


Types of Astroturfing campaigns of disinformative and polarised content in times of pandemic in Spain

Tipos de campaña Astroturfing de contenidos desinformativos y polarizados en tiempos de pandemia en España

Tipos de campanhas Astroturfing de conteúdo desinformativo e polarizado em tempos de pandemia em Espanha

Sergio Arce-García^{1*} 

Elías Said-Hung^{2**} 

Daria Mottareale-Calvanese^{2***} 

¹ School of Engineering and Technology (ESIT), Universidad Internacional de la Rioja, Spain

² Faculty of Education at the Universidad Internacional de la Rioja, Spain

* Associate professor PhD at the School of Engineering and Technology (ESIT) at the Universidad Internacional de la Rioja (UNIR)

** Professor of the Faculty of Education at the Universidad Internacional de la Rioja

*** Researcher and teacher of the Faculty of Education at the Universidad Internacional de la Rioja

Received: 04/04/2022; Revised: 19/05/2022; Accepted: 14/07/2022; Preprint: 16/08/2022; Published: 01/01/2023

Translation to English: AJE (American Journal Experts)



To cite this article: Arce-García, Sergio; Said-Hung, Elías; & Mottareale-Calvanese, Daria. (2023). Types of Astroturfing campaigns of disinformative and polarised content in times of pandemic in Spain. *ICONO 14. Scientific Journal of Communication and Emerging Technologies*, 21(1). <https://doi.org/10.7195/ri14.v21i1.1890>

Abstract

The paper seeks to determine the application of Astroturfing strategies on Twitter in Spain during the COVID-19 pandemic in the spring of 2020. Statistical analysis, network analysis and machine learning techniques are used to evaluate approximately 32,527 messages published from the state of alarm decree in Spain (14 March, 2020) until the end of May of the same year, associated with eight tags that address issues related to misleading content identified by two of the main fact-checking projects (Maldito Bulo and Newtral). Data allow us to observe the participation of users (not bots) who play the role of influencers despite having an average profile or a profile that is far from being considered a public personality. The application of Astroturfing can be seen as a communication strategy used to position issues on social networks through the distribution, amplification and flooding of disinformation. The scenario allows us to verify the presence of a digital communication scenario that would favour a framework difficult to detect, from strategies such as the one studied, aimed at breaking the echo chamber and filter bubble of social networks, with the aim of positioning issues at the level of public opinion.

Keywords: disinformation; hoaxes; Astroturfing; political communication; digital media; social networks

Resumen

Este trabajo busca determinar la aplicación de estrategias de Astroturfing en Twitter, a nivel español, durante el periodo de pandemia a causa de la covid-19, en la primavera del año 2020. Se aplica análisis estadístico, análisis de redes y técnicas de machine learning, en 32.527 mensajes publicados a partir del decreto de estado de alarma en España (14 de marzo de 2020) hasta finales de mayo del mismo año, asociados ocho etiquetas que abordan temas vinculados a contenidos desinformativos identificados por dos de los principales proyectos de fact-checking (Maldito Bulo y Newtral). Los datos permiten observar la participación de usuarios (no bots), que ejercen un rol de influenciadores pese a que cuentan con un perfil promedio o alejado a ser considerados como personalidad pública. Se aprecia la aplicación del Astroturfing como estrategia de comunicación empleada para posicionar temas en las redes sociales a través de la distribución, amplificación y la inundación de contenidos desinformativos. El escenario permite comprobar la presencia de un escenario comunicativo digital que favorecería un marco difícil de detección, desde estrategias como la estudiada, orientadas a romper el efecto campana y filtro de burbuja de redes sociales. Todo con el fin de posicionar temas a nivel de opinión pública.

Palabras clave: contenidos desinformativos; bulos; Astroturfing; comunicación política; medios digitales; redes sociales

Resumo

Este documento procura determinar a aplicação de estratégias de Astroturfing no Twitter, a nível espanhol, durante o período pandémico devido à covid-19, na Primavera de 2020. Análise estatística, análise de rede e técnicas de aprendizagem de máquinas são aplicadas a 32.527 mensagens publicadas desde o decreto do estado de alarme em Espanha (14 de Março de 2020) até ao final de Maio de 2020, associadas a oito etiquetas que abordam tópicos relacionados com conteúdos desinformativos identificados por dois dos principais projectos de verificação de factos (Maldito Buló e Newtral). Os dados permitem-nos observar a participação dos utilizadores (não dos bots), que desempenham o papel de influenciadores apesar de terem um perfil médio ou um perfil que está longe de ser considerado uma personalidade pública. A aplicação de Astroturfing pode ser vista como uma estratégia de comunicação utilizada para posicionar questões sobre redes sociais através da distribuição, amplificação e inundação de conteúdo desinformativo. O cenário permite-nos verificar a presença de um cenário de comunicação digital que favoreceria um quadro difícil de detectar, a partir de estratégias como a que foi estudada, visando quebrar o efeito sino e filtrar a bolha das redes sociais. Tudo com o objectivo de posicionar as questões ao nível da opinião pública.

Palavras-chave: conteúdo desinformativo; hoaxes; Astroturfing; comunicação política; meios de comunicação digitais; redes sociais

1. Introduction

The growth in the phenomenon of the transmission of false information from social networks has generated an exponential increase in academic studies around the three main concepts that are usually used to address this phenomenon (Boididou et al., 2018), without an apparent clear demarcation or debate on the differences that these may have between them: false information, disinformation and misleading content (Said-Hung et al., 2021). Defining these terms is not an easy task since the similarities, differences and juxtapositions between them are very subtle and there are not enough mechanisms to identify them (Estrada-Cuzcano et al., 2020). Fake news, considered an intentional disinformation strategy with political and social objectives (Magallón, 2019), has begun to spread rapidly, thanks to digital media, replacing traditional hoaxes. The latter is defined as "all those false contents that reach public dissemination, intentionally fabricated for multiple reasons, which can range from simple jokes or parody, to ideological controversy, to economic fraud" (Salaverría et al., 2020, p. 4).

The study of the dissemination of misleading content has focused on understanding the phenomenon from the influence it has on electoral processes and public opinion (Van-der-Linden, 2017; Zerback & Töpfl, 2021) and the impact that social networks have on what is known as the echo chamber effect by amplifying ideas and beliefs, from the constant transmission and repetition of messages (not necessarily truthful) to the censorship of differentiated views (Flaxman et al., 2016; Guess et al., 2018). It has also focused on the filter bubble effect promoted by the overabundance of information in current digital scenarios (Flaxman et al., 2016).

Studies on the production and dissemination of fake news on the internet, defined as "viral publications based on fictitious accounts and made to look like real news" (Tandoc et al., 2018, p. 2), have focused in recent years on the construction of models that help to detect and identify the dissemination strategies of this type of content through computational techniques, network analysis and semiotics (Zheng et al., 2017; Howard et al., 2017; Zhao et al., 2020).

