Comunicar, nº 71, v. XXX, 2022 | Revista Científica de Educomunicación | ISSN: 1134-3478; e-ISSN: 1988-3293
www.revistacomunicar.com | www.comunicarjournal.com

Cuando la negatividad es el combustible. Bots y polarización política en el debate sobre el COVID-19

When negativity is the fuel. Bots and political polarization in the COVID-19 debate

Dr. José-Manuel Robles. Profesor Titular, Departamento de Sociología Aplicada, Universidad Complutense de Madrid (España) (jmrobles@ucm.es) (https://orcid.org/0000-0003-1092-3864)
Dr. Juan-Antonio Guevara. Investigador Posdoctoral, Departamento de Sociología Aplicada, Universidad Complutense de Madrid (España) (juanguev@ucm.es) (https://orcid.org/0000-0003-3946-3910)
Dra. Belén Casas-Mas. Profesora Ayudante Doctor, Departamento de Sociología, Universidad Complutense de Madrid (España) (bcasas@ucm.es) (https://orcid.org/0000-0001-8329-0856)
Dr. Daniel Gómez. Profesor Titular, Departamento de Estadística y Ciencia de los Datos, Universidad Complutense de Madrid (España) (dagomez@estad.ucm.es) (https://orcid.org/0000-0001-9548-5781)

RESUMEN

Los contextos de polarización social y política están generando nuevas formas de comunicar que inciden en la esfera pública digital. En estos entornos, distintos actores sociales y políticos estarían contribuyendo a extremar sus posicionamientos, utilizando «bots» para crear espacios de distanciamiento social en los que tienen cabida el discurso del odio y la «incivility», un fenómeno que preocupa a científicos y expertos. El objetivo principal de esta investigación es analizar el rol que desempeñaron estos agentes automatizados en el debate en redes sociales sobre la gestión del Gobierno de España durante la pandemia global de COVID-19. Para ello, se han aplicado técnicas de «Social Big Data Analysis»: algoritmos de «machine learning» para conocer el posicionamiento de los usuarios; algoritmos de detección de «bots»; técnicas de «topic modeling» para conocer los temas del debate en la red, y análisis de sentimiento. Se ha utilizado una base de datos compuesta por mensajes de Twitter publicados durante el confinamiento iniciado a raíz del estado de alarma español. La principal conclusión es que los «bots» podrían haber servido para diseñar una campaña de propaganda política iniciada por actores tradicionales con el objetivo de aumentar la crispación en un ambiente de emergencia social. Se sostiene que, aunque dichos agentes no son los únicos actores que aumentan la polarización, sí coadyuvan a extremar el debate sobre determinados temas clave, incrementando la negatividad.

ABSTRACT

The contexts of social and political polarization are generating new forms of communication that affect the digital public sphere. In these environments, different social and political actors contribute to extreme their positions, using bots to create spaces for social distancing where hate speech and incivility have a place, a phenomenon that worries scientists and experts. The main objective of this research is to analyze the role that these automated agents played in the debate on social networks about the Spanish Government's management of the global COVID-19 pandemic. For this, "Social Big Data Analysis" techniques were applied: machine learning algorithms to know the positioning of users; bot detection algorithms; "topic modeling" techniques to learn about the topics of the debate on the web, and sentiment analysis. We used a database comprised of Twitter messages published during the confinement, as a result of the Spanish state of alarm. The main conclusion is that the bots could have served to design a political propaganda campaign initiated by traditional actors with the aim of increasing tension in an environment of social emergency. It is argued that, although these agents are not the only actors that increase polarization, they do contribute to deepening the debate on certain key issues, increasing negativity.

PALABRAS CLAVE | KEYWORDS

COVID-19, bots políticos, polarización política, propaganda digital, opinión pública, análisis de redes sociales.
COVID-19, political bots, political polarization, digital propaganda, public opinion, social networks analysis.

Recibido: 2021-10-04 | Revisado: 2021-10-19 | Aceptado: 2021-12-03 | OnlineFirst: 2022-01-15 | Publicado: 2022-04-01
DOI https://doi.org/10.3916/C71-2022-05 | Páginas: 63-75

1. Introducción y estado de la cuestión

La polarización que se produce en los debates sociales y políticos que tienen lugar en redes como Twitter o Facebook se ha transformado en un fenómeno cada vez más relevante para las ciencias sociales. No solo porque puede generar una brecha entre las partes que participan en el debate público, sino porque dicha brecha se produce como consecuencia de estrategias que, como la «incivility» o el «flaming», están basadas en el odio, el descrédito de una de las partes, el insulto, etc. En definitiva, la polarización no es solo relevante por sus consecuencias, sino por la emergencia de «modos de comunicar» que pueden generar un estado de «comunicación fallida».

En términos generales, la polarización política se ha asociado a procesos de exposición selectiva a la información. Se entiende que dicha limitación favorece el desarrollo de valores, actitudes y posiciones políticas extremas o, al menos, basadas en el reforzamiento de las posiciones previas de los agentes involucrados en el debate (Prior, 2013). En este sentido, las redes sociales digitales favorecerían la polarización al permitir un mayor control sobre el tipo y las fuentes de información (Sunstein, 2001; 2018). En este contexto, los expertos en polarización han alertado sobre la importancia que adquieren distintos actores y agentes sociales y políticos en el desarrollo de procesos de polarización. Así, se ha evidenciado que, posiciones extremas adoptadas por líderes políticos, generan un efecto de propagación que termina afectando a la posición de sus seguidores. En la medida en que los líderes sesgan o parcializan la información o la interpretación de un acontecimiento mediático, dicho sesgo aleja a sus seguidores de la posibilidad de entenderse con los representantes de opiniones opuestas (Boxell et al., 2017).

Nuestro trabajo pretende indagar en este terreno al centrarnos en el rol de un agente que puede ser clave en los procesos de polarización: los «bots». Aunque la literatura sobre la injerencia de este tipo de agentes no humanos en los procesos sociales, políticos y electorales es amplia (Ferrara et al., 2016; Howard et al., 2018; Keller & Klinger, 2019), no lo es tanto el estudio sobre su rol en los procesos de polarización. Cabe plantearse la pregunta de si los «bots» estarían contribuyendo a polarizar el debate político en situaciones de convulsión social. En este sentido, nuestro principal objetivo es estudiar en qué medida y mediante qué estrategias estos agentes crean o extreman procesos de polarización política en debates que tienen lugar en redes sociales digitales. Para alcanzar este objetivo, tomaremos como caso de estudio el debate público que tuvo lugar en redes sociales sobre la gestión que el Gobierno de España realizó durante los primeros meses de la crisis sanitaria generada por la pandemia global de COVID-19.

