

# La presencia de los *bots* políticos en Twitter durante la crisis de la COVID-19 en España

*The Presence of Political Bots on Twitter during the COVID-19 Crisis in Spain*

Ángela Martínez Torralba, Juan Antonio Guevara Gil y Aitor Jiménez de la Fuente

## Palabras clave

*Astroturfing*

- Bots políticos
- COVID-19
- Debate
- Incivility
- Polarización
- Propaganda
- Twitter

## Key words

*Astroturfing*

- Political Bots
- COVID-19
- Debate
- Incivility
- Polarization
- Propaganda
- Twitter

## Resumen

La crisis sanitaria provocada por la pandemia de COVID-19 ha conllevado el decreto del estado de alarma en España y un severo confinamiento domiciliario que se fue suavizando en fases de desescalada. Durante este periodo, las redes se utilizaron aún más como herramienta de debate. Lo que se pretende con esta investigación es averiguar si se utilizaron herramientas de IA para que interviniesen en el debate político de Twitter. Para ello, se aplicaron algoritmos que determinan la presencia de *bots* en la conversación, sus roles comunicativos y su relación con los principales partidos políticos. Los datos extraídos demuestran que se crearon campañas de desinformación ejecutadas por *bots* con el objetivo de manipular la opinión pública.

## Abstract

The crisis caused by the COVID-19 pandemic has led to the decree of the State of Alarm in Spain and a severe home confinement that was softened in phases of de-escalation. During this period, social networks were used as a discussion tool. With this research we intend to find out if artificial intelligence tools were used in the political debate on Twitter. To achieve this, algorithms that determine the presence of bots in the conversation, their communicative roles and their relationship with the main political parties, were applied. The results show that disinformation campaigns were created by bots with the aim of manipulating public opinion.

## Cómo citar

Martínez Torralba, Ángela; Guevara Gil, Juan Antonio; Jiménez de la Fuente, Aitor (2023). «La presencia de los *bots* políticos en Twitter durante la crisis de la COVID-19 en España». *Revista Española de Investigaciones Sociológicas*, 181: 61-80. (doi: 10.5477/cis/reis.181.61)

La versión en inglés de este artículo puede consultarse en <http://reis.cis.es>

**Ángela Martínez Torralba:** Universidad Complutense de Madrid | [angela19@ucm.es](mailto:angela19@ucm.es)

**Juan Antonio Guevara Gil:** Universidad Complutense de Madrid | [juanguiev@ucm.es](mailto:juanguiev@ucm.es)

**Aitor Jiménez de la Fuente:** Universidad Complutense de Madrid | [aitorj01@ucm.es](mailto:aitorj01@ucm.es)

## INTRODUCCIÓN<sup>1</sup>

En muchos sentidos, la incorporación de Twitter en la vida diaria de los ciudadanos ha supuesto una revolución para los espacios de discusión clásicos. Las estrategias políticas de mejora de imagen y posicionamiento de ciertos principios ideológicos en la conversación pública de las redes sociales no han parado de evolucionar. Entre estas nuevas estrategias, una de las más extendidas durante los últimos años es el uso de herramientas de inteligencia artificial (IA en adelante) diseñadas para alterar y distorsionar el impacto de determinados mensajes. Esta investigación está centrada en analizar cómo se han utilizado estas tecnologías y cuál ha sido su papel durante la circunstancia de mayor relevancia en la historia reciente de España: la pandemia de COVID-19 y la consiguiente declaración del estado de alarma en todo el país.

La declaración del estado de alarma trajo consigo un severo confinamiento domiciliario que tuvo lugar entre el 15 de marzo y el 28 de abril de 2020. A partir de ese momento se entró en un periodo de desescalada por fases con las que se comenzaron a relajar las medidas de contención del virus hasta el 21 de junio de 2020.

En este momento crucial, los medios sociales cobraron especial relevancia. Por ello, generar ventaja con respecto al posicionamiento a favor o en contra de los usuarios de Twitter en el debate público en torno a ciertas ideas políticas se ha convertido en uno de los objetivos principales de los partidos (Howard *et al.*, 2017; Bradshaw y Howard, 2018). En los últimos años se ha podido observar un mayor uso de *bots*, para distorsionar la realidad, conseguir beneficios e influenciar a los ciudadanos (Neyazi, 2019; Santana y

Huerta, 2019; Yan *et al.*, 2020; Bradshaw y Howard, 2020).

Twitter se ha convertido en un espacio aprovechable para la implantación de estas herramientas. Su estructura propicia una esfera en la que se llega a generar polarización y separación entre grupos enfrentados (Neyazi, 2019). Además, en algunos casos, se crean momentos de especial tensión entre grupos que tienden a utilizar un lenguaje inapropiado, lo que ha sido denominado como *incivility* (Papacharissi, 2004).

En el marco de esta situación cabe preguntarse qué suponen los *bots* para el debate público que albergan las redes sociales. ¿Cuál es su rol principal?, ¿qué relación tienen con los partidos políticos?, ¿puede tener algunas consecuencias?

El asentamiento del uso de estas herramientas como prácticas de propaganda empleadas en todo el mundo (Bradshaw y Howard, 2020) nos lleva a plantear la hipótesis de que durante el periodo del estado de alarma en España se utilizaron algoritmos de IA en Twitter para condicionar la conversación política. Lo que se pretende conocer en concreto es si se crearon campañas de desinformación ejecutadas por grupos de *bots* organizados estratégicamente y vinculados con los agentes sociopolíticos tradicionales cuyo objetivo es el de generar una falsa sensación de apoyo o desaprobación de los temas políticos discutidos durante el periodo señalado.

Para ello se detectará la presencia de *bots* políticos en Twitter mediante algoritmos diseñados con este fin durante los meses de confinamiento y las fases de la desescalada. Posteriormente, se definirá el rol comunicativo que ocupan los *bots* en el debate político de la red social y se determinará la relación de los *bots* con los principales actores políticos.

En este trabajo se hace, en primer lugar, una introducción teórica sobre el nuevo contexto digital y la esfera política en la

<sup>1</sup> Financiación: Esta investigación está parcialmente apoyada por el Gobierno de España, Plan Nacional de I+D+i, PID2019-106254RB-I00.

que se enmarca nuestro objeto de estudio. También se repasa la literatura existente sobre el uso de *bots* en este nuevo entorno digital y los principales usos de las inteligencias artificiales en campañas de desinformación política. Después se muestra la metodología utilizada para llevar a cabo la investigación empírica. Por último, se exponen los resultados obtenidos sobre la existencia de campañas de desinformación en Twitter durante el periodo del estado de alarma, así como la conclusión final del estudio.

## EL NUEVO ENTORNO DIGITAL EN LA ESFERA POLÍTICA

Desde los últimos años del siglo xx se comenzó a advertir que las tecnologías de la información y la comunicación (TIC) y, más concretamente, las tecnologías web o redes sociales, se convirtieron en uno de los grandes catalizadores de la opinión pública además de un elemento de socialización imprescindible (Gallardo y Enguix, 2016). Este sistema de «autocomunicación de masas» (Castells, 2012) permite que cualquier actor esté capacitado para decidir el mensaje a emitir, designar los posibles receptores y seleccionar qué mensajes quiere recuperar y así informar, movilizar, apoyar o debatir sin ser necesariamente un participante directo (Castells, 2012; Dader y Domínguez, 2017).

A la vista de semejantes posibilidades comenzó a surgir un halo de revitalización para los sistemas democráticos que imaginaban que sectores cada vez más amplios y diversos de la población podrían acceder con mayor facilidad al debate público (Winner, 2003). En cierto sentido, se entendieron las redes como canal primordial de difusión de noticias y deliberación democrática donde el rol cada vez más relevante de líderes de opinión y ciudadanos pudiera permitir enmarcar temas y condi-

cionar la agenda de los medios convencionales (Bennett y Segerberg, 2012; Neuman *et al.*, 2014; Bode *et al.*, 2015; Stier *et al.*, 2018; Yan *et al.*, 2020).

Desde esta perspectiva la ciudadanía puede desafiar el monopolio que los agentes sociopolíticos tradicionales mantienen sobre los medios de comunicación convencionales (Rheingold, 2009; Castells, 2012). Los ciudadanos, de forma individual o colectiva, son los que ganan entidad frente a los actores convencionales y son capaces de generar una esfera pública híbrida (Benkler, 2006; Santana y Huerta, 2019).

Sin embargo, y a pesar de las investigaciones que afirman que existe una correlación positiva entre el uso de redes sociales y el compromiso cívico y político de la ciudadanía (Boulianne y Theocharis, 2018; Boulianne, 2018), hay otras corrientes que apuestan por la teoría de que aquellas personas dispuestas a participar políticamente *online* solo son aquellas que también lo están en el mundo real. Según estas investigaciones, se entiende que Internet no es todavía un factor de transformación colectiva en el que los instrumentos digitales sean la base de la organización social (Margolis y Resnick, 2000). El papel de la política es el de siempre, pero trasladado a otros escenarios en los que, además, y como se mostrará más adelante, su propia dinámica puede facilitar ciertos comportamientos que enturbian la visión horizontal e incluyente de esta herramienta (Margolis y Resnick, 2000).

Actualmente, la confrontación entre unas enormes expectativas de partida con una realidad que no para de atenuar estos avances ha desembocado en un prolífico campo de investigación. Se pueden observar tres grandes líneas de investigación al respecto (Ward y Gibson, 2009):

En primer lugar, y siendo probablemente la perspectiva más conservadora de las tres, está la teoría de la «normali-

zación». Dicha teoría asume que existen evidencias suficientes como para afirmar que los patrones comunicativos del mundo *offline* se están trasladando al mundo *online* (Margolis y Resnick, 2000). Por lo tanto, los sujetos políticos que tradicionalmente han mantenido una posición privilegiada continúan manteniendo esta ventaja competitiva en Internet bajo el argumento de «*political as usual*» (Margolis y Resnick, 2000; Molyneux y Mourão, 2017). En segundo lugar, y partiendo de unos supuestos totalmente distintos, surge la teoría de la «desintermediación». Esta interpretación asume que las nuevas tecnologías fomentan la producción y distribución de contenidos y mensajes dentro de una audiencia potencialmente global (Benkler, 2006). Su principal propiedad es la capacidad para reducir el coste de participar, haciendo decrecer el poder de los mediadores clásicos y fomentando un modelo de comunicación más horizontal (Benkler, 2006; Robles y Córdoba, 2019). Por último, surge un modelo híbrido conocido como «igualación» o «nivelación». Esta teoría se fundamenta en una simbiosis entre un sistema comunicativo tradicional (basado en la verticalidad y la estructura convencional de poder) y un sistema digital (basado en la horizontalidad y la acción conectiva) (Benkler, 2006; Balcells y Cardenal, 2013). Esto quiere decir que la lógica de Internet puede contribuir a nivelar las diferencias tanto de recursos como de alcance social entre ciudadanos, medios de comunicación y partidos o líderes políticos, asimilando sus niveles de visibilidad (Balcells y Cardenal, 2013).

Es dentro de las teorías de la hibridación donde podríamos situar las estrategias de carácter político entre las que sobresale el uso masivo en redes sociales de herramientas de IA y *bots*. A pesar de que las redes sociales son un instrumento que garantizan la escucha activa de voces que de otra forma no serían escuchadas y que ponen

en jaque el sistema de mediaciones tradicionales, lo cierto es que cada vez más los agentes sociopolíticos tienden a desplazar las formas comunicativas y de propaganda existentes en el mundo real a la esfera *online* (Margolis y Moreno-Riaño, 2009).

## LOS *BOTS* POLÍTICOS EN EL NUEVO ESCENARIO DIGITAL

En términos de participación, el uso masivo de las redes sociales ha provocado un incentivo en los usuarios de tomar partido en el debate público (Cáceres, Brändle y Ruiz, 2017). Son el centro del nuevo sistema de comunicación híbrido, que bebe de los medios tradicionales y que extiende el ámbito político hacia ecologías multimedia en las que la atención a las noticias suele ser personalizada y cualquier usuario puede convertirse en creador de contenido (Bennett, Segerberg y Yang, 2018; Santana y Huerta, 2019). De hecho, hay investigadores que se inclinan hacia la idea de que las redes sociales son el barómetro de la opinión pública y que actúan como amplificadoras de lo recogido por los medios convencionales (Neyazi, 2019).

Este acontecimiento, sumado al hecho de que contar con presencia *online* puede darse por un coste económico mínimo, ha traído consigo un considerable aumento del aprovechamiento de este instrumento de comunicación social por parte de los poderes políticos convencionales para desplegar sus distintas estrategias (Pastor *et al.*, 2020). El uso de herramientas de IA y de microsegmentación que explotan y utilizan la información privada de los usuarios es parte de este abanico de nuevas tácticas propagandísticas. Se define la propaganda como el uso intencionado de información, verídica o falsa, que influye en la actitud ciudadana ante cierto asunto (Persily, 2017). Esto, extrapolado al ámbito digital, se puede entender como el uso de instrumentos infor-

máticos deliberadamente creados con la intención de manipular la opinión de los usuarios (Neyazi, 2019).

Los *bots* y otras herramientas de IA juegan un papel esencial en cuanto al desarrollo de las principales estrategias propagandísticas en redes sociales. Pero en realidad no todos los *bots* son diseñados con fines maliciosos. Están aquellos que son meros difusores de información, los *bots* sociales, encargados de publicar contenido periódico e interactuar con usuarios (Yan *et al.*, 2020). Sin embargo, como contrapunto, encontramos a los *bots* políticos. Son cuentas automatizadas creadas a raíz de algoritmos que reproducen el comportamiento humano y cuya misión es la de generar falsa sensación de apoyo hacia un partido, candidato o causa política, además de difamar a sus oponentes y a aquellos que expongan opiniones contrarias, afectando a los ritmos naturales del debate en redes sociales (Yan *et al.*, 2020).

En comparación con los *bots* sociales, los políticos destacan por la difusión masiva de noticias falsas (Shao *et al.*, 2018; Yan *et al.*, 2020), exhibir mensajes con sentimiento negativo (Stella, Ferrara y De Domenico, 2018; Neyazi, 2019; Yan *et al.*, 2020) y por explotar la información privada de los usuarios con fines partidistas (Boshmaf *et al.*, 2013; Yan *et al.*, 2020). Hwang y Wooley (2016) clasifican los *bots* políticos en dos categorías: los controladores, que son los que se encargan de manipular y engrosar o bloquear la conversación en torno a un tema; y los facilitadores, que comparten y difunden contenido con el objetivo de que los mensajes alcancen un gran número de visualizaciones y lleguen a convertirse en virales.

