UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE RELAÇÕES INTERNACIONAIS

CHRISTOPHE MATTHIEU ERWAN MARRET

The impact of social media on the polarization of the political debate: three case studies in the European context

São Paulo
2022

CHRISTOPHE MATTHIEU ERWAN MARRET

The impact of social media on the polarization of the political debate: three case studies in the European context

Versão Corrigida

Dissertação apresentada ao Programa de Pós-Graduação em Relações Internacionais do Instituto de Relações Internacionais da Universidade de São Paulo, para a obtenção do título de Mestre em Ciências.

Orientadora: Profa. Dra. Marislei Nishijima

São Paulo
2022

Dissertação de autoria de Christophe Matthieu Erwan Marret, sob o título "**The impact of social media on the polarization of the political debate: three case studies in the European context**", apresentada ao Instituto de Relações Internacionais da Universidade de São Paulo, para obtenção do título de Mestre em Relações Internacionais pelo Programa de Pós-graduação de Relações Internacionais.

Prof. Dra. Marislei Nishijima

Instituição: Instituto de Relações Internacionais IRI –Universidade de São Paulo USP

Profa. Dra. Daniela Osvald Ramos

Instituição: Escola de Comunicações e Artes ECA - Universidade de São Paulo USP

Profa. Dra. Regina Célia Cati

Instituição: Departamento de Economia - Universidade de Boston

Abstract

MARRET, Christophe Matthieu Erwan. **The impact of social media on the polarization of the political debate**: three case studies in the European context. 2022. 91 p. Dissertation (Master of International Relations) – Institute of International Relations, University of São Paulo, São Paulo, 2022.

Social media platforms like Facebook are taking an increasingly important place in political and public opinion debates, and some commentators argue that they have a significant impact on the results of elections (Brexit and Trump election, for instance). This study, conducted through three academic articles, aims to assess the impact of social media on the polarization of the political debate through three major European case studies. To this finality, the research performs a literature review based on key studies besides investigating and grabbing institutional information. The first article explores how the new versions of Facebook's algorithm accentuate the social bubble effect, weakening the confrontation of the plurality of opinions necessary for the democratic debate. The literature shows that these social bubbles are a risk for democracies, but that they are inherent to the human psychological condition (the algorithms are just amplifying an existing phenomenon by reinforcing our persistent individual beliefs): the real issue seems to be our human capacity to question the veracity of the message we receive (hence the need for regulation of the social media platforms to minimize the risk of manipulations). Faced with these new challenges, the second article assesses how the institutions of the European Union sought to protect the parliamentary elections in May 2019 against the viral proliferation of fake news and against the unlawful use of citizens' data to target specific groups of strategic voters. This research focuses on the European General Data Protection Regulation (GDPR) mechanisms to avoid fake news and microtargeting effects on the electoral context. Preliminary results suggest that GDPR reduced social media's effects on the 2019 European Parliament elections by adopting a public and private co-regulation approach, which avoids counterproductive reinforcement of the auto-persuasion power of fake news. Finally, the article discusses the difficulties of applying the European data protection law in a new digital era to conclude that it is essential to continue developing alternative measures. To finish with, the third article discusses to what extent the concepts developed in the two previous articles can be applied to the Yellow Vests' social manifestation on France's streets from November of 2018 to the end of the year 2019, and specifically how the new algorithm introduced by Facebook in January 2018 (with the creation of the controversial "Facebook groups" which encourage the information exchanged between users sharing common points at the expense of information coming from "official pages" of traditional media or political figures) facilitated the emergence of this movement. The research explains how this algorithm played a key role in transforming digital anger into a radical social movement in the streets and investigates how the lack of governance on the social networks could lead to violent real-life manifestations, resulting in negative net gains in the long run and a high potential to destabilize democracies. The final considerations of this dissertation address the limits of the European model of protection (GDPR), open avenues for reflection on its adaptation, and present the new European legislation (the Digital Markets Act (DMA) and the Digital Services Act (DSA)) which should come into force in 2023 and whose general principles are to strengthen competition law in the face of the GAFAM "monopoly", and to condemn in digital life everything that is reprehensible in real life.

Keywords: democracy, algorithm, social media, fake news, filter bubble

Resumo

Plataformas de mídia social como o Facebook ocupam um lugar cada vez mais importante nos debates políticos, e alguns especialistas analisam que estas têm tido um impacto significativo nos resultados das eleições (Brexit e eleição de Trump, por exemplo). Este estudo, realizado através de três artigos acadêmicos busca avaliar o impacto das redes sociais sobre a polarização do debate político através de três estudos de caso europeus. Para tanto, a pesquisa realiza uma revisão de literatura baseada em estudos-chave, além de investigar informações institucionais. O primeiro artigo explora como as novas versões do algoritmo do Facebook acentuam o efeito de "filter bubble", enfraquecendo a pluralidade de opiniões necessária ao debate democrático. A literatura mostra que essas bolhas sociais são um risco para as democracias, mas que são inerentes à condição psicológica humana (os algoritmos estão apenas amplificando um fenômeno existente reforçando crenças individuais persistentes): o verdadeiro problema parece ser a capacidade humana a questionar a veracidade das mensagens que recebemos (sugerindo então a necessidade de regulamentação das plataformas de mídia social para minimizar o risco de manipulações). Diante destes novos desafios apresentados, o segundo artigo avalia como as instituições da União Europeia procuraram proteger as eleições parlamentares de maio de 2019 contra a proliferação viral de notícias falsas e contra a utilização ilícita de dados de cidadãos para atingir grupos específicos de eleitores estratégicos. Esta pesquisa foca em avaliar os mecanismos da European General Data Protection Regulation (GDPR) para evitar *fake n*ews e efeitos de *microtargeting* no contexto eleitoral. Os resultados preliminares sugerem que o GDPR reduziu os efeitos das mídias sociais nas eleições para o Parlamento Europeu de 2019 ao adotar uma abordagem de co-regulação pública e privada, o que teria evitado o efeito contraproducente do poder de autopersuasão das notícias falsas. O artigo então discute as dificuldades de aplicação da lei europeia de proteção de dados na era digital para concluir que é essencial continuar desenvolvendo medidas alternativas. Por fim, o terceiro artigo discute em que medida os conceitos desenvolvidos nos dois artigos anteriores podem ser aplicados à manifestação social dos Coletes Amarelos nas ruas da França de novembro de 2018 ao final de 2019, e mais precisamente como o novo algoritmo introduziu pelo Facebook em janeiro de 2018 (com a criação dos polêmicos "grupos do Facebook" que incentivam a troca de informações entre usuários que compartilham pontos em comum em detrimento de informações de "páginas oficiais" da mídia tradicional ou figuras políticas) facilitou o surgimento desse movimento. A pesquisa explica como esse algoritmo desempenhou um papel fundamental na transformação da raiva digital em um movimento social radical nas ruas e investiga como a falta de governança nas redes sociais pode levar a manifestações violentas na vida real, resultando em ganhos negativos no longo prazo e alto potencial de desestabilização das democracias. As considerações finais desta dissertação abordam os limites do modelo europeu de proteção (GDPR), abrem caminhos para a reflexão sobre a sua adaptação, e apresentam a nova legislação europeia (the Digital Markets Act (DMA) and the Digital Services Act (DSA)) que deve entrar em vigor em 2023 e cujos princípios gerais são fortalecer o direito da concorrência face ao "monopólio" dos GAFAM, e condenar na vida digital tudo que é repreensível na vida real.