1.1. La polarización política digital

Ciertamente, no existe una definición de polarización universalmente aceptada. Abramowitz (2010) la ve como un proceso que dota de mayor consistencia y fortaleza las actitudes y opiniones de ciudadanos y partidos; y otros como Fiorina y Abrams (2008) enfatizan el distanciamiento de los puntos de vista de las personas en un contexto político determinado. Igualmente, los expertos están divididos entre aquellos que plantean la centralidad de una polarización ideológica y otros que proponen la emergencia de una polarización afectiva (Lelkes, 2016). No obstante, existen procesos de polarización que podrían ser «esperables» (Sartori, 2005), por ejemplo, en contextos bipartidistas especialmente si el sistema electoral es presidencialista. Esto es, cuando dos candidatos se enfrentan en un proceso electoral, es esperable que sus seguidores se estructuren en torno a dos polos (el que generan a su alrededor cada uno de los candidatos). Lo mismo puede suceder cuando el debate surge en torno a un tema que plantea dos posiciones claramente delimitadas (e.g. a favor o en contra). Los estudios en polarización han centrado su atención recientemente, no tanto en el estudio de la formación de los polos, sino en aquellos aspectos que llevan a esa estructura hacia un proceso negativo para el debate público. Así, numerosos trabajos sugieren que, el hecho de no estar en contacto con información plural y/o fidedigna, intencionalmente negativizada y/o incompleta es lo que genera la transformación de la polarización en un proceso negativo.

La tesis de autores como Prior (2013) es que el efecto que media entre la exposición selectiva a la información y la polarización negativa es el enrocamiento o la radicalización de las posiciones de partida de los seguidores que articulan ambos polos. Los expertos han señalado distintos mecanismos para explicar este proceso. Desde un punto individualista metodológico y experimental, Taber y Lodge (2006) iniciaron un ámbito de análisis sobre los componentes psicológicos y racionales para la selección de aquellas piezas

de información que mejor coincidían con las concepciones previas que tienen los individuos. En este sentido, los humanos tendríamos la tendencia a filtrar la información de forma que se identifiquen como más relevantes o veraces aquellas que coinciden con las concepciones o disposiciones emotivas previas. Por su parte, Sunstein (2001; 2018) señala el relevante rol que tiene Internet y las redes sociales digitales en los procesos de polarización. Desde este punto de vista, Internet ofrece la posibilidad de seleccionar de forma más precisa las fuentes de información a la que los ciudadanos están expuestos, así como las personas con las que deciden debatir.

Por último, distintos autores muestran cómo los principales actores políticos son los agentes clave para la polarización al generar un efecto de propagación de sus mensajes potencialmente sesgados y/o negativos (Allcott & Gentzhow, 2017). Sabemos que los niveles de polarización de los ciudadanos están estrechamente relacionados con la pertenencia a determinados grupos sociales y al consumo de cierta información política (Boxell et al., 2017). Esta circunstancia puede ser consecuencia de la polarización que reciben de los principales agentes políticos: partidos, organizaciones, medios, etc., con lo que nos encontraríamos ante un fenómeno de contagio. El peligro que tiene este efecto contagio es lo que Lelkes (2016) llama «affective polarization», proceso en el que los ciudadanos tienden a radicalizar sus emociones y/o afectos sobre diversos temas, así como a enrocarse, siguiendo los discursos polarizados de partidos políticos y representantes públicos. En esta línea, Calvo y Aruguete (2020: 60-70) consideran que la polarización afectiva en las redes constituye «una defensa encendida de creencias propias ante los objetivos comunicacionales del otro» y afirman que «odiar las redes es un acto afectivo, cognitivo y político».

Estudios recientes (Mueller & Saeltzer, 2020) también se refieren a este contagio en las redes originado por los mensajes que evocan emociones negativas y apuntan a que la comunicación afectiva negativa surge de manera voluntaria como resultado de campañas estratégicas (Martini et al., 2021). En este contexto, la «incivility» surge como una estrategia comunicativa que, gracias a la generación de emociones negativas mediante el insulto o el desprestigio social, trata de excluir al adversario del debate público (Papacharissi, 2004). Nuestro trabajo se enmarca en este último contexto teórico en el que la polarización es entendida como proceso de extensión de las actitudes de líderes políticos o, como es el caso, de herramientas digitales que adquieren roles centrales en el debate.

1.2. El rol de los «bots» en la propaganda digital

Los avanzados sistemas de IA y de microsegmentación consiguen movilizar a los usuarios a través de las redes sociales. Concretamente, en comunicación política digital estas tácticas de propaganda van más allá de las noticias falsas con ánimo de lucro y las teorías de la conspiración, ya que consisten en el uso deliberado de información errónea para influir en las actitudes sobre un tema o hacia un candidato (Persily, 2017). Entre estas estrategias propagandísticas se encuentran los «bots» políticos, que son cuentas de redes sociales controladas total o parcialmente por algoritmos informáticos. Crean contenido automáticamente para interactuar con usuarios, a menudo haciéndose pasar por humanos o imitándolos (Ferrara et al., 2016).

La finalidad principal es romper los flujos de debate en las redes mediante la difamación de los oponentes o de aquellos usuarios que expongan opiniones opuestas (Yan et al., 2020). Algunos autores señalan que el uso de «bots» no siempre va ligado a fines maliciosos, presentando la utilidad en Twitter de informar a la población sobre los riesgos de la pandemia de COVID-19, por ejemplo, para difundir noticias veraces de última hora o para instar a los ciudadanos a quedarse en casa (Al-Rawi & Shukla, 2020). Sin embargo, pese a que el uso de los «bots» en las pandemias todavía es un campo de estudio poco investigado, ya hay constancia empírica de que han servido para promover dinámicas de conspiración en la esfera política multimedia relativas a la difusión de mensajes polémicos y polarizados (Moffit et al., 2021). En contraposición a los «bots» sociales de carácter inclusivo, diversas investigaciones destacan los «bots» políticos cuya función es difundir mensajes que implican emociones negativas (Stella et al., 2018; Neyazi, 2019; Yan et al., 2020; Adlung et al., 2021), extender masivamente noticias falsas (Shao et al., 2018; Shu et al., 2020; Yan et al., 2020) y utilizar la información privada de los usuarios con fines políticos partidistas (Boshmaf et al., 2013; Persily, 2017; Yan et al., 2020). Diversos autores advierten que, en un entorno de redes cada vez más polarizado y convulsionado, los «bots» políticos incrementan

