Deepfakes on Twitter: Which Actors Control Their Spread?

Pérez Dasilva, Jesús; Meso Ayerdi, Koldobika; Mendiguren Galdospin, Terese

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Pérez Dasilva, J., Meso Ayerdi, K., & Mendiguren Galdospin, T. (2021). Deepfakes on Twitter: Which Actors Control Their Spread? *Media and Communication*, *9*(1), 301-312. https://doi.org/10.17645/mac.v9i1.3433

COGITATIO

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 1, Pages 301–312
DOI: 10.17645/mac.v9i1.3433

Article

Deepfakes on Twitter: Which Actors Control Their Spread?

Jesús Pérez Dasilva *, Koldobika Meso Ayerdi and Terese Mendiguren Galdospin

Department of Journalism II, University of the Basque Country, 48940 Leioa, Spain;
E-Mails: jesusangel.perez@ehu.eus (J.P.D.), koldo.meso@ehu.eus (K.M.A.), terese.mendiguren@ehu.eus (T.M.G.)

* Corresponding author

Submitted: 4 July 2020 | Accepted: 24 August 2020 | Published: 3 March 2021

Abstract
The term deepfake was first used in a Reddit post in 2017 to refer to videos manipulated using artificial intelligence techniques and since then it is becoming easier to create such fake videos. A recent investigation by the cybersecurity company Deeptrace in September 2019 indicated that the number of what is known as fake videos had doubled in the last nine months and that most were pornographic videos used as revenge to harm many women. The report also highlighted the potential of this technology to be used in political campaigns such as in Gabon and Malaysia. In this sense, the phenomenon of deepfake has become a concern for governments because it poses a short-term threat not only to politics, but also for fraud or cyberbullying. The starting point of this research was Twitter's announcement of a change in its protocols to fight fake news and deepfakes. We have used the Social Network Analysis technique, with visualization as a key component, to analyze the conversation on Twitter about the deepfake phenomenon. NodeXL was used to identify main actors and the network of connections between all these accounts. In addition, the semantic networks of the tweets were analyzed to discover hidden patterns of meaning. The results show that half of the actors who function as bridges in the interactions that shape the network are journalists and media, which is a sign of the concern that this sophisticated form of manipulation generates in this collective.

Keywords
cybersecurity; deepfake; fake news; NodeXL; social media; Social Network Analysis; Twitter

Issue
This article is part of the issue "Disinformation and Democracy: Media Strategies and Audience Attitudes" edited by Pere Masip (University Ramon Llull, Spain), Bella Palomo (University of Málaga, Spain) and Guillermo López (University of Valencia, Spain).

1. Introduction

The recent upsurge in artificial intelligence (AI), along with image processing and machine learning, have made deepfake production possible. A video scarcely a minute long that featured Barack Obama spouting harsh criticism against current US President, Donald Trump, went viral in early April 2018 (Fagan, 2018). In fact, the previous US leader had said nothing, although it was his image that appeared in the video. The person who made it was actor Jordan Peele. He sought to sound the alarm on how dangerously easy it was to use new technologies to manipulate and falsify someone's identity. Deepfake videos entail risks, and can potentially undermine truth, confuse citizens and falsify reality. With the arrival of social media, the spread of this sort of content seems to be unstoppable. Potentially, it may worsen issues related to disinformation and conspiracy theories (Hasan & Salah, 2019). They could even be weaponized to unleash national or international crises (Stover, 2018).

The firstly widely known examples of deepfakes appeared in November 2017, when a Reddit user called Deepfakes uploaded a series of videos with the faces of famous actresses, including Gal Gadot and Scarlett Johansson, over the faces of pornographic actresses (Rense, 2018). Since then, the media and the general

public have begun using the term deepfakes to refer to this sort of video made with AI, where one person's face can be confused with another's.

When the computer code used to make the fakes was shared, it sparked great interest in the Reddit community and the amount of fake content spread and increased. The fakes' initial targets were celebrities, including actors, singers and politicians. There are two possible main reasons that they were successful: accessibility and credibility (Kietzmann, Lee, McCarthy, & Kietzmann, 2020), since we tend to trust more in voices we know and in videos we see (Brucato, 2015).

2. State of Play

Manipulating photographs and videos, altering the reality of the recorded moment, came before the Internet. Different countries have carried out propaganda campaigns since World War II (Rutenberg, 2017). However, deepfakes account for a fundamental paradigm shift in how the world will operate online (Chesney & Citron, 2019).

Driven by the latest technological progress in AI and machine learning, there is a growing number of tools that make it possible for any unqualified individual to relatively easily create fake content that is increasingly more difficult to detect. In 2018, the popular face-swap program FakeApp required huge amounts of input data to create fakes (Robertson, 2018). One year later, similar applications like Zao, Doublicat and DiffSnap were more accessible and less demanding (Mehta, 2020).

This technical resource is being widely used in action films to replace actors with digital avatars in certain dangerous scenes, or even to digitally resurrect actors who have passed away (Atkin, 2019). However, when we observe their use in information systems, there are a great number of dangers and ethical challenges (Sora, 2018).

While there are those who mention the entertaining and even positive side of fakes (Kietzmann et al., 2020), some works address the use of deepfakes in online disinformation campaigns to manipulate public opinion (Riechmann, 2018). Many authors warn of the important repercussions that failure to curb their spread may pose, both to the population (Newman et al., 2015) and to democracy (Bennett & Livingston, 2018; Chadwick, Vaccari, & O'Loughlin, 2018; Rojecki & Meraz, 2016; Waisbord, 2018). There are even those who state that their fast and widespread dissemination can lead to great economic loss or national security risks (Yadlin-Segal & Oppenheim, 2020). In parallel fashion, if deepfakes contribute to increased uncertainty regarding the information they contain, another one of the risks of their use is reduced trust in the news media (Fletcher, 2018; Vaccari & Chadwick, 2020). Credibility in the news is falling around the world (Hanitzsch, Van Dalen, & Steindl, 2018) and trust in social media news is now less than news accessed through other channels (Newman, Fletcher, Kalogeropoulos, Levy, & Nielsen, 2018).

Given that there is a great deal at stake, automatic detection of deepfakes is an important problem, although difficult to undertake. Some argue that they can be fought through legislation and regulations, company policies, education and training (Westerlund, 2019). There are others who advocate for developing technology to detect deepfakes and to authenticate content and for prevention. In fact, many tools have been created to automatically detect deepfakes. To date, methods to detect these digital manipulations were based on intrinsic contradictions in the algorithm synthesis. For example, a lack of actual eye blinking (Li, Chang, & Lyu, 2018), or mismatching lip movement with speech (Korshunov & Marcel, 2018). There are systems that use a Convolutional Neural Network that extracts frame-level features that are then used to train a Recurrent Neural Network that learns to determine whether or not a video has been manipulated (Güera & Delp, 2018; Li & Lyu, 2018). There are even those who suggest tracking and monitoring the source and history of content to the origin, based on the principle that if it is reliable or prestigious, then the content can be real and authentic (Hasan & Salah, 2019).