Es posible que los instrumentos de IA puedan operar e interactuar con usuarios humanos debido a la falta de filtros individuales (Yan *et al.*, 2020; Price, Priisalu y Nomin, 2019). A pesar de que gran parte de

las investigaciones se centran en la creación de algoritmos computacionales de detección de *bots* (Yan *et al.*, 2020; Neyazi, 2019), se profundiza poco en el asunto de la distinción humana entre cuentas reales o artificiales. Yan *et al.* (2020) dan a conocer en su estudio que las redes sociales están siendo masivamente adulteradas por algoritmos maliciosos de manipulación política porque los humanos carecemos de instrumentos de diferenciación entre usuario *bot* o persona. También llegaron a la conclusión de que el sesgo político juega un papel más que relevante en cuanto a la percepción de este asunto. Las cuentas automáticas cuya afinidad tiende hacia partidos de derecha o extrema derecha son más efectivas no solo porque sean más difíciles de detectar por las personas, sino porque los usuarios reales conservadores son más vulnerables ante la difusión de noticias falsas (Yan *et al.*, 2020).

Respecto de lo anterior, existen taxonomías que clasifican el tipo de comportamiento de las personas en relación con los *bots*. Por un lado, están los usuarios activos, que son aquellos que siguen a la cuenta automática sin haber sido seguidos antes por ella; en segundo lugar, encontramos a los reactivos, que interaccionan ante las acciones del *bot*; y, por último, los usuarios inactivos que ni responden ni siguen a este tipo de cuentas (Fazil y Abulaish, 2017).

Es posible que los niveles de vulnerabilidad aumenten ante este tipo de propaganda debido a que los ataques de los *bots* políticos están estratégicamente cada vez mejor organizados y dirigidos a audiencias más segmentadas en un entorno polarizado y crispado como el de las redes sociales (Stella, Ferrara y De Domenico, 2018; Yan *et al.*, 2020). Además, los algoritmos son cada vez más capaces de modificar o mejorar su comportamiento con el propósito de no ser descubiertos por las diferentes herramientas de detección (Price, Priisalu y Nomin, 2019).

La irrupción en la Era de la Posverdad (Keyes, 2004), un escenario en el que la falsedad es asimilada por la ciudadanía y aceptada como verdad (Murolo, 2019; Orbegozo, Morales y Larrondo, 2020), es el caldo de cultivo idóneo para la injerencia de *bots* políticos. En este sentido, el concepto de posverdad y el uso de herramientas de IA están estrechamente ligados a la desinformación en redes sociales, ya que es aquí donde las noticias falsas y la propaganda política encuentran su brecha de expansión (Price, Priisalu y Nomin, 2019).

## **ASTROTURFING Y CAMPAÑAS DE DESINFORMACIÓN**

La campaña electoral de las elecciones presidenciales estadounidenses de 2016, así como la del referéndum británico para su consiguiente salida de la Unión Europea, supusieron un cambio de mentalidad en cuanto a la conciencia global de la existencia de manipulación de redes sociales a escala mundial (Keller *et al.*, 2019; Neyazi, 2019). La interferencia rusa en ambos procesos, dos de los más relevantes de los últimos años, ha generado una creencia ciudadana acerca del peligro democrático que supone la intromisión de instrumentos de IA en el debate en redes (Keller *et al.*, 2019). Más de un 66 % de los participantes en un estudio realizado durante la campaña a las elecciones estadounidenses de 2016 (Stoking y Sumida, 2018) afirmó conocer y sentirse intranquilo con respecto a este asunto y un 80 % era consciente de sus efectos negativos.

Aunque existe cierta ambigüedad sobre el concepto de desinformación y las categorías que entrarían dentro de sus parámetros —noticias falsas, *misinformation* o *malinformation*— podemos entenderla como una consecuencia perjudicial a la que tienen que enfrentarse los ciudadanos en las redes sociales. Asimismo, el hecho de en-

gañar a la audiencia *online* volcando en la web infinidad de algoritmos maliciosos de manipulación política es desinformación, aunque la información que se difunda sea veraz, puesto que erosiona e interfiere en los ciclos de debate democrático entre personas reales (Keller *et al.*, 2019). Teniendo esto en cuenta, el *astroturfing* puede ser considerado como una campaña de desinformación.

Con este término nos referimos a aquellas campañas de redes sociales que parecen ser llevadas a cabo por activistas de base o *grass-roots* a favor o en contra de una causa política y que en realidad son un amplio conjunto de cuentas automáticas que actúan de forma estratégicamente organizada emulando el comportamiento humano (Howard, 2006; Walker, 2014). Con estas operaciones lo que se pretende es generar artificialmente una falsa sensación de apoyo o desacuerdo en torno a una idea en la que no existe consenso alguno con el fin de influenciar a las audiencias de las redes sociales con propósitos políticos y propagandísticos (Santana y Huerta, 2019). En casi toda la literatura en torno al tema existen diferentes términos para referirse a este mismo concepto, como *botnets* (Boshmaf *et al.*, 2013), *cibertropas* (Bradshaw y Howard, 2018) o *Twitter bombs* (Pastor *et al.*, 2020).

En la práctica, estos grupos coordinados de robots pueden hacer lo mismo que un movimiento social solo que más rápido, a escala masiva y a menor coste (Yan *et al.*, 2020). No obstante, a diferencia de los *grass-roots* que no cuentan con una organización centralizada y, por lo tanto, sus consignas son más variadas y se expanden más prolongadamente en el tiempo, las *cibertropas* se definen por publicar de forma simultánea mensajes casi idénticos o por difundir conjuntamente aquellos que son congruentes con el objetivo de la campaña que se esté llevando a cabo (Keller *et al.*, 2019).

Por otro lado, estas maniobras suelen funcionar mediante el hostigamiento y la exclusión del debate de grupos de opinión minoritarios o contrarios. La información falsa, los mensajes de odio y el uso de un lenguaje incívico son características frecuentes en este tipo de campañas (Keller *et al.*, 2019; Santana y Huerta, 2019). Tal y como apuntan Santana y Huerta (2019), el *astroturfing* está estrechamente vinculado con la teoría de la espiral del silencio según la cual los individuos omiten dar su opinión cuando se perciben en minoría (Noelle-Neuman, 1984).

Existen ejemplos de la incidencia de *cibertropas* en todo el mundo. Según el «Inventario Global de la Manipulación Organizada de Redes Sociales» del año 2019 (Bradshaw y Howard, 2020), se han detectado campañas organizadas de manipulación en 70 países de todo el mundo en las que intervienen herramientas de IA. En 2018 el mismo estudio las descubrió en 48 países y en 28 países el año anterior. Este fenómeno no solo es una tendencia al alza a nivel mundial, sino que en cada una de las naciones investigadas existe al menos una formación política que hace uso de esta estrategia de forma premeditada y organizada para manipular la opinión pública (Bradshaw y Howard, 2020). Sin embargo, su uso es diferente dependiendo del país del que se trate, aunque cada vez es más sofisticada no solo su detección, sino también sus niveles de interacción con las cuentas humanas (Santana y Huerta, 2019).

Según Pastor *et al.* (2020), durante la campaña electoral de las elecciones generales de 2019 en España apareció un número de *bots* significativamente alto. Además, según su comportamiento, no solamente eran capaces de generar un sentimiento negativo, sino que se pudieron apreciar evidencias de organización en la difusión y propagación de noticias falsas y mensajes de odio en los grupos de cuentas automáticas asociadas a cada partido.

Este tipo de fenómenos no solo son perjudiciales para el transcurso del debate en redes sociales, sino que su consecuencia más inmediata puede ser que la conversación política desemboque en terrenos altamente polarizados. De acuerdo con Sartori (2005), la polarización no tiene por qué tener connotaciones negativas al contemplar dos tipos de polarización: la polarización centrípeta es aquella donde los grupos opuestos están abiertos al diálogo y se agrupan en torno al centro; sin embargo, con la centrífuga las comunidades se dividen y sitúan en torno a los extremos, evitando el diálogo y produciendo consecuencias democráticas nocivas. Este último modelo de polarización no solo es el más probable, sino que teniendo en cuenta los niveles de *incivility* (Rowe, 2015) y de sentimientos negativos con los que inundan la red, también se podría convertir en la más peligrosa.

Sumado a esto, está el asunto de que la detección de las campañas de *astroturfing* y de cuentas falsas depende por entero del criterio de los usuarios (Yan *et al.*, 2020). La falta de gestión a este respecto de los CEO de los principales medios sociales (Yan *et al.*, 2020), así como el papel de espectadores que juegan los Estados frente a la desinformación orquestada por los actores políticos, no hacen más que crear en la ciudadanía sentimientos extremos y de impunidad ante los mensajes de odio (Neyazi, 2019).

Teniendo en cuenta la inesperada crisis sanitaria por la pandemia de la COVID-19, nos instalamos en un escenario en el que las herramientas maliciosas de IA implantadas en el debate de Twitter pueden generar un ambiente aún más nocivo de polarización y de *incivility*. Por este motivo, nos planteamos estudiar la intervención de *bots* en la conversación política durante el estado de alarma y destacamos sus consecuencias.

## METODOLOGÍA

Teniendo en cuenta lo expuesto anteriormente, las hipótesis que se plantean son:

H1: Durante el periodo del confinamiento domiciliario y las fases de la desescalada en España se utilizaron algoritmos de IA en Twitter para condicionar la conversación política.

HS1: Se crearon campañas de desinformación ejecutadas por grupos de *bots* organizados estratégicamente y vinculados a los agentes sociopolíticos tradicionales cuyo objetivo era el de generar una falsa sensación apoyo o desaprobación de los temas políticos discutidos durante el periodo señalado.

Para resolverlas, los objetivos de investigación que se proponen son:

- Conocer la presencia de *bots* políticos en la red social Twitter durante la crisis del coronavirus en España.
- Definir el papel de los *bots* en la red social, así como su relación con los principales partidos políticos de España.
- Describir cuáles podrían ser las posibles consecuencias de la presencia de *bots* en el debate público *online* que tiene lugar en Twitter durante el periodo estudiado.

### Caso de estudio y fuentes de datos

Desde el 15 de marzo hasta el 21 de junio de 2020 en España se instaura el estado de alarma a causa de la pandemia de la COVID-19, lo que conlleva un severo confinamiento domiciliario y una posterior desescalada por fases. Esta situación social de excepción ha supuesto para todas las autoridades del mundo, y en concreto para el recién formado gobierno de coalición español, una prueba de administración sanitaria y política de una envergadura desconocida hasta la fecha. Bajo esta coyuntura, el debate público se volcó de manera más

acentuada en las redes sociales. No solo se convirtieron en un espacio en el que la conversación giraba en torno a la situación sanitaria, sino que la gestión política por parte del nuevo gobierno fue uno de los asuntos más recurrentes.

El caso de estudio de este trabajo se centra en aquellos mensajes publicados en Twitter cuya temática aborda el debate político en relación a la pandemia de COVID-19. El periodo de tiempo seleccionado es el de la duración del confinamiento domiciliario y las fases de la desescalada en España.

Los datos se descargaron de la API de Twitter a través de R-Studio, usando el paquete «rtweet» (Kearney, 2019). En total, se realizaron cinco tandas de descargas, de una semana de duración cada una, a lo largo de todo el periodo seleccionado con el fin de abarcar, de forma general, todo el proceso:

- 1.<sup>a</sup> tanda de descarga: «2020-03-16» - «2020-03-23».
- 2.<sup>a</sup> tanda de descarga: «2020-04-14» - «2020-04-21».
- 3.<sup>a</sup> tanda de descarga: «2020-05-15» - «2020-05-22».
- 4.<sup>a</sup> tanda de descarga: «2020-06-08» - «2020-06-15».
- 5.<sup>a</sup> tanda de descarga: «2020-06-21» - «2020-06-29».

Los mensajes se descargaron de acuerdo con un conjunto de etiquetas clave congruentes con el objeto de estudio tales como «gobierno», «españa» y «estadodealarma», así como las cuentas de los principales partidos políticos (@PSOE; @populares; @PODEMOS; @CiudadanosCs; @vox\_es) y sus máximos representantes (@sanchezcastejon; @pablocasado.; @PabloIglesias; @InesArriamas; @santi\_ABASCAL).

Finalmente, y tras el proceso de descarga, se obtuvieron un total de 4 895 747 tuits.



## Limpieza de datos

A continuación, se inició la fase de codificación manual con el fin de hacer un filtrado de todos aquellos mensajes que realmente son congruentes con el objeto de estudio. Esto permite la aplicación de algoritmos de «aprendizaje automático supervisado» para realizar una codificación automática a toda la base de datos y abarcarla en su totalidad de acuerdo con los objetivos planteados. Se realizó un muestreo aleatorio simple de 1500 tuits por tanda para obtener la muestra de codificación manual. Es importante resaltar la necesidad de la estratificación por tandas debido a que el contenido del debate público-político no se mantiene estático y hay que atenderse a la evolución de los acontecimientos.

De entre todos los clasificadores automáticos para entrenar, se escogieron las máquinas de soporte vectorial lineales (*linear-SVM* en inglés) por ser reconocidos como uno de los algoritmos que mejor re-

sultados ofrecen en categorización de texto (Wang *et al.*, 2006). Para aplicar los algoritmos de aprendizaje automático, se llevó a cabo el procesado del texto para su correcta utilización. En primer lugar, se inició el proceso de *tokenización*, que consiste en la separación de un determinado tuit en todas las palabras que contiene. Posteriormente, se aplicó un filtrado de palabras conocido como *stopwords*, que elimina palabras que carecen de importancia como determinantes, preposiciones, etc. Luego se transformaron los datos en matrices *tf-idf* (*term frequency – inverse document frequency*). Cada una de sus filas representa un determinado tuit y sus columnas representan todas las palabras que aparecen en el *corpus*. Finalmente se eliminaron los mensajes incongruentes con la investigación usando la librería «e1071» (Meyer *et al.*, 2019), restando únicamente aquellos que pertenecen a la temática de estudio. A continuación, se muestran los resultados obtenidos en la clasificación:

**TABLA 1.** Resultados del clasificador SVM-lineal para las variables «TOPIC» y «POSITION»

Tanda	Variable	Precisión	Sensibilidad	Especificidad	Kappa	F-Score	AUC
1	TOPIC	0,8017	0,9322	0,3846	0,3670	0,8620	0,6583
2	TOPIC	0,8167	0,5476	0,9213	0,5077	0,6556	0,7344
3	TOPIC	0,8267	0,7027	0,8995	0,6187	0,7596	0,8010
4	TOPIC	0,7867	0,7090	0,8494	0,564	0,7457	0,7791
5	TOPIC	0,7659	0,8758	0,6377	0,5216	0,8171	0,7567

Fuente: Elaboración propia.

## Detección de bots y su relación con los partidos políticos españoles

Con la base de datos limpia, se contaron un total de 1 208 631 tuits procedentes de 445 230 usuarios. Posteriormente, se realizó la identificación de *bots* a través de la librería «tweetbotornot2» (Kearney, 2020). Para proceder con la clasificación, y debido a la cantidad de usuarios, así como de las

restricciones de la API de Twitter, se aplicó el algoritmo de detección «FAST - *gradient boosted*», proporcionado por «tweetbotornot2». Manteniendo un criterio conservador para reducir las posibilidades de detectar como *bot* a un usuario real, se codificó como *bot* aquellas cuentas que presentaron una probabilidad de ser *bot* situadas por encima del cuartil más alto de la probabilidad ( $P > 0,9754$ ).