Palavras-chave: democracia, algoritmo, Facebook, GDPR, coletes amarelos

Figures

List of Abbreviations

GDPR General Data Protection Regulation

DMA Digital Markets Act

DSA Digital Services Act

GAFAM Google-Apple-Facebook-Amazon-Microsoft

US/USA United States of America

USD US Dollar

CEO Chief Executive Officer

UK United Kingdom

EU European Union

EC/EUC European Commission

ENISA European Union Agency for Cybersecurity

DPA Data Protection Authority

EDPB European Data Protection Board

EUR Euro

KPI Key Indicator of Performance

NGO Non-Governmental Organization

AI Artificial Intelligence

INSEE Institut National de la Statistique et des Étude Économique

VPN Virtual Private Network

WWII 2nd World War

IGPN Inspection Générale de la Police Nationale

GDP Gross Domestic Product

LREM La République en Marche (political party founded by French president Macron)

MIP Most Important Problem

Index

1. Introduction

As Allcott and Gentzkow (2017) remind us, new media technologies have repeatedly affected our democracies in the last centuries: print newspapers in the 19th century, radio and television in the 20th century. As every new technology spreads in society, the advent of each new media industry raised concerns about its impact on democracies with its potential mass influence and manipulation of public opinion. While they have allowed democratization of access to information, new media technologies have always been accused of reducing the specter and quality of political debate. In the early 2000s, the profusion of sources of information made possible by the Internet raised the question of the "echo chambers" and "filter bubbles": the excess diversity of available viewpoints (the premises of a sort of personalization of the delivered message according to the viewer's political behavior) would isolate the citizen from contrary perspectives.

Nowadays, the concern has shifted to social media (Facebook, Twitter, …) as more and more citizens get informed, abandoning the traditional media. Social media platforms have developed a dramatically different structure of information transmission than previous media technologies (ALLCOTT 2017). We are rapidly shifting from a model where one identified - and the official source is used to transmit information to various receptors without reciprocity - to a new model where "content can be relayed among users with no significant third-party filtering, fact-checking, or editorial judgment" (ALLCOTT 2017). In this model, everyone can potentially speak to everyone without any filter: in brief, from a vertical (centralized) to a horizontal (decentralized) model of information transmission. The 2016 US Elections have raised two new specific concerns for the democratic process: (i) the growing amount of "fake news" circulating on social media, and (ii) the (illegal) use of sensitive personal data of citizens

to incentive them to vote for a specific candidate/program thanks to personalized marketing campaigns (Cambridge Analytica case[1]).

The literature review recognizes the positive global impact of these new ways of "direct" and "personalized" communication on the political debate as it can enable a rise in the interest of some parts of the population who had moved away from it (rise in abstention in most of the European countries for instance). These social media platforms allow the political parties and candidates to communicate "directly" with the citizens, discussing topics that interest them, without using traditional media (radio, television, newspapers). To some extent, they allow a "democratization" of the political debate widening the audience. In the same way, they allow the creation of "communities" of citizens who share common points of view and the conversation between them (to some extent, the development of violent and hateful conversations with those who do not belong to the community). However, at the same time, they allow the spread of unverified fake news, with the risk of personalized manipulation. Their algorithms may lock the users in an intellectually comfortable "social bubble" from where it is difficult to escape and to get other points of view. One of the pillars of democracy is the pacific confrontation of ideas; these social media platforms represent both an opportunity and a danger for our modern democracies, even more nowadays with the rise of radical political parties and alternative models (Russia, China, for instance).

This study focuses on the dangers of these social media platforms to democracies and how societies are answering to mitigate them. Based on academic literature, laws, statistics, and anecdotal information on websites, this study performs three recent European cases to

---

[1] In the 2010s, personal data belonging to millions of Facebook users was collected without their consent by British consulting firm Cambridge Analytica, predominantly to be used for political advertising, with supposed impact on the 2016 UK referendum for Brexit, and the 2017 US Presidential elections (see section 3.3)

investigate social media dangers to democracies and study possible solutions. The study was conducted through three academic articles.

Currently, social media regulations have been conducted for their own companies, which is nonsense. Since the new media emerged under the monopolies of billionaire companies, it is important for societies to discuss and find alternative answers to potential threats to democracies.

Besides this introduction, section 2 describes the impacts of the new version of Facebook's algorithm implemented in January 2018 for the users (for archival purposes of the history of Facebook's algorithms evolutions), and explores how this new algorithm accentuated the social bubble effect, weakening the confrontation of the plurality of opinions necessary for the democratic debate. Section 3 assesses how the institutions of the European Union sought to protect the last parliamentary elections in May 2019 against the viral proliferation of fake news and against the unlawful use of citizens' data to target specific groups of strategic voters. Finally, section 4 discusses the psychological and political impacts of the new algorithm introduced by Facebook in January 2018 (which highlights and encourages the information exchanged between users of the same "groups" of people sharing common points -for example, political behavior, hobbies, ethnical or social origins, localization- at the expense of information coming from "official pages" of traditional media or political figures) through the case study of the Yellow Vests' social manifestation on France's streets from November of 2018 to the end of the year 2019: this section studies to what extent the concepts developed in the two previous sections can be applied to the emergence of this movement.