Deepfakes promise to be one of the greatest challenges for social media platforms in 2020. Some, like Facebook and Adobe, raised policies to detect and fight deepfakes. The latest was Twitter, which announced a new policy in February to fight the impact of manipulated content (Robertson, 2020). Moreover, Google has also decided to take action to limit their reach by creating an algorithm to detect and automatically delete deepfakes uploaded to YouTube and other Google services. It also created a tool called Assemble that helps journalists to identify manipulated images (Alba, 2020).

Although deepfakes have become a topic of debate, academic research has only just recently begun addressing digital disinformation on social media (Anderson, 2018), which can be dangerous to the public sphere given the potential to create states of false opinion (Pennycook, Cannon, & Rand, 2018). In this regard, this study contributes to this debate by analyzing the conversation on Twitter about the deepfake phenomenon and which type of actors are most referenced and made viral by users, all after the news that Twitter was going to double down in its efforts to fight fake news and videos.

3. Objectives

This study's generic objective is to analyze the conversation and the structure of relationships on the net arising around the term deepfake on Twitter by means of the social network analysis technique. Deepfakes is still a relatively new and 'fluid' phenomenon in the making. This article may help people understand how different actors try to shape and 'crystalize' our understanding of the emerging issue, as well as mapping the most important actors in this debate. It contains the following specific objectives: 1) Identify the main actors and research

which ones hold a greatest advantage when controlling the spread of messages—all actors who posted messages containing the term deepfake or who were replied to or mentioned in those messages have been examined; and 2) analyze the semantic network arising around the search term deepfake and discover which content predominates in messages.

4. Hypothesis

The following hypothesis are formulated:

> H1: Politicians and the media are amongst the actors who are most referenced and made viral by third parties when speaking of deepfakes on Twitter (related to the first objective).

Politicians, because they often become the involuntary protagonists of videos which, with a humorous tone, form part of disinformation campaigns that affect their image and credibility. The news media, because they are worried about the consequences that improper use of this face-swapping technology may have for governments, companies and the general population.

> H2: The most relevant topics when users discuss deepfakes on Twitter (related to the second objective) are related to politics and concern over the growing difficulty in distinguishing between reality and fiction in the near future.

It is important to examine whether the content about deepfakes also relates to politics because "political deepfakes are an important product of the Internet's visual turn. They are at the leading edge of online, video-based disinformation and, if left unchallenged, could have profound implications for journalism, citizen competence, and the quality of democracy" (Vaccari & Chadwick, 2020, p. 2). In this sense, according to Maddocks (2020), most of the deepfakes that are spread on the Internet today are pornographic in nature, but public attention is typically focused on political deepfakes. Often simulating the image of high-profile politicians, these videos spread hoaxes and lead to political instability.

5. Methodology

Using the Social Network Analysis technique (Borgatti, Mehra, Brass, & Labianca, 2009; Freeman, 2004; Otte & Rousseau, 2002; Wasserman & Faust, 1994), this article studies the structure of networked relationships woven around the term deepfake on the social media platform Twitter. This platform was selected because it is open and creates a huge amount of interpersonal interactions that can be collected by academic researchers to study processes of how information is spread on social networks (Benson, 2016; Boyd, 2014; Brubaker & Wilson, 2018; Evans, 2016; Tolson, 2010).

To explore the properties of the net (relevance of actors and information flows), open-source software NodeXL Pro was used, one of the programs to study digital networks that is most used by the scientific community (Hansen, Shneiderman, & Smith, 2010; Ricaurte & Ramos-Vidal, 2015; Smith et al., 2010). This tool was used in different works of research, such as the one analyzing connections between politicians and journalists in Holland (Verweij, 2012), the use of hashtags and trending topics (Dossis, Amanatidis, & Mylona, 2015; Wukich & Steinberg, 2013), news-spreading processes (Ahmed & Lugovic, 2019), the spread of hoaxes regarding the coronavirus (Pérez-Dasilva, Meso-Ayerdi, & Mendiguren-Galdospín, 2020), and more.

The software captured a network of 15,885 actors who posted messages containing the term deepfake or who were replied to or mentioned in those messages. The sample was taken from a data set limited to a maximum of 18,000 tweets (formal limits of the NodeXL software sample universe). The database was obtained through Twitter's streaming API February 28th, 2020, at 09:41 UTC. The reason for this choice is that on February 5th, the platform created by Jack Dorsey announced a new policy to fight content manipulation like fake news and fake videos (Robertson, 2020). The collected tweets were posted between February 7, 2018, at 11:17 UTC and February 28, 2020, at 09:28 UTC. Users were grouped by hierarchical conglomerates (or cluster analysis; Kaleel & Abhari, 2015; Paolillo, 2008), using the algorithm by Clauset, Newman, and Moore (2004). To visualize the network, Harel and Koren's (2000) multiscale design algorithm was used, which facilitates identification of actors and their links. Analysis was based on directed and weighted edges. The weight reflected the number of times that actors posted messages containing the term deepfake or who were replied to or mentioned in those messages.

To analyze the role held by actors and the relationships that occur between them on the network revolving around the term deepfake, two of the most common centrality indicators in the SNA were used: in-degree and the degree of betweenness. The in-degree is the number of interactions an actor has received from other users forming the structure (Aguilar-Gallegos et al., 2016; Fernández, 2019). Actors with the highest numbers were the most-referenced and made most-viral, so their content is the most influential. On the other hand, the degree of betweenness is the capacity to control spreading of a message (Gibbs & McKendrick, 2015; Hansen et al., 2010). Users with the highest numbers acted as bridges over which relevant information flowed, and they contributed to spreading or blocking it for other parts of the network. The color and size of the nodes show the most relevant accounts, and the strength of the bonds between them was shown with the intensity of the lines joining them.

To study the semantic network created around the search term deepfake, words such as conjunctions and

prepositions, which are not relevant, were eliminated. Next, a data-mining strategy based on word-matching was applied (on nouns, verbs, adverbs and adjectives) to identify the strongest connections (Seo, Kim, Kim, Kim, & Kim, 2019) and its presence in each message was studied from a relational perspective. These data were interpreted as non-directed graphs.

6. Results

Structure indicators or cohesion measurements, such as density or reciprocity, that analyze the complete network's properties were 0.006393862 for the ratio of reciprocal vertex pairs and 5.981029 for the average geodesic distance. The first data indicates that 6 of every 100 users held mutual communication during the period of study, and the second that the actor was located at almost six steps on average from any other in the analyzed structure. Moreover, density or cohesion was 6.86199%. These data indicate that this is a dense network, where nodes are not very far from one another and with a high speed of information transmission.

Centrality indicators, which show the position a node or actor holds on the network, also bore interesting results. Figure 1 shows the existence of different communities. Of note are eight large-sized clusters (light-blue, orange, red, green, dark green, yellow, light green and maroon), followed by ten moderately-sized groupings.