el nivel de vulnerabilidad de los usuarios porque tienen mayor capacidad de segmentarlos y orientar su propaganda (Stella et al., 2018; Yan et al., 2020). Price et al. (2019) añaden la dificultad que tienen las diversas herramientas de detección de «bots» debido al desarrollo constante en su capacidad de mejora y modificación de comportamiento. El aumento de la crispación política en la esfera digital pública también se vincula a campañas de desinformación en las redes sociales como el «astroturfing»: una actividad iniciada por actores políticos en Internet, fabricada de forma estratégica de arriba hacia abajo, imitando la actividad de abajo hacia arriba por parte de individuos autónomos (Kovic et al., 2018: 71). Consisten en un conjunto de robots coordinados por activistas de base que emulan ser ciudadanos comunes y actúan de forma independiente. Tienen el potencial de influir en los resultados electorales y en el comportamiento político posicionándose a favor o en contra de diversas causas (Howard, 2006; Walker, 2014; Keller et al., 2019). Conocidos también como «Twitter bombs» (Pastor-Galindo et al., 2020) o «cibertroops» (Bradshaw & Howard, 2019), operan difundiendo comentarios muy semejantes entre sí y coherentes con el objetivo de la campaña propagandística (Keller et al., 2019). Frecuentemente, los rasgos que caracterizan este tipo de mensajes robotizados son el uso de información falsa, del lenguaje incívico y de los mensajes de odio contra grupos minoritarios o de opinión opuesta, a los que se hostiga e intenta excluir del debate (Keller et al., 2019; Santana & Huerta-Cánepa, 2019).

Más allá del perjuicio que supone esta práctica para el flujo natural de las conversaciones en las redes sociales, el problema principal es que suelen derivar en procesos de fuerte polarización. Cuando estas dinámicas implican posiciones extremas que impiden el diálogo se denominan polarización centrífuga (Sartori, 2005) y pueden suponer una amenaza para la democracia (Morgan, 2018). Papacharissi (2004) se refiere a este tipo de polarización como «incivility» y específica que implica el uso de un lenguaje inapropiado, insultante o denigrante. Mensajes que invaden el debate polarizado de la red y que atentan contra las libertades personales o de determinados grupos sociales (Rowe, 2015). Por su parte, Sobieraj y Berry (2011) se refieren a la indignación («outrage») como un tipo de discurso político cuyo fin es provocar respuestas viscerales en la audiencia como el miedo o la indignación moral, mediante el uso de exageraciones, sensacionalismo, mentiras, información inexacta o verdades parciales que afectan a individuos, organizaciones o grupos específicos.

La literatura revisada apunta a que el discurso del odio en el ciberespacio puede ser alimentado por actores no humanos como los «bots» sociales, lo que conlleva un problema global agravado por la actual crisis del COVID-19 (Uyheng & Carley, 2020). Entendemos que este uso de herramientas nocivas de IA polariza y aumenta la «incivility» en el debate político generado en torno a la pandemia. La polarización derivada de esta crisis sanitaria ha sido objeto de estudio en redes como Youtube (Serrano-Contreras et al., 2020; Luengo et al., 2021). Entendemos que la presencia de «bots» en los debates durante esta crisis sanitaria es un objeto de estudio en el que hay que profundizar. Nuestra hipótesis es que, dichos agentes, no son los únicos actores clave en los procesos de polarización, pero sí utilizan situaciones polarizadas para extremar el debate incluyendo mayor grado de negatividad, especialmente, en determinados temas clave. Por ello consideramos necesario investigar la participación de estos agentes y sus efectos durante el Estado de Alarma en Twitter.

2. Material y métodos

Nuestra base de datos se compone de mensajes descargados («tweets») de la red social Twitter a lo largo de todo el confinamiento impuesto con la instauración del estado de alarma en España. Para llevar a cabo nuestros objetivos, se aplicaron técnicas de «Social Big Data Analysis», como algoritmos de «machine learning» para conocer la posición de los usuarios en la red hacia el Gobierno de España, algoritmos de detección de «bots», técnicas de «topic modeling» para conocer los temas de los que se compone el debate en la red social y análisis de sentimiento.

2.1. Fuente de datos y limpieza

Los datos fueron descargados de la API de Twitter a través de R-Studio con la librería «rtweet» (Kearney, 2019). Se descargaron de acuerdo con un conjunto de palabras clave, compuesto por los nombres de las cuentas de los principales partidos políticos de España, las de sus líderes políticos, además

de incluir las palabras «estado de alarma», «coronavirus» o «COVID». La base de datos que se generó abarca un período comprendido entre el 16 de marzo de 2020 al 29 de junio de 2020. La descarga de datos se realizó en 5 tandas diferentes, durante la primera semana de cada una de las fases del estado de alarma, con el fin de abarcar el período en su totalidad. Finalmente, se recolectaron 4.895.747 mensajes.

Se llevó a cabo una limpieza de los datos con el fin de eliminar aquellos mensajes descargados que no pertenecían al objetivo de estudio. Para ello, se aplicó la metodología «machine learning». Así mismo, como consecuencia del dinamismo del debate en las redes, se entrenaron tantos algoritmos como diferentes tandas de descarga. Para ello, se generó un muestreo aleatorio simple de 1.500 «tweets» por tanda para una codificación manual previa por un experto entrenado previamente. Esta codificación consistió en etiquetar los mensajes como «pertenece» o «no pertenece» al objetivo de estudio. Tras este proceso, se aplicaron algoritmos de aprendizaje automático, siendo las máquinas de soporte vectorial lineal (Support Vector Machines–SVM) aquellas que mostraron un mejor rendimiento, encontrando una precisión media entre tandas de 0,8 y una medida F media de 0,768. Para llevar a cabo esta tarea, se aplicó un procesamiento del texto para su correcta compatibilidad con algoritmos de aprendizaje automático. Primero, se procedió a la «tokenización» del contenido, separando un determinado «tweet» en todas las palabras que incluye. Segundo, se eliminaron aquellas palabras cuyo contenido no aporta información relevante «stopwords», como determinantes, preposiciones, etc. Para finalizar, se construyó una matriz tf-idf («term frequency – inverse document frequency») como input para los algoritmos de aprendizaje automático, donde cada fila representa un «tweet» y las columnas representan todas las palabras que aparecen en el corpus. Tras la aplicación del SVM-lineal, restaron un total de 1.208.631 mensajes que se corresponden con el objetivo de estudio publicados por 469.616 usuarios.

2.2. Detección de «bots»

Para detectar y clasificar los usuarios como «bots» o «no bots», se aplicó el algoritmo propuesto por Kearney (2019) denominado «tweetbotornot» en su versión «FAST–gradient boosted», incorporado en el paquete estadístico de R-Cran «tweetbotornot». Así mismo, con el fin de mantener un criterio conservador, etiquetar únicamente como «bots» aquellos usuarios que presenten una probabilidad de ser «bots» situada en el cuartil más alto. Consideramos de vital importancia mantener este criterio con el fin de, a riesgo de detectar menos «bots», no incluir ningún usuario real en la categoría «bot».