**TABLA 2.** Resultados del clasificador SVM-lineal para la variable «PARTIDO POLÍTICO»

Variable	Precisión	Sensibilidad	Especificidad	Kappa	F-Score
GENERAL	0,9683			0,9557	0,9632
PODEMOS		0,9490	0,9939		
PSOE		0,9828	0,9794		
CIUDADANOS		0,8591	0,9974		
PP		0,9472	0,9953		
VOX		0,9887	0,9935		
OTRO		0,9587	0,9964		

Fuente: Elaboración propia.

Para poder identificar cada *bot* con un determinado partido político, se llevó a cabo la asignación de acuerdo con la similitud encontrada en los estilos de comunicación de los tuits. Se aplicó la metodología de aprendizaje automático para entrenar a un clasificador con los estilos de comunicación de los principales partidos políticos de España. De esta forma, se contrastó el uso del lenguaje de cada *bot* con los diferentes estilos comunicativos de los principales partidos de España para detectar con cual presentaba mayor similitud. Para ello, se seleccionaron un total de 10 264 tuits de sus cuentas oficiales, así como aquellos mensajes generados por los 350 representantes políticos en el Congreso de los Diputados. Por último, se aplicó un SVM-lineal para entrenar al clasificador con el estilo de comunicación que lo vinculaba con una formación.

### Análisis de sentimiento

Por otro lado, se aplicó un análisis de sentimiento a los tuits propagados por las cuentas automáticas con la finalidad de conocer el tono emocional de su contenido. Para ello, se usó el diccionario «afinn» que se compone de un total de 2477 palabras. Todas ellas tienen una puntuación que va de -5 a 5: cuanto menor es la puntuación, más negativo es el sentimiento y viceversa

(Hansen *et al.*, 2011). Este tipo de pruebas entran dentro de los modelos teóricos dimensionales. Es decir, se estudian las emociones a través de unas dimensiones concretas y basadas en la división dicotómica entre positividad y negatividad. Con este análisis también se muestran las palabras más usadas en los mensajes que han generado los *bots*. A este diccionario se añadió la palabra «coronavirus» con una puntuación de -5, con el fin de registrar el término debido a su importancia en los mensajes. El valor de la puntuación del término «coronavirus» viene justificado por las connotaciones negativas que presenta durante el periodo de estudio, tanto en el ámbito sanitario y económico, como en el social y político.

Simultáneamente, y siguiendo el mismo procedimiento, se realizó un análisis de sentimiento de los mensajes lanzados por las cuentas de las formaciones políticas seleccionadas. Así, se puede observar si se asemejan o no con los niveles de negatividad o positividad con los que proceden de usuarios artificiales.

### Detección de comunidades y análisis de redes

Para conocer desde un punto de vista topológico la actividad de los usuarios de Twitter se ha representado gráficamente la red me-

diante los retuits. A causa de la gran envergadura de la base de datos, se utilizó la aplicación *Gephi* (Bastian, Heymann y Jacomy, 2009) para su representación gráfica. En cuanto a la detección de comunidades, se ha aplicado el algoritmo estándar de Blondel *et al.* (2008) junto con el algoritmo «ForceAtlas2» para la mejora de visualización de la red.

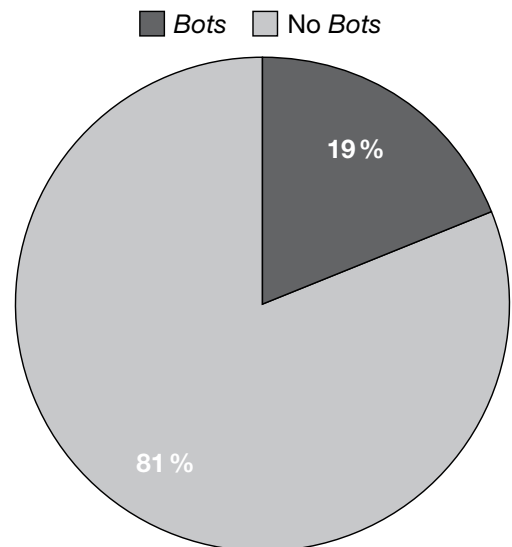
En primer lugar, se procedió a la representación de la red de aquellos usuarios codificados como «no bots» para conocer la estructura topográfica de la red conformada por aquellos usuarios no automatizados. Esta red se compone de un total de 194 992 nodos y 419 130 aristas. Dado que se trabaja con una red de retuits, es importante señalar que, a pesar de tener un total de 359 709 usuarios «no bots», únicamente 194 992 de estos usuarios pudieron ser incorporados a la red específica del análisis.

En segundo lugar, la red de bots se compone de un total de 79 246 nodos, compuestos por los usuarios bots y los usuarios con los que estos interactuaron, así como de 118 260 aristas. Para esta red, se añadió como información extra el partido de pertenencia de cada uno de los nodos. Dado que son los mensajes los que fueron codificados respecto con un partido político, para aquellos casos donde un bot únicamente había realizado un retuit, se le codificó con el partido al que correspondía el contenido del mensaje. Para aquellos que habían retuiteado varios mensajes, se le calculó la moda y se le asignó el partido político que más veces aparecía en la lista de retuits de dicho bot. Finalmente, se obtuvo una lista de atributos conformada por un partido político correspondiente a cada nodo o bot. Para facilitar la visualización de las redes (no bots —gráfico 6— y bots —gráfico 7—), los usuarios con un rango menor a 8 fueron excluidos de la visualización.

## RESULTADOS

Durante la vigencia del confinamiento domiciliario y las fases de la desescalada se ha estimado en esta investigación que un 19,20% de los usuarios de Twitter que participaron en el debate político eran robots. Es de gran importancia señalar que, de todos los mensajes seleccionados para la muestra, un 12,91% fue creado y lanzado a la red por algoritmos artificiales. Además, la actividad principal y predominante de los usuarios automáticos fue la de emitir y difundir contenido, pues el 80,84% de su contenido compartido fueron retuits a otras cuentas.

GRÁFICO 1. Proporción de bots y cuentas reales

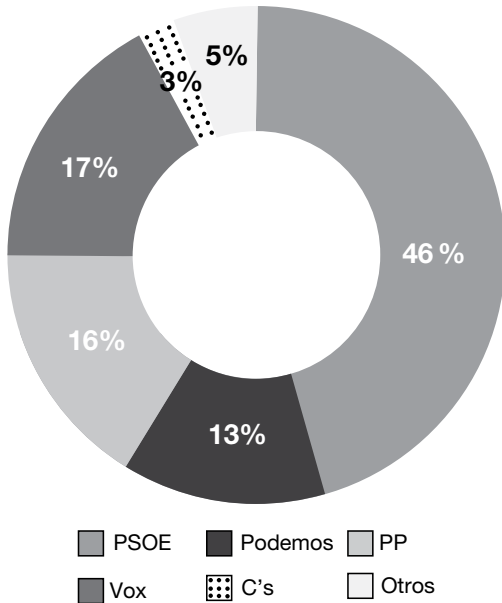


Fuente: Elaboración propia.

En cuanto a la identificación de las cuentas automáticas con los principales partidos políticos, en el gráfico 2 se puede observar la cantidad de publicaciones que han sido producidas por cuentas automáticas a lo largo del periodo seleccionado, codificadas conforme a los estilos de comunicación de cada partido. En este sentido, se puede apreciar una predominan-

cia de mensajes del PSOE producidos por cuentas artificiales, seguido por el ala derecha de la política, PP y Vox, precedidos por Podemos y encontrando un menor número de mensajes pertenecientes a los estilos de comunicación de Ciudadanos y Otros.

**GRÁFICO 2.** *Tuits producidos por bots y su identificación con cada partido*



Fuente: Elaboración propia.

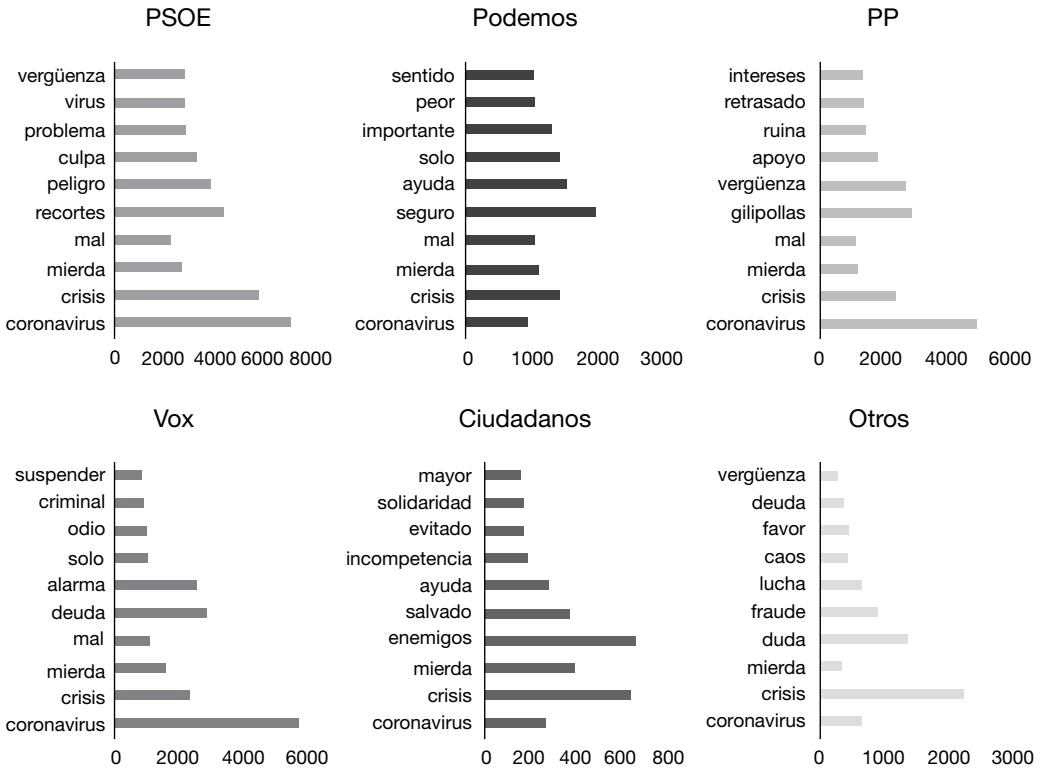
Por otro lado, se puede advertir que, según el análisis de contenido, los mensajes lanzados por los *bots* giran en torno a tres ejes principales (gráfico 3). Uno de los más recurrentes fue el de la salud, con palabras repetidas como «sanidad», «salud» o «covid». Otro asunto frecuente en

los tuits de cuentas automáticas es el de la economía, con palabras entre las que destacan las de «deuda», «crisis» o «recortes». Finalmente, encontramos el tema de la política, en el que se han agrupado todos aquellos mensajes que hacen referencia o mencionan a los partidos políticos y sus representantes.

Según se muestra en el análisis de sentimiento, el lenguaje incívico y hostil es el más presente en el diálogo de los *bots* (gráfico 4). Los porcentajes de mensajes con sentimientos negativos muestran una mayor proporción de aparición que la de los positivos. Tal y como se puede apreciar, los tuits de cuentas falsas con un estilo de comunicación que se asemeja al de Podemos presentan una mayor proporción de sentimientos positivos respecto al resto de partidos. Se encuentran proporciones similares entre PSOE, Ciudadanos y PP, mientras que los tuits que se corresponden con el estilo de Vox muestran los mayores niveles de sentimientos negativos.

Para completar el análisis de sentimiento, se expone un gráfico (gráfico 5) en el que se reflejan las palabras con connotaciones negativas más usadas por los *bots* en función de su afinidad con cada partido. Se puede apreciar que, en cada uno de los gráficos, aparecen los tres temas más recurrentes en el discurso en Twitter de las cuentas automáticas (salud, economía y política). Sin embargo, entre las palabras más usadas se observan diferencias sustanciales dependiendo del partido del que se trate. Los *bots* afines al Partido Popular y Vox hacen un mayor uso del lenguaje incívico, con palabras malsonantes y hostiles.



**GRÁFICO 5.** Palabras con connotaciones negativas por partido

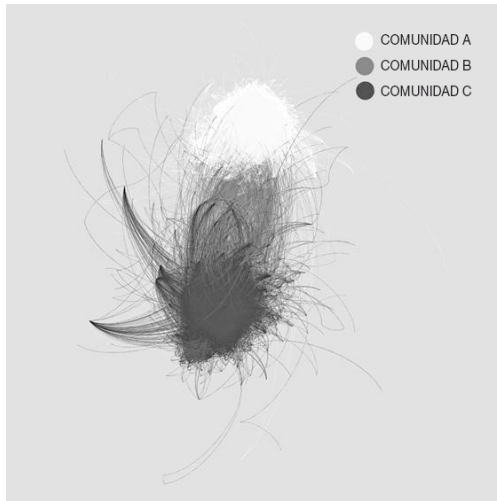
Fuente: Elaboración propia.

Tal y como se ha expuesto en el apartado anterior, se realizó un análisis de detección de comunidades individual tanto para los usuarios *bots* como para las cuentas reales. En el gráfico 6 se representa la red de usuarios reales o «no *bots*».

Tras aplicar el algoritmo de detección de comunidades, se obtuvo una red de «no *bots*» con una modularidad de 0,538 y 582 comunidades. Se entiende por modularidad, en términos generales, como la medida que ofrece información sobre la calidad de las comunidades detectadas en la red, encontrando su valor máximo en 1. Entre todas las comunidades detectadas, hay tres grupos que son los predominantes y que suponen un 77,73 % de todos los usuarios. La comunidad A representa al espec-

tro de la izquierda política con un 40,01 % de usuarios de la red, entre los que destacan las cuentas de Pedro Sánchez o Pablo Iglesias, y las de los partidos Izquierda Unida, Podemos y PSOE; por otro lado, las comunidades B y C (de tamaño 29,03 % y 8,69 %, respectivamente) simbolizan el clivaje de la derecha y ultraderecha ideológica. Las cuentas más características de esta parte del espectro gráfico son las de los partidos PP y Vox, acompañadas de las de sus respectivos líderes Pablo Casado y Santiago Abascal. Finalmente, se puede observar cómo estas tres comunidades conforman dos grupos distantes y divididos entre sí, observando una polarización estructural de la red.

**GRÁFICO 6.** Red de usuarios reales

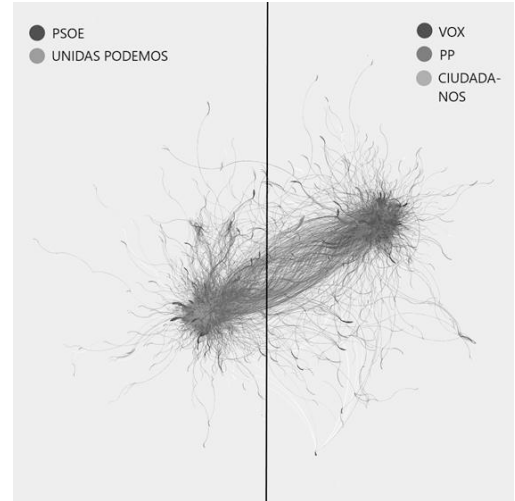


Fuente: Elaboración propia.

