al verdadero y falso, lo suficiente para sospechar que quizás sea una categoría que valga la pena considerar al realizar este tipo de estudios, ya que generalmente la información suele considerarse como verdadera o falsa, dejando de lado las medias verdades.

5.1. Trabajo Futuro

Esta investigación tiene diversas limitaciones que podrían superarse a futuro. Una de ellas es la cantidad de datos obtenidos, se lograron recolectar aproximadamente mil noticias chilenas a lo largo del periodo de investigación las cuales se reparten en tres medios verificadores. Sería bastante beneficioso para este tipo de análisis el poder tener una base más poblada (sobre todo de contenido compartido en redes sociales), para eso sería bueno lograr recopilar contenido verificado de más medios verificadores chilenos (como por ejemplo, el polígrafo de El Mercurio).

Por otra parte, los arboles de retweets resultaron ser una aproximación al comportamiento real de la desinformación, pero en general no hay ninguna seguridad que en esa aproximación no se haya perdido información estructural importante, por lo que una ventana de mejora es desarrollar alguna metodología que permita reconstruir los arboles de forma completa.

Otro aspecto interesante que no se indagó en esta investigación fue el comportamiento de la información en medio más tradicionales o en otros medios que no sean RRSS clásicas, este tipo de análisis suelen ser más complejos debido al difícil acceso a la información pero quizás a partir de la información recopilada en el dataset creado se tenga un comienzo.

Un evento interesante derivado en de estudio es que al considerar las medias verdades podemos notar que, por si solas, tienen características lo suficientemente representativas y distintas para demostrar que su comportamiento es diferente a los otros tipos de veracidades, por lo que se estima necesario comenzar a plantearse el análisis independiente de las medias verdades a la hora de realizar estudios relacionados a la desinformación en medios de comunicación.

Finalmente, cada métrica expuesta se podría indagar más con el fin de detallar de mejor manera el comportamiento de la desinformación/verificaciones, por ejemplo, el análisis de sentimiento a la respuestas podría ampliarse a realizar una análisis de postura de los comentarios con respecto al contenido verificado y buscar si existe una relación de la postura con la veracidad del contenido. Otro aspecto que podría ser interesante es caracterizar a los usuarios presentes en los arboles de propagación, quizás detectando el tipo de persona, posición política o rango etario e investigar si existe una relación de estas características y el contenido que comentan/comparten. De esta manera, lo que se quiere es recalcar el dataset tiene distintos vértices de los cuales puede ser explotado, por lo tanto las posibilidades son inmensas.

REFERENCIAS BIBLIOGRÁFICAS

- [Allcott y Gentzkow, 2017] Allcott, H. y Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36.
- [Alonso et al., 2021] Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., y Vilares, J. (2021). Sentiment analysis for fake news detection. *Electronics*, 10(11).
- [Amazeen, 2015] Amazeen, M. A. (2015). Revisiting the epistemology of fact-checking. *Critical Review*, 27(1):1–22.
- [Augenstein et al., 2019] Augenstein, I., Lioma, C., Wang, D., Lima, L. C., Hansen, C., Hansen, C., y Simonsen, J. G. (2019). Multifc: A real-world multi-domain dataset for evidence-based fact checking of claims. *arXiv preprint arXiv:1909.03242*.
- [BABACAN y TAM, 2022] BABACAN, K. y TAM, M. S. (2022). Sosyal medyanın bilgi savaşı rolü: Rusya – ukrayna savaşında sahte haberler. *Erciyes İletişim Dergisi*, (3):75–92.
- [Bastick, 2021] Bastick, Z. (2021). Would you notice if fake news changed your behavior? an experiment on the unconscious effects of disinformation. *Computers in Human Behavior*, 116:106633.
- [Brashier et al., 2021] Brashier, N. M., Pennycook, G., Berinsky, A. J., y Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, 118(5):e2020043118.
- [Cresci, 2020] Cresci, S. (2020). A decade of social bot detection. *Communications of the ACM*, 63(10):72–83.
- [Goswami, 2018] Goswami, M. (2018). *Fake News and Cyber Propaganda: A Study of Manipulation and Abuses on Social Media*, pp. 535–544.
- [Himelein-Wachowiak et al., 2021] Himelein-Wachowiak, M., Giorgi, S., Devoto, A., Rahman, M., Ungar, L., Schwartz, H. A., Epstein, D. H., Leggio, L., y Curtis, B. (2021). Bots and misinformation spread on social media: A mixed scoping review with implications for COVID-19 (preprint). *Journal of Medical Internet Research*.
- [Küçük y Can, 2020] Küçük, D. y Can, F. (2020). Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.
- [Lim, 2018] Lim, C. (2018). Checking how fact-checkers check. *Research & Politics*, 5(3):2053168018786848.
- [Mendoza et al., 2023] Mendoza, M., Valenzuela, S., Núñez-Mussa, E., Padilla, F., Providel, E., Campos, S., Bassi, R., Riquelme, A., Aldana, V., y López, C. (2023). A study on information disorders on social networks during the chilean social outbreak and covid-19 pandemic. *Applied Sciences*, 13(9).

- [Nuñez y Fernández, 2019] Nuñez, E. y Fernández, F. (2019). Chile. *Digital News Report*, p. 126–127.
- [Nuñez y Fernández, 2020] Nuñez, E. y Fernández, F. (2020). Análisis digital news report 2020: Chile. <https://bit.ly/3oqrm3L>.
- [Pierri y Ceri, 2019] Pierri, F. y Ceri, S. (2019). False news on social media: a data-driven survey. *ACM Sigmod Record*, 48(2):18–27.
- [Porter y Wood, 2021] Porter, E. y Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in argentina, nigeria, south africa, and the united kingdom. *Proceedings of the National Academy of Sciences*, 118(37):e2104235118.
- [Providel y Mendoza, 2021] Providel, E. y Mendoza, M. (2021). Misleading information in spanish: a survey. *Social Network Analysis and Mining*, 11(1):1–26.
- [Ribeiro et al., 2017] Ribeiro, M. H., Calais, P. H., Almeida, V. A., y Meira Jr, W. (2017). “everything i disagree with is# fakenews”: Correlating political polarization and spread of misinformation. *arXiv preprint arXiv:1706.05924*.
- [Rocha et al., 2021] Rocha, Y. M., de Moura, G. A., Desidério, G. A., de Oliveira, C. H., Lourenço, F. D., y de Figueiredo Nicolete, L. D. (2021). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. *Journal of Public Health*, 31(7):1007–1016.
- [Shahi y Nandini, 2020] Shahi, G. K. y Nandini, D. (2020). Fakecovid—a multilingual cross-domain fact check news dataset for covid-19. *arXiv preprint arXiv:2006.11343*.
- [Sharma et al., 2019] Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., y Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(3):1–42.
- [Swire-Thompson y Lazer, 2019] Swire-Thompson, B. y Lazer, D. (2019). Public health and online misinformation: Challenges and recommendations. *Annual Review of Public Health*, 41:433–451.
- [Varol et al., 2017] Varol, O., Ferrara, E., Davis, C., Menczer, F., y Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. 11(1).
- [Vosoughi et al., 2018] Vosoughi, S., Roy, D., y Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.
- [Yang et al., 2022] Yang, K.-C., Ferrara, E., y Menczer, F. (2022). Botometer 101: social bot practicum for computational social scientists. *Journal of Computational Social Science*, 5(2):1511–1528.

[Zhou y Zafarani, 2020] Zhou, X. y Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.