One of the focuses of the study of disinformation is the presence and/or use of bots or "social bots" (Ferrara et al., 2016; Allem, & Ferrara, 2018; Luceri et al., 2019), defined as "automated social network accounts operated automatically by malicious actors intending to manipulate public opinion" (Gallwitz & Kreil, 2021, p. 1). However, the propagation of misleading content comes in different forms (Kucharski, 2016); it should not be limited to the study of bots or the search for individual accounts with pronounced "robotic" behaviour (Keller et al., 2019).

Studies by authors such as Zheng et al. (2017), Howard et al. (2017) and Zhao et al. (2020) point to different propagation strategies, which vary significantly from truthful content. Messages that do not necessarily come from bots but from organic (real) users coordinate around a single purpose: to position a topic at a social level from social networks, around a certain event and for a certain period of time. Zheng et al. (2017) call these "Internet water armies" or "elite sibylline troops", as they are at the service of a political or economic group interested in "flooding" social networks and the internet with intentional content, trying to make the most of collective communicative action, typical of current digital scenarios.

As Guess et al. (2018) point out, the strategy applied by political movements from social networks is carried out with the support of websites created to give a sense of reality to the different arguments and messages created and redistributed (Zhao et al., 2020) or distributed, amplified and flooded by users (soldiers) far from the profile of influencers (many followers and high activity). A strategy where these users (micro- and nano-influencers, users with up to 100,000 and 10,000 followers or less, respectively) appear to be neither organised nor related to each other but who act in coordination, avoiding, from their "apparent anonymity", the suspicions that a particular leadership from clearly identifiable political and social actors may have given their actions relevance at a social level (Howard et al., 2017; Ong et al., 2019). A strategy called Astroturfing has been used in politics, public relations and advertising for decades (Sorensen et al., 2017; Keller et al., 2019) and, since the current digital boom, has gained increasing use and interest for the dissemination of misleading content (Elmas et al., 2021). This, from the use of marketing techniques such as the so-called Thunderclap, where people, considered normal and identifiable by the rest of the population as their peers, introduce a certain message within their respective fields (Sorensen et al., 2017). Using the Thunderclap technique, lead users (alphas) do not follow the client account (influencers) to avoid detection. In addition, the alphas have a team of users (not necessarily bots, called beta) who act in an itinerant way and who have several accounts on social networks (in our case Twitter) and are responsible for 1) answering the different interactions generated by the disseminated

messages and 2) passing the misleading content transmitted by the leading users (alpha) to influencers (e.g., journalists and media) with the purpose of generating trends (trending topics). These tasks serve to monitor the effectiveness rate of the campaign carried out by both profiles (alphas and betas), as pointed out by authors such as Pérez (2020).

To study Astroturfing, it is not enough to focus on finding bots since not all accounts are bots. Keller et al. (2019) recommend an identification strategy based on coordination patterns, arguing that "similar behaviour among a group of human-managed accounts is a stronger signal of a disinformation campaign than individual 'bot-like' behaviour" (p. 2). The current scientific challenge is to identify patterns in campaigns and attacks rather than in the behaviour of individual actors. This requires longitudinal observations as well as data from multiple social media and online platforms for the analysis of the spatial dimension (Grimme et al., 2018).

The study of Astroturfing takes three approaches (Table 1), which help to understand it and adapt its application from social media: 1) from the importance of internet posts to create and cocreate shared value among users (Sorensen et al., 2017); 2) from common characteristics that authors such as Keller et al. (2019) have detected in different Astroturfing campaigns in different parts of the world; and 3) from more "ephemeral" typologies of online Astroturfing (Elmas et al., 2021) based on Trending Topics (TT) and on the capacity of this type of strategy in conditioning public opinion.

Table 1. Theoretical conceptualisations of Astroturfing

Authors	Theoretical conceptualisations	Characteristic features of the theory
Sorensen et al. (2017)	Cocreation of value through social media postings	<p>Social media posts offer opportunities for the cocreation of the associated value.</p> <p>By conveying information to other users, it shares, learns and generates consumer engagement through a click (a retweet) rather than a textual narrative from a supporter.</p>
Keller et al. (2019)	<i>Political Astroturfing as an online coordination strategy</i>	<p>In social media, political Astroturfing takes the form of a coordinated and organised campaign in which participants pose as "ordinary users" who post opinions favourable to the instigators of such a campaign, distracting others with negative news, drawing attention to divisive political issues or attack opponents and/or critics. This process is known as cotweeting, i.e., the posting of the same message by accounts within a very short window of time (e.g., a few minutes). It is this type of coordination that best distinguishes Astroturfing, as ordinary accounts rarely post a similar original message at the same time.</p>
Elmas et al. (2021)	<i>Ephemeral Astroturfing and Ephemeral Trending Topics (TTs) on Twitter</i>	<p>A keyword or a chosen topic is promoted artificially through coordinated and inauthentic activities. This type of action is very frequent in Twitter Trending Topics.</p> <p>Ephemeral Astroturfing outperforms other approaches in three key aspects: (i) it allows the use of active, compromised accounts as sources of fake interactions; (ii) it evades detection by users, the platform and academic studies; and (iii) it prevents users from reporting malicious activity, as traditional spam classifiers are unable to prevent future attacks.</p>

Source: Own elaboration

This phenomenon would be linked to Granovetter's (1973) sociological theory of the strength of weak links, where contacts far from the central group have great influence by contributing new ideas about the group. In this way, messages from users far from the centrality of the network may be the most suitable for introducing new messages and achieving that they reach further afield (Ribera, 2014).

2. Materials and methods

This paper aims to determine the use of political online Astroturfing strategies in Spain through the study of messages linked to topics or tags that have been trending on Twitter during the COVID-19 pandemic period in the spring of 2020. The objective is to:

- Establish how the accounts or users involved in the messages analysed behave.
- Identify the features that characterise the accounts that disseminate misleading content.
- Determine the temporal evolution of dissemination.
- Estimate the probability of the presence of bots in viralisation.

The main hypothesis that is expected to be confirmed or not in this work is the following:

H1: T Astroturfing strategy is being applied around the Spanish political debate on Twitter for the dissemination of misleading content in the terms exposed by authors such as Sorensen et al. (2017), Keller et al. (2019) and Elmas et al. (2021).

The analysis will focus on Astroturfing detection in terms of behaviour, structure and content (Chen et al., 2021). The analysed data are collected through the Twitter API and programming in R, using the RTweet library (Kearney, 2019) on different campaigns. The collection process was carried out in a geolocalised manner, in parallel to the global capture, using the coordinates of longitude, latitude and radius of influence of the area of interest. Although this technique has worldwide reliability of 77.84% and 88.15% in Europe (Van-der-Veen et al., 2015), it allows for the detection of possible geographical interferences. The countries chosen for the search are those identified by Bradshaw et al. (2021) as places where disinformation is broadcast, as well as several provincial capitals in Spain.