2. The relative impact of Facebook's algorithm on the creation of social bubbles

Abstract: "Filter bubble" and "echo chambers" have recently been accused by many commentators of being responsible for the polarization of public opinion, especially after the surprising 2016 Brexit and US presidential elections results; by isolating citizens in a form of tunnel vision overwhelmed by information reinforcing their own cultural or ideological opinions thanks to personalized algorithms, social networks seem to be a risk for our democracies, based on the plurality of points of views. In January 2018, the tech company Facebook (currently Metaverse) even tented to react and implemented a predictive news feed algorithm that should "bring back people closer together" promoting posts of the users' family and friends first. Thanks to an extensive literature review, we will see in the first part of this paper that the new algorithm of Facebook is responsible for the increase of "social bubbles" through personalized, oriented, and exclusive news feeds; and the possibility to get involved with huge groups of like-minded people. However, the second part of this paper suggests that the "technological" filter (algorithm) is just one part of the intellectual and social confinement problem: we will see that it would be more adequate to speak about a "triple-filter-bubble" model constituted of the individual and social filter as well. We highlight the role of the user's actions and behavior on this social network platform and the cognitive formation of our individual beliefs. Finally, we will see that these "social bubbles" are a risk for our democracies, but that they are inherent to the human psychological condition: in a way, the algorithms of the social networks are just amplifying a phenomenon already existing, by reinforcing our already existing and persistent individual beliefs The real problem seems to be our human capacity to question the veracity of the message we receive.

2.1. Introduction

In 2001, Sunstein (2001) popularized the theory of "echo chamber", an "information environment populated by social recommendations coming from overwhelmingly like-minded others" (Beam, Hutchens, and Hmielowski 2018). According to Bakir and McStay (2018), Pariser (2011) introduced the concept of "filter bubbles". Pariser (2011) describes filter bubbles as echo chambers created by algorithms applied to online content, selecting gauged information a user wants to see, based on his information, connections, browsing history, purchases, and posts and searches. Recently, filter bubbles and echo chambers "have both been linked by commentators to rapid societal changes such as Brexit and the polarization of the US American society in the course of Donald Trump's election campaign" (Geschke, Lorenz, and Holtz 2018).

In January 2018, in reaction to these critics made against the social network, particularly after the "Cambridge Analytica" case, Mark Zuckerberg (Facebook's co-CEO and co-founder) announced the progressive implementation of an innovative news feed ranking algorithm aiming to "make sure the time we all spend on Facebook is time well spent" and to bring "people closer together" (Zuckerberg 2018). According to Zuckerberg, the new algorithm reduces the weight of public content like posts from businesses, brands, and media to improve "meaningful social interactions" with family and friends, and important moments in the world (Zuckerberg 2018). To do so, the social platform would employ artificial intelligence through new algorithms to predict which posts a person might want to interact with friends to show them higher in feed (Mosseri 2018). Based on an extensive literature review, this paper investigates the importance of these new algorithms in the creation of "social bubbles," homogeneous social environments (either on- or off-line) so intellectually close to one's way of thinking that it is tough to escape and to get other points of views for a specific issue.

The first part of this paper reminds the time-line of the important changes in the algorithm of Facebook and explains how, paradoxically, these changes have contributed to the enhancement of the "filter bubble" social confinement effect, despite the unprecedented opportunity open to users and citizens to get a window on a massive plurality of points of views. We will also see that IT programmers are developing new algorithm models to counterpart this direct "filter bubble" effect on the users. However, we will see in the second part of this paper that filter bubbles and echo chambers are only one part of the problem of polarization of the public opinion.

Thanks to social psychology theory and semiotic perspectives of the philosopher Charles Sanders Peirce, all of them revisited by the contemporaneous philosopher Byung-Chul Han, we will balance the weight of the technology (and of social networks' algorithms) in the formation of our individual beliefs and highlight the role of individual users' actions in this process (authority, a priori and scientific methods) (Monteiro Borges and Rampazzo Gambarato 2019): we will study the "triple-filter-bubble" model constituted of individual, social and technological filters (Geschke, Lorenz, and Holtz 2018).

Finally, we will see that these "social bubbles" are a risk for our democracies but that they are inherent to the human psychological condition: in a way, the algorithms of the social networks are just amplifying a phenomenon already existing by reinforcing our already existing and persistent individual beliefs (through personalized, oriented and exclusive news feeds, and the possibility to get involved with huge groups of like-minded people).

2.2. The new versions of the Facebook's algorithm and the social bubble effect

In this section, we study how the new versions of Facebook's algorithm accentuate the social bubble effect, weakening the confrontation of the plurality of opinions necessary for the democratic debate.

2.2.1. Facebook's current panorama

Facebook Inc. is today one of the biggest and most powerful companies in the world: since its creation in 2004 in Cambridge, Massachusetts, US, by Mark Zuckerberg (and four other campus students), the original social network Facebook has experienced tremendous growth in the number of active users (Figure 1). Considering a world population estimation of 7,80bi people (World Population Clock_ 7.8 Billion People (2020) - Worldometer 2020), we can deduce that one-third of the humanity (all ages, countries, social conditions, …) actively connect monthly to this platform. Considering an estimation of 4,57bi current active internet users in the world (Clement 2020), we can deduce that 57% of the internet users in the world actively interact on Facebook monthly.
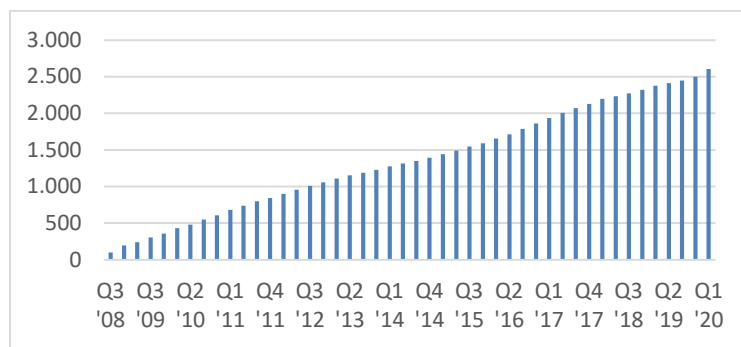
Figure 1 Facebook: number of monthly active users worldwide 2008-2020, number of monthly active Facebook users worldwide as of 1st quarter 2020 (in millions). Source: Statista.com

Facebook Inc. (Metaverse currently) is now a tech giant with a 2019 revenue of more than USD 70bi, net income of more than USD 18bi, around 50k employees worldwide, and a quotation on the US stock exchanges (Facebook, Inc. Class A Common Stock (FB) Financials _ Nasdaq 2019). Facebook Inc. owns various social network platforms, including Facebook,

Instagram, and WhatsApp. In September 2019, the platforms owned by Facebook Inc. represented a market share of 55% of the audience of all the social media monthly used in the US (Figure 2). This study will focus on the proper platform Facebook which still seems to be the most influential in public opinion.