En lo que concierne al análisis de comunidades de bots, se puede decir que su red consta de un total de 4601 grupos con una modularidad de 0,766, siendo esta mayor para la red de bots que para la de usuarios no automatizados. Al igual que en la red de usuarios reales, se detectan muy pocas comunidades principales; en este caso, dos de ellas suman un total de 41,35 % de las cuentas automáticas totales (Comunidad A = 21,83 % y Comunidad B = 19,52 %). A diferencia del caso anterior, se detectan quince comunidades pequeñas (entre el 5,77 % y el 1 % del total), restando una gran cantidad de comunidades de tamaño no significativo. En el gráfico 7 se representa la red de bots con sus dos comunidades principales en las que los nodos están coloreados en función del partido político con el que son afines (véase gráfico 2). Desde este punto de vista, aunque una comunidad pueda estar compuesta por diferentes partidos, atendiendo al color de los partidos cuyo estilo de comunicación se asemeja al del bot, se puede observar cómo las cuentas falsas se agrupan en torno a dos discursos políticos opuestos, encontrando el discurso de izquierdas (PSOE y Unidas Po-

demos), por un lado, y en contraposición el discurso de derechas (colores Vox, PP y, en menor medida, Ciudadanos).

**GRÁFICO 7.** Red de cuentas falsas



Fuente: Elaboración propia.

Es relevante señalar que, en lo que respecta a los resultados del análisis de detección de comunidades tanto de usuarios reales como de bots, se observa una división en dos grupos principales que indica la posible existencia de altos niveles de polarización política. Tanto las comunidades principales como las de menor tamaño se dividen claramente en torno a un eje. Estas dos mitades representan el clivaje político tradicional de izquierda y derecha, en el que la escasez de centralidad y puentes entre comunidades es síntoma de la división ideológica en el actual escenario político.

## CONCLUSIONES

La inesperada crisis sanitaria provocada por la COVID-19 y la consiguiente imposición del estado de alarma, sumadas al desequilibrio que atraviesan las democracias occidentales, han supuesto la entrada a un es-

cenario de vulnerabilidad con respecto a la construcción de la opinión pública. Sin embargo, las características comunicativas y deliberativas que ofrecen las redes sociales, vistas desde una perspectiva *ciberoptimista*, resultan en un gran beneficio democráticamente hablando. El hecho de contar con múltiples canales de debate puede llegar a fomentar la participación cívica real en asuntos políticos (Santana y Huerta, 2019).

No obstante, existen varias circunstancias que muestran la evidencia de la vigente resistencia al cambio de orden social que podrían suponer las redes. La injerencia de instrumentos de IA en el debate político de Twitter es una prueba de lo anterior. Durante el periodo de confinamiento domiciliario y las fases de desescalada se utilizaron cuentas automáticas que intervinieron en la conversación política de Twitter: en esta investigación se demuestra que 19,20 % de los usuarios participantes fueron *bots*. Este hecho, junto a que la estructura misma de la red facilita esta circunstancia, evidencia la oposición a que las redes se conviertan en un espacio para la deliberación democrática horizontal.

Teniendo en cuenta la consonancia de las cuentas falsas con los principales partidos políticos del país que demuestra el análisis, se puede validar la hipótesis de que se crearon campañas de desinformación ejecutadas por grupos de *bots* coordinados estratégicamente y vinculados a los agentes sociopolíticos tradicionales con el objetivo de generar una falsa sensación de apoyo o en contra de los temas principales discutidos en Twitter. Los roles comunicativos con los que se ha programado a las cuentas automáticas, así como las consecuencias que provocan en el debate, son propios de las campañas de *astroturfing*.

Una vez sabido esto, en la investigación se puede observar la existencia de un escenario en el que los *bots* están diseñados como megáfonos de contenido sesgado y

congruente con los objetivos ideológicos y propagandísticos de sus partidos afines. Sus patrones conductuales se corresponden con los que Hwang y Wooley (2016) definen como *bots facilitadores*: difundir mensajes para darles el mayor alcance posible. Con esta estrategia se pretende engordar la conversación en torno a ciertos supuestos, no solo para crear la ilusión de que esa es la postura mayoritaria, sino también para difuminar el alcance y la predominancia de las posiciones contrarias.

Que el debate de los *bots* gire en torno a los tres asuntos más discutidos en Twitter (salud, política y economía) constata la lucha por la hegemonía de posicionamientos ideológicos entre partidos y entre bloques ideológicos rivales. Sin embargo, además de intentar copar e inflar falsamente los temas de conversación, los mensajes con connotaciones negativas, así como el uso de un lenguaje incívico y de hostigamiento, son las características más presentes en el diálogo generado por las cuentas falsas.

Los datos del análisis de sentimiento se muestran, en este caso, congruentes con la realidad: los *bots* afines a Podemos, que forma parte del Gobierno, difunden mensajes con un sentimiento proporcionalmente más positivo que el resto. Esto se puede traducir en la adopción de estrategias que creen la sensación de apoyo y legitimación de las medidas y políticas adoptadas durante el estado de alarma por el Poder Ejecutivo. Sin embargo, los *bots* que se asemejan en estilo al partido de ultraderecha Vox y al conservador Partido Popular, formaciones que se encuentran dentro del grupo de la oposición y que se han mostrado críticas con las decisiones del Gobierno, mantienen una postura claramente negativa y con niveles significativamente más altos de *incivility*.

En este sentido, la dimensión que alcanzan los niveles de sentimiento nocivo de los mensajes indica que las estrategias políticas



que se diseñan para que sean ejecutadas por los *bots* se fundamentan en el ataque al contrario. Esta conducta puede generar el arraigo de partidismo negativo entre los usuarios reales de Twitter. Estas muestras se ven reflejadas en la fractura estructural de la red observada para la presente investigación. La división de los usuarios de Twitter, tanto en la red de cuentas reales como en las falsas, que gira en torno a dos ideologías enfrentadas, es síntoma de la posible existencia de polarización centrífuga. Esos dos grandes grupos representan el clivaje tradicional del eje de izquierda y derecha. No obstante, las comunidades reflejadas en el estudio se ven afectadas por el partidismo negativo y por el hostigamiento mediante el lenguaje, lo que no hace otra cosa sino romper los puentes y alejar cada vez más del centro a los usuarios de la red.

En definitiva, la difusión masiva de propaganda partidista y sesgada, camuflada bajo la forma y estructura de activistas sociales, forma parte de las campañas de desinformación que se utilizan con el fin de manipular y convencer políticamente a la opinión pública. Estas condiciones, que complican la legitimidad de los procesos participativos y comprometen los valores democráticos (Santana y Huerta, 2019), han llevado a una mayor ruptura del consenso entre la población en un escenario de por sí polarizado.

Lo que se evidencia en este estudio es que la propagación de mensajes que son parte de una campaña negativa por parte de las cuentas automáticas ha creado en Twitter una red de confrontación, no de diálogo. La consiguiente fractura en la red de usuarios reales, tal y como se observa en el análisis, es un reflejo de la situación política tradicional española que, lejos de haber sido superada gracias a las posibilidades que ofrecen las redes, se ha trasladado al espacio *online*.

Es importante resaltar que el potencial ofrecido por las TIC permite construir una

sociedad más horizontal, inclusiva y deliberativa en favor de una ciudadanía empoderada frente a los actores convencionales. Sin embargo, este posible empoderamiento social se ve coartado por los intereses de los agentes sociopolíticos tradicionales resistentes al cambio y en perjuicio de la salud democrática.

Para concluir, las limitaciones que se observan en este estudio tienen que ver con el ámbito metodológico. Debido a las grandes dimensiones de la muestra, el análisis de sentimiento se ejecutó de forma automática. A pesar de que se añadieron palabras que se creyeron relevantes para la investigación a la herramienta, el hecho de que funcione con un diccionario preestablecido pudo restar riqueza a los resultados obtenidos. En este sentido, la realización de un análisis de contenido cualitativo podría aportar resultados más detallados sobre el significado de los mensajes.

Por otro lado, en cuanto a investigaciones futuras que pueden derivar de este trabajo, cabe resaltar el interés teórico que supone un estudio en profundidad sobre la posible responsabilidad de los partidos y organizaciones en el uso de *bots* políticos. Es de relevancia mencionar que la naturaleza del presente estudio es correlacional, no causal. Por lo tanto, aun teniendo en cuenta los resultados obtenidos, no nos es posible establecer una relación de causa-efecto que vaya más allá de la consonancia lingüística. También merece consideración la aparente polarización encontrada en los resultados en el debate público *online*, así como sus principales consecuencias.

## BIBLIOGRAFÍA

- Balcells, Joan y Cardenal, Ana S. (2013). «Internet y la competición electoral: el caso de Esquerra Republicana de Catalunya»/«Internet and Electoral Competition: The Case of Esquerra Republicana de Catalunya». *Revista Española de Investigaciones Sociológicas*, 141: 3-28.

- Bastian, Mathieu; Heymann, Sebastien y Jacomy, Mathieu (2009). «Gephi: An Open Source Software for Exploring and Manipulating Networks». *Third International ICWSM Conference*, 3(1): 361-362. doi: 10.13140/2.1.1341.1520
- Benkler, Yochai (2006). «*The Wealth of Networks: How Social Production Transforms Markets and Freedom*». New Haven: Yale University Press.
- Bennet, W. Lance y Segerberg, Alexandra (2012). «The Logic of Connective Action. Digital Media and the Personalization of Contentious Politics». *Information, Communication & Society*, 15(5): 739-768. doi: 10.1080/1369118X.2012.670661
- Bennet, W. Lance; Segerberg, Alexandra y Yang, Yunkang (2018). «The Strength of Peripheral Networks: Negotiating Attention and Meaning in Complex Media Ecologies». *Journal of Communication*, 68(4): 659-684. doi: 10.1093/joc/jqy032
- Blondel, Vincent D.; Guillaume, Jean-Loup; Lambiotte, Renaud y Lefebvre, Etienne (2008). «Fast Unfolding of Communities in Large Networks». *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): 1-12. doi: 10.1088/1742-5468/2008/10/p10008
- Bode, Leticia; Hanna, Alexander; Yang, Junghwan y Shah, Dhavan V. (2015). «Candidate Networks, Citizen Clusters, and Political Expression: Strategic Hashtag Use in the 2010 Midterms». *The ANNALS of the American Academy of Political and Social Science*, 659: 149-165. doi: 10.1177/0002716214563923
- Boshmaf, Yazan; Musluhkhov, Ildar; Beznosov, Konstantin y Ripeanu, Matei (2013). «Design and Analysis of a Social Botnet». *Computer Networks*, 57(2): 556-578. doi: 10.1016/j.comnet.2012.06.006
- Boulianne, Shelley (2018). «Twenty Years of Digital Media Effects on Civic and Political Participation». *Communication Research*, 47(7): 947-966. doi: 10.1177/0093650218808186
- Boulianne, Shelley y Theocharis, Yannis (2018). «Young People, Digital Media, and Engagement: A Meta-Analysis of Research». *Social Science Computer Review*, 38(2): 111-127. doi: 10.1177/0894439318814190
- Bradshaw, Samantha y Howard, Phillip N. (2018). *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*. Disponible en: <http://comprop.oii.ox.ac.uk/wpcontent/uploads/sites/93/2018/07/ct2018.pdf>, acceso 15 de diciembre de 2020.
- Bradshaw, Samantha y Howard, Phillip N. (2020). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. Oxford: Project on Computational Propaganda.
- Cáceres Zapatero, María D.; Brändle, Gaspar y Ruiz San Román, José A. (2015). «Hacia la construcción de una ciudadanía digital». *Prisma Social*, 15: 643-684.
- Castells, Manuel (2012). «*Networks of Outrage and Hope: Social Movements in the Internet Age*». Cambridge: Polity.
- Dader, José L. y Campos Dominguez, Eva (2017). «*La búsqueda digital del voto: Ciber campañas electorales en España 2015-2016*». Valencia: Tirant Lo Blanch.
- Fazil, Mohd y Abulaish, Mohammad (2017). «Identifying Active, Reactive, and Inactive Targets of Socialbots in Twitter». *Proceedings of the International Conference on Web Intelligence*: 573-580.
- Gallardo, Beatriz y Enguix, Salvador (2016). *Pseudo-política: el discurso político en las redes sociales*. Valencia: Universitat de València.
- Habermas, Jürgen (2006). «Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension The Impact of Normative Theory on Empirical Research». *Communication Theory*, 411-426. doi: 10.1111/j.1468-2885.2006.00280.x
- Hansen, Lars K.; Arvidsson, Adam; Nielsen, Finn A.; Colleoni, Elanor y Etter, Michael (2011). «Good Friends, Bad News-affect and Virality in Twitter». *Future Information Technology*: 34-43. doi: 10.1007/978-3-642-22309-9\_5
- Howard, Phillip N. (2006). *New Media Campaigns and the Managed Citizen*. Cambridge: Cambridge University Press.
- Howard, Phillip N.; Savage, Saiph; Flores, Claudia; Toxtli, Carlos y Monroy-Hernández, Andrés (2017). «Social Media, Civic Engagement, and the Slacktivism Hypothesis: Lessons From Mexico's "El Bronco"». *Journal of International Affairs*, 70(1): 55-73. Disponible en: <https://www.jstor.org/stable/90012597?seq=1>, acceso 15 de diciembre de 2020.
- Hwang, Tim y Wooley, Samuel (2016). *How Politicians Should or Shouldn't Use Twitter Bots*. Disponible en: <https://slate.com/technology/2016/03/how-politicians-should-use-twitter-bots.html>, acceso 15 de diciembre de 2020.
- Kearney, Michael W. (2020). *tweetbotornot2: Detect Twitter Bots. R package version 0.0.1*. Disponible en: <https://github.com/mkearney/tweetbotornot2>, acceso 15 de diciembre de 2020.