2.3. Medición de la polarización

Para la medición de la polarización en redes sociales se ha aplicado la medida propuesta por Guevara et al. (2020) basada en la lógica difusa, denominada JDJ. Los autores se basan en la premisa de que la realidad no es nítidamente de una forma o de su contraria, sino que existen diferentes matices en las actitudes de las personas. Se entiende que, si bien un determinado individuo puede ser, por ejemplo, partidario de un partido político, esto no implica que no pueda estar de acuerdo con algunas propuestas de otros partidos políticos diferentes. De esta forma, en vez de contemplar de forma nítida la puntuación actitudinal de una persona, se computa el grado de pertenencia (o cercanía) que tiene la actitud de esa persona a los polos del eje actitudinal que se está midiendo. De esta forma, se contempla de forma simultánea la posición de un individuo hacia los extremos de una variable. Así, se entiende que existe riesgo de polarización entre un individuo «i» y un individuo «j» como la consideración conjunta de los siguientes escenarios:

- Cómo de cerca está un individuo «i» al polo A y cómo de cerca está un individuo «j» al polo B.
- Cómo de cerca está un individuo «i» al polo B y cómo de cerca está un individuo «j» al polo A.

De esta forma, la polarización total de un conjunto poblacional es el sumatorio de todas las comparaciones posibles entre los individuos que la componen.

- Dada una variable X.
- Cada individuo $i \in N$.
- X_A,X_B son los polos de X y $\mu X_A \mu X_B$ las funciones de pertenencia de un individuo a los polos: $\mu X_A \mu X_B$:$N \rightarrow [0,1]$ son funciones, y para cada $i \in N$ $\mu {X_A}^{(i)}$ y $\mu {X_B}^{(i)}$ y son las funciones de pertenencia de un individuo «i» a ambos polos.

$$JDJ(X) = \sum_{i,j \in N, i \leq j} \varphi\left(\phi\left(\mu_{X_A}(i), \mu_{X_B}(j)\right), \phi\left(\mu_{X_B}(i), \mu_{X_A}(j)\right)\right)$$

Donde Φ es un operador de agregación de «overlapping» y φ es la función de agrupación. En este estudio, se ha utilizado como operador de «overlapping» el producto y como función de agrupación el máximo. Esta medida presenta su valor máximo cuando el 50% de la población tiene un grado de pertenencia máximo al polo A y un grado de pertenencia nulo al polo B, y el otro 50% de la población presenta un grado de pertenencia máximo al polo B y un grado de pertenencia nulo al polo A. Por otro lado, se encuentra un nivel nulo de polarización no solo cuando el 100% de la población presenta el mismo nivel de actitud, sino cuando este valor está situado en torno a un extremo, siendo este escenario el que presenta una mayor distancia con el valor máximo de polarización. Dado que la ecuación anterior presenta como resultado la suma del riesgo de polarización para todas las combinaciones de pares de individuos posibles, con el fin de facilitar su interpretación se calcula la siguiente operación:

$$JDJ\_INDEX = \frac{JDJ}{N} * 2$$

Donde N es el número total de individuos. De esta forma, la medida muestra su valor mínimo en 0 y su máximo en 1. En Guevara et al. (2020) se observa detalladamente una comparación de esta propuesta con otras medidas de la literatura.

2.4. Detención de topics

Para la detección de temáticas en el discurso presente en los mensajes descargados, se hizo uso del algoritmo Latent Dirichlet Allocation (LDA) presente en el paquete de R llamado «topicmodels» (Grün & Hornik, 2011). Este algoritmo se basa en crear distancias entre palabras de acuerdo con la aparición conjunta de las mismas entre sí. El algoritmo presenta la particularidad de tener que indicar el número de «topics» a priori, por lo que para la correcta determinación del número de temáticas se pueden aplicar medidas de coherencia semántica por «topics». En este sentido, se recomienda adicionalmente la exploración del contenido por parte de un experto para determinar el número de «topics» adecuado.

2.5. Análisis de sentimiento

Se aplicaron diccionarios de análisis de sentimiento con el fin de detectar la cantidad de contenido negativo o positivo presente en el debate digital estudiado. Para ello, se hizo uso del diccionario Afinn (Hansen et al., 2011), compuesto por 2.477 palabras, puntuadas de mayor negatividad a mayor positividad, en una escala de uno a cinco.

3. Análisis y resultados

3.1. Clasificación de mensajes como «a favor» o «en contra» hacia el Gobierno

En primer lugar, se aplicaron los algoritmos de aprendizaje automático para codificar un determinado mensaje como «a favor» o «en contra» hacia el Gobierno. En este caso, también las máquinas de soporte vectorial mostraron mejores resultados (Tabla 1). Como se puede observar, los resultados indican niveles de rendimiento satisfactorios que permiten la correcta clasificación automática de la totalidad de los mensajes presentes en la base de datos.

Tabla 1. Resultados del clasificador SVM-lineal para la Codificación de mensajes como «a favor» o «en contra» del Gobierno de España

Medidas de precisión					
Tanda	**Precisión**	**Sensibilidad**	**Kappa**	**F-Score**	**AUC**
1	0,8492	0,9854	0,4816	0,9122	0,6950
2	0,8960	0,9619	0,7761	0,9277	0,8780
3	0,8392	0,8488	0,6675	0,8439	0,8366
4	0,9133	0,9048	0,8225	0,9090	0,9121
5	0,8318	0,8600	0,6638	0,8456	0,8335

3.2. Detección de «bots»

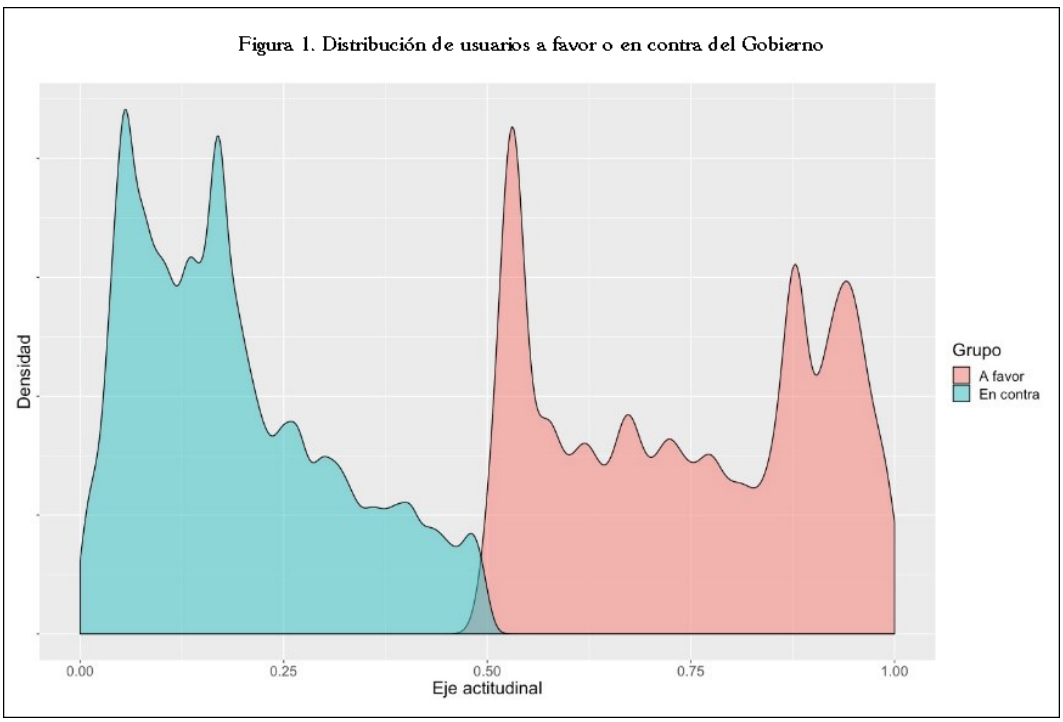
Seguidamente, se aplicó el algoritmo de detección de bots para identificar como «bots» aquellas cuentas automatizadas. El criterio llevado a cabo se corresponde con el de clasificar como «bot» aquellos usuarios cuya probabilidad de serlo está situada en el cuartil más alto de dicha probabilidad, siendo la misma de > 0,975. Al aplicar el algoritmo a los 469.616 usuarios de la base de datos, se detectaron 69.033 cuentas definibles como «bots». Esto supone un total del 15% de todas las cuentas presentes en el debate digital. Igualmente, los 69.033 «bots» publicaron 172.704 mensajes de los 1.208.631 que presenta la base de datos filtrada, suponiendo un 14,28% del total.