According to the nodes' in-degree, of the first 20, a group of 10 accounts from India stands out, followed by eight profiles from the US or headquartered in the US (such as Mashable or Elon Musk), plus another one from Brazil and another from France (Figures 2 and 3). The one with the highest value is @techreview, MIT's magazine. This actor's contents are amongst the most-referenced and most virally spread by third parties. One of the messages reports on the purchase by Fintech Square, headed by Twitter CEO Jack Dorsey, of the AI research company Dessa, a company known for its deepfake-detection software. Its technology became known thanks to the deepfake on Joe Rogan, a mixed martial arts commentator and one of the most popular podcasters in the world on May 17, 2019 (Vincent, 2019).

Moreover, another post under the same profile is of note, which went widely viral and reported on the use of a deepfake to win voters, used by Indian politician Manoj Tiwari, president of the Bharatiya Janata party (BJP). According to MIT's magazine, this was the first time in the world that a political party used a deepfake for an electoral campaign. The controversy arose when

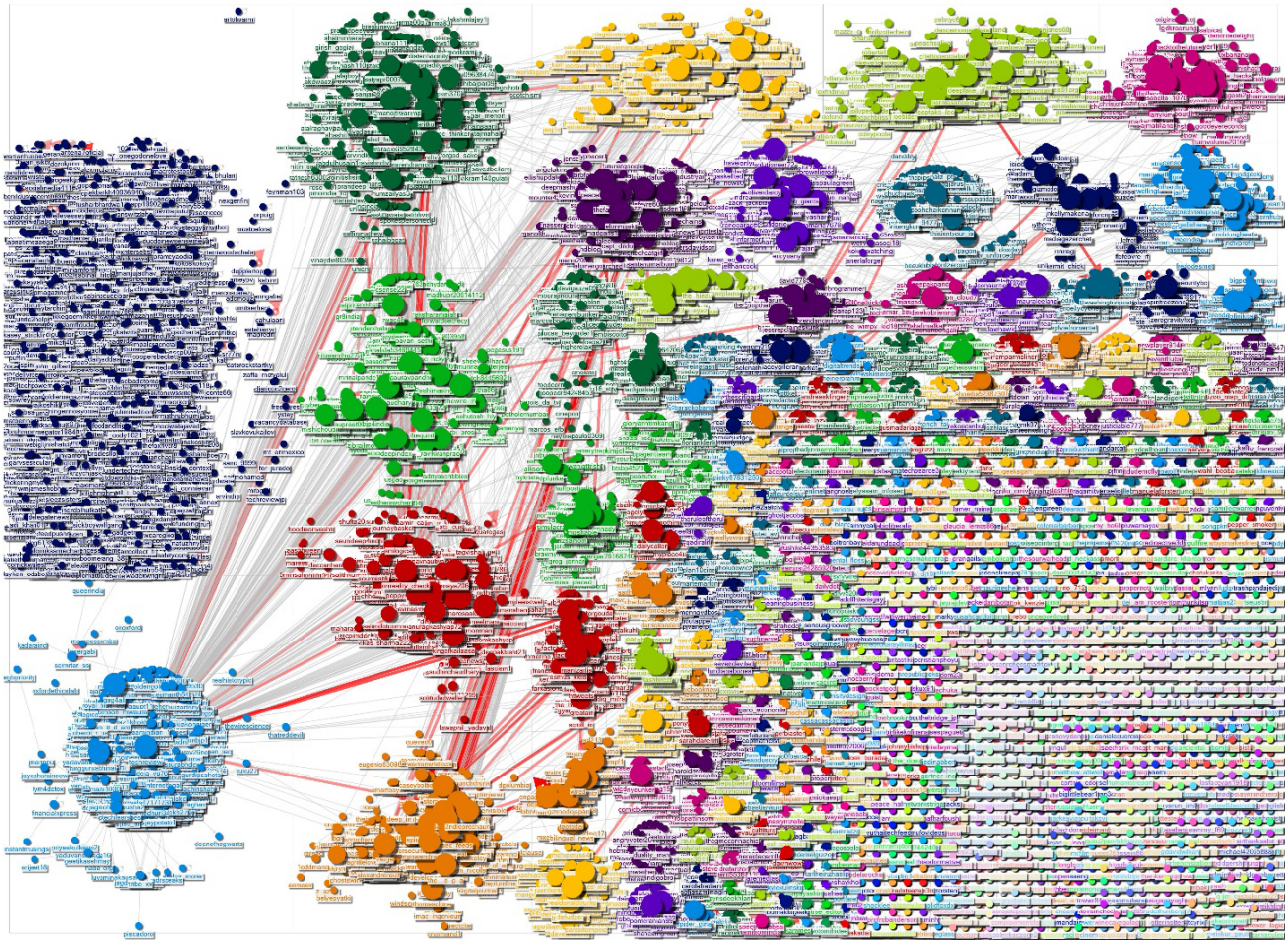

Figure 1. Illustration of the network around the term deepfake.