The work employs a quantitative approach through statistical analysis, networks and machine learning techniques, which help to analyse the data generated around the tags and messages linked to them on Twitter from the publication of the State of Alarm Act in Spain in March 2020 until the end of May.

The selection of tags taken into account as a case study was made, taking into consideration 1) the selection of misleading content collected in Maldito Bulo and/or Newtral based on political topics and 2) tags that became trends through messages with a clear partisan political profile.

The tags collected for the analysis were as follows: "Pedro Duque Denia", "Pedro Sanchez Huete", "respiradores Granada Madrid", "Muerte al Rey" (Death to the King), "#SilenciaelPaís" (Silence El País newspaper), "#Cacerolada21 h" (pot-and-pan protest), "#Pedroelsepulturero" (Pedro the gravedigger) and "#Elgobiernotemiente" (The government lies to you). The analysis was based on the collection made during the first

moments of dissemination of the linked messages. Thus, it is possible to identify the pattern of the disseminated and related messages, as well as the application or absence of Astroturfing strategies in them, from an apparent spontaneity of opinions and content by nonprominent profiles.

Of all the preselected tags, the analysis focused on three: "Pedro Duque Denia", "#SilenciaelPaís" and "Muerte al Rey", because they imply different situations of interest (unlike the rest of the tags), around the general objective set: 1) an unsuccessful campaign without a subsequent flood of messages in a very specific time, after being quickly identified by other users as a hoax (disproved in maldita.es: <https://bit.ly/3Fr1pJr>); 2) with a flood, after a campaign has achieved a certain success; and 3) with a double flood in time, due to the success of the campaign's scope.

The final sample analysed consisted of a total of 32,527 messages, to which the following were applied.

- Network analysis, through the Gephi programme (version 0.9.2), is used to determine the existing connections between users through their representation through points, and the size and thickness of lines. The Open Ord algorithm was used to separate the different groups (Martin et al., 2011) and achieve better visualisation, in addition to modularity analysis (Blondel et al., 2008) and closeness centrality, which measures the position of each node concerning others and gives an idea of who is connected or acting individually. Thus, a weak connection, close to zero, would account for peripheral nodes and a value close to one for being very close to others (Hansen et al., 2020).
- Analysis of bot behaviour was performed through the R package tweetbotornot by Kearney (2018), which, according to its author, presents a 93.53% hit rate for bot classification and 95.32% for nonbots (average 93.8% hit rate). This algorithm, although it has discrepancies with other methods, is considered among the best and most widely used in the social sciences (Martini et al., 2021).
- Closeness centrality analysis between nodes in a network (Twitter accounts) helps to identify influencers in a community (Pozzi et al., 2017).
- Statistical analysis, using R software, was used to observe the behaviour between variables and time.

3. Results

3.1. Campaign: Pedro Duque in Denia

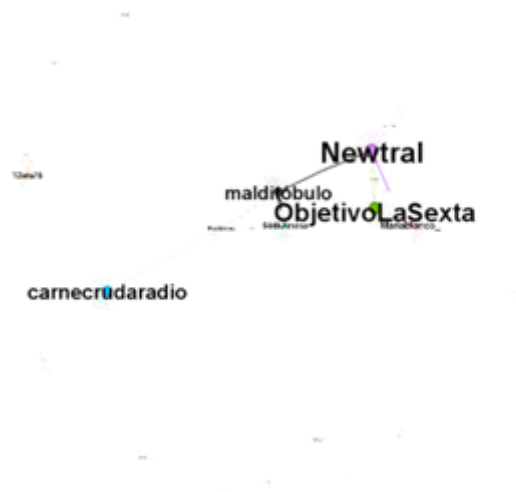
The study of the topic associated with the admission of the Spanish Minister of Science (Pedro Duque) to a hospital in Denia (Valencia) offers a total of 370 messages. Only messages containing the words "Pedro Duque" or "minister" and the town of "Denia"

were considered for the study. The first two tweets analysed (IDs 1241509359722598406 and 1241644721975549952) come from 21 and 22 March 2020 at 23:37 and 08:35 (GMT), respectively, mentioning that the Minister has been admitted. From 10:48 on 22/03/2020, tweets began to be sent during the distribution period, all with the same message (although with different texts, such as: "Can anyone verify this news? What is Buzz Light Year doing in Denia? Is not the movement of people prohibited? @astro_duque, where are you? Pedro Duque is admitted to the Hospital de Denia") and based on information published that day on the website espanaesvoz.es. During the viralisation of these messages, the period of amplification begins, with requests for corroboration of the fact made to other websites (e.g., "@okdiario Pedro Duque (the astronaut minister) admitted to the hospital in Denia... he had gone to Jávea for the weekend!!!! A lot of irresponsible imbeciles. Is this news true? Investigate") but also others that link this topic with other related tags (e.g., #PedroElSepulturero).

The messages continue until the Portal Newtral (dedicated to fact-checking) publishes the denial at 22:24, and the Minister himself appears on a Spanish free-to-air television channel (La Sexta). From that moment on, the distribution of the misleading content almost disappeared, and the portal that disseminated this fact issued a statement acknowledging the falsehood of this event while eliminating all references to it on its website.

The network analysis, Figure 1, shows 330 user nodes and the existence of accounts that are considered influential (according to the size of the node and the size of the account name), whose messages are followed by many others. The figure only shows accounts that disseminated disinformation content from the media and journalists, redisseminated by their followers, but there are no disinformation content disseminators in the graph shown. The cluster analysis detects 62 user communities, with a certain number of users who acted by disseminating disinformation at the beginning but who are not circumscribed to any important group, unlike the influencers who appear later in the disinformation, such as @Newtral, @ObjetivoLaSexta, @malditobulo and @carnecrudaradio. Therefore, there are a significant number of messages that do not appear to be important, nor do they stand out in the network analysis.