In parallel with the growing influence of the platform Facebook in the public debate over the last two decades, more and more commentators pointed out the risks that this platform posed to our democracies in terms of personal data privacy (for instance, the Cambridge Analytica case in 2016) (Marret 2020), spread of "fake news" (Marret 2020), or polarization of the political debate (see the surprising results of the Brexit poll and the 2016 US presidential elections) (Beam, Hutchens, and Hmielowski 2018). The specialists scrutinize any modification of the Facebook's algorithm because the latter is the "black and opaque box" of the platform, as it selects and prioritizes the information posted on our news feed based on predictions made on our personal data (likes, friendships, views, history searches, etcetera).
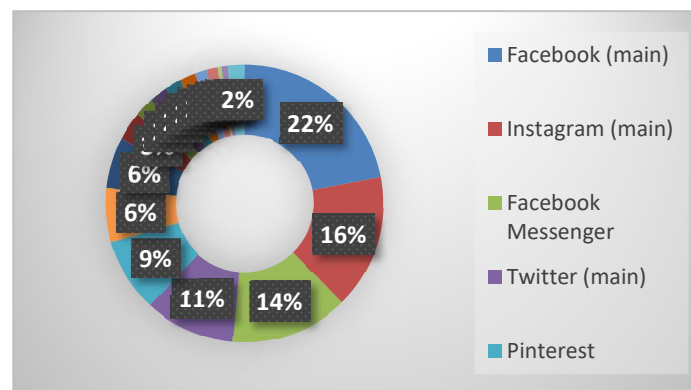
Figure 2. Most popular social media apps in the U.S. 2019 by the audience and monthly users (in millions). Source: Statista.com

2.2.2. A brief history of the Facebook algorithms

Justifying his decision to implement a new algorithm based on "more meaningful social interactions," the CEO and president of Facebook Inc. declared in 2018 that he felt "a responsibility to make sure our services aren't just fun to use, but also good for people's well-being" (Zuckerberg 2018). Like any private company, Facebook Inc.'s goal is to maximize

profits for its shareholders, with the interested contribution of various stakeholders: as we can see in Figure 2.4, the engagement rate (which measures the interactions on the platform such as likes, reactions, messages, and comments) had drastically dropped in the first two months of 2018 vs. 2017. This indicator is very important for Facebook and its advertisers because it is a sign of actual activity (time spent) of the platform's users and incentivizes the user's contacts to discover what sparked their interest (and to spend more time on the platform). More time spent on the platform signifies more advertising seen by the users and more revenue for Facebook.

According to Cooper, 2020), following the company's objective and politics, the algorithm is regularly "enhanced", see Figure 3. The social network was created in 2004, the feed news appeared in 2006 (reverse chronological order), and the bouton "like" was launched in 2007. The first algorithm of prioritization of the news feed seemed to have been implemented in 2009, based on the "popularity" of each post. From 2009 to 2014, many improvements were made by the programmers to turn the algorithm always more efficient and personalized so that Facebook could also personalize the advertising seen. In 2015, Facebook allowed the users to inform the pages they wanted to see in the news feed (and so, Facebook was able to catch many private details on the user's preferences and build a massive personal data bank). In 2016, the algorithm began considering the time spent on each publication to deduce its "value." It gave a better exposure to posts from friends and family members, "informative" or "entertaining" content, and live streaming videos. In 2017, the algorithm began to consider the user's "reaction extensively" (thanks to the bouton "I love," "Grr" for instance) rather than the simple appropriation with the traditional bouton "like." In 2018, Facebook implemented the latest version of its algorithm step by step with a simple objective: to prioritize posts that generate the highest "engagement."
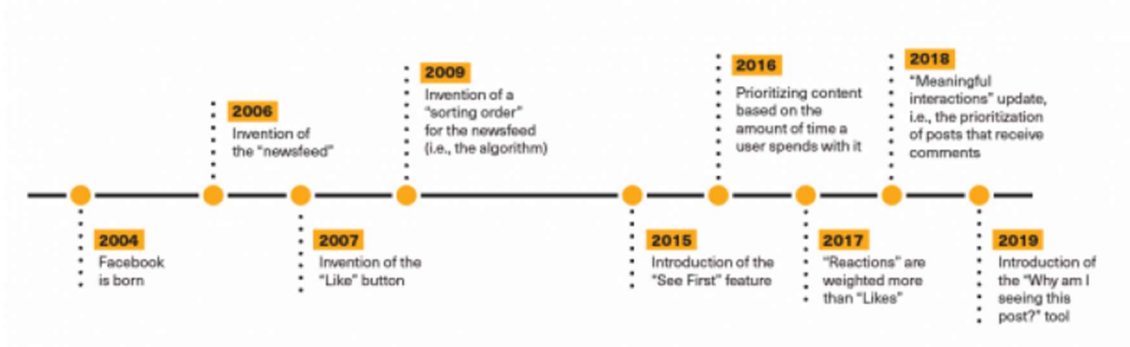
Figure 3. Key moments in the history of the Facebook algorithm. Source: Paige Cooper, Hootsuite