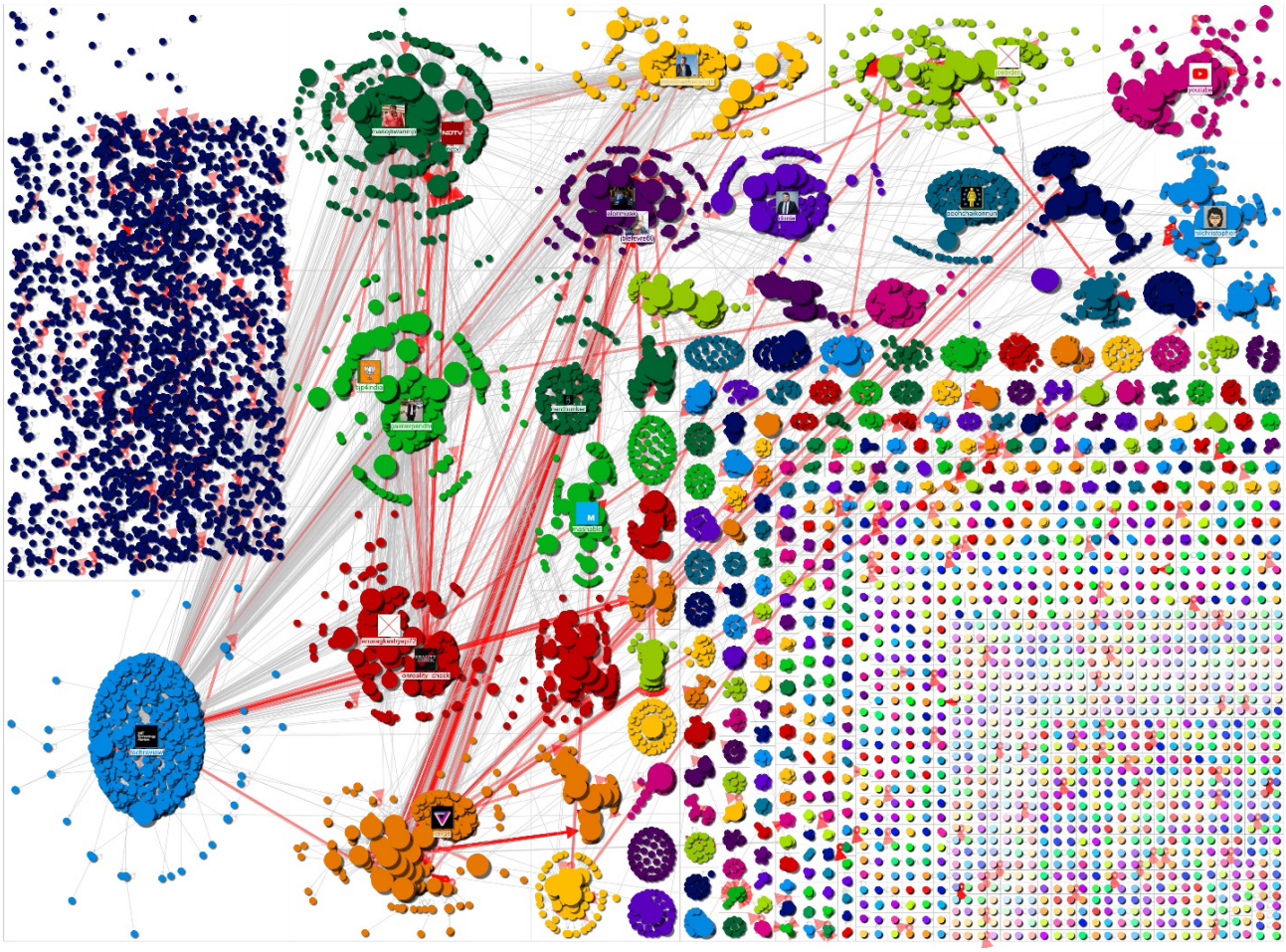

Figure 2. Illustration of the 20 actors most-referenced by users.

Tiwari manipulated one of his electoral videos with deepfake technology to simulate that he was speaking a Hindi dialect, and thus reach millions of voters that would have been unreachable otherwise, since they only speak this dialect. According to the party itself, they hired the company Ideaz Factory to create deepfakes to reach voters in the 22 different languages and 1,600 dialects used in India.

An examination of the content around the 20 most-referenced actors detects three aspects. On the one hand, the presence of politics when speaking of deepfakes must be mentioned. Throughout the period of study, ten of the actors most-referenced by third parties are related to Manoj Tiwari (the magazine @techreview, activist @GuaravPandhi, @ManojTiwariMP, journalist @UmashankarSingh, TV program @OnReality_Check, politician @amitmalviya, film director @anuragkashyap72, the party @BJP4India, journalist @NilChristopher and television channel @ndtv). In addition to this group is US politician Joe Biden, another node to which a huge number of users go in an attempt to generate a direct link with him. Former vice president Biden became a protagonist based on a video related to the Democratic Primary debate of 2020. The original recording of the debate in Nevada was edited by Mike Bloomberg, one of the participants, to improve his image since he did not appear in a flattering light. The billionaire modified the audio and order of scenes in the video and included grasshopper sounds when his adversaries responded. The video obtained 4,2 million views. Shortly thereafter, Twitter announced that it would sanction the video for violating its new media manipulation policy.

Another topic revolves around film. Six of the most viralized actors are related to two manipulations of popular films using AI to swap the faces of movie stars in one or several iconic scenes. One of the deepfakes, with almost half a million views, fakes a moment from the well-known science fiction series *Star Trek*. The video was made by The Fakening, a famous YouTube channel owned by programmer Paul Shales, devoted to creating fake videos with AI. This face-swapping technology places Elon Musk and Jeff Bezos in the role of the series' actors, and is associated with the media profiles @verge, owner of Tesla @elonmusk, owner of Amazon @JeffBezos and French influencer @jblefevre60. The other manipulated film, the work of YouTuber EZRyderX47 with almost nine million views, is *Back to the Future*. Thanks to its quality, it is referenced by commentator @PoohChaikonNun, entertain-

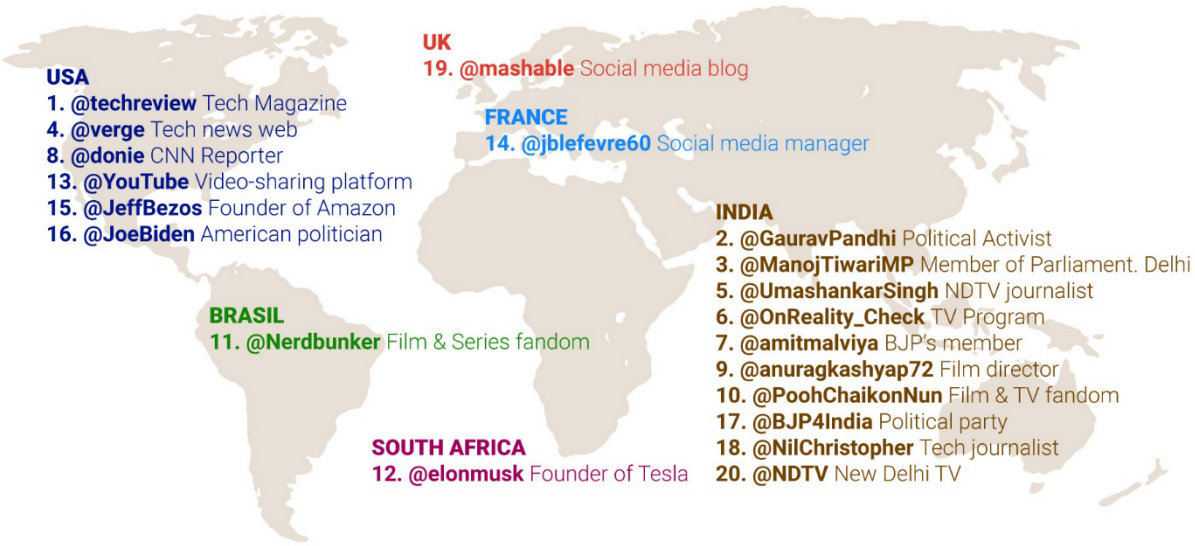

Figure 3. In-degree: Actors most-referenced and viralized by other users.

ment website @Nerdbunker and another media outlet @mashable.

Lastly, the third topic area has to do with the very technology used to produce deepfakes. In this area, five profiles are of note whose contents are amongst the most-referenced and made most-viral by third parties. As mentioned previously, one of them is @techreview, MIT's magazine, reporting on Jack Dorsey's purchase of the deepfake software company Dessa. Another is *CNN* journalist Donie O'Sullivan (@donie), author of the highly viral news piece on the dangers associated with improper use of this technology. This information is also related to @NilChristopher, another one of the actors with a high in-degree. Within this scope, we also see French influencer @jblefevre60 holding one of the top positions. He is mentioned in a very widespread tweet by global influencer Spiros Margaris, explaining how video faking technology works. The fifth actor is @Youtube, since there are plenty of videos about the dangers of deepfakes that end with the phrase "via @Youtube" to indicate the platform from whence the content was obtained.