Figure 1. Network analysis of Minister Pedro Duque's disinformation in Denia

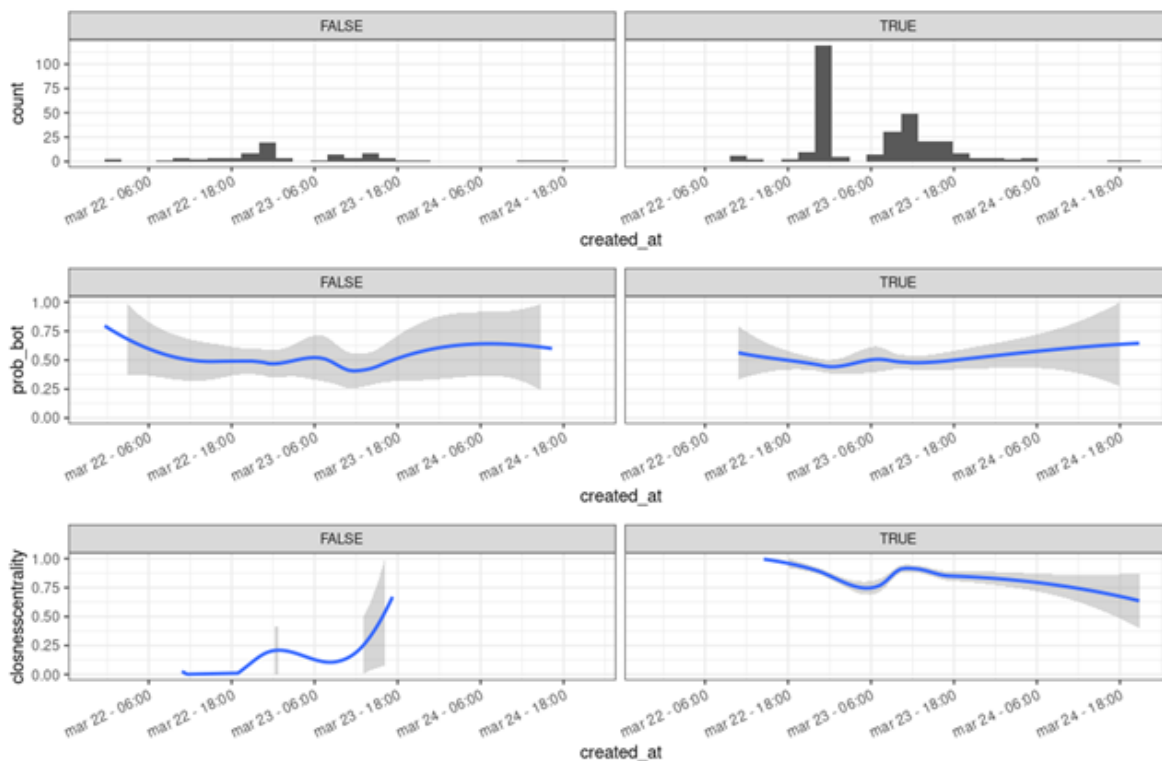


Source: Own elaboration.

The study of the data shown in Figure 2, which represents the probability of being a bot (Kearney, 2018), shows the presence of 330 accounts, which disseminated a total of 370 tweets (most of which were not retweeted). The data shown allow us to observe that all the accounts that disseminate misleading content in the first few minutes present an average closeness value around zero. Therefore, there is no relationship with other users, and they are distanced from any network. The accounts that disseminate, despite having different probabilities of being a bot, have an average 50% probability of being a bot, with a value at the beginning of more than 75% in the first messages.

Figure 2. Number of tweets, average bot probability and average closeness centrality in uninformative content about Pedro Duque in Denia

Note: Direct tweets are marked as "False", and retweets are marked as "True".



Source: Own elaboration.

Table 2 shows accounts created more than five years ago on Twitter (=2013 and Me=2012) but with a high level of activity, as the average and the median number of tweets issued exceeds several tens of thousands (Me=8 tweets and established almost 4 favourites per account daily). The network of followers, friends and primary contacts of this type of user detected are not very extensive, but neither are they scarce, standing at Me= 557 followers and following 770 accounts. The contact capacity of this type of identified user, together with a closeness centrality value of almost zero, suggests that messages are not being sent to the same set of contacts but rather to different groups of users who contact the senders of these misleading contents. This could correspond to the cotweeting technique of Keller et al. (2019) cited above, where unconnected accounts spread the same topic (distribution) in a few minutes, or messages that mention certain media or important users to wait for their dissemination, such as Luis del Pino (@ldpsincomplejos), OkDiario (@okdiario) or Caso Aislado (@CasoAislado_Es) in the process of amplification. As the disinformation was quickly cut off, no subsequent flooding process was observed; instead, there was a process of countering the disinformation.

The analysis of the profile of the disseminating accounts shows that they use descriptions aimed at being identified by other Twitter members as "equals" or users with an "average profile": housewives, warehouse workers, journalists, biologists, lawyers, and freelancers, among others. It can also be seen that, in the descriptions used, reference is made to ideological values (e.g., pro-family, anti-feminist, anti-communist, heterosexual, patriotic and liberal) and national values (being Spanish), which serve to establish a connection with other like-minded users.

Table 2. Data on the accounts that disseminated misleading content about Pedro Duque in Denia

	Followers	Following	Tweets written	Favourites	Date created
Minimum	1	0	43	20	13/12/2007
First quartile	243	370,2	8.152	1.903	21/05/2011
Median	557	770	23.572	11.325	11/08/2012
Media	1.470	1.439,6	49.478	27.591	06/08/2013
Third quartile	2.027	1.968,8	50.236	32.868	10/04/2015
Maximum	15.773	10.975	511.989	342.166	25/02/2020

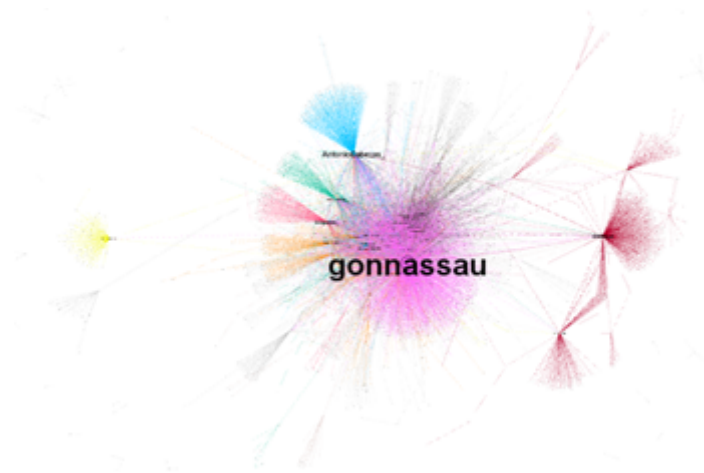
Own elaboration from 162 tweets spreading disinformation.

3.2. Campaign: #SilenciaElPaís

The hashtag #SilenciaElPaís was launched on 16 April 2020 at 14:20 hours (GMT). From the creation of the first tweet, the dissemination of messages, both direct and retweets (RT), on the subject began. A process of dissemination in which the opposition of a group of accounts with messages related to a left-wing ideological orientation that did not manage to stop the initial dissemination can be seen. This was done through the leadership of the user @03690jul, who acted as an opposing influencer within this tag with 10.72% of the traffic generated. Another user (@mgonzalezelpais) took on the role of influencer and opponent of the misleading content transmitted around the hashtag #SilenciaElPaís, who generated 2.56% of the traffic and whose profile identified him as "Foreign Policy and Defence Journalist for EL PAÍS". I report on Vox, Vox does not want me to report on him, what should I do?