- Keller, Franziska B.; Schoch, David; Stier, Sebastian y Yang, JungHwan (2019). «Political Astrourfing on Twitter: How to Coordinate a Disinformation Campaign». *Political Communication*, 37(2): 256-280. doi: 10.1080/10584609.2019.1661888
- Keyes, Ralph (2004). *The Post-truth Era: Dishonesty and Deception in Contemporary Life*. New York: Macmillan.
- Margolis, Michael y Resnick, David (2000). *Politics as Usual: The Cyberspace «Revolution»*. Thousand Oaks: Sage Publications.
- Margolis, Michael y Moreno-Riaño, Gerson (2009). *The Prospect of Internet Democracy*. London: Routledge.
- Meyer, David; Dimitriadou, Evgenia; Hornik, Kurt; Weingessel, Andreas; Leisch, Friedrich; Chang, Chih-Chung y Lin, Chih-Chen (2019). «e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-3». *Cran.r-project.org*. Disponible en: <https://cran.r-project.org/web/packages/e1071/>, acceso 15 de diciembre de 2020.
- Molyneux, Logan y Mourao, Rachel (2017). «Political Journalists' Normalization of Twitter: Interaction and New Affordances». *Journalism Studies*, 20(1): 1-19. doi: 10.1080/1461670X.2017.1370978
- Murolo, Leonardo (2019). La posverdad es mentira. Un aporte conceptual sobre *fake news* y periodismo. En: R. Aparici y D. García-Marín (coords.). *La Posverdad. Una cartografía de los medios, las redes y la política*. Barcelona: Gedisa.
- Neuman, Russell W.; Guggenheim, Lauren; Jang, S. Mo y Bae, Soo Y. (2014). «The Dynamics of Public Attention: Agenda-setting Theory Meets Big Data». *Journal of Communication*, 64(2): 193-214. doi: 10.1111/jcom.12088
- Neyazi, Tabereh A. (2019). «Digital Propaganda, Political Bots and Polarized Politics in India». *Asian Journal of Communication*, 30(1): 39-57. doi: 10.1080/01292986.2019.1699938
- Noelle-Neumann, Elisabeth (1984). *The Spiral of Silence: Public Opinion, Our Social Skin*. Chicago: University of Chicago Press.
- Orbegozo Terradillos, Julen; Morales i Gras, Jordi y Larrondo Ureta, Ainara (2020). «Desinformación en redes sociales: ¿compartimentos estancos o espacios dialécticos? El caso Luther King, Quim Torra y *El Confidencial*». *Revista Mediterránea de Comunicación*, 11(2): 55-69. doi: 10.14198/ME-DCOM2020.11.2.2
- Papacharissi, Zizi (2004). «Democracy Online: Civility, Politeness and the Democratic Potential of Online Political Discussion Groups». *New Media and Society*, 6(2): 259-283. doi: 10.1177/1461444804041444
- Pastor-Galindo, Javier; Zago, Mattia; Nespolei, Pantaleone; López Bernal, Sergio; Huertas Celdrán, Alberto; Gil Pérez, Manuel; Ruipérez-Valiente, José A.; Martínez Pérez, Gregorio y Gómez Marmol, Félix (2020). «Spotting Political Social Bots in Twitter: A Use Case of the 2019 Spanish General Election». *IEEE Transactions on Network and Service Management*, 17(4): 2156-2170. doi: 10.1109/tns.2020.3031573
- Persily, Nathaniel (2017). «The 2016 U.S. Election: Can Democracy Survive the Internet?». *Journal of Democracy*, 64(2): 64-76. doi: 10.1353/jod.2017.0025
- Price, Kristopher R.; Priisalu, Jaan y Nomin, Sven (2019). «Analysis of the Impact of Poisoned Data within Twitter Classification Models». *IFAC-PapersOnLine*, 52(19): 175-180. doi: 10.1016/j.ifacol.2019.12.170
- Rheingold, Howard (2009). *La comunidad virtual: una sociedad sin fronteras*. Barcelona: Gedisa.
- Robles, José M. y Córdoba, Ana M. (2019). *Digital Political Participation: Disintermediation in the Era of Web 2.0*. London: Palgrave.
- Rowe, Ian (2015). «Civility 2.0: A Comparative Analysis of Incivility in Online Political Discussion». *Information, Communication and Society*, 18(2): 121-138. doi: 10.1080/1369118X.2014.940365
- Santana, Luis E. y Huerta Cánepa, Gonzalo (2019). «¿Son bots? Automatización en redes sociales durante las elecciones presidenciales de Chile 2017». *Cuadernos.info*, 44: 61-77. doi: 10.7764/cdi.44.1629
- Sartori, Giovanni (2005). *Parties and Party Systems: A Framework for Analysis*. Colchester: ECPR press.
- Shao, Chengcheng; Ciampaglia, Giovanni L.; Varol, Onur; Yang, Kai-Cheng; Flammini, Alessandro y Menczer, Filippo (2018). «The Spread of Low-credibility Content by Social Bots». *Nature Communication*, 9(1): 1-10. doi: 10.1038/s41467-018-06930-7
- Stella, Massimo; Ferrara, Emilio y De Domenico, Manlio (2018). «Bots Increase Exposure to Negative and Inflammatory Content in Online Social Systems». *Proceedings of the National Academy of Sciences*, 115(49): 12435-12440. doi: 10.1073/pnas.1803470115
- Stier, Sebastian; Bleier, Arnim; Lietz, Haiko y Strohmaier, Markus (2018). «Election Campaigning on

- Social Media: Politicians, Audiences and the Mediation of Political Communication on Facebook and Twitter». *Political Communication*, 35(1): 50-74. doi: 10.1177/1461444817709282
- Stoking, Galen y Sumida, Nami (2018). «Social Media Bots Draw Public's Attention and Concern». *Pew Research Center*. Disponible en: <https://www.journalism.org/2018/10/15/social-media-bots-draw-publics-attention-and-concern/>, acceso 15 de diciembre de 2020.
- Walker, Edward T. (2014). *Grassroots for Hire: Public Affairs Consultants in American Democracy*. New York: Cambridge University Press.
- Wang, Zi-qiang; Sun, Xia; Zhang, De-xian y Li, Xin (2006). «An Optimal SVM-Based Text Classification Algorithm». *International Conference on Machine Learning and Cybernetics: 1378-1381*. doi: 10.1109/ICMLC.2006.258708
- Ward, Stephen y Gibson, Rachel (2009). European Political Organizations and the Internet. En: A. Chadwick y P. N. Howard (eds.). *Routledge Handbook of Internet Politics* (pp. 25-39). Abingdon: Routledge.
- Winner, Langdon (2003). «Internet y los sueños de una renovación democrática». *Isergoría*, 28: 55-71. doi: 10.3989/isergoria.2003.i28.506
- Yan, Harry Y.; Yang, Kai-Cheng; Menczer, Filippo y Shanahan, James (2020). «Asymmetrical Perceptions of Partisan Political Bots». *New Media & Society*: 1-22. doi: 10.1177/1461444820942744

**RECEPCIÓN:** 09/02/2021

**REVISIÓN:** 11/06/2021

**APROBACIÓN:** 23/11/2021

# The Presence of Political Bots on Twitter during the COVID-19 Crisis in Spain

*La presencia de los bots políticos en Twitter durante la crisis de la COVID-19 en España*

Ángela Martínez Torralba, Juan Antonio Guevara Gil and Aitor Jiménez de la Fuente

## Key words

- Astroturfing
- Political Bots
  - COVID-19
  - Debate
  - Incivility
  - Polarization
  - Propaganda
  - Twitter

## Palabras clave

- Astroturfing*
- *Bots* políticos
  - COVID-19
  - Debate
  - *Incivility*
  - Polarización
  - Propaganda
  - Twitter

## Abstract

The crisis caused by the COVID-19 pandemic has led to the decree of the State of Alarm in Spain and a severe home confinement that was softened in phases of de-escalation. During this period, social networks were used as a discussion tool. With this research we intend to find out if artificial intelligence tools were used in the political debate on Twitter. To achieve this, algorithms that determine the presence of bots in the conversation, their communicative roles and their relationship with the main political parties were applied. The results show that disinformation campaigns were created by bots with the aim of manipulating public opinion.

## Resumen

La crisis sanitaria provocada por la pandemia de COVID-19 ha conllevado el decreto del estado de alarma en España y un severo confinamiento domiciliario que se fue suavizando en fases de desescalada. Durante este periodo, las redes se utilizaron aún más como herramienta de debate. Lo que se pretende con esta investigación es averiguar si se utilizaron herramientas de IA para que interviniesen en el debate político de Twitter. Para ello, se aplicaron algoritmos que determinan la presencia de *bots* en la conversación, sus roles comunicativos y su relación con los principales partidos políticos. Los datos extraídos demuestran que se crearon campañas de desinformación ejecutadas por *bots* con el objetivo de manipular la opinión pública.

## Citation

Martínez Torralba, Ángela; Guevara Gil, Juan Antonio; Jiménez de la Fuente, Aitor (2023). "The Presence of Political Bots on Twitter during the COVID-19 Crisis in Spain". *Revista Española de Investigaciones Sociológicas*, 181: 61-80. (doi: 10.5477/cis/reis.181.61)

**Ángela Martínez Torralba:** Universidad Complutense de Madrid | [angela19@ucm.es](mailto:angela19@ucm.es)

**Juan Antonio Guevara Gil:** Universidad Complutense de Madrid | [juanguev@ucm.es](mailto:juanguev@ucm.es)

**Aitor Jiménez de la Fuente:** Universidad Complutense de Madrid | [aitorj01@ucm.es](mailto:aitorj01@ucm.es)

## INTRODUCTION<sup>1</sup>

The addition of Twitter to everyday life has created a revolution in the classic discussion forums. Political strategies implemented to improve the image and positioning of ideological principles in public conversations held over the social networks continue to evolve. Of the new strategies implemented, one of the most extensive is the use of artificial intelligence (AI) tools, designed to alter and distort the impact of specific messages.

This study analyzes how these artificial technologies have been used and their role during the one of the most significant recent events in Spain, the COVID-19 pandemic and the resulting countrywide State of Alarm.

The declaration of a State of Alarm in Spain resulted in a strict at-home lockdown taking place from March 15<sup>th</sup> to April 28<sup>th</sup> of 2020. As of this moment, the country underwent a de-escalation process until June 21<sup>st</sup> of 2020, during which the lockdown measures applied to contain the virus were gradually relaxed in phases.

During this crucial time, social media took on a special relevance. One of the main objectives of the political parties was to gain an advantage amongst Twitter users, with respect to positioning (for or against) in the political debate (Howard *et al.*, 2017; Bradshaw and Howard, 2018). Recently, an increased use of bots has been evident. These bots are used to distort reality, gaining support and influencing citizens (Neyazi, 2019; Santana and Huerta, 2019; Yan *et al.*, 2020; Bradshaw and Howard, 2020).

The Twitter network has been increasingly used for the implementation of these

tools. Its structure offers a sphere in which polarization and separation of opposing groups may be increased (Neyazi, 2019). And, in some cases, increased tension can be created between groups through the use of inappropriate language: so-called “incivility” (Papacharissi, 2004).

Within the framework of this situation, one might ask: What do bots mean in terms of the political debate being carried out across the social networks? What is their main role? What relationship do they have with the political parties? Do they have any real consequences?

The increased use of these tools as global propaganda practices (Bradshaw and Howard, 2020) leads to our hypothesis that during the State of Alarm period in Spain, AI algorithms were used in Twitter to condition political conversation. Specifically, we attempt to determine whether or not disinformation campaigns were conducted by groups of bots that were strategically organized and identified with traditional socio-political participants in order to generate a false sense of support or disapproval with the relevant political topics of this period.

To do so, the presence of political bots in Twitter based on algorithms designed with this purpose during the Spanish lockdown months and de-escalation phases will be detected. Subsequently, the communicative role played by the bots in the political debate carried out over the social networks will be defined and the relationship between the bots and the main political participants will be determined.

This study initially offers a theoretical introduction to the new digital context and political sphere in which our study objective is framed. It also reviews the literature existing on the use of bots in this new digital environment and the main uses of artificial intelligence in political disinformation campaigns. Then, the methodology used to carry out the empirical study will be de-

<sup>1</sup> Financing: This study has been partially supported by the Spanish government, National Plan of RandD+i, PID2019-106254RB-I00

tailed. Finally, the results will be presented with regard to the existence of disinformation campaigns in Twitter during the Spanish State of Alarm period. A final conclusion will then be offered.

## THE NEW DIGITAL ENVIRONMENT IN THE POLITICAL SPHERE

Since the final years of the 20th century, warnings have been made with regard to Information and Communication Technology (ICT), and more specifically, web and social network technologies, which were becoming some of the major catalysts of public opinion and essential socialization elements (Gallardo and Enguix, 2016). This system of “mass self-communication” (Castells, 2012) enables participants to decide what message is sent, designating potential recipients and selecting the messages that they want to recover. It also allows them to inform, mobilize, support or debate without necessarily being a direct participant (Castells, 2012; Dader and Domínguez, 2017).

With these possibilities, democratic systems became increasingly invigorated, permitting more diverse and extensive sectors of the population to access public debate with greater ease (Winner, 2003). Thus, social networks are considered a primary channel of dissemination of news and democratic deliberation, offering an increasingly relevant role to leaders of opinion and citizens who can decide on topics and condition the agenda of the conventional media (Bennett and Segerberg, 2012; Neuman *et al.*, 2014; Bode *et al.*, 2015; Stier *et al.*, 2018; Yan *et al.*, 2020).

From this perspective, citizens can challenge the monopoly held by traditional socio-political participants over the conventional mass media (Rheingold, 2009; Castells, 2012). Citizens, individually or collectively, win ground against the conven-

tional participants and are capable of creating a hybrid public sphere (Benkler, 2006; Santana and Huerta, 2019).

However, despite studies affirming a positive correlation between the use of social networks and civic and political commitment of citizens (Boulianne and Theocharis, 2018; Boulianne, 2018), other streams of thought suggest that those willing to participate politically online are the same individuals who will do so in the real world. According to these studies, the Internet is yet to be a factor of collective transformation by which digital instruments are the foundation of social organization (Margolis and Resnick, 2000). The politician’s role remains unchanged, although it has been displaced to other scenarios. And as discussed below, the very dynamic of this environment may facilitate certain behaviors that will blur the horizontal and inclusive possibilities of this setting (Margolis and Resnick, 2000).

The existing confrontation between enormous initial expectations with a reality that does not cease to reduce these advances has created a prolific field of study. Three major lines of research have emerged (Ward and Gibson, 2009):

First, and perhaps the most conservative of the three, is the theory of “normalization”. This theory assumes that sufficient evidence exists to affirm that the communication patterns of the off-line world are entering the on-line environment (Margolis and Resnick, 2000). Therefore, political participants who traditionally have maintained a privileged position will continue to hold this competitive advantage online in the Internet, in line with the “politics as usual” argument (Margolis and Resnick, 2000; Molyneux and Mourão, 2017). Second, and based on some completely distinct assumptions, there is the “disintermediation” theory. This interpretation assumes that the new technologies

will promote the production and distribution of content and messages with a potentially global audience (Benkler, 2006). Its main property is the ability to reduce participation costs, decreasing the power of the classic mediators and promoting a more horizontal communication model (Benkler, 2006; Robles and Córdoba, 2019). Finally, a hybrid model exists, the so-called “equalization” or “levelling” model. This theory is based on the symbiosis that exists between the traditional communicative system (based on verticality and the conventional power structure), and the digital system (based on horizontality and collective action) (Benkler, 2006; Balcells and Cardenal, 2013). This permits the logic of the Internet to help balance the differences existing in resources and social scope of citizens, means of communication and political parties or leaders, adapting their levels of visibility (Balcells and Cardenal, 2013).

Political strategies, including the massive use of AI tools and bots in the social networks, may be situated within the theories of hybridization. Although social networks are instruments that ensure the active listening of voices that might otherwise not be heard and that jeopardize the traditional system of mediation, an increasing number of socio-political participants are tending to change from real world to online means of communication and propaganda (Margolis and Moreno-Riaño, 2009).