Regarding the degree of betweenness (Figures 4 and 5), we observe that 13 of the actors mentioned in the section above appear again in the 20 top positions. They have the highest values, which means that these nodes are intermediaries through which relevant information related to deepfakes is spread. These users are the ones who contribute the most to spreading or blocking messages to other people that give shape to the structure. In this regard, it is interesting to highlight that, of the 20 actors with the most favorable positions, there are six media outlets (@techreview, @verge, @mashable, @OnReality_Check, @CNN, @YouTube) and three journalists (@donie, @UmashankarSingh, @NilChristopher) who act as bridges in interactions giving shape to the network. In this regard, also in analyzing the role played by certain actors in configuring the structure, we must make special mention of @thefakening (15th position), because this YouTube channel creates a good portion of the most popular deepfakes spread amongst social media. Barring exceptions, many of his fake videos garner no more than 25,000 views, but the fake with Elon Musk and Jeff Bezos as actors on *Star Trek* reached almost half a million views and became his most viral video.

Regarding semantic analysis of the network, the most relevant conversation threads revolve around the video of Indian politician Manoj Tiwari, highlighting that this is the first time in the world that a political party used deepfake technology to conduct an electoral campaign (Figure 6). Moreover, it is rated as "dangerous and illegal" (Pandhi, 2020). The second most-significant association has to do with the deepfake of *Star Trek*, with the owners of Amazon and Tesla. In third position, we find references to the *Back to the Future* video.

7. Conclusions

This study shows the network woven around the term deepfake after Twitter's announcement that it was tightening its protocols to fight fake news and videos. The data indicate that this is a dense network with high connectivity where information on deepfakes quickly spreads. Although reports state that 96% of these fakes are non-consensual pornography (Patrini, 2019), this piece of research observes that in the microblogging network, the most important topics are not related to pornographic content. The nodes with the most favorable positions in the structure converse on fake videos related to politicians (H1). This coincides with studies such as those by Maddocks (2020) which explain that, although most of the deepfakes that spread over the

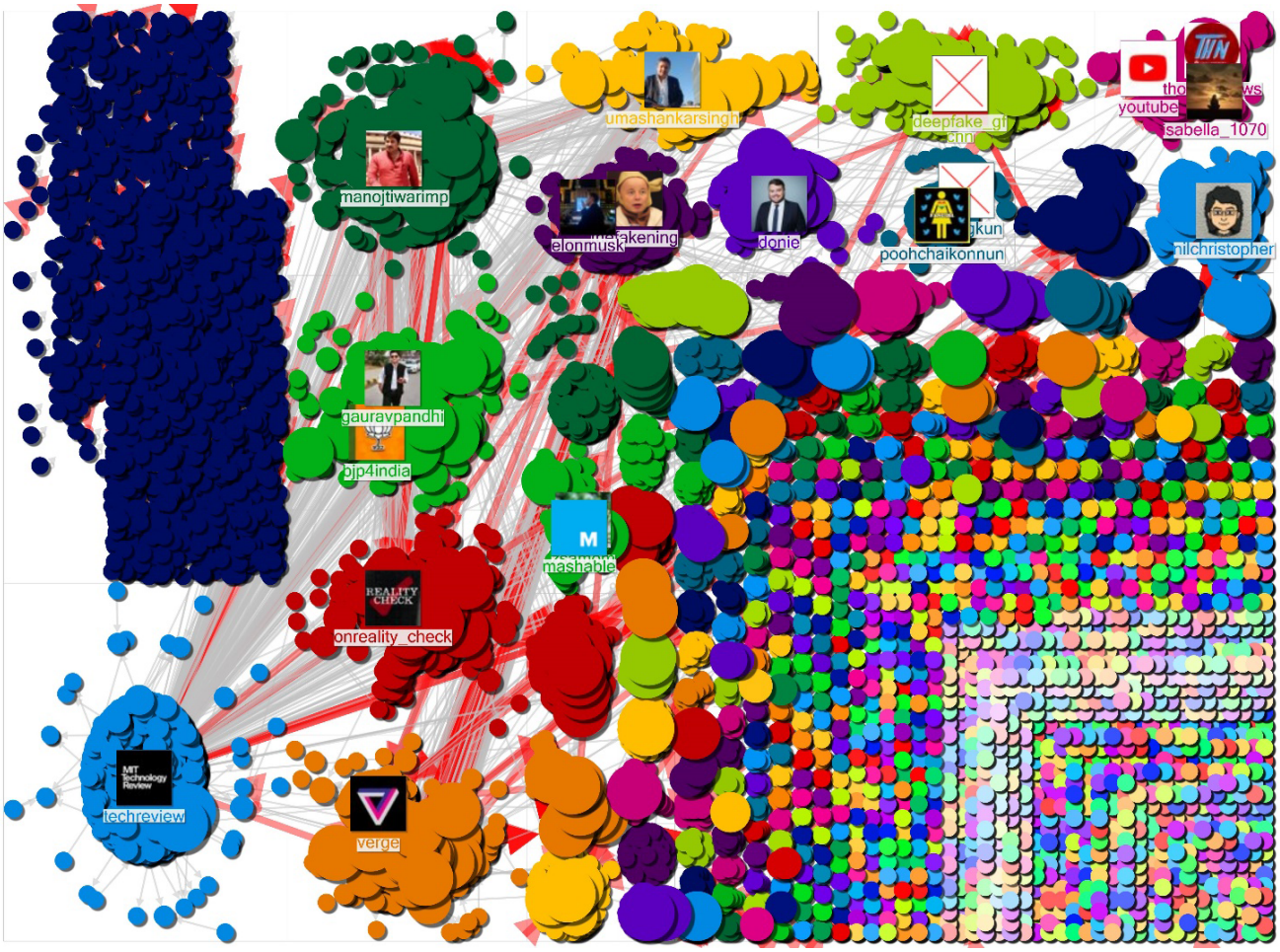

Figure 4. Illustration of the main actors control the flow of information on the network.