Around the analysed tag, 15,977 tweets were generated, of which 2,771 were direct and 13,209 were RT. Figure 3 shows the influencers in each of the identified nodes, in which the account @gonnasau stands out, with 36.03% of the RT traffic, followed by others such as @AntonioCabezas_, @ElPais_Today, or @ElAguijon_, among others, which become new influencers over time on the main group. In the analysis of the linked messages, there are almost no connections between groups in favour of the content disseminated against their opponents.

Figure 3. Network analysis of the #SilenciaElPaís campaign

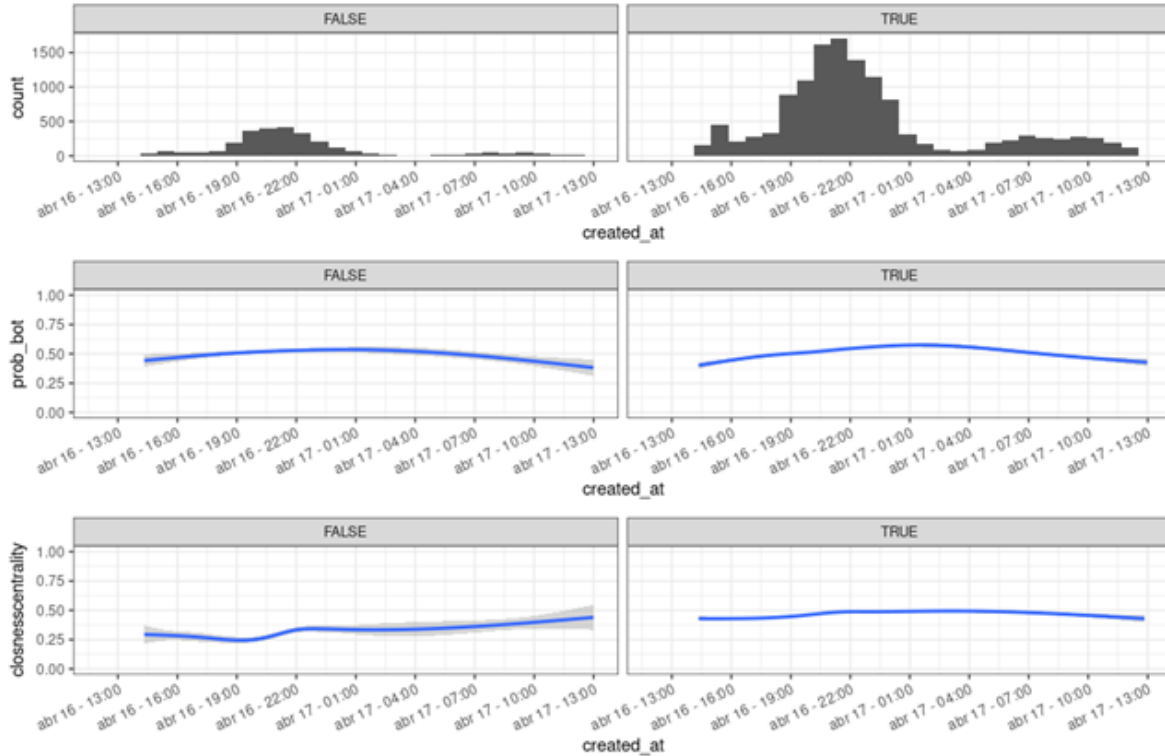


Source: Own elaboration

Figure 4 shows how from the beginning of the dissemination of messages with the hashtag #SilenciaElPaís, several tweets with a low centrality of proximity are produced in the first moments, which emit the same message (although not literally) without connection to other accounts (e.g., distribution: "Let us start this evening at 20:00 h. Please circulate. #SilenciaElPaís <https://t.co/ctGYI4VQ06>, amplification e.g., "@Santi_ABASCAL @javiernegre10 #silenciaelpais <https://t.co/G5hfj2bLnE>, with calls to the leader of Vox and the founder of EDATV), and where the average probability of being a bot is 50%. No major changes in closeness centrality values can be seen in the accounts that issue direct messages, with average values below 0.5 but rising over time, while those that are retweeted have somewhat higher closeness centrality values. In this case, the first phase of amplification and distribution from numerous independent accounts can be seen, arriving at approximately 8 pm on 16 April 2020 (18:00 GMT) in the flood phase, in which there is a significant increase in tweets, with a slight increase in the centrality of closeness and an increase in the average bot probability above 50%. On the evening of 16 April 2020, the messages linked to this hashtag disappeared, only to return in the early morning of 17 April 2020, with the same pattern of operation, but with much less flooding.

Figure 4. Number of tweets, average bot probability and average closeness centrality in the #SilenciaElPaís campaign

Note: Direct tweets are marked as "False", and retweets are marked as "True".



Source: Own elaboration

As was observed in the study of the misleading content about Pedro Duque (Table 2), we can find users who follow a greater number of users (followings) than followers (followers), in addition to highlighting the intensive activity they have on Twitter, writing several thousand tweets, with one account reaching more than a million messages (Table 3).

Table 3. Accounts disseminating the #SilenciaElPaís campaign

	Followers	Following	Tweets written	Favourites	Date created
Minimum	0	0	1	0	07/03/2007
First quartile	77	179	1.847	1.349	21/11/2011
Median	280	494	8.142	6.469	09/09/2014
Media	1.425	1.152	26.648	21.886	11/02/2015
Third quartile	967	1.306	26.931	22.911	12/09/2015
Maximum	1.315.139	98.157	1.149.572	866.264	17/04/2020

Source: Own elaboration.

An analysis of the dates of creation of the accounts shows that 2014 and 2015 were the periods where the average number of these profiles were opened on Twitter but also that most of them are concentrated in very specific periods and are not spread over time (2011, autumn 2017 and the beginning of 2020).

3.3. Campaign: "Death to the King"

The viralisation of content, resulting from a video recorded in Alcorcón (Madrid) on 19 May 2020, during altercations between different demonstrators, where a person with his face covered and holding a Republican flag can be heard saying "Death to the King and his daughters", begins on the same day at 19:19 (GMT). In the messages linked to this event on Twitter, leftist groups are accused of wanting to kill the head of state and his daughters, a fact echoed by several media outlets (González, 2020) (e.g., the distribution "Death to the king and his daughters" these democrats say. I do not think they are more than 18 years old. I apologise for these images; the left has been working since ZP to return to the Spain of the II Republic. These images are very similar. I do not like it. <https://t.co/VSi8r0iUzh>, amplification: "@Malumarquez4 @abc_es @policia @guardiacivil He said "death to the king and his daughters").