## **POLITICAL BOTS IN THE NEW DIGITAL SCENARIO**

As for participation in the massive use of social networks, users are offered increasing incentives to take part in public debates (Cáceres, Brändle and Ruiz, 2017). They are the focus of a new, hybrid communication system that feeds off of the traditional media and extends the political environment

into multi-media settings where the attention to news is personalized and any user can become a content creator (Bennett, Segerberg and Yang, 2018; Santana and Huerta, 2019). In fact, some researchers support the idea that the social networks are a barometer of public opinion and act as amplifiers of that which is reported by the conventional communication media (Neyazi, 2019).

This, in addition to the minimal costs of an online presence, has led to a major increase in the use of this social communication instrument by conventional political powers for the promotion of their distinct strategies (Pastor *et al.*, 2020). The use of AI tools and micro-segmentation that exploits and uses private user information forms a part of these new propaganda tactics. Propaganda is defined as the intentional use of true or false information, influencing citizen actions on a specific issue (Persily, 2017). This, extrapolated to the digital environment, may be considered to be the use of computerized instruments that were deliberately created to manipulate user opinion (Neyazi, 2019).

Bots and other AI tools play an essential role in the development of the main propaganda strategies in the social networks. But not all bots are designed with malicious intentions. Some are merely used to spread information. This is the case with “social” bots, used to publish periodic content and interact with users (Yan *et al.*, 2020). In contrast, “political” bots are automated accounts created from algorithms that reproduce human behavior and are intended to create a false sense of support for a political party, candidate or cause, while also defaming their opponents and those expressing contrasting opinions. In this way, they alter the natural rhythms of debate on the social networks (Yan *et al.*, 2020).

Unlike social bots, political bots are used to massively disseminate fake news (Shao *et al.*, 2018; Yan *et al.*, 2020), negative mes-



sages (Stella, Ferrara and De Domenico, 2018; Neyazi, 2019; Yan *et al.*, 2020) and private information of users for partisan purposes (Boshmaf *et al.*, 2013; Yan *et al.*, 2020). Hwang and Wooley (2016) classified political bots into two categories: controllers, those that are used to manipulate or block conversation on a certain topic; and facilitators, which share and disseminate content so that the messages will receive a large number of visualizations and become viral.

AI instruments may operate and interact with human users given the lack of individual filters (Yan *et al.*, 2020; Price, Priisalu and Nomin, 2019). Although numerous studies have focused on the creation of computational algorithms to detect bots (Yan *et al.*, 2020; Neyazi, 2019), few have examined the issue of human distinction between real and artificial accounts. Yan *et al.*, (2020), suggested that social networks are being massively tainted by malicious algorithms of political manipulation since individuals lack the necessary tools to differentiate between users as bots or real people. It was also concluded that political bias plays a more than relevant role in terms of the perception of this issue. Automatic accounts identified with right-wing or far-right parties are more effective, not only because they are more difficult to detect, but also because conservative users tend to be more vulnerable to fake news dissemination (Yan *et al.*, 2020).

In this regard, certain taxonomies have classified types of human behavior with respect to bots. On the one hand, active users tend to follow the automatic account without having been followed previously by said account; secondly, reactive users tend to interact in response to actions of the bot; and finally, inactive users tend to refrain from responding to or following these accounts (Fazil and Abulaish, 2017).

Degrees of vulnerability may increase in the face of this propaganda, given that attacks by political bots are increasingly better

organized in terms of strategy and tend to be directed at more segmented audiences in a polarized and tense environment, such as that of the social networks (Stella, Ferrara and De Domenico, 2018; Yan *et al.*, 2020). Furthermore, algorithms are increasingly more capable of modifying or improving their behavior in order to remain undiscovered by detection tools (Price, Priisalu and Nomin, 2019).

The irruption of the Post-Truth Era (Keyes, 2004), a scenario in which fake news is assimilated by citizens and accepted as the truth (Murolo, 2019; Orbegozo, Morales and Larrondo, 2020), is a breeding ground for interference by political bots. Thus, the concept of post-truth and the use of AI tools are closely related to disinformation in the social networks, since it is in these networks where fake news and political propaganda have ample area to be disseminated (Price, Priisalu and Nomin, 2019).

## **ASTROTURFING AND DISINFORMATION CAMPAIGNS**

The electoral campaign prior to the 2016 US presidential elections and the UK referendum to exit the European Union (Brexit) led to a change in mentality with regard to international awareness of the existence of manipulation of the social networks on a global scale (Keller *et al.*, 2019; Neyazi, 2019). Russian interference in both processes, two of the most relevant events of recent years, generated beliefs regarding the democratic dangers resulting from the use of AI instruments in online debates (Keller *et al.*, 2019). In a study conducted during the 2016 US elections (Stoking and Sumida, 2018), over 66 % of the participants affirmed their knowledge and discomfort regarding this issue, with some 80 % being aware of the negative effects of the same.

Although a certain ambiguity exists regarding the concept of disinformation and

the categories within its parameters– fake news, “misinformation” or “bad information” – it may be considered a harmful consequence that social network users must deal with. Likewise, disinformation is the act of deceiving the online audience by dumping infinite malicious algorithms into the network for purposes of political manipulation. Although the information may be truthful, it is used to erode and interfere with the democratic debate cycles between real individuals (Keller *et al.*, 2019). Along these lines, “astroturfing” is considered a disinformation campaign.

This term refers to social network campaigns that appear to be carried out by grass roots activists or groups, in favor or against a certain political cause, but which, in fact, are a large set of automatic accounts acting strategically to emulate human behavior (Howard, 2006; Walker, 2014). Through these operations, an attempt is made to generate a false sense of support or disagreement regarding an idea that lacks consensus, to influence social network users with political and propaganda-based purposes (Santana and Huerta, 2019). In almost all of the literature on this topic, distinct terms are used to refer to this concept: “botnets” (Boshmaf *et al.*, 2013), “cybertroopers” (Bradshaw and Howard, 2018) or “Twitter bombs” (Pastor *et al.*, 2020).

In practice, these coordinated groups of bots act like a social movement, only more rapidly, on a massive scale and at a lower cost (Yan *et al.*, 2020). However, unlike the grass roots efforts, which do not have a centralized organization and therefore have more varied followers and extend more slowly over time, “cybertroopers” are defined by the fact that they publish almost identical messages simultaneously or collectively disseminate congruent messages for the campaign being carried out (Keller *et al.*, 2019).

On the other hand, these maneuverings tend to function through the harassment and exclusion of minority and contradictory opinions or group debate. Fake news, hate messages and the use of uncivil language are common characteristics of this type of campaigns (Keller *et al.*, 2019; Santana and Huerta, 2019). As suggested by Santana and Huerta Cánepa (2019), “astroturfing” is closely linked to the “spiral of silence” theory, according to which individuals tend to refrain from giving their opinion when they perceive themselves to be in the minority (Noelle-Neuman, 1984).

There are examples of the global incidence of cyber troopers. According to the “Global Inventory of Organised Social Media Manipulation” from 2019 (Bradshaw and Howard, 2020) organized manipulation campaigns have been detected in at least 70 countries across the globe in which AI tools are used. In 2018, the same study discovered them in 48 countries and the previous year, they were found in 28 countries. This phenomenon is not only an increasing trend on a global level, but furthermore, in each of the examined countries, there is at least one political group that makes use of the premeditated and organized strategy to manipulate public opinion (Bradshaw and Howard, 2020). However, their use differs depending on the specific country, although their detection and levels of interaction with human accounts is increasingly sophisticated (Santana and Huerta, 2019).

According to Pastor *et al.*, (2020), during the 2019 general election campaign in Spain, a significantly high number of bots appeared. Furthermore, depending on their behavior, they were capable of generating not only negative sentiment, but also of diffusing and disseminating fake news and hate messages in the automatic accounts identified with each party.

This type of phenomenon is not only harmful for the course of the debate in social media, but it may also have the more immediate consequence of spreading political conversation in highly polarized areas. According to Sartori (2005), polarization does not necessarily have negative connotations since two types exist: centripetal polarization is that in which opposing groups are open to dialogue and group together around the center; however, with the centrifuge, communities are divided and situate around the edges, avoiding dialogue and having harmful democratic consequences. This last model of polarization is not only the most likely, but, given the high levels of incivility (Rowe, 2015) and negative sentiment inundating the internet, it may also be the most dangerous.

In addition, there is the fact that the detection of these astroturfing campaigns and fake accounts depends entirely on user criteria (Yan *et al.*, 2020). The lack of CEO management of the main social media (Yan *et al.*, 2020) as well as the spectator role played by the states in response to the disinformation orchestrated by the political participants creates extreme citizen sentiment with a sense of impunity in the face of these hate messages (Neyazi, 2019).

Given the unexpected health crisis caused by the COVID-19 pandemic, we are faced with a scenario in which malicious AI tools implemented in the Twitter debate may create an even more hazardous environment of polarization and incivility. Therefore, we have proposed a study of the intervention of bots in the political conversation during the state of alarm in Spain, highlighting their consequences.

## METHODOLOGY

Given the previous, the following hypotheses have been proposed:

H1: During the lockdown period and de-escalation phases in Spain, AI algorithms were used in Twitter to condition the political conversation.

HS1: Disinformation campaigns were created by groups of bots that were strategically organized and linked to traditional socio-political participants with the objective of generating a false sense of support or disapproval of political topics discussed during the referred period.

To resolve these hypotheses, the following research objectives have been proposed:

- To determine the presence of political bots in the Twitter social network during the coronavirus crisis in Spain.
- To define the role of bots in the social network, as well as their relationship with the main political parties in Spain.
- To discover the potential consequences of the presence of bots in online public debate taking place in Twitter during the examined period.

## Case study and data source

From March 15<sup>th</sup> to June 21<sup>st</sup> of 2020, a State of Alarm was decreed in Spain in response to the COVID-19 pandemic, resulting in a harsh at-home lockdown and a subsequent period of de-escalation in phases. This exceptional social situation was an unprecedented administrative, healthcare and political test for all of the global authorities, especially for the newly-formed Spanish coalition government. At this juncture, the public debate took a marked turn towards the social networks. Not only were these networks a space where conversation revolved around the healthcare situation, they also focused on the political management by the new government, one of the most recurrent issues.

The case study of this work focuses on the messages published in Twitter, discuss-

ing the political debate with regard to the COVID-19 pandemic. The selected time period is that of the duration of the at-home lock down and the de-escalation phases in Spain.

Data were downloaded from the Twitter API using R-Studio, with the “rtweet” package (Kearney, 2019). In all, five download batches were used, having a weekly duration each, across the entire period selected, in order to consider the entire process:

- 1<sup>st</sup> download batch: “2020-03-16” - “2020-03-23”.
- 2<sup>nd</sup> download batch: “2020-04-14” - “2020-04-21”.
- 3<sup>rd</sup> download batch: “2020-05-15” - “2020-05-22”.
- 4<sup>th</sup> download batch: “2020-06-08” - “2020-06-15”.
- 5<sup>th</sup> download batch: “2020-06-21” - “2020-06-29”.

Messages were downloaded according to a set of key labels congruent with the study objective, such as “government”, “Spain” and “State of Alarm”, as well as the accounts of the main Spanish political parties (@PSOE; @populares; @PODEMOS; @CiudadanosCs; @vox\_es) and their highest level representatives: (@sanchezcastejon; @pablocasado\_; @PabloIglesias; @InesArrimadas; @santi\_ABASCAL).

Finally, after the downloading process, a total of 4 895 747 tweets were obtained.

### Data cleansing

Below, a phase of manual coding was initiated in order to filter all of the messages that were congruent with the study objective. This permitted the application of “supervised automatic learning” algorithms to carry out automatic coding on the database and consider it in its entirety, in accordance with the proposed study objectives. Simple random sampling was carried out on 1500 tweets by batch, to obtain the manual coding sample. It should be highlighted that stratification by batches was necessary since the content of the public-political debate was not static and it was necessary to respect the evolving events.

Of all of the automatic classifiers to be used, linear SVM machines were selected, since they have been found to offer some of the best results in text categorization (Wang *et al.*, 2006). To apply the automatic learning algorithms, text processing was carried out to ensure proper use. First, the tokenization process was initiated. This consists of the separation of a specific tweet into all of the words that it contains. Then, a filter of known words “stopwords” was applied to eliminate words lacking importance, such as determinants, prepositions, etc. Next,

**TABLE 1.** Results of the linear SVM classifier for the “TOPIC” and “POSITION” variables

Batch	Variable	Precision	Sensitivity	Specificity	Kappa	F-Score	AUC
1	TOPIC	0.8017	0.9322	0.3846	0.3670	0.8620	0.6583
2	TOPIC	0.8167	0.5476	0.9213	0.5077	0.6556	0.7344
3	TOPIC	0.8267	0.7027	0.8995	0.6187	0.7596	0.8010
4	TOPIC	0.7867	0.7090	0.8494	0.564	0.7457	0.7791
5	TOPIC	0.7659	0.8758	0.6377	0.5216	0.8171	0.7567

Source: Own elaboration.

the data were transformed into *tf-idf* matrixes “term frequency – inverse document frequency”. Each of their rows represents a specific tweet and their columns represent all of the words appearing in the *corpus*. Finally, messages that were incongruent with the study were eliminated using the “e1071” package (Meyer *et al.*, 2019), with only those belonging to the study topic remaining. Below, the results obtained from the classification are presented:

**Detection of bots and their relationship with the Spanish political parties**

With the clean database, a total of 1 208 631 tweets were found, coming from 445 230 users. Subsequently, bot identification was performed using the “tweetbotornot2” package (Kearney, 2020). To proceed with this classification, and given the quantity of users as well as the restrictions of the Twitter API, a “FAST - gradient boosted” detection algorithm was used, provided by “tweetbot-

ornot2”. Using a conservative criterion to reduce the possibilities of detecting a real user as a bot, those accounts presenting a probability of being a bot that was over the highest quartile of the probability ( $P > 0.9754$ ) were classified as a bot.

In order to identify each bot with a specific political party, assignment was carried out according to the similarity found in communication styles of the tweets. Automatic learning methodology was applied to train a classifier with the communication styles of the main Spanish political parties. In this way, the language use of each bot was contrasted with the distinct communicative styles of the main political parties to determine which had the greatest similarity. For this, a total of 10 264 tweets were selected from their official accounts, as well as those messages generated from the 350 political representatives in the Spanish Parliament. Finally, a linear SVM was applied to train the classifier with the communication style linked to the political party.

**TABLE 2.** Results of the linear SVM classifier for the “POLITICAL PARTY” variable

Variable	Precision	Sensitivity	Specificity	Kappa	F-Score
GENERAL	0.9683			0.9557	0.9632
PODEMOS		0.9490	0.9939		
PSOE		0.9828	0.9794		
CIUDADANOS		0.8591	0.9974		
PP		0.9472	0.9953		
VOX		0.9887	0.9935		
OTHER		0.9587	0.9964		

Source: Own elaboration.