2.2.3 The new 2018 Facebook algorithm: "Bringing people closer together"

In his presentation post of January 2018, Mark Zuckerberg justifies this new algorithm by the necessity "to make sure the time we all spend on Facebook is time well spent" (Zuckerberg 2018). He pretends that "the research shows that when we use social media to connect with people we care about, it can be good for our well-being. We can feel more connected and less lonely, and that correlates with long term measures of happiness and health" (Zuckerberg 2018). He expects that the users "will see less public content like posts from businesses, brands, and media" so that "it should encourage meaningful interactions between people" (Zuckerberg 2018). Facebook's head of newsfeed explains better the new algorithm's philosophy: "we will predict which posts you might want to interact with your friends about, and show these posts higher in feed. These are posts that inspire back-and-forth discussion in the comments and posts that you might want to share and react to – whether that's a post from a friend seeking advice, a friend asking for recommendations for a trip, or a news article or video prompting lots of discussions" (Mosseri 2018). As we can see in Figure 4, the impact of this modification of the algorithm was quantitatively very effective.
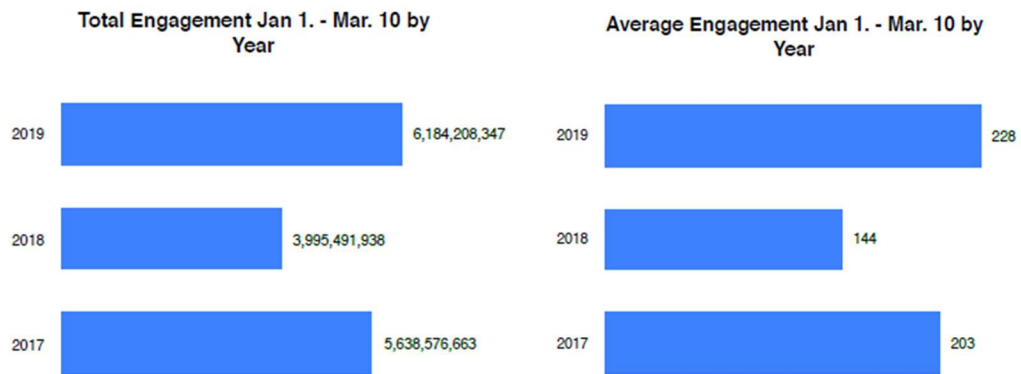
Figure 4. Engagement on Facebook 2017-2019. Source: Newswhip.com, 2019 Guide to Publishing on Facebook

To understand how the new algorithm prioritizes the publications, it is important to understand the notion of "ranking signals": currently, the algorithm ranks the posts that each user sees in the order that is most likely to please him. The rank is based on several factors, known as ranking signals. Ranking signals are data points that inform the behavior of a given user, but also that of all other users of the platform. Examples: are users sharing this post with their friends? How often do you like your boss' posts? Do you often watch live video? What is your favorite Facebook group? In theory, how many posts are you seeing right now? Are they recent? Facebook reports three main categories of ranking signals: (i) the people with whom the user generally interacts, (ii) the media type of the post (video, link, photo, etc.), (iii) the popularity of the publication (Figure 5).
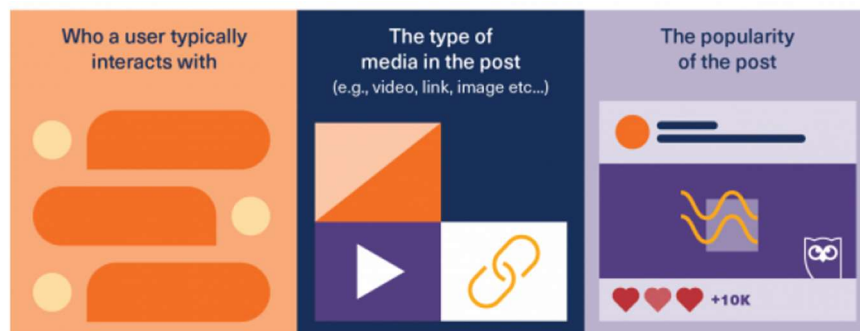
Figure 5. Facebook algorithm ranking signals in 2020. Source: Paige Cooper, Hootsuite

2.2.4. The paradoxical impact of this new 2018 algorithm

We could think like Mark Zuckerberg that this new algorithm would have helped "bringing people together", and reduced the polarization of the political debate, refocusing the platform's conversations on the everyday life of its users and close contacts (Mosseri 2018). Paradoxically, the first studies show that the changes made to the algorithm favor dissensions and scandals, promoting publications that arise the passions of the users. In the US, the Fox News channel which broadcasts controversial opinions has become the leading Facebook press editor in terms of engagement. The new algorithm has pushed up articles on "divisive topics like abortion, religion, and guns", and "the 'angry' reaction dominates many pages with 'Fox News' driving the angriest reactions of anyone, with nearly double that of anyone else" (Owen 2019). Instead of "bringing people closer together", it seems that "it's often an angry, reactive place where people go to get worked up and to get scared" (Owen 2019). Owen cites the example of the two most-shared Facebook stories of 2019 (Figure 6):

PUBLISHER	HEADLINE	FB SHARES
US105FM.COM	Suspected Human Trafficker, Child Predator May Be In Our Area	806,051
LIFENEWS.COM	You Can't Give a Lethal Injection to Murderers in New York, But You Can Give One to an Unborn Baby I LifeNews.com	484,138

Figure 6. Two most-shared Facebook stories of 2019. Source: Newswhip.com, 2019 Guide to Publishing on Facebook

Politics issues are very discussed on the network, as shown in Figure 7. And as Owen notes, "'angry' is the top reaction when it comes to politics content", from all the parts of the political chessboard (Owen 2019).
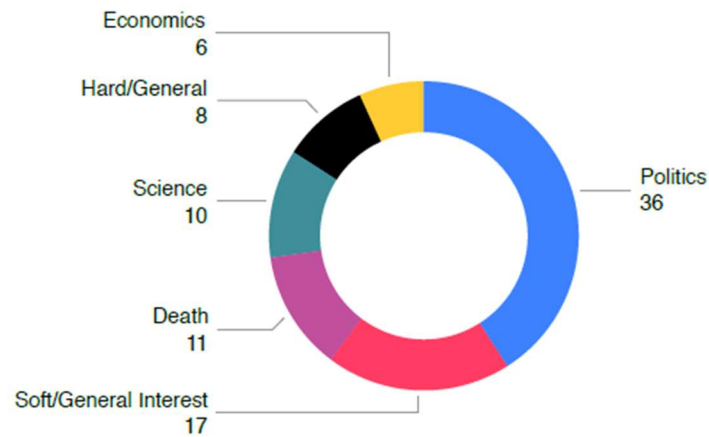
Figure 7. Percentage of top 100 stories of 2019 by genre. Source: Newswhip.com, 2019 Guide to Publishing on Facebook

At the same time, the consequence of this new algorithm is to value the propagation of "fake news", published by unreliable sources who know how to foil the algorithm control system. Owen notes that "PolitiFact rates the "New York abortion up to birth" story -which includes 4 of the 15 stories on the list in one form or another- as False. Henry Winkler is in fact not dead, not 77, and not a "Huge Trump Supporter"; Conservative Tears is a fake news site" (Owen 2019) (Figure 8).