The study of networks from 19 May until midday on 20 May offers a total of 16,180 tweets, with 1,970 direct tweets and 14,480 RT, as shown in Figure 5. The analysis shows that the related messages are widely dispersed among a multitude of small users who play the role of influencers in the viralisation of this issue on Twitter, without any opposition to their messages.

Figure 5. Network analysis of the "Death to the King" campaign



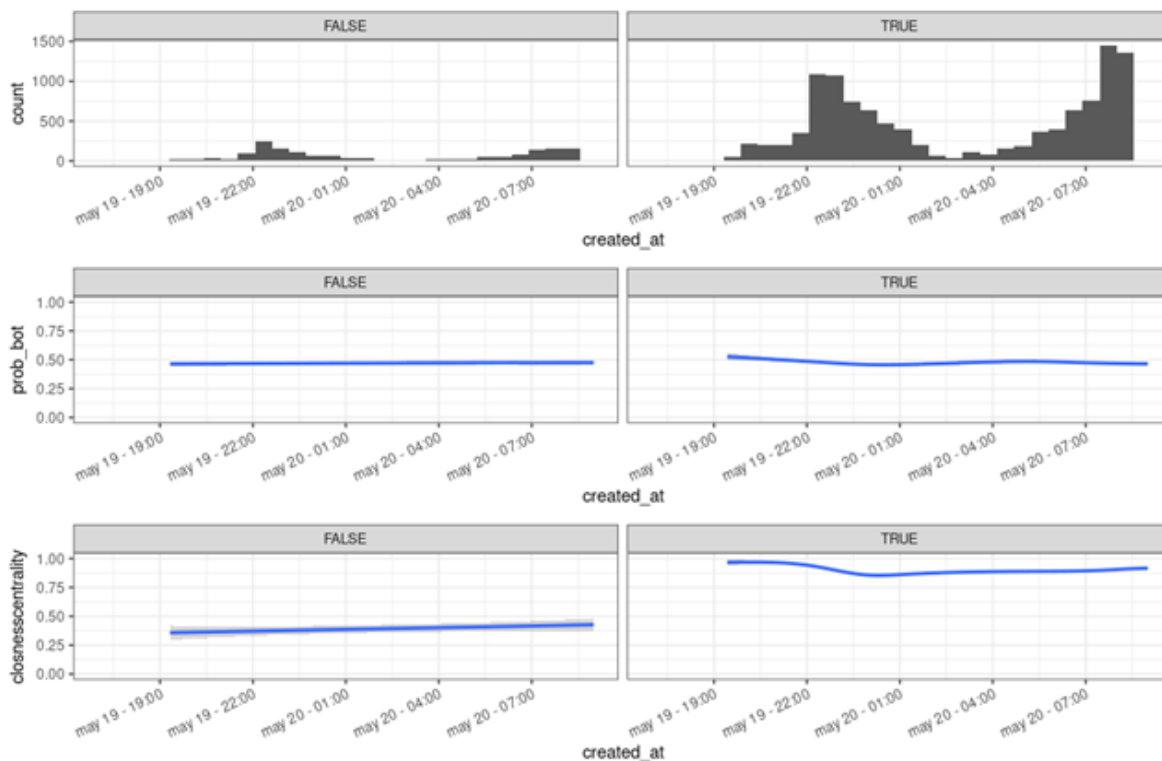
Source: Own elaboration.

Figure 6 shows how, during the period in which content related to the topic went viral, different accounts directly published messages following a clear pattern of amplification and distribution of disseminated content until the arrival of the flooding phase, which occurs approximately two hours after the first messages were published on Twitter. The user accounts that initiate the viralisation (discussion) show small levels of average closeness centrality. The average bot probability is nearly always 50%, as well as for accounts that retweet. The centrality and proximity of the retweeting accounts are very high, close to unity, which shows a very important connection between them, demonstrating a large partisan following, which would link to the coretweeting phase of Keller et al. (2019).

Unlike the rest of the cases, a feature that distinguishes the content dissemination strategy around the topic of "Death to the King" is the presence of a second flood at the beginning of the following morning (20 May 2020), in addition to a slight increase in the presence of bots during this process, although the tweets with the highest number of favourite marks come from accounts (users) with little likelihood of being favourites.

Figure 6. Number of tweets, average bot probability and average closeness centrality in the "Death to the King" campaign

Note: Direct tweets are marked as "False", and retweets are marked as "True".



Source: Own elaboration.

Table 4 allows us to observe what has already been indicated in the cases previously addressed in this work (e.g., the case of Pedro Duque or the hashtag #SilenciaElPaís): influencer users with more followers than followers (follower Me of 367 and 229 followers), with a Me of 8,282 tweets written, and 7,242 marked as favourites. In this case, the most striking feature is the presence of an account that manages to write almost one and a half million tweets and accounts created on the same day of the campaign (19 and 20 May 2020).

Table 4. Data on the accounts disseminating the campaign “Death to the King”

	Followers	Following	Tweets written	Favourites	Date created
Minimum	0	0	1	0	25/03/2007
First quartile	74,25	151	2156	1.844	03/02/2012
Median	229	367	8282	7.242	28/10/2014
Media	1.025,17	909,1	24.788	21.531	22/03/2015
Third quartile	648,75	908,8	25.874	23.024	12/08/2018
Maximum	269.860	50.850	1.459.308	563.868	20/05/2020

Own elaboration.

The data collected show a nonuniform distribution on the date of creation of the authors of these contents, where three moments in which these users were created on Twitter can be observed (2011, autumn 2017 and early 2020).

If we look at the location of the 11,698 self-proclaimed participating users around the topic, we can see accounts from all the autonomous communities and many Spanish cities in this campaign, highlighting Spain (1,064), Madrid (959), Barcelona (197), Seville (173), Valencia (152), Malaga (72), and Zaragoza (49). A total of 7,179 accounts do not state where they are located in their profile, and there are hardly any accounts from outside Spain. Therefore, the origin of all the messages, and the debate generated around the topic analysed (“Death to the King”), would be national, with special incidence in Madrid or Spain in general.

By capturing the geolocation coordinates at a global level on the Twitter API, different results were found: Spain (10,521), the Philippines (2,688), the United States (292) and Venezuela (73). The rest of the tweets could not be traced, or their exact origin could be found through geolocation. Users from Asturias (northern Spain), the United States and Venezuela have different accounts sending the same message in a short period within the flooding period. In this way, certain messages are observed to be very widespread from each of these places.

In the case of the messages sent from the Philippines and the United States, some groups that sent alternating messages every few seconds in confrontation can be detected, where there are accounts of profiles ideologically identified as left-wing, which manage to be followed by other politically related accounts, in an attempt to ensure that the impact of the disseminated content reaches more people and more areas. In both cases, there is a very low probability of a bot presence during the amplification and distribution phase, although during the flooding phase, the probability of this type of user in the debate increases.