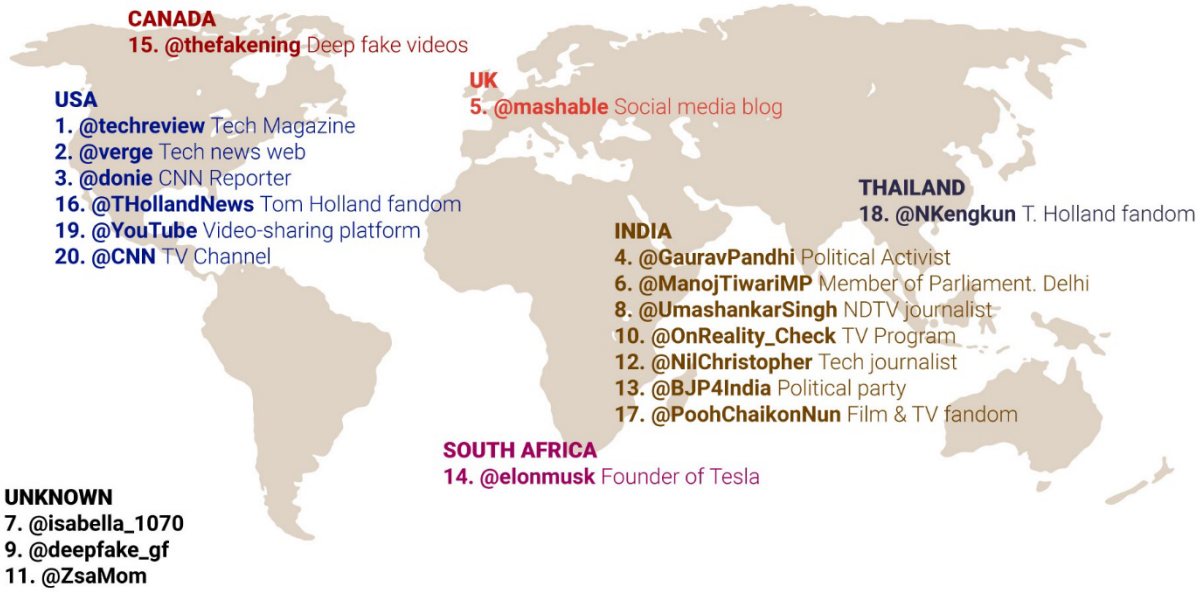

Figure 5. Betweenness degree: Actors who act as bridges in the interactions that shape the network.

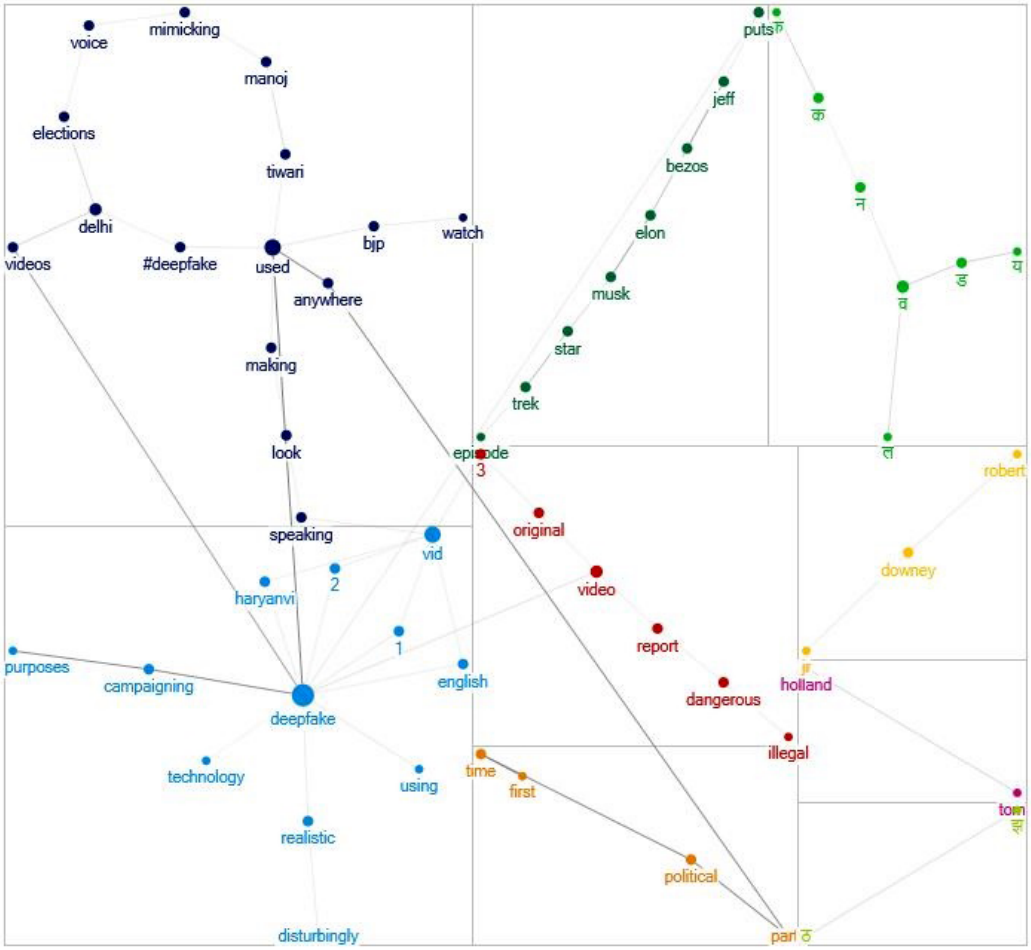

Figure 6. Illustration of the most relevant conversation topics regarding deepfakes.

Internet are pornographic in nature, public attention is focused above all on political deepfakes because of their ability to generate political instability. In contrast, other authors such as Westerlund (2019) conclude that the reason many counterfeits focus on celebrities and politicians is basically because they are public figures who have a large number of free videos and photos on the Internet and it is an easy way to train an AI deepfake system. According to these two recent studies, this is only the beginning. There are going to be more and more deepfakes that use AI to spread fake political videos tailored to the preferences of social media users.

Secondly, the nodes that have the greatest structural advantage in the network also refer to satirical videos of famous films where face-swapping technology is used on the actors in one or several iconic scenes from said films.

The result of this is that if we consider the network according to the in-degree, the most-referenced and viralized users are celebrities (politicians, businessmen or businesswomen, singers, athletes and more) that see how they become the target of manipulations. Thirdly, and in relation to the above two points, we also observe concern, especially with news media, for the consequences that improper use of this AI technology may have for citizens, companies and governments (H2). In all their tweets, analyzed in this study, news media talk about the potential danger of this technology. This research coincides with recent studies such as those by Yadlin-Segal and Oppenheim (2020, p. 1) because it "shows how journalists frame deepfakes as a destabilizing platform that undermines a shared sense of social and political reality." On the other hand, if we consider the network in terms of degree of betweenness, we observe that half the actors with the greatest capacity to control the spread of messages on deepfakes are also journalists or news media (H1). In this study, although most of the videos are entertaining and easy to spot, these professionals are clearly concerned and have the responsibility of discrediting these fake videos and avoiding the manipulation of public opinion: "Authentication of video is especially important to news media companies who have to determine authenticity of a video spreading in a trustless environment, in which details of the video's creator, origin, and distribution may be hard to trace" (Westerlund, 2019, p. 46)

In this work, attention has been focused on the conversation about deepfakes, about the way users talk about the subject, and it has been shown who are the most referenced actors and whose contents are the most viralized by users. As mentioned above in the arti-

cle, deepfakes is still a relatively new phenomenon and the purpose of this manuscript has been to help understand how different actors try to shape and crystalize our understanding of the emerging issue, as well as mapping the most important actors in this debate. Current research could be extended to include the study of the spread of deepfakes. Work is continuing in this area in order to overcome limitations connected to this study.