PUBLISHER	HEADLINE	FB COMMENTS
WWW1.CBN.COM	A Sad and Evil Day': New York Legalizes Abortion Up to Baby's Birth Day on Anniversary of Roe v. Wade	1,129,222
TMZ.COM	Beverly Hills, 90210' Star Luke Perry Dead at 52 After 'Massive' Stroke	640,815
CBSNEWS.COM	Trump backs push for Bible classes in schools	621,254
HUFFINGTONPOST. COM	Viral Video Shows MAGA Hat-Wearing Teens Harassing Native American Vietnam Veteran	585,939
DAILYWIRE.COM	HORROR: New York Set To Legalize Abortion Up To Birth	469,924
CONSERVATIVETEA RS.COM	Henry Winkler Dead at 77 – A Huge Trump Supporter and Lifelong Gun Lover	457,825
LIFENEWS.COM	New York Senate Passes Bill Legalizing Abortions Up to Birth I LifeNews.com	396,146
INFORMATIONLIBER ATION.COM	HATE HOAX: Native American Activist Approached Chanting Covington Catholic Teens, Got In Their Face	384,296
NEONNETTLE.COM	Democrats Pass Bill to Fund College for Illegal Immigrants with Taxpayer Money	370,311
LIFENEWS.COM	You Can't Give a Lethal Injection to Murderers in New York, But You Can Give One to an Unborn Baby I LifeNews.com	368,444
FOXNEWS.COM	Dems to strike 'so help me God' from oath taken in front of key House committee, draft shows	368,031
USATODAY.COM	New Jersey to require schools to teach LGBT history	353,499
UNILAD.CO.UK	Woman Smears Menstrual Blood On Face To Show Periods Are 'Beautiful and Powerful'	344,752
SCARYMOMMY.COM	Gillette's New Ad On Toxic Masculinity Is Breathtaking — And Necessary	341,080
MIRROR.CO.UK	Momo challenge is 'hacking Peppa Pig, Fortnite and YouTube Kids' warns school	328,686

Figure 8. Most commented Web content in 2019. Source: Newswhip.com, 2019 Guide to Publishing on Facebook

Facebook is so aware of the bias that the company declared in 2020 that one its priorities would be to help the users understand the algorithm and to be able to personalize some of its functionalities (Cooper 2020). For instance, the company introduced a function in March 2019 allowing the user to better understand why a particular message has been posted on the news feed: this function allows the user to inform the algorithm that he/she finds a publication important (want to see more of the same kind), or on the contrary useless or irritating (Cooper 2020).

2.2.5. Polarization in social networks: from echo chambers to filter bubbles

Since its release in the early 2000s, Facebook (and generally speaking the social network's market) has been "linked to an increased diversity of information and perspectives for users" (Chitra and Musco, 2019). Paradoxically, recent studies show that the same social networks have contributed to a severe rise in the user's polarization on many issues. As these authors note, "a popular theory for this phenomenon centers on the concept of 'filter bubbles': by automatically recommending content that a user is likely to agree with, social network algorithms create echo chambers of similarly-minded users that would not have arisen otherwise" (Chitra and Musco 2019). Developing a "mathematical framework" the authors show the emergence of filter bubbles in real-world networks and provide theoretical support for "why social networks are so vulnerable to outside actors" (Chitra and Musco 2019). Outside actors here would be the administrator of the social network, i.e., the programmer of the algorithm.

The tech activist Eli Pariser introduced the theory of the "filter bubble" in 2011. According to Pariser (2011), the social network platforms use one of our human beings psychological biases: our wish to validate our own beliefs by receiving information confirming it: "preferential attention to viewpoints similar to those already held by an individual is explicitly encouraged by social media companies: to increase metrics like engagement and ad revenue, recommendation systems tend to connect users with information already similar to their current beliefs" (Chitra and Musco 2019). Those recommendations can be direct (friends or follow suggestions) or more subtle: the algorithm's artificial intelligence can now predict which publication each user is statistically most likely to interact with and post it at the top of the news feed. "By recommending such content, social network companies create 'echo chambers' of similar-minded users. Owing to their root cause -the external filtering of content shown to a user- Pariser called these echo chambers filter bubbles" (Chitra and Musco 2019).

The concept of "echo chamber" was developed by Cass Sunstein in 2001, before the rise of the social networks and was much more linked to the "social bubbles" created in the "real world" (Sunstein 2001). With his theory of "filter bubble," Eli Pariser updated the same concept applied to an automatized environment with a powerful algorithm able to filter and organize the information each user will get, most of the time validating their own beliefs and making possible the virtual meeting of a massive number of people sharing the same points of views (Pariser 2011). Using the examples of the Brexit debate and the 2016 US presidential elections campaign, Chitra and Musco note that "in each of these incidents, instead of bringing diverse groups of users together, social media has reinforced differences between groups and wedged them apart" (Chitra and Musco 2019).

2.3. From the online to the off-line real world: an opportunity as well as a risk for democracies

The open political debate and the confrontation of ideas and points of view are inherent to democracies. We cannot deny that social networks encourage the public debate, and personal positions, and in a way, being more accessible, allows the participation of citizens who had felt excluded from this debate for economic, social, and cultural reasons (Mcclurg,2003).

As Seargeant and Tagg (2019) note, "in the days before algorithmic personalization became commonplace". Sunstein (2007) argued that online communities resulted in people cutting themselves off from opinions and information that challenged their belief systems and that this was likely to harm democratic debate. However, the development of algorithms has led to a new situation in which people's actions are increasingly shaped by processes hidden from most users. Although this applies to all information that is shared on Facebook, it also includes opinions and expressions about social or political values as well as news stories, which,

so the argument goes, results in a newsfeed filled predominantly with opinions with which the user agrees – a phenomenon which Jones and Hafner (2012) refer to as the 'ghettoization' of the internet. The significance of this, according to Pariser (2011), is that 'democracy requires a reliance on shared facts: instead, we're being offered parallel but separate universes. Pariser's warning relates to the way that civic debate is not best served by intellectual segregation, and leads more readily to extremism than to consensus. Research shows that when people discuss issues with those who share their opinion, this leads to more polarized attitudes towards the topic (Stinchcombe, 2010), whereas exposure to diversity increases people's tolerance for those with different or opposing views (Garrett and Resnick, 2011) (Seargeant and Tagg 2019).