4. Conclusions

The data analysed allow us to confirm the main hypothesis of this study by observing a strategy on Twitter around the cases studied, based on the dissemination of misleading content that went viral through the Astroturfing strategy (Elmas, 2019) and led by users whose profiles on social networks are not associated with those considered up to now as influencers (characterised by having a large number of followers and a prominent or differentiating social profile). This makes it increasingly difficult to distinguish the particular political or interest group promoting these types of debates analysed, except for the fact that they have in common: 1) a critical attitude, from different perspectives, toward the national government; 2) the presence of very low-profile alpha users who lead the debate in an initial phase of amplification and distribution far from the centre of the network (taking advantage of Granovetter's (1973) theory of weak link strength), and beta users, who act in the terms set out by Sorensen et al. (2017), Keller et al. (2019) and Elmas et al. (2021), i.e., replying to and disseminating the misleading content disseminated by alpha users; 3) with a relatively low likelihood of the presence of bots during the early stages of viralisation of these topics on Twitter supported by others who clearly are; and 4) support from created web pages or content taken out of context to give 'credibility' to the content viralised by them in the initial amplification phase (Guess et al., 2018; Zhao et al., 2020).

The content dissemination scenario analysed would confirm the presence of "soldiers", in the terms described by Zheng et al. (2017), who use bots to reinforce messages (early stages of viralisation) or flood the digital scenario where misleading content is disseminated (later stages of viralisation). Users who not only assume profiles ideologically aligned with the misleading content disseminated to condition public opinion through social networks (in this case, Twitter) but also take on opposing positions (in a false flag operation) to attract the attention of a greater number of users who are not necessarily ideologically aligned. All of this is in line with Pérez-Curiel & Limón (2019), and the use of anti-system messages and the recurrent self-positioning as outsiders, when spreading messages where the use of the national sense (Spanish) and qualifiers against being left-wing, right-wing, fascist, Marxist or communist. In the Spanish case, few studies focus on this strategy, which is very important to be able to move forward with new analyses that allow us to understand it more deeply, given the rise of ideologically extremist or populist groups and the permanent perception of conflict and populist communication promoted by social networks. These are aspects highlighted by authors such as Mazzoleni & Bracciale (2018), where the appeal of emotions and beliefs reduces the influence capacity of objectifiable facts.

The data shown in this paper also help to confirm the findings of authors such as Howard et al. (2017), i.e., the implementation of political online Astroturfing. This is carried out in three clearly demarcated stages: distribution, amplification and flooding of misleading content on social networks. All of this is done with the use of nano-influencers, who participate with different levels of intensity or levels, depending on whether the misleading content is verified or not by traditional media and professional journalistic projects oriented to such purposes, for example. This favours a framework that is more difficult to detect,

with strategies aimed at breaking the bell and filtering the bubble effect of social networks, such as Twitter, to position issues at the level of public opinion (Flaxman et al., 2016).

Authors' contribution

Sergio Arce-Garcia: Conceptualization, Data curation, Methodology, Software, Writing-original draft, Formal analysis, Investigation, and Writing-review and editing. **Elias Said-Hung:** Conceptualisation, Writing-original draft, Visualisation, Formal analysis, and Writing-review and editing. **Daria Mottareale-Calvanese:** Writing-original draft, Visualisation, Conceptualisation, Formal analysis, and Writing-review and editing. All authors have read and accepted the published version of the manuscript. Conflict of interest: The authors declare no conflict of interest.

References

- Allem, Jon-Patrick; & Ferrara, Emilio (2018). Could social *bots* pose a threat to public health? *American journal of public health*, 108(8), 1005-1006. <https://doi.org/10.2105/AJPH.2018.304512>
- Blondel, Vincent; Guillaume, Jean-Lup; Lambiotte, Renaud; & Lefebvre, Etienne (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10). <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Boididou, Christina; Middleton, Stuarthe-E.; Jin, Zhiwei; Papadopoulos, Symeon; Dang-Nguyen, Duc-Tien; Boato, Giulia; & Kompatsiaris, Yiannis (2018). Verifying information with multimedia content on Twitter. *Multimedia tools and applications*, 77(12), 15545-15571. <https://doi.org/10.1007/s11042-017-5132-9>
- Bradshaw, Samantha; Bailey, Hanah; & Howard, Philip-N. (2021). *Industrialized disinformation. 2020 Global inventory of organized social media manipulation*. Working Paper 2021.1. Project on Computational Propaganda. <https://cutt.ly/VOgTtjO>
- Chen, Tong; Liu, Jiqiang; Wu, Yalun; Tian, Yunzhe; Tong, Endong; Niu, Wenjia, Li, Yike, Xiang, Yingxiao; & Wang, Wei (2021). Survey on Astroturfing Detection and Analysis from an Information Technology Perspective. *Security and Communication Networks*, 2021, 3294610. <https://doi.org/10.1155/2021/3294610>
- Elmas, Tugrulcan (2019). *Lateral Astroturfing Attacks on Twitter Trending Topics*. AMLD EPFL. Lausanne. <https://cutt.ly/4yGaj5L>
- Elmas, Tugrulcan; Overdorf, Rebekah; Özkalay, Ahmed-Furkan; & Aberer, Karl (2021). *Ephemeral Astroturfing Attacks: The Case of Fake Twitter Trends*. arXiv preprint arXiv:1910.07783 <https://arxiv.org/abs/1910.07783>.
- Estrada-Cuzcano, Alonso; Alfaro-Mendives, Karen; & Saavedra-Vásquez, Valeria (2020). Desinformación y desinformación, Posverdad y noticias falsas: precisiones conceptuales, diferencias, similitudes y yuxtaposiciones. *Información, cultura y sociedad*, (42), 93-106. <https://doi.org/10.34096/ics.i42.7427>
- Ferrara, Emilio; Varol, Onur; Davis, Clayton; Menczer, Filippo; & Flammini Alessandro (2016). The rise of social *bots*. *Communications of the ACM* 59(7), 96–104. <https://doi.org/10.1145/2818717>