3.3. Medición de la polarización

En primer lugar, es importante recordar que para el cálculo de la polarización se usan las probabilidades de estar «a favor» o «en contra» del Gobierno ofrecidas por los algoritmos de aprendizaje automático como los grados de pertenencia de los usuarios a ambos polos –favor/contra–, por lo que para un determinado individuo «i» se tienen dos valores: 1) su probabilidad de estar a favor del Gobierno y 2) su probabilidad de estar en contra del Gobierno. Sin embargo, el «input» de los clasificadores automáticos son «tweets», siendo el objetivo calcular la polarización por usuarios. Para ello, se calculó para cada usuario la media de las probabilidades de estar a favor y en contra del Gobierno de todos sus mensajes publicados. Así, para cada usuario se obtienen los dos grados de pertenencia necesarios que se usarán para calcular la medida de polarización. Por otro lado, debido a los costes computacionales de aplicar la medida a 469.616 usuarios, lo que supone la comparación de todos los usuarios entre ellos, sería necesario realizar un total de $(469.616^2)/2=(469\ 616^2)/2=110.269.593.728$ cálculos de JDJ. Así, se procede a calcular el índice JDJ como la media de 1.500 iteraciones de JDJ para una muestra aleatoria simple de N=200 usuarios por iteración.

En primer lugar, se midió la polarización para la muestra general (no «bots» y «bots»), obteniendo un nivel de $(JDJ\_mean)_{1500}=0,76$; sd=0,027, siendo un nivel alto de polarización JDJ_mean→[0,1]. En la Figura 1 se muestra la representación gráfica de la distribución de los usuarios a estar a favor o en contra del Gobierno, donde una probabilidad 0,5 supone estar en contra del Gobierno, mientras que >0,5 estar a favor.

Figura 1. Distribución de usuarios a favor o en contra del Gobierno



Así mismo, se calcularon los niveles de polarización para 1) aquellos usuarios que no son «bots» y 2) para los «bots». Así, JDJ mostró una media del nivel de polarización para las 1.500 iteraciones de $JDJ\_mean\_no\_bots_{1500}=0{,}761$; sd=0,026. Por su parte, para el grupo de «bots» se encontró que $JDJ\_mean\_bots_{1500}=0{,}765$; sd=0,026. Finalmente, se llevó a cabo la aplicación de la prueba Mann-Witney para dos variables independientes con el fin de conocer si las diferencias encontradas en las medias son estadísticamente significativas. Así, con un estadístico U=1021715, se encontró un nivel de significación <0,000, encontrando un mayor nivel de polarización para los usuarios «bots» respecto a los no «bots».

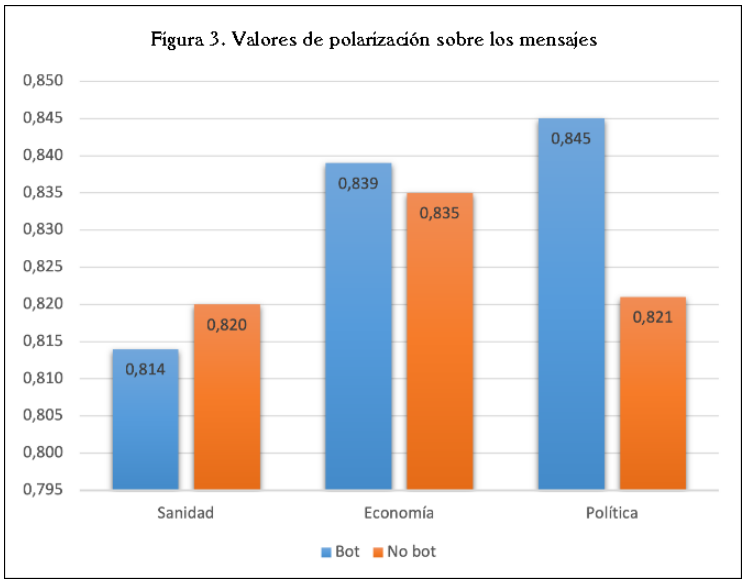
3.4. Detección de topics y polarización

En primer lugar, se aplicó el algoritmo LDA de detección de topics. Dado que los grupos se establecen a priori, se aplicó el algoritmo reiteradas veces cambiando el parámetro «número de grupos» de uno a diez con el fin de calcular la coherencia para cada número de «clusters». Es importante recordar que este índice de coherencia se basa en la similitud semántica entre palabras por lo que, tras la supervisión de un experto de las diferentes propuestas del algoritmo, se decidió optar por la detección de tres temáticas (coherencia de 0,42) en el diálogo ya que, como se puede observar en la Figura 2, se encuentran bien delimitadas (sanidad, economía y política). Por otro lado, cabe mencionar que debido a los costes computacionales del algoritmo LDA se pudo acceder a la totalidad de los mensajes publicados por «bots» (N=172.704), mientras que se aplicó sobre una muestra aleatoria simple de N=200.000 a los 1.035.927mensajes del grupo de no «bots» De acuerdo a las fórmulas de cálculo de tamaño muestral (Fernández, 1996), con una población de N=1.035.927, un nivel de confianza del 99% y un margen de error del 1%, la muestra necesaria para que sea representativa del total es de 16.378 mensajes, por lo que consideramos suficiente una muestra aleatoria simple N=200.000 para conseguir la representatividad del resto de mensajes de usuarios no «bots».