**Sentiment analysis**

On the other hand, a sentiment analysis was carried out on the tweets sent by the automatic accounts in order to determine the emotional tone of their content. To do so, the “afinn” dictionary was used, containing of a

total of 2477 words. They all have a score ranging from -5 to 5: the lower the score, the more negative the sentiment and vice versa (Hansen *et al.*, 2011). This type of tests forms part of the dimensional theoretical models. That is, they study emotions through concrete dimensions based on the dichotomous

division between positivity and negativity. With this analysis, the most frequently used words in the messages generated by the bots are also shown. The word “coronavirus” was added to this dictionary with a score of -5, in order to register the term given its importance in the messages. The value of the score of the coronavirus term is justified by the negative connotations that it had during the study period, in healthcare, economic, social and political contexts.

Simultaneously and following the same procedure, a sentiment analysis was carried out for the messages sent from the selected political party accounts. In this way, it was possible to observe whether or not they resemble the levels of negativity or positivity of the artificial users.

### Detection of communities and analysis of networks

To determine Twitter user activity from a topological perspective, the network has been graphically represented through re-tweets. Due to the large scale of the database, the *Gephi* (Bastian, Heymann and Jacomy, 2009) application was used for its graphic representation. As for the detection of communities, the standard algorithm of Blondel *et al.* (2008) was applied, together with the “ForceAtlas2” algorithm for the improvement of network visualization.

First, the representation of the network for those users codified as “no bots” was conducted to determine the topographic structure of the network, made up of the non-automated users. This network consisted of a total of 194 992 nodes and 419 130 edges. Given that a network of re-tweets is considered, it is important to note that, despite having a total of 359 709 “no bot” users, only 194 992 of these users can be included in the specific network being analyzed.

Second, the network of bots consisted of a total of 79 246 nodes, made up of the

“bot” users and the users with whom they interacted, as well as 118 260 edges. For this network, as extra information, the item “belonging to each of the nodes” was added. Given that these are the messages that were coded with respect to a political party, for those cases in which a bot had only made a re-tweet, it was coded with the party to which the message content corresponds. For those who had re-tweeted various messages, the mode was calculated and they were assigned the political party that appeared the most times in the list of re-tweets of said bot. Finally, a list of attributes was obtained, created for a political party corresponding to each node or bot. To facilitate the visualization of the networks (no bots – Graph 6 and bots – Graph 7), users with a range of lower than 8 were excluded from the visualization.

## RESULTS

During the period of at-home lockdown, and the de-escalation period, it was estimated in this study that approximately 19.20 % of the Twitter users who participated in the political debate were robots. It is of great importance to note that, of all of the messages selected for the sample, 12.91 % were created and launched on the network by artificial algorithms. In addition, the main and predominant activity of the automatic users was to issue and disseminate content, since 80.84 % of the shared content was sent from re-tweets to other accounts.

As for the identification of the automatic accounts with the main political parties, in Graph 2 it is possible to observe the quantity of publications that have been produced in automatic accounts during the selected period, coded according to the communication styles of each party. Thus, it is possible to appreciate a predominance of messages from the PSOE party, created by artificial accounts, followed by conservative parties, PP and Vox,

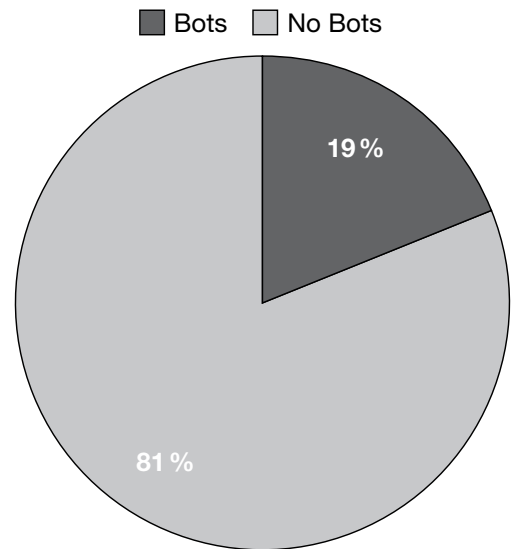
followed by Podemos and with a minor number of messages having the communication styles of Ciudadanos and Others.

On the other hand, it may be noted that, according to the content analysis performed, the messages sent by the bots revolve around three main axes (Graph 3). One of the most recurrent was that of health, including repeated words such as “healthcare”, “health” or “Covid”. Another frequent issue in the automatic account tweets is that of economics, with frequently used terms such as “debt”, “crisis” or “cutbacks”. Finally, there is the topic of politics, in which all of the messages referring to or mentioning the political parties and their representative have been grouped together.

As shown in the sentiment analysis, uncivil and hostile language is most often present in the dialogue of the bots (Graph 4). The percentage of messages having negative sentiments is greater than the percentage revealing positive sentiments. The tweets from the fake accounts that use a communication style similar to that of Podemos have more positive sentiments, as compared to those of the other political parties. Similar percentages are found for PSOE, Ciudadanos and PP, while the tweets corresponding to the style of Vox demonstrate the highest levels of negative sentiments.

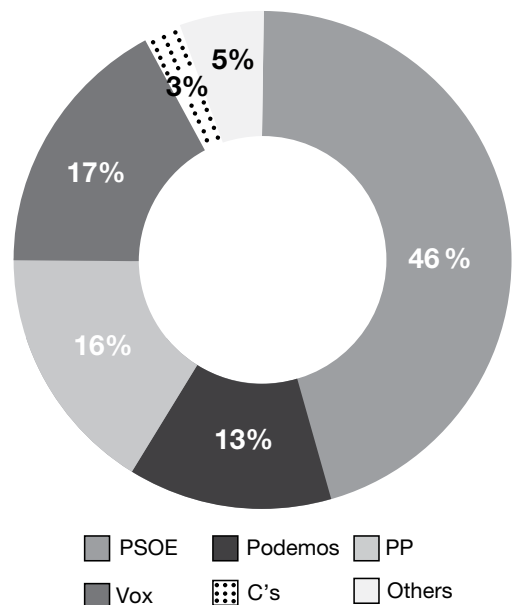
To complete the sentiment analysis, a graph is presented (Graph 5) reflecting the words with negative connotations that are most frequently used by the bots based on their affinity with each party. It is evident that, in each of the graphs, three recurrent topics appear in the Twitter discourse from the automatic accounts (health, economics and politics). However, in terms of the most commonly used words, substantial differences are observed depending on the political party at hand. The bots similar to the Partido Popular (PP) and Vox make greater use of uncivil language, using hostile and negative words.

**GRAPH 1.** Percentage of bots and real accounts



Source: Own elaboration.

**GRAPH 2.** Tweets created by bots and their identification with each party

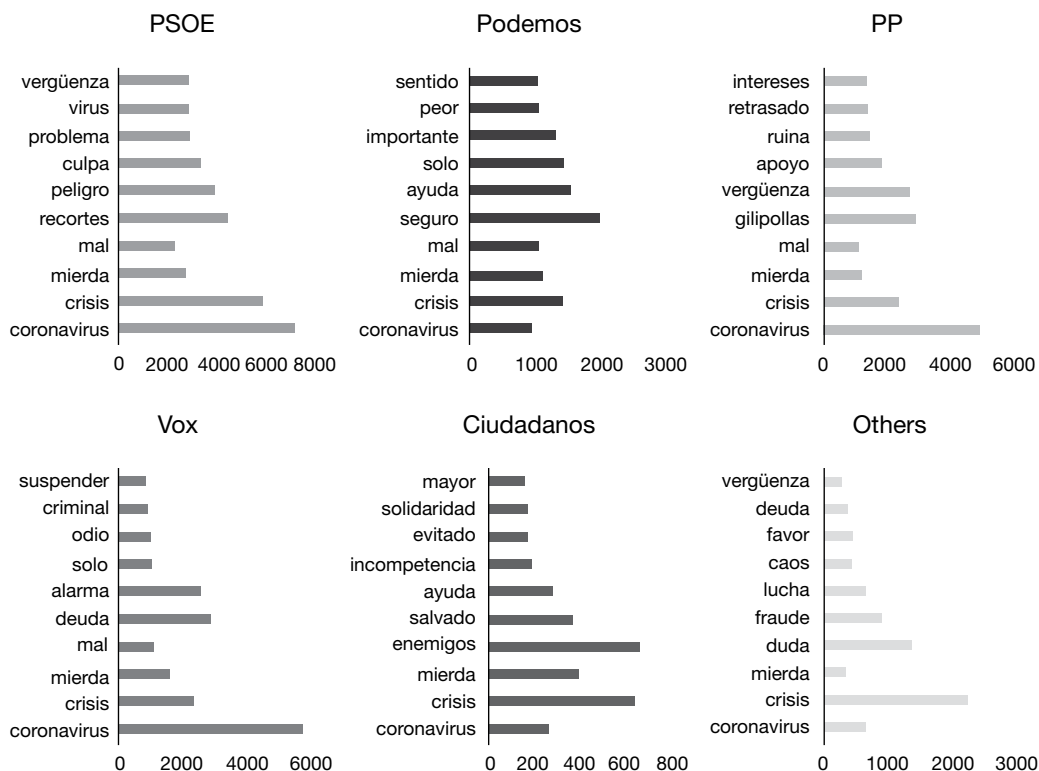


Source: Own elaboration.





**GRAPH 5.** Terms having negative connotations by party

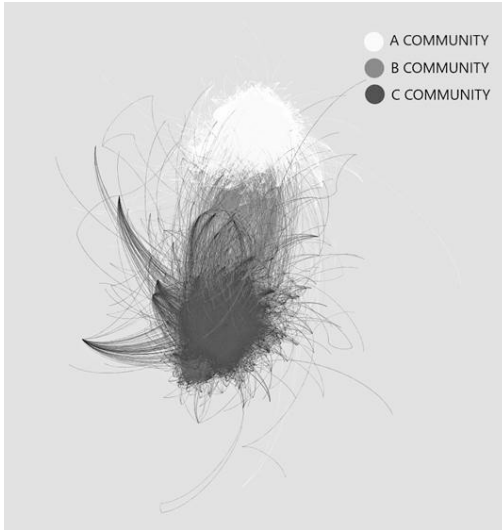


Source: Own elaboration.

As discussed in the previous section, an analysis was conducted to detect individual communities, both for the bot users as well as the real accounts. In Graph 6, the network of real users (or not bots) is represented.

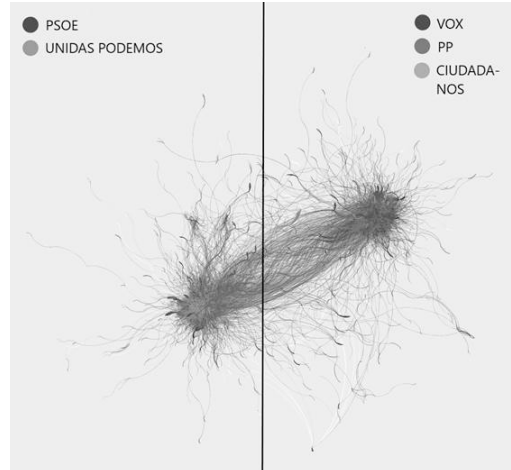
After applying the algorithm of community detection, a network of “no bots” was obtained, with a modularity of 0.538 and 582 communities. Modularity, in general terms, is a measure offering information on the quality of the communities detected in the network, with a maximum value of 1. Of all of the detected communities, three groups predominate, making up 77.73 % of all of the users. The Community A represents the leftist political spec-

trum, with 40.01% of the network users, with the accounts of Pedro Sánchez and Pablo Iglesias being especially noteworthy, as well as those of the following political parties: Izquierda Unida, Podemos and PSOE. On the other hand, Communities B and C, making up 29.03% and 8.69% respectively, symbolize the heart of the right and far right ideologies. The most characteristic accounts of this part of the graphic spectrum are those of the following political parties: PP and Vox, accompanied by those of their respective leaders, Pablo Casado and Santiago Abascal. Finally, it may be seen that these three communities make up two distinct and divided groups, observing a structural polarization of the network.

**GRAPH 6.** *Network of real users*

Source: Own elaboration.

Regarding the analysis of bot communities, it may be said that their network includes a total of 4601 groups with a modularity of 0.766, more in the network of bots than in the network of non-automated users. As is the case for the network of real users, very few main communities are detected. In this case, two of these make up a total of 41.35 % of the total automated accounts (Community A = 21.83 % and Community B = 19.52 %). Unlike the previous case, here fifteen small communities are detected (between 5.77 % and 1 % of the total), with a large number of insignificantly sized communities remaining. In Graph 7, the network of bots is represented, with its two main communities in which the nodes are colored based on the political party with which they are associated (see Graph 2). From this perspective, even though a community may be made up of distinct parties, considering the color of the parties whose communication style is similar to that of the bot, it is possible to see how the fake accounts group together around two opposing political discourses, with the discourse of the left-wing (PSOE and Unidas Podemos) on one side, and that of the right

**GRAPH 7.** *Network of fake accounts*

Source: Own elaboration.

(Vox, PP and, to a lesser degree, Ciudadanos) on the other side.

Here, it should be noted that, with respect to the results of the analysis of detection of communities, a division into two main groups was detected for both the real users and the bots, indicating the potential existence of high levels of political polarization. Both the main communities as well as the smaller ones were clearly divided around an axis. These two halves represent the traditional political break between left and right, in which the lack of centrality and bridges between communities is a symptom of the ideological division existing in the current political scenario.

## CONCLUSIONS

The unexpected health crisis caused by the COVID-19 pandemic and the resulting State of Alarm increased the imbalance already existing in western democracies, presenting a scenario of vulnerability with respect to the creation of public opinion. However, the communicative and deliberative charac-

teristics offered by the social media, viewed from a cyber-optimistic perspective, offer great benefits, in democratic terms. The fact that multiple channels of debate exist may increase real civic participation in political affairs (Santana and Huerta, 2019).

However, various circumstances offer evidence of the resistance to change in social order that may arise in the social networks. Interference caused by AI instruments in the political debate carried out over Twitter is proof of this. During the lockdown period and the de-escalation phases, automatic accounts were used to intervene in political conversation in Twitter: in this study, it is found that 19.20 % of the participating users were, in fact, bots. This is facilitated by the very structure of the network, suggesting an opposition to the use of these networks as spaces for horizontal democratic deliberation.

Given the consonance of the fake accounts with the main Spanish political parties, as demonstrated by the analysis, the hypothesis that those creating disinformation campaigns carried out by coordinated bot groups that are linked to traditional socio-political participants have the objective of creating a false sense of support or disagreement with the main principles discussed in Twitter. The communicative roles with which the automatic accounts have been programmed, as well as the results that they cause in the debate are characteristic of “astroturfing” campaigns.

Once this has been determined, it is apparent that in this scenario, the bots act as bullhorns of biased content, congruent with the ideological and propaganda-based objective of their relevant parties. Their behavior patterns correspond to what Hwang and Woolley (2016) defined as facilitator bots: spreading messages to give them the greatest reach possible. This strategy attempts to increase conversation regarding certain situations, not only to create the illusion that this is the majority positioning, but

also to blur the scope and predominance of opposing positions.