2.3.1. A Swiss experimentation on the 2017 French presidential elections

In the run-up to the 2017 French presidential elections, Nathalie Pignard-Cheynel (2017) and co-author organized an experiment with students to understand how Facebook's news feed algorithm works and its impact on the feed information feed on the platform. During six weeks, the students were separated into eight groups which each embodied a profile, through a "fake" account (temporary), corresponding to an orientation close to a candidate of the campaign (from all the political spectrum). Each account followed the same list of fifty media (of all kinds), was a "friend" of all the other groups (but was prohibited any form of interaction), and had to interact every day with publications corresponding to its political behavior (through likes, reactions, comments, shares, …). The main results were: (i) publications of groups and friends had more exposition than the media; (ii) the other control group which supported the same candidate had also much more exposition than the other groups; (iii) the political behavior of each media was much more important than the content of the publication; (iv) the selection is severe and concentrates on ten out of the fifty media for each account; (v) the algorithm was not very "subtle" in the sense that it considered only the engagement (the fact that the account

reacted -in a good or bad way-) to increase the publications associated on the news feed; (vi) it was very easy to create a fake account and to broadcast fake news (the commentary -based on fake information- of one of account had been liked more than 1300 times. This experience, despite the fact that it occurred in 2017, i.e., before the implementation of the latest news feed ranking algorithm, demonstrates the effect of the "filter bubble" in practice, and the risk of ideological confinement run by citizens, in particular during the electoral periods.

2.3.2 New algorithm models to counterpart the effect of the "filter bubble"

The literature review showed that many studies try to propose technological solutions to "break free of the filter bubble" (Milan and Agosti, 2019). For instance, one experiment "propose[s] a new strategy for improving the socially relevant properties of a recommender system: adding antidote data." (Rastegarpanah, Gummadi, and Crovella 2019); another project "aim[s] to train different chat-bots to become highly opinionated on specific topics (e.g., pro/against gun laws) and provide the opportunity to discuss political views with a highly biased opponent" (Dingler, Choudhury, and Kostakos 2018) to "trick" the Facebook's news feed ranking algorithm. Other researchers ask the Facebook users to "send" their profile data for analysis so that they can work on a massive amount of cases to better understand the logic of the current algorithm and propose useful enhancements (Milan and Agosti, 2019).

Indeed, it may seem attractive to "fight fire with fire" to invent technological responses or fixes to a technological issue that impacts real-world democracies. However, it is here important to point out that "other studies have suggested a different phenomenon, whereby the extensiveness of online networks means that a small but significant fraction of ties is with people with different political outlooks, which increases exposure to different opinions" (Sharad et al., 2010) (Seargeant and Tagg 2019). Axel Bruns notes that "there is scant empirical

evidence for their existence, or for the related concept of 'echo chambers': indeed, search and social media users generally appear to encounter a highly centrist media diet that is, if anything, more diverse than that of non-users. However, the continued use of these concepts in mainstream media and political debates has created a discursive reality that continues to impact materially on societal institutions, media and communication platforms, and ordinary users themselves (Bruns 2019a). Media, citizens, and researchers can blame an algorithm, but "filter bubbles" seem to be just one side of a much wider problem: the construction of individual beliefs and the user's actions on social networks. It seems important now to redirect scholarly attention to "far more critical areas of inquiry" (Bruns 2019a).

2.3.3. The construction of individual beliefs based on partial information and its impact on our "social bubbles"

The question is not to minimize the impact of the platform's algorithmic "filter bubbles" but to replace this issue with a much broader perspective that is not intellectually comfortable for a citizen. The literature review shows that "users' actions also play a key role in how the site operates as a forum for debate" (Seargeant and Tagg, 2019). "The strategies people use to navigate the complex social space contribute to the polarizing of debate, as they seek to avoid conflict with the diverse members of their network" (Seargeant and Tagg 2019). The same study points out the paradox that even though platform's users are pushed towards "ideological segregation" in terms of the information they are exposed to -apparently- by the algorithms, this is not necessary because they do not have contact with people holding divergent points of views, but more probably because of their actions, behavior, and decisions on the social network. As Bruns argues, "echo chambers and filter bubbles principally constitute an unfounded moral panic that presents a convenient technological scapegoat (search and social platforms and their affordances and algorithms) for a much more critical problem: growing

social and political polarization. But this is a problem that has fundamentally social and societal causes, and therefore cannot be solved by technological means alone" (Bruns 2019b). Provocatively, it seems that our shared strong and persistent belief about the impact of the "filter bubbles" of the social networks' platforms, is in itself the fruit of a conformist self-persuasion, due to our "social bubble" where this theory is largely dominant.

2.3.3.1 The role of the user's decisions on the platforms

2015 research concluded that "Compared with algorithmic ranking, individuals' choices played a stronger role in limiting exposure to cross-cutting content […] Of course, we do not pass judgment on the normative value of cross-cutting exposure. Although normative scholars often argue that exposure to a diverse "marketplace of ideas" is key to a healthy democracy, a number of studies have found that exposure to cross-cutting viewpoints is associated with lower levels of political participation. Regardless, our work suggests that the power to expose oneself to perspectives from the other side in social media lies first and foremost with individuals" (Bakshy, Messing, and Adamic 2015). "As evidenced by Bakshy et al. (2015), even though algorithms play a crucial role in manipulating the content prioritized in Facebook's news feed and potentially influence the circulation and spread of fake news reports, the online behaviors and beliefs of the users themselves directly influence the way these algorithms perform" (Monteiro Borges and Rampazzo Gambarato 2019).

2.3.3.2. A new model with three filters

The most innovative concept of this literature review is developed in a study by Daniel Geschke, Jan Lorenz and Peter Holtz Daniel (2018), the "triple-filter-bubble model". The research refers to filters "in a very general way as processes that lead to a limitation of information that is available to individuals" (Geschke, Lorenz, and Holtz 2018) and shows that the technological filter, i.e. the algorithms in our case, is just one part of the problematic of the

construction of our individual beliefs based on partial information. Their model takes into account 3 types of limitations, of filters: (i) the individual filters, (ii) the social filters and (iii) the technological filters.

2.3.3.2.1 The individual filters - Ipsedixitism

This first group of filters comes under "cognitive motivational processes" and "has been studied extensively in cognitive and social psychology [;;;]: individuals are to different extents cognitively motivated to search for and add fitting bits of information and to ignore or deny conflicting ones" (Geschke, Lorenz, and Holtz 2018). This "confirmation bias" has been studied by Nickerson (1998) with the theory of the "pre-existing attitudes", by Festinger (1957) with the concept of avoidance of cognitive dissonance, and again by Brewer (1991) with the theory of social identity boost. "In all these cases, filtering refers to selective exposure due to an individual's information search, processing, and memory" (Geschke, Lorenz, and Holtz 2018). Under this bias, we apply Ipsedixitism (from Latin ipse dixit -he himself said it-) in the construction of our beliefs: we no longer exercise our critical mind when the advanced thesis, without being argued, pleases us more than its opposite. And the consequences can be very harmful today with the dissemination power of social networks (hoax, virus, fake news…).