Acknowledgments

The authors would like to thank the three anonymous reviewers as well as the Academic Editors for their valuable feedback on the manuscript. This research was supported by the Spanish Ministry of Science, Innovation and Universities ("News, Networks and Users in the Hybrid Media System: Shared Creation and Dissemination of News in Online Media," RTI2018-095775-B-C41). It was carried out within the Consolidated Research Group 'Gureiker' (A) (IT1112-16), funded by the Basque Government.

Conflict of Interests

The authors declare no conflict of interests.

References

Aguilar-Gallegos, N., Martínez-González, E. G., Aguilar-Ávila, J., Santoyo-Cortés, H., Muñoz-Rodríguez, M., & García-Sánchez, E. I. (2016). Análisis de redes sociales para catalizar la innovación agrícola: De los vínculos directos a la integración y radialidad [Social network analysis for catalysing agricultural innovation: From direct ties to integration and radiality]. *Estudios Gerenciales*, *32*(140), 197–207. https://doi.org/10.1016/j.estger.2016.06.006

Ahmed, W., & Lugovic, S. (2019). Social media analytics: Analysis and visualisation of news diffusion using NodeXL. *Online Information Review*, *43*(1), 149–160. https://doi.org/10.1108/OIR-03-2018-0093

Alba, D. (2020, February 4). Tool to help journalists spot doctored images is unveiled by jigsaw. *The New York Times*. Retrieved from https://www.nytimes.com/2020/02/04/technology/jigsaw-doctored-images-disinformation.html

Anderson, K. E. (2018). Getting acquainted with social networks and apps: Combating fake news on social media. *Library Hi Tech News*, *35*(3), 1–6. https://doi.org/10.1108/LHTN-02-2018-0010

Atkin, M. (2019, March 27). Human assets. *Medium*. Retrieved from https://immerse.news/human-assets-624f3066c2ce

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Benson, P. (2016). *The discourse of YouTube: Multimodal text in a global context*. London: Routledge. https://doi.org/10.4324/9781315646473

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895. https://doi.org/DOI:10.1126/science.1165821

Boyd, M. S. (2014). (New) participatory framework on YouTube? Commenter interaction in US political speeches. *Journal of Pragmatics*, *72*, 46–58. https://doi.org/10.1016/j.pragma.2014.03.002

Brubaker, P. J., & Wilson, C. (2018). Let's give them something to talk about: Global brands' use of visual content to drive engagement and build relationships. *Public Relations Review*, *44*(3), 342–352. https://doi.org/10.1016/j.pubrev.2018.04.010

Brucato, B. (2015). Policing made visible: Mobile technologies and the importance of point of view. *Surveillance & Society*, *13*(3/4), 455–473. https://doi.org/10.24908/ss.v13i3/4.5421

Chadwick, A., Vaccari, C., & O'Loughlin, B. (2018). Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing. *New Media & Society*, *20*(11), 4255–4274. https://doi.org/10.1177/1461444818769689

Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*. Retrieved from https://www.questia.com/magazine/1P4-2161593888/deepfakes-and-the-new-disinformation-war-the-coming

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, *70*(6). https://doi.org/10.1103/PhysRevE.70.066111

Dossis, M., Amanatidis, D., & Mylona, I. (2015, November 9). Mining Twitter data: Case studies with trending hashtags. In M. Mokryš & Š. Badura (Eds.), *Proceedings in ARSA-advanced research in scientific areas* (pp. 242–246). Zilina: EDIS Publishing Institution of the University of Žilina.

Evans, M. (2016). Information dissemination in new media: YouTube and the Israeli-Palestinian conflict. *Media, War & Conflict*, *9*(3), 325–343. https://doi.org/10.1177/1750635216643113

Fagan, K. (2018, April 17). A viral video that appeared to show Obama calling Trump a "dips" shows a disturbing new trend called "deepfakes." *Business Insider*. Retrieved from https://www.businessinsider.com/obama-deepfake-video-insulting-trumpp.2018--4

Fernández, D. (2019, August 19). Análisis de grafos en redes sociales: Medidas de centralidad [Analysis of networks in social networks: Measures of centrality]. *Datahack*. Retrieved from https://www.datahack.es/grafos-redes-sociales-centralidad

Fletcher, J. (2018). Deepfakes, artificial intelligence, and some kind of dystopia: The new faces of online post-fact performance. *Theatre Journal*, *70*(4), 455–471.

https://doi.org/10.1353/tj.2018.0097
Freeman, L. (2004). *The development of social network analysis: A study in the sociology of science*. Vancouver: Empirical Press.
Gibbs, W. J., & McKendrick, J. (Eds.). (2015). *Contemporary research methods and data analytics in the news industry*. Pennsylvania, PA: IGI Global.
Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. In *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)* (pp. 1–6). Los Alamitos, CA: IEEE Computer Society.
Hanitzsch, T., Van Dalen, A., & Steindl, N. (2018). Caught in the nexus: A comparative and longitudinal analysis of public trust in the press. *The International Journal of Press/Politics*, *23*(1), 3–23. https://doi.org/10.1177/1940161217740695
Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Burlington, MA: Morgan Kaufmann.
Harel, D., & Koren, Y. (2000). A fast multi-scale method for drawing large graphs. In S. Levialdi, V. Di Gesu, & L. Tarantino (Eds.), *Proceedings of the working conference on advanced visual interfaces: AVI '00* (pp. 282–285). New York, NY: ACM Press. https://doi.org/10.1145/345513.345353
Hasan, H. R., & Salah, K. (2019). Combating deepfake videos using blockchain and smart contracts. *IEEE Access*, *7*, 41596–41606. https://doi.org/10.1109/ACCESS.2019.2905689
Kaleel, S. B., & Abhari, A. (2015). Cluster-discovery of Twitter messages for event detection and trending. *Journal of Computational Science*, *6*, 47–57. https://doi.org/10.1016/j.jocs.2014.11.004
Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006
Korshunov, P., & Marcel, S. (2018). Speaker inconsistency detection in tampered video. In *2018 26th European signal processing conference (EUSIPCO)* (pp. 2389–2393). Los Alamitos, CA: IEEE Computer Society.
Li, Y., Chang, M.-C., & Lyu, S. (2018). *In ictu oculi: Exposing AI generated fake face videos by detecting eye blinking*. ArXiv. http://arxiv.org/abs/1806.02877
Li, Y., & Lyu, S. (2018). *Exposing deepfake videos by detecting face warping artifacts*. ArXiv. https://arxiv.org/abs/1811.00656
Maddocks, S. (2020). A deepfake porn plot intended to silence me: Exploring continuities between pornographic and 'political' deep fakes. *Porn Studies*. https://doi.org/10.1080/23268743.2020.1757499
Mehta, I. (2020, January 13). New deepfake app pastes your face onto GIFs in seconds. *The Next Web*. Retrieved from https://thenextweb.com/artificial-intelligence/2020/01/13/new-deepfake-app-pastes-your-face-onto-gifs-in-seconds
Newman, E. J., Garry, M., Unkelbach, C., Bernstein, D. M., Lindsay, D. S., & Nash, R. A. (2015). Truthiness and falsiness of trivia claims depend on judgmental contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1337–1348. https://doi.org/10.1037/xlm0000099
Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2018*). Reuters Institute digital news report 2018*. Oxford: Reuters Institute for the Study of Journalism. Retrieved from http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf
Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, *28*(6), 441–453.
Pandhi, G. [GauravPandhi]. (2020, February 19). Watch how BJP used #deepfake in Delhi elections where a voice mimicking Manoj Tiwari is used, making it look like he is speaking. Report: https://t.co/8zcKwnSxbL This is dangerous, should be illegal! https://t.co/WEXb0zaXdl [Tweet]. Retrieved from https://twitter.com/gauravpandhi/status/1229966733370257411
Paolillo, J. C. (2008). Structure and network in the YouTube core. *Proceedings of the 41st annual Hawaii international conference on system sciences (HICSS 2008)* (pp. 156–156). Los Alamitos, CA: IEEE Computer Society. https://doi.org/10.1109/HICSS.2008.415
Patrini, G. (2019, October 7). Mapping the deepfake landscape. *Deeptrace*. Retrieved from https://deeptracelabs.com/mapping-the-deepfake-landscape
Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465
Pérez-Dasilva, J.-A., Meso-Ayerdi, K., & Mendiguren-Galdospín, T. (2020). Fake news y coronavirus: Detección de los principales actores y tendencias a través del análisis de las conversaciones en Twitter [Fake news and coronavirus: Detecting key players and trends through analysis of Twitter conversations]. *El Profesional de la Información*, *29*(3). https://doi.org/10.3145/epi.2020.may.08
Rense, S. (2018, February 12). What are 'deepfakes,' and why are Pornhub and Reddit banning them? *Esquire*. Retrieved from https://www.esquire.com/lifestyle/sex/a17043863/what-are-deepfakes-celebrity-porn
Ricaurte, P., & Ramos-Vidal, I. (2015). Investigación en redes sociales digitales: Consideraciones metodológicas desde el paradigma estructural [Research on digital social networks: methodological considerations from the structural paradigm]. *Virtualis: Revista de Cultura Digital*, *6*(11), 165–194.
Riechmann, D. (2018). I never said that! High-tech deception of 'deepfake' videos. *WCJB*. Retrieved from https://www.wcjb.com/content/news/I-never-