The debate revolving around bots focuses on the three most widely discussed issues in Twitter: health, politics and economics, confirming the fight for the hegemony of ideological positioning between parties and rival ideological blocks. However, in addition to attempting to falsely occupy and inflate the topics of conversation, negative messages, uncivil language and harassment are some of the most common characteristics found in the dialogue generated by the fake accounts.

Data from the sentiment analysis is congruent with reality: the bots that are identified with those of Podemos, a party that forms a part of the current government, tend to disseminate messages having more positive messages, as compared to the others. This may translate into the adopting of strategies that create a sense of support and legitimation of the measures and policies adopted by the government during the State of Alarm. However, the bots that are identified with those of the ultra-right party, Vox and the conservative party, Partido Popular (political parties that are currently part of the opposition government and that are critical of the government’s decisions), tend to maintain a clearly negative posture, with significantly higher levels of “incivility”.

Thus, the dimension demonstrating harmful levels of sentiment in the messages indicates that the political strategies designed for the bots is based on an attack of the opposition. This behavior may create roots of negative partisanship between the real users of Twitter. These samples are reflected in the structural break of the network observed in this study. The division of the Twitter, in both the real and fake accounts of the networks, revolving around two opposing ideologies, is a symptom of the potential existence of centrifugal polarization. These two large groups represent the traditional break between the left and

the right. However, the communities reflected in this study are affected by negative partisanship and bullying through language that serves only to destroy bridges and distance the network users even further from the center.

Clearly, the massive dissemination of partisan and biased propaganda, camouflaged by the shape and structure of social activists, forms part of the disinformation campaigns used to politically manipulate and sway public opinion. These conditions, which complicate the legitimacy of the participative processes and compromise democratic values (Santana and Huerta, 2019), have led to a greater break in consensus in an already polarized scenario.

This study reveals that the dissemination of messages within a negative campaign by automatic accounts has created a network of confrontation in Twitter, as opposed to one of dialogue. The resulting fracture in the network of real users, as observed in the analysis, is a reflection of the traditional Spanish political situation which, far from being overcome thanks to the possibilities offered by the networks, has merely moved to the online space.

It should be highlighted that ICT has the clear potential to help create a more horizontal, inclusive and deliberating society with empowered citizens. However, this potential social empowerment is being suppressed by the interests of the traditional socio-political participants who resist change and thereby act to jeopardize democratic wellbeing.

In conclusion, the limitations of this study relate to the methodology used. Due to the large sample dimensions, the sentiment analysis was carried out automatically. Even though words were added as considered relevant for the study of the tool, the fact that it works with a pre-established dictionary may impoverish the results obtained. Therefore, a qualitative content analysis

may be useful in order to offer more details on the meaning of the messages.

On the other hand, as for future studies that may be derived from this work, future works may consider the potential responsibility of political parties and organizations with respect to the use of political bots. It should be noted that this is a correlational study, as opposed to a causal one. Therefore, even considering the results that have been attained, it is not possible to establish a cause-effect relationship that extends beyond the linguistic consonance. The apparent polarization found in the online political debate should also be considered, as well as the main consequences of the same.

## BIBLIOGRAPHY

- Balcells, Joan and Cardenal, Ana S. (2013). "Internet and Electoral Competition: The Case of Esquerra Republicana de Catalunya"/"Internet y la competición electoral: el caso de Esquerra Republicana de Catalunya". *Revista Española de Investigaciones Sociológicas*, 141: 3-28.
- Bastian, Mathieu; Heymann, Sebastien and Jacomy, Mathieu (2009). "Gephi: An Open Source Software for Exploring and Manipulating Networks". *Third International ICWSM Conference*, 3(1): 361-362. doi: 10.13140/2.1.1341.1520
- Benkler, Yochai (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.
- Bennet, W. Lance and Segerberg, Alexandra (2012). "The Logic of Connective Action. Digital Media and the Personalization of Contentious Politics". *Information, Communication and Society*, 15(5): 739-768. doi: 10.1080/1369118X.2012.670661
- Bennet, W. Lance; Segerberg, Alexandra and Yang, Yunkang (2018). "The Strength of Peripheral Networks: Negotiating Attention and Meaning in Complex Media Ecologies". *Journal of Communication*, 68(4): 659-684. doi: 10.1093/joc/jqy032
- Blondel, Vincent D.; Guillaume, Jean-Loup; Lambiotte, Renaud and Lefebvre, Etienne (2008). "Fast Unfolding of Communities in Large Networks". *Journal of Statistical Mechanics: Theory and Ex-*

- periment*, 2008(10):1-12. doi: 10.1088/1742-5468/2008/10/p10008
- Bode, Leticia; Hanna, Alexander; Yang, Junghwan and Shah, Dhavan V. (2015). "Candidate Networks, Citizen Clusters, and Political Expression: Strategic Hashtag Use in the 2010 Midterms". *The ANNALS of the American Academy of Political and Social Science*, 659: 149-165. doi: 10.1177/0002716214563923
- Boshmaf, Yazan; Muslukhov, Ildar; Beznosov, Konstantin and Ripeanu, Matei (2013). "Design and Analysis of a Social Botnet". *Computer Networks*, 57(2): 556-578.
- Boulianne, Shelley (2018). "Twenty Years of Digital Media Effects on Civic and Political Participation". *Communication Research*, 47(7): 947-966. doi: 10.1177/0093650218808186
- Boulianne, Shelley and Theocharis, Yannis (2018). "Young People, Digital Media, and Engagement: A Meta-Analysis of Research". *Social Science Computer Review*, 38(2): 111-127. doi: 10.1177/0894439318814190
- Bradshaw, Samantha and Howard, Phillip N. (2018). *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*. Available at: <http://comprop.oii.ox.ac.uk/wpcontent/uploads/sites/93/2018/07/ct2018.pdf>, acces December 15, 2020.
- Bradshaw, Samantha and Howard, Phillip N. (2020). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. Oxford: Project on Computational Propaganda.
- Cáceres Zapatero, María D.; Brändle, Gaspar and Ruiz San Román, José A. (2015). "Hacia la construcción de una ciudadanía digital". *Prisma Social*, 15: 643-684.
- Castells, Manuel (2012). *Networks of Outrage and Hope: Social Movements in the Internet Age*. Cambridge: Polity.
- Dader, José L. and Campos Dominguez, Eva (2017). "La búsqueda digital del voto: Ciber campañas electorales en España 2015-2016". Valencia: Tirant Lo Blanch.
- Fazil, Mohd and Abulaish, Mohammad (2017). "Identifying Active, Reactive, and Inactive Targets of Socialbots in Twitter". *Proceedings of the International Conference on Web Intelligence*: 573-580.
- Gallardo, Beatriz and Enguix, Salvador (2016). *Pseudopolítica: el discurso político en las redes sociales*. Valencia: Universitat de València.
- Habermas, Jürgen (2006). "Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension The Impact of Normative Theory on Empirical Research". *Communication Theory*, 411-426. doi: 10.1111/j.1468-2885.2006.00280.x
- Hansen, Lars K.; Arvidsson, Adam; Nielsen, Finn A.; Colleoni, Elanor and Etter, Michael (2011). "Good Friends, Bad News-affect and Virality in Twitter". *Future Information Technology*: 34-43. doi: 10.1007/978-3-642-22309-9\_5
- Howard, Phillip N. (2006). *New Media Campaigns and the Managed Citizen*. Cambridge: Cambridge University Press.
- Howard, Phillip N.; Savage, Saiph; Flores, Claudia; Toxtli, Carlos and Monroy-Hernández, Andrés (2017). "Social Media, Civic Engagement, and the Slacktivism Hypothesis: Lessons From Mexico's 'El Bronco'". *Journal of International Affairs*, 70(1): 55-73. Available at: <https://www.jstor.org/stable/90012597?seq=1>, acces December 15, 2020.
- Hwang, Tim and Wooley, Samuel (2016). *How Politicians Should or Shouldn't Use Twitter Bots*. Available at: <https://slate.com/technology/2016/03/how-politicians-should-use-twitter-bots.html>, acces December 15, 2020.
- Kearney, Michael W. (2020). *tweetbotnot2: Detect Twitter Bots. R package version 0.0.1..* Available at: <https://github.com/mkearney/tweetbotnot2>, acces December 15, 2020.
- Keller, Franziska B.; Schoch, David; Stier, Sebastian and Yang, JungHwan (2019). "Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign". *Political Communication*, 37(2): 256-280. doi: 10.1080/10584609.2019.1661888
- Keyes, Ralph (2004). *The Post-truth Era: Dishonesty and Deception in Contemporary Life*. New York: Macmillan.
- Margolis, Michael and Resnick, David (2000). *Politics as Usual: The Cyberspace "Revolution"*. Thousand Oaks: Sage Publications.
- Margolis, Michael and Moreno-Riaño, Gerson (2009). *The Prospect of Internet Democracy*. London: Routledge.
- Meyer, David; Dimitriadou, Evgenia; Hornik, Kurt; Weingessel, Andreas; Leisch, Friedrich; Chang, Chih-Chung and Lin, Chih-Chen (2019). "e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-3". *Cran.r-project*.

- org. Available at: <https://cran.r-project.org/web/packages/e1071/>, acces December 15, 2020.
- Molyneux, Logan and Mourao, Rachel (2017). "Political Journalists' Normalization of Twitter: Interaction and New Affordances". *Journalism Studies*, 20(1): 1-19. doi: 10.1080/1369118X.2012.670661
- Murolo, Leonardo (2019). La posverdad es mentira. Un aporte conceptual sobre *fake news* y periodismo. In: R. Aparici and D. García-Marín (coords.). *La Posverdad. Una cartografía de los medios, las redes y la política*. Barcelona: Gedisa.
- Neuman, Russell W.; Guggenheim, Lauren; Jang, S. Mo and Bae, Soo Y. (2014). "The Dynamics of Public Attention: Agenda-setting Theory Meets Big Data". *Journal of Communication*, 64(2): 193-214. doi: 10.1111/jcom.12088
- Neyazi, Tabereh A. (2019). "Digital Propaganda, Political Bots and Polarized Politics in India". *Asian Journal of Communication*, 30(1): 39-57. doi: 10.1080/01292986.2019.1699938
- Noelle-Neumann, Elisabeth (1984). *The Spiral of Silence: Public Opinion, Our Social Skin*. Chicago: University of Chicago Press.
- Orbegozo Terradillos, Julen; Morales i Gras, Jordi and Larrondo Ureta, Ainara (2020). "Desinformación en redes sociales: ¿compartimentos estancos o espacios dialécticos? El caso Luther King, Quim Torra y *El Confidencial*". *Revista Mediterránea de Comunicación*, 11(2): 55-69. doi: 10.14198/MEDCOM2020.11.2.2
- Papacharissi, Zizi (2004). "Democracy Online: Civility, Politeness and the Democratic Potential of Online Political Discussion Groups". *New Media and Society*, 6(2): 259-283. doi: 10.1177/1461444804041444
- Pastor-Galindo, Javier; Zago, Mattia; Nespole, Pantaleone; López-Bernal, Sergio; Huertas-Celdrán, Alberto; Gil-Pérez, Manuel; Ruipérez-Valiente, José A.; Martínez-Pérez, Gregorio and Gómez-Mármol, Félix (2020). "Spotting Political Social Bots in Twitter: A Use Case of the 2019 Spanish General Election". *IEEE Transactions on Network and Service Management*, 17(4): 2156-2170. doi: 10.1109/tnsm.2020.3031573
- Persily, Nathaniel (2017). "The 2016 U.S. Election: Can Democracy Survive the Internet?". *Journal of Democracy*, 64(2): 64-76. doi: 10.1353/jod.2017.0025
- Price, Kristopher R.; Priisalu, Jaan and Nomin, Sven (2019). "Analysis of the Impact of Poisoned Data within Twitter Classification Models". *IFAC-PapersOnLine*, 52(19): 175-180. doi: 10.1016/j.ifacol.2019.12.170
- Rheingold, Howard (2009). *La comunidad virtual: una sociedad sin fronteras*. Barcelona: Gedisa.
- Robles, José M. and Córdoba, Ana M. (2019). *Digital Political Participation: Disintermediation in the Era of Web 2.0*. London: Palgrave.
- Rowe, Ian (2015). "Civility 2.0: A Comparative Analysis of Incivility in Online Political Discussion". *Information, Communication and Society*, 18(2): 121-138. doi: 10.1080/1369118X.2014.940365
- Santana, Luis E. and Huerta Cánepa, Gonzalo (2019). "¿Son bots? Automatización en redes sociales durante las elecciones presidenciales de Chile 2017". *Cuadernos.info*, 44: 61-77. doi: 10.7764/cdi.44.1629
- Sartori, Giovanni (2005). *Parties and Party Systems: A Framework for Analysis*. Colchester: ECPR press.
- Shao, Chengcheng; Ciampaglia, Giovanni L.; Varol, Onur; Yang, Kai-Cheng; Flammini, Alessandro and Menczer, Filippo (2018). "The Spread of Low-credibility Content by Social Bots". *Nature Communication*, 9(1): 1-10. doi: 10.1038/s41467-018-06930-7
- Stella, Massimo; Ferrara, Emilio and De Domenico, Manlio (2018). "Bots Increase Exposure to Negative and Inflammatory Content in Online Social Systems". *Proceedings of the National Academy of Sciences*, 115(49): 12435-12440. doi: 10.1073/pnas.1803470115
- Stier, Sebastian; Bleier, Arnim; Lietz, Haiko and Strohmaier, Markus (2018). "Election Campaigning on Social Media: Politicians, Audiences and the Mediation of Political Communication on Facebook and Twitter". *Political Communication*, 35(1): 50-74. doi: 10.1177/1461444817709282
- Stoking, Galen and Sumida, Nami (2018). "Social Media Bots Draw Public's Attention and Concern". *Pew Research Center*. Available at: <https://www.journalism.org/2018/10/15/social-media-bots-draw-publics-attention-and-concern/>, acces December 15, 2020.
- Walker, Edward T. (2014). *Grassroots for Hire: Public Affairs Consultants in American Democracy*. New York: Cambridge University Press.
- Wang, Zi-quiang; Sun, Xia; Zhang, De-xian and Li, Xin (2006). "An Optimal SVM-Based Text Classification Algorithm". *International Conference on Machine Learning and Cybernetics*: 1378-1381. doi: 10.1109/ICMLC.2006.258708

Ward, Stephen and Gibson, Rachel (2009). European Political Organizations and the Internet. In: A. Chadwick and P. N. Howard (eds.). *Routledge Handbook of Internet Politics* (pp. 25-39). Abingdon: Routledge.

Winner, Langdon (2003). "Internet y los sueños de una renovación democrática". *Isegoría*, 28: 55-71. doi: 10.3989/isegoria.2003.i28.506

Yan, Harry Y.; Yang, Kai-Cheng; Menczer, Filippo and Shanahan, James (2020). "Asymmetrical Perceptions of Partisan Political Bots". *New Media and Society*: 1-22. doi: 10.1177/1461444820942744

**RECEPTION:** February 9, 2021

**REVIEW:** June 11, 2021

**ACCEPTANCE:** November 23, 2021