- Flaxman, Seth; Goel, Sharad; & Rao, Justin-M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, 80, 298–320. <https://doi.org/10.1093/poq/nfw006>
- Gallwitz, Florian; & Kreil, Michael (2021). *The Rise and Fall of "Social Bot"*. Research (March 28, 2021). <https://ssrn.com/abstract=3814191>
- González, Fernán (2020, 20 de mayo). Manifestantes de extrema izquierda gritan "¡Muerte al Rey y a sus hijas!". *Ok Diario*. <https://cutt.ly/CyVjTUR>.
- Granovetter, Mark (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360-1380.
- Grimme, Christian; Assenmacher, Dennis; & Adam, Lena (2018). Changing Perspectives: Is It Sufficient to Detect Social Bots?. In Meiselwitz G. (eds.) *Social Computing and Social Media. User Experience and Behavior* (pp. 445-461). Lecture Notes in Computer Science, vol. 10913. Springer, Cham. https://doi.org/10.1007/978-3-319-91521-0_32
- Guess, Andrew; Nyhan, Brendan; & Reifler, Jason (2018). Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. Presidential campaign. *European Research Council*. <https://cutt.ly/FOgUe1R>.
- Hansen, Derek-L.; Shneiderman, Ben, Smith, Marc-A.; & Himerlboim, Itai (2020). *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Elsevier. <https://doi.org/10.1016/C2018-0-01348-1>
- Howard, Philip-N.; Bolsover, Gillian; Kollanyi, Bence; Bradshaw, Samantha; & Neudert, Lisa-Maria (2017). *Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter?* Computational Propaganda Project-Oxford Internet Institute, Data Memo, 1. <https://cutt.ly/kRihRoY>
- Kearney, Michael-W. (2018). Tweetbotornot: An R package for classifying Twitter accounts as bot or not. <https://github.com/mkearney/tweetbotornot>
- Kearney, Michael-W. (2019). Rtweet: Collecting y analyzing Twitter data. *Journal of Open Source Software*, 4(42), 1829. <https://doi.org/10.21105/joss.01829>
- Keller, Franziska-B.; Schoch, David; Stier, Sebastian; & Yang, Jung-Hwan (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>
- Kucharski, Adam (2016). Study epidemiology of fake news. *Nature*, 540(525). <https://doi.org/10.1038/540525a>
- Luceri, Luca; Deb, Ashok; Badawy, Adam; & Ferrara, Emilio (2019). *Red bots do it better: Comparative analysis of social bot partisan behavior*. In Companion Proceedings of the 2019 World Wide Web Conference, 1007-1012. <https://arxiv.org/abs/1902.02765>
- Martin, Shawn; Brown, W.-Michael; Klavans, Richard; & Boyack, Kevin-W. (2011). *OpenOrd: An Open-Source Toolbox for Large Graph Layout*. In Proc. SPIE, Visualization and Data Analysis 2011. San Francisco, Estados Unidos. <https://doi.org/10.1117/12.871402>
- Martini, Franziska; Samula, Paul; Keller, Tobias-R., & Klinger, Ulrike (2021). Bot, or not? Comparing three methods for detecting social bots in five political discourses. *Big Data & Society*, 8(2), 1-13. <https://doi.org/10.1177/20539517211033566>
- Mazzoleni, Gianpietro; & Bracciale, Roberta (2018). Socially mediated populism: the communicative strategies of political leaders on Facebook. *Palgrave Communications*, 4(50). <https://doi.org/10.1057/s41599-018-0104-x>
- Magallón, Raúl (2019). *Unfaking News. Cómo combatir la desinformación*. Pirámide.

- Ong, Jonathan-Corpus; Tapsell, Ross; & Curato, Nicole (2019) Tracking Digital Disinformation in the 2019 Philippine Midterm Election. *New Mandala*. <https://cutt.ly/6RhPHt4>
- Pérez-Curiel, Concha; & Limón, Pilar (2019). Political *influencers*. A study of Donald Trump's personal brand on Twitter and its impact on the media and users. *Comunicación y Sociedad*, 32(1), 57-75. <https://doi.org/10.15581/003.32.1.57-75>
- Pérez, Jordi (2020, 21 de may). "Yo fui un bot": las confesiones de un agente dedicado al engaño en Twitter. *El País*. <https://cutt.ly/wRihXGu>
- Pozzi, Federico-Alberto; Fersini, Elisabetta; Messina, Enza; & Liu, Bing (2017). *The aim of Sentiment Analysis*. Elsevier. <https://doi.org/10.1016/C2015-0-01864-0>
- Ribera, Carles-Salom (2014). Estrategia en redes sociales basada en la teoría de los vínculos débiles. *Más poder local*, 19, 23-25. <https://dialnet.unirioja.es/download/articulo/4753468.pdf>
- Said-Hung, Elías; Merino-Arribas, Adoración; & Martínez, Javier (2021). Evolución del debate académico en la Web of Science y Scopus sobre unfaking news (2014-2019). *Estudios sobre el Mensaje Periodístico*, 27(3), 961-971. <https://doi.org/10.5209/esmp.71031>
- Salaverría, Ramón; Buslón, Nataly; López-Pan, Fernando; León, Bienvenido; López-Goñi, Ignacio; & Erviti, María-Carmen (2020). Desinformación en tiempos de pandemia: tipología de los bulos sobre la covid-19. *El profesional de la información*, 29(3). <https://doi.org/10.3145/epi.2020.may.15>
- Sorensen, Anne; Andrews, Lynda; & Drennan, Judy (2017). Using social media posts as resources for engaging in value co-creation: The case for social media-based cause brand communities. *Journal of Service Theory and Practice*, 27(4), 898-922. <https://doi.org/10.1108/JSTP-04-2016-0080>
- Tandoc, Edson-C.; Lim, Zheng-Wei; & Ling, Richard (2018). Defining "fake news" A typology of scholarly definitions. *Digital journalism*, 6(2), 137-153. <https://doi.org/10.1080/21670811.2017.1360143>
- Van-der-Linden, Sander; Maibach, Edward; Cook, John; Leiserowitz, Anthony; & Lewandowsky, Stephan (2017). Inoculating Against Misinformation. *Science*, 358(6367), 1141-1142. <https://doi.org/10.1126/science.1257533>
- Van-der-Veen, Han; Hiemstra, Djoerd; Van-den-Broek, Tijs; Ehrenhard, Michel; & Need, Ariana (2015). Determine the User Country of a Tweet. *Social and Information Networks*. <https://arxiv.org/abs/1508.02483>
- Zerback, Thomas; & Töpfl, Florian (2021). Forged Examples as Disinformation: The Biasing Effects of Political Astroturfing Comments on Public Opinion Perceptions and How to Prevent Them. *Political Psychology*, 43(3), 399-418. <https://doi.org/10.1111/pops.12767>
- Zhao, Zilong; Zhao, Jichang; Sano, Yukie; Levy, Orr; Takayasu, Hideki; Takayasu, Misako; Li, Daqing; Wu, Junjie; & Havlin, Shlomo (2020). Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*, 9(7). <https://doi.org/10.1140/epjds/s13688-020-00224-z>
- Zheng, Haizhong; Xue, Minhui; Hao, Lu; Hao, Shuang; Zhu, Haojin; Liang, Xiaohui; & Ross, Keith (2017). Smoke Screener or Straight Shooter: Detecting Elite Sybil Attack. *Social and Information Networks*. <https://arxiv.org/abs/1709.06916>