said-that-High-tech-deception-of-deepfake-videos-487147011.html

Robertson, A. (2018, February 11). I'm using AI to face-swap Elon Musk and Jeff Bezos, and I'm really bad at it. *The Verge*. Retrieved from https://www.theverge.com/2018/2/11/16992986/fakeapp-deepfakes-ai-face-swapping

Robertson, A. (2020, February 4). Twitter will ban 'deceptive' faked media that could cause 'serious harm.' *The Verge*. Retrieved from https://www.theverge.com/2020/2/4/21122661/twitter-deepfake-manipulated-media-policy-rollout-date

Rojecki, A., & Meraz, S. (2016). Rumors and factitious informational blends: The role of the web in speculative politics. *New Media & Society*, *18*(1), 25–43. https://doi.org/10.1177/1461444814535724

Rutenberg, J. (2017, September 13). RT, Sputnik and Russia's new theory of war. *The New York Times*. Retrieved from https://www.nytimes.com/2017/09/13/magazine/rt-sputnik-and-russias-new-theory-of-war.html

Seo, S., Kim, J.-K., Kim, S.-I., Kim, J., & Kim, J. (2019). Semantic hashtag relation classification using co-occurrence word information. *Wireless Personal Communications*, *107*(3), 1355–1365. https://doi.org/10.1007/s11277-018-5745-y

Smith, M., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., & Dunne, C. (2010). *NodeXL: A free and open network overview, discovery and exploration add-in for Excel 2007/2010*. Redwood City, CA: Social Media Research Foundation.

Sora, C. (2018). El futuro digital de los hechos [The digital future of facts]. *Hipertext.net: Revista Académica sobre Documentación Digital y Comunicación Interactiva*, *0*(17), 1–10. https://doi.org/10.31009/hipertext.net.2018.i17.01

Stover, D. (2018). Garlin Gilchrist: Fighting fake news and the information apocalypse. *Bulletin of the Atomic Scientists*, *74*(4), 283–288. https://doi.org/10.1080/00963402.2018.1486618

Tolson, A. (2010). A new authenticity? Communicative practices on YouTube. *Critical Discourse Studies*, *7*(4), 277–289. https://doi.org/10.1080/17405904.2010.511834

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408

Verweij, P. (2012). Twitter links between politicians and journalists. *Journalism Practice*, *6*(5/6), 680–691. https://doi.org/10.1080/17512786.2012.667272

Vincent, J. (2019, May 17). This AI-generated Joe Rogan fake has to be heard to be believed. *The Verge*. Retrieved from https://www.theverge.com/2019/5/17/18629024/joe-rogan-ai-fake-voice-clone-deepfake-dessa

Waisbord, S. (2018). Truth is what happens to news: On journalism, fake news, and post-truth. *Journalism Studies*, *19*(13), 1866–1878. https://doi.org/10.1080/1461670X.2018.1492881

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge University Press.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, *9*(11), 39–52. https://doi.org/10.22215/timreview/1282

Wukich, C., & Steinberg, A. (2013). Nonprofit and public sector participation in self-organizing information networks: Twitter hashtag and trending topic use during disasters: Self-organizing information networks. *Risk, Hazards & Crisis in Public Policy*, *4*(2), 83–109. https://doi.org/10.1002/rhc3.12036

Yadlin-Segal, A., & Oppenheim, Y. (2020). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence: The International Journal of Research into New Media Technologies*. https://doi.org/10.1177/1354856520923963

About the Authors

Jesús Pérez Dasilva is Professor of Online Journalism at the University of the Basque Country. His research focuses on digital journalism and social networks. He has published widely on these topics in peer-reviewed international journals included in listings as JCR or Scopus. Nowadays is member of the research project "News, Networks and Users in the Hybrid Media System: News Creation and Sharing in Online Media," funded by Ministry of Science, Innovation and Universities of Spain.

Koldobika Meso Ayerdi is a Senior Lecturer at the Department of Journalism II of the University of the Basque Country. He has researched online journalism and blogging and organizes the annual International Conference on Online Journalism and Web 2.0 at the University of the Basque Country in Bilbao, Spain.

Terese Mendiguren Galdospin is a Lecturer at the Department of Journalism II at the University of the Basque Country. Author of works on cyberjournalism, she investigates trends in contemporary journalism such as citizen journalism or the use of social networks in the dissemination of information. She is Member of the consolidated investigation group Gureiker. In her professional role, she has worked as an Editor and Program Coordinator for the Basque television channel (EITB) and Bilbovision Channel.