Figura 2. Temáticas en el diálogo

El número de temáticas encontradas fueron iguales para ambos grupos («bots» y no «bots»), encontrando un dominio del debate en los usuarios no «bots» de la temática economía (53,48%), seguido por sanidad (25,95%) y política (20,55%), mientras que para los «bots» la temática mayoritaria es la política (48,36%), seguida por la sanidad (36,04%) y economía (15,58%). Una vez identificadas las temáticas, se procedió a calcular los niveles de polarización con JDJ. Para ello, se siguió el mismo procedimiento de calcular la media de polarización sobre 1.500 iteraciones para una muestra aleatoria de N=200 por iteración. A diferencia del cálculo JDJ del apartado anterior, aquí se calculó la polarización sobre los

mensajes y no sobre usuarios. Como se puede ver en la Figura 3, los valores más altos de polarización se encontraron en los mensajes producidos por «bots», más concretamente los que hablan sobre política ($JDJ\_mean_{1500}=0{,}845$) y economía ($JDJ\_mean_{1500}=0{,}839$), seguidos por los mensajes publicados por los no «bots» que hablan de economía ($JDJ\_mean_{1500}=0{,}835$).



Para conocer si las diferencias en niveles de polarización son estadísticamente significativas, se realizó un ANOVA de 2 factores–user («bot» o no «bot») x topic (salud, economía y política) (Tabla 2). Dados los niveles de significación encontrados ($p<0{,}00$), cabe concluir que tanto el tipo de usuario como la temática afectan a los niveles de polarización. Además, se puede observar que el efecto de la interacción también es significativo, por lo que se concluye que los niveles de polarización encontrados en la variable «topic» están condicionados por ser, o no, un «bot».

Tabla 2. Pruebas de efectos inter-sujetos

Variable dependiente: Polarización					
Origen	**Tipo III de suma de cuadrados**	**gl**	**Media cuadrática**	**F**	**Sig.**
Modelo corregido	1.123[a]	5	,225	327.641	,000
Intersección	6.194,468	1	6.194,468	9.032.376,674	,000
USER	,121	1	,121	176.075	,000
TOPIC	,668	2	,334	486.967	,000
USER * TOPIC	,335	2	,167	244.100	,000
Error	6.168	8.994	,001		
Total	6201.759	9.000			
Total corregido	7.292	8.999			

a. R al cuadrado=,154 (R al cuadrado ajustada=,154).

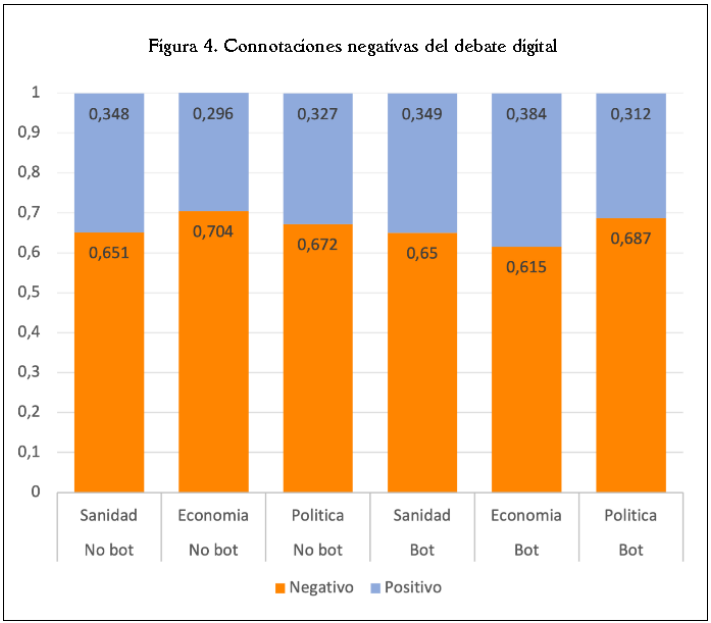
Finalmente, se creó una nueva variable fruto de la combinación («user» x «topic») con seis niveles para determinar cuál de estos escenarios presenta un mayor nivel de polarización. Así, se aplicó un ANOVA de un factor, mostrando un estadístico $F_{(5)}=327.641$, $p<0{,}000$, donde se realizaron comparaciones múltiples con la prueba de Tukey, encontrando diferencias estadísticamente significativas entre todos los niveles salvo para los niveles «no bot» – sanidad y «no bot» – política. Cabe asumir que se encontraron los tres mayores niveles de polarización del discurso en la temática política del debate de los «bots», seguidos por la temática economía en el debate de «bots» y el topic economía en los usuarios no «bots» (Figura 3).

3.5. Análisis de sentimiento y topics

Finalmente, se aplicó el diccionario Afinn para el análisis de sentimiento a cada una de las temáticas detectadas. Como se puede observar en la Figura 4, las palabras con connotaciones negativas predominan

en todo el debate digital, encontrando una mayor presencia en los mensajes que hablan sobre economía para el grupo de no «bots» (0,704%), seguido por la temática política en el discurso de los «bots» (0,687%) y política en no «bots» (0,672%).

Figura 4. Connotaciones negativas del debate digital



4. Discusión y conclusiones

En este artículo nos propusimos analizar en qué medida y con qué estrategias participan los «bots» políticos en procesos de discusión pública a través de las redes sociales digitales. En concreto, nos centramos, como caso ejemplar, en el debate a través de Twitter sobre la gestión del Gobierno de España durante la Pandemia global de COVID-19. El trasfondo teórico de esta investigación es avanzar en el conocimiento de un proceso comunicacional complejo y potencialmente dañino. Esto es, la polarización política y el surgimiento de un escenario de «comunicación fallida». En definitiva, hablaríamos de situaciones en las que la comunicación entre los participantes tiende a enrocarse en torno a posiciones fuertes y a ignorar, cuando no atacar («incivility»), a aquellos que piensan diferente.

Son muchos los expertos que señalan la presencia de la polarización política en procesos de debate público como el que aquí analizamos. No obstante, la propia definición de polarización apuntada en este artículo nos advierte de que, dicho fenómeno, no depende únicamente de la existencia de dos o más polos contrapuestos, sino de una tendencia hacia el aislamiento y ruptura de la comunicación entre las distintas partes (polos). Una dinámica derivada de las denominadas «echo chambers» en la red (Colleoni et al., 2014), en los que los intercambios de información se producen principalmente entre individuos con preferencias ideológicas similares, especialmente cuando se trata de cuestiones políticas (Barberá et al., 2015). Así, hemos mostrado (Figura 1) cómo el debate en torno a la gestión del Gobierno estuvo fuertemente polarizado. De hecho, nuestros datos apuntan a que los «bots» políticos identificados en el análisis tienen una mayor tendencia a polarizar la opinión pública que las cuentas que no son «bots». Este hallazgo es la antesala, causa necesaria pero no suficiente, para encontrar el tipo de polarización que nos alerta y que favorece una ruptura comunicativa.

El factor clave (razón suficiente) para esta ruptura es, desde nuestro punto de vista, la identificación de una estrategia de alejamiento y enrocamiento de las partes participantes en el debate. Más allá de la acción de los líderes políticos, mediáticos o de algunos líderes de opinión, en este trabajo, y gracias al análisis de sentimientos, hemos identificado una estrategia de negativización del debate que está más presente entre los «bots» que entre las cuentas no «bots». Esta estrategia busca, según nuestra interpretación, sesgar la opinión sobre la gestión del Gobierno de España. Este tipo de sesgos son los que señalan el camino

de la polarización en su sentido más negativo de ruptura y alejamiento. Una polarización afectiva que puede tener graves repercusiones, especialmente en momentos de agitación política (Iyengar, 2019) como la provocada por el inicio de la pandemia.

Dicha estrategia queda aún más clara cuando analizamos las temáticas centrales de nuestro caso de estudio. Aquí observamos cómo los «bots» tienden a centrar el debate, más que en temas económicos o sanitarios (temas de debate potencialmente más sujetos a criterios científicos y objetivables), en el terreno de la política. Se trata de un campo en el que es más fácil atacar a una o a varias figuras en vez de hablar de temas generales (como hemos visto la literatura apunta a que esta estrategia es una fuente de polarización). Es decir, es un terreno en el que es más accesible desarrollar estrategias «ad hominem», representada en la Figura 3 por las constantes referencias al Presidente del Gobierno de España, claramente más sesgadas y centradas en los defectos y circunstancias negativas que rodean a la persona. Es decir, una estratégica personalista y centrada en la «incivility» como forma de gestión de la comunicación.

Los temas de naturaleza política son, además, los más polarizados en este debate. Adicionalmente, son los que muestran una mayor diferencia entre «bots» y no «bots». Estos últimos presentan un debate más polarizado sobre la gestión de la pandemia en términos políticos siendo, de entre las tres temáticas, la más negativizada por parte de los «bots». En línea con el concepto de indignación propuesto por Sobieraj y Berry (2011) las temáticas encontradas en los bots refuerzan el melodrama y los pronósticos improbables fruto de una fatalidad inminente atribuidos a las decisiones tomadas por el Gobierno. Un discurso que se distingue por las tácticas utilizadas para provocar la emoción, más que por evocar la emoción en la arena política. Por ello, el uso de «bots» no parece estar orientado a informar a la sociedad sobre los riesgos de la pandemia o promover dinámicas de prevención (Al-Rawi & Shukla, 2020), sino mayoritariamente centrado en movilizar, negativizando, la opinión pública en contra del Gobierno.

Como apuntaban Howard (2006), Walker (2014) o Keller et al. (2019), los «bots» tienen el potencial de influir en el posicionamiento político de los usuarios en la red porque, según nuestros resultados, emulan ser ciudadanos comunes preocupados por cuestiones meramente sanitarias. Se pueden considerar «bots» políticos y no sociales porque difunden mensajes con sentimientos negativos (Stella el al., 2018; Neyazi, 2019; Yan et al., 2020; Adlung et al., 2021), concretamente hacia el Gobierno. Estos robots podrían haber sido diseñados para poner en marcha una campaña de propaganda política de «astroturfing» iniciada por actores tradicionales con el objetivo de aumentar la crispación en un contexto de emergencia social. Iyengar et al. (2019) advertían que se trata de un tipo estrategia estrechamente ligada a la teoría de la espiral del silencio porque, en un contexto de incertidumbre y frustración generalizada, dificulta que los usuarios expresen opiniones favorables con alguna de las medidas sanitarias tomadas por el Gobierno.

Nuestra impresión es que, el binomio polarización-negativización, es el combustible elegido por este tipo de cuentas para alejar y enfrentar las partes que participan en este debate público, así como para el surgimiento de un entorno de crispación, falta de civismo y ataques al que piensa diferente. En un contexto de por sí ya polarizado, sea por la acción de otros agentes (políticos, sociales o mediáticos) o por la propia situación de excepcionalidad e incertidumbre, el combustible utilizado por los «bots» en este debate ha consistido en extremar las posiciones de partida. Se trata de un hallazgo que puede servir de base para investigaciones futuras que puedan contrastarlo en diversos casos de estudio con características similares o distintas al aquí realizado.

Contribución de Autores

Idea, JMR, JAG, BC-M.; Revisión de literatura (estado del arte), JMR, BC-M; Metodología, JAG, DG.; Análisis de datos, JMR, JAG, BC-M, DG.; Resultados, JMR, JAG, BC-M.; Discusión y conclusiones, JMR, BC-M.; Redacción (borrador original), JMR, JAG, BC-M.; Revisiones finales, JMR, JAG, BC-M; Diseño del Proyecto y patrocinio, JMR.

Apoyos

Grupo de investigación Data Science and Soft Computing for Social Analytics and Decision Aid. Esta investigación ha sido apoyada por los proyectos de investigación nacional financiados por el Gobierno de España, con referencia I+D+i, PID2019-106254RB-100 financiación: MINECO (Duración: 2020-2024) y PGC2018-096509B-I00.

Referencias

Abramowitz, A.I. (2010). *The disappearing center*. Yale University Press. https://bit.ly/3s7UlwC

Adlung, S., Lünenborg, M., & Raetzsch, C. (2021). Pitching gender in a racist tune: The affective publics of the# 120decibel campaign. *Media and Communication*, *9*, 16-26. https://doi.org/10.17645/mac.v9i2.3749

Al-Rawi, A., & Shukla, V. (2020). Bots as active news promoters: A digital analysis of COVID-19 tweets. *Information*, *11*(10), 461-461. https://doi.org/10.3390/info11100461

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), 211-247. https://doi.org/10.1257/jep.31.2.211

Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*, 1531-1542. https://doi.org/10.1177/0956797615594620

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2013). Design and analysis of a social botnet. *Computer Networks*, *57*(2), 556-578. https://doi.org/10.1016/j.comnet.2012.06.006

Boxell, L., Gentzkow, M., & & J Shapiro (2017). *Is the internet causing political polarization? Evidence from demographics*. National Bureau of Economic Research. https://doi.org/10.3386/w23258

Bradshaw, S., & Howard, P.N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation*. Oxford Internet Institute. https://acortar.link/puyazU

Calvo, E., & Aruguete, N. (2020). *Fake News, trolls y otros encantos. Cómo funcionan (para bien y para mal) las redes sociales*. Siglo XXI. https://doi.org/10.22201/fcpys.24484911e.2020.29.76061

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*(2), 317-332. https://doi.org/10.1111/jcom.12084

Fernández, P. (1996). Determinación del tamaño muestral. *Cad Aten Primaria*, *3*, 1-6. https://bit.ly/3DYcijz

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96-104. https://doi.org/10.1145/2818717

Fiorina, M.P., & Abrams, S.J. (2008). Political polarization in the American public. *Annual Review of Political Science*, *11*, 563-588. https://doi.org/10.1146/annurev.polisci.11.053106.153836

Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, *40*(13). https://doi.org/10.18637/jss.v040.i13

Guevara, J.A., Gómez, D., Robles, J.M., & Montero, J. (2020). Measuring polarization: A fuzzy set theoretical approach. In M. Lesot, S. Vieira, M. Reformat, J. Carvalho, A. Wilbik, & B. B.-M. R. Yager (Eds.), *Information processing and management of uncertainty in knowledge-based systems* (pp. 510-522). Springer. https://doi.org/10.1007/978-3-030-50143-3_40

Hansen, L.K., Arvidsson, A., Nielsen, F.A., Colleoni, E., & Etter, M. (2011). Good friends, bad news-affect and virality in Twitter. In *Future information technology* (pp. 34-43). Springer. https://doi.org/10.1007/978-3-642-22309-9_5

Howard, P.N. (2006). *New media campaigns and the managed citizen*. Cambridge University Press. https://doi.org/10.1080/10584600701641532

Howard, P.N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the U.S. 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, *15*(2), 81-93. https://doi.org/10.1080/19331681.2018.1448735

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S.J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*(1), 129-146. https://doi.org/10.1146/annurev-polisci-051117-073034

Kearney, M.W. (2019). Rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software*, (42), 4-4. https://doi.org/10.21105/joss.01829

Keller, F.B., Schoch, D., Stier, S., & Yang, J.H. (2019). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, *37*(2), 256-280. https://doi.org/10.1080/10584609.2019.1661888

Keller, T.R., & Klinger, U. (2019). Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, *36*(1), 171-189. https://doi.org/10.1080/10584609.2018.1526238

Kovic, M., Rauchfleisch, A., Sele, M., & Caspar, C. (2018). Digital astroturfing in politics: Definition, typology, and countermeasures. *Studies in Communication Sciences*, *18*, 69-85. https://doi.org/10.24434/j.scoms.2018.01.005

Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, *80*(1), 392-410. https://doi.org/10.1093/poq/nfw005

Luengo, O., García-Marín, J., & De-Blasio, E. (2021). COVID-19 on YouTube: Debates and polarisation in the digital sphere. [COVID-19 en YouTube: Debates y polarización en la esfera digital]. *Comunicar*, *69*, 9-19. https://doi.org/10.3916/C69-2021-01

Martini, F., Samula, P., Keller, T.R., & Klinger, U. (2021). Bot, or not? Comparing three methods for detecting social bots in five political discourses. *Big Data & Society*, *8*(2). https://doi.org/10.1177/20539517211033566

Moffitt, J.D., King, C., & Carley, K.M. (2021). Hunting conspiracy theories during the COVID-19 pandemic. *Social Media+Society*, *7*(3). https://doi.org/10.1177/20563051211043212

Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. *Journal of Cyber Policy*, *3*(1), 39-43. https://doi.org/10.1080/23738871.2018.1462395

Mueller, S.D., & Saeltzer, M. (2020). Twitter made me do it! Twitter's tonal platform incentive and its effect on online campaigning. Information. *Communication & Society*, (pp. 1-26). https://doi.org/10.1080/1369118X.2020.1850841

Neyazi, T.A. (2019). Digital propaganda, political bots and polarized politics in India. *Asian Journal of Communication*, *30*(1), 39-57. https://doi.org/10.1080/01292986.2019.1699938

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & society*, *6*(2), 259-283. https://doi.org/10.1177/1461444804041444

Pastor-Galindo, J., Nespoli, P., Gómez-Mármol, F., & Mártinez-Pérez, G. (2020). Spotting political social bots in Twitter: A use case of the 2019 Spanish general election. *IEEE Transactions on Network and Service Management*, *8*, 10282-10304. https://doi.org/10.1109/access.2020.2965257

Persily, N. (2017). The 2016 U.S. election: Can democracy survive the internet. *Journal of Democracy*, *28*(2), 63-76. https://doi.org/10.1353/jod.2017.0025

Price, K.R., Priisalu, J., & Nomin, S. (2019). Analysis of the impact of poisoned data within twitter classification models. *IFAC-PapersOnLine*, *52*(19), 175-180. https://doi.org/10.1016/j.ifacol.2019.12.170

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, *16*, 101-127. https://doi.org/10.1146/annurev-polisci-100711-135242

Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, *18*(2), 121-138. https://doi.org/10.1080/1369118X.2014.940365

Santana, L.E., & Huerta-Cánepa, G. (2017). ¿Son bots? Automatización en redes sociales durante las elecciones presidenciales de Chile 2017. *Cuadernos.info*, *44*, 61-77. https://doi.org/10.7764/cdi.44.1629

Sartori, G. (2005). *Parties and party systems: A framework for analysis*. ECPR press.

Serrano-Contreras, I.J., García-Marín, J., & Luengo, O.G. (2020). Measuring online political dialogue: Does polarization trigger more deliberation? *Media and Communication*, *8*, 63-72. https://doi.org/10.17645/mac.v8i4.3149

Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A., Menczer, F., & Yang, K.C. (2018). The spread of low-credibility content by social bots. *Nature Communication*, *9*(1), 1-10. https://doi.org/10.1038/s41467-018-06930-7

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, *8*, 171-188. http://doi.org/10.1089/big.2020.0062

Sobieraj, S., & Berry, J.M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, *28*(1), 19-41. https://doi.org/10.1080/10584609.2010.542360

Stella, M., Ferrara, E., & De-Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. In J. Kleinberg (Ed.), *Proceedings of the National Academy of Sciences*, volume 115 (pp. 12435-12440). https://doi.org/10.1073/pnas.1803470115

Sunstein, C.R. (2001). *Designing democracy: What constitutions do?* Oxford University Press.

Sunstein, C.R. (2018). *#Republic*. Princeton university press.

Taber, C.S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, *50*(3), 755-769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Uyheng, J., & Carley, K.M. (2020). Bots and online hate during the COVID-19 pandemic: Case studies in the United States and the Philippines. *J Comput Soc Sc*, *3*, 445-468. https://doi.org/10.1007/s42001-020-00087-4

Walker, E.T. (2014). *Grassroots for hire: Public affairs consultants in American democracy*. Cambridge University Press. https://doi.org/10.1017/CBO9781139108829

Yan, H.Y., Yang, K., Menczer, F., & Shanahan, J. (2020). Asymmetrical perceptions of partisan political bots. *New Media & Society*, *23*(10), 3016-3037. https://doi.org/10.1177/1461444820942744