Simulations ran by Geschke, Lorenz and Holtz (2018) show that, "even without any social or technological filters, echo chambers emerge as a consequence of cognitive mechanisms, such as confirmation bias, under conditions of central information propagation through channels reaching a large part of the population […] What is important is that, because of their intradiverse networks, users are coming across divergent opinions and alternative viewpoints, as well as information they deem to be inaccurate or false, despite the Facebook personalization algorithm working to feed them posts which are likely to appeal most to them – and which in media commentary is often interpreted as being those which consist of news stories, opinions and values with which they are likely to agree" (Geschke, Lorenz, and Holtz

2018). To resume: "we may believe what we are inclined to believe. Stories that conform to what we already believe are taken as true, regardless of whether they correspond to experience or not. They eliminate the uncertainty of the uncomfortable state of doubt. This is very similar to what happens with filter bubbles" (Pariser, 2011) and echo chambers on social media in general and on Facebook in particular" ,(Monteiro Borges and Rampazzo Gambarato 2019). Applied to the phenomenon of disinformation, Borges and Gambarato (2019) note that "even though […] there is a manipulative role played by the configuration of algorithms on Facebook, our own cognitive preferences lead us to develop contexts in which filter bubbles and echo chambers are stimulated, creating a fertile terrain for fake news to grow" (Monteiro Borges and Rampazzo Gambarato 2019).

The South Korean-born German philosopher Byung-Chul Han stated that "global communication and likes only tolerate more of the same; the same does not hurt!" (Geli 2018). "Byung-Chul Han's perspective coincides with the XIXth century's Charles Sanders Peirce's a priori method of fixing beliefs, in which beliefs that are agreeable to reason and eliminate doubt are fixed, so they do not "hurt." Stories that conform to what we already believe eliminate the discomfort, the irritation, of doubt. Thus, our own cognitive preferences expressed in digital environments offer a fertile terrain for fake news to grow" (Monteiro Borges and Rampazzo Gambarato 2019).

2.3.3.2.2 The social filters – Conformism, homophily

Human beings have a persistent tendency to form social links preferably with people with whom they share "sociodemographic, behavioral, and intrapersonal characteristics" (McPherson, Smith-Lovin, and Cook 2001): what is called "homophily" in Social Sciences. On the social networks, this theory means that users make virtual "friendships" with like-minded people, and create groups of shared interests. Furthermore, they frequently go away of groups or terminate virtual "friendships" when views are too different on conflictual topics (John and

Dvir-Gvirsman 2015). Hence, homogeneous network structures can potentially limit the width of information a social media user is exposed" (Geschke, Lorenz, and Holtz 2018).

When it comes to users holding conservative or extremist views, this tendency for homophily appears to be even stronger (Boutyline and Willer, 2017), and so is the risk for the emergence of group polarization among them (Geschke, Lorenz, and Holtz 2018).

In every social group exists what is called "conformism": when the judgment of individual conflicts with that of the group, they tend to conform their judgment to that of the group. Furthermore, their judgment consistent with that of the group tends to "internalize," i.e., the individual will persist in their misjudgment even in the absence of the group. This is the result of the famous Asch experience in the 1950s. The main objective of conformist comportment is to escape the discomfort of the group's judgment. In these social bubbles (Facebook groups), it is difficult to convey a discordant idea under the social pressure of the group and the fear of being rejected. And this is why in these groups, conformism can lead to extremism, in a kind of emulation of the most extremist ideas, without social fear of being judged (because every member "speaks" the same language): "in other words, the media is mediated by the friendships you have with the people in your network" (Seargeant and Tagg 2019).

2.3.3.2.3 The technological filters – Algorithms

The last group of filters is the algorithms that are operated by the platforms to deliver individually select media offers to each and every user all over the world, and at any time (Pariser 2011): this selection and organization of the published information are based on the "individual users' assumed wants and needs" (Geschke, Lorenz, and Holtz 2018; Pariser 2011). The ultimate objective of this algorithm is to maximize the time spent by every user on the platform, in order to deliver more well-paid personalized advertising. For instance, none of the users gets the same ranked results on a Google Search, because this information is processed

and organized by the algorithm based on your past behaviors, preferences, searches, etc… on the Google Chrome navigator. According to the literature review, it is difficult to conclude whether the algorithm restricts the user's sources of information or on the contrary, increases its spectrum as showed by Herlocker et al. (2004), etc…: according to their study, the platforms would "constantly confront the user with novel not yet consumed information to maximize click-through rates, thereby potentially increasing the exposure to different points of views" (Herlocker et al. 2004). "Therefore, an alternative assumption is that, in spite of the filtering processes […], online media increase the spectrum of attitudes that are cognitively available and salient in individuals" (Monteiro Borges and Rampazzo Gambarato 2019).

Monteiro Borges and Rampazzo Gambarato (2019) remind us that already in the XIXth century, the American philosopher, logician, mathematician and scientist Charles Sanders Peirce (the "father" of the pragmatic school) had presented "four methods for the fixation of beliefs: (i) the method of tenacity, (ii) the method of authority, (iii) the a priori method, and (iv) the scientific method. In the method of tenacity, there is no room for raising doubts: once a belief has been established, people will constantly reiterate it for themselves, and no contrary opinion posed by others will affect their confidence in it. The same situation occurs in polarized discussion on Facebook. The method of authority is more common in groups, and institutions where there is a strong leader with a specific dogma which it is necessary to perpetuate and to prohibits other ideas to "contaminate" the community. It is often the case in religious communities, and in closed groups where the members are "discouraged from thinking for themselves and questioning the beliefs propagated by the institutions. On Facebook, this kind of belief is frequently found in interactions related to religious debates and political viewpoints (Franco, 2017). Another pertinent example is the infamous case when President Trump dismissed "CNN as 'fake' and Fox News as 'real' expressing the correlation of his political interests with the interpretation of the facts represented by media outlets" (Monteiro Borges

and Rampazzo Gambarato 2019). The third method is the a priori method (already studied under the name of "confirmation bias"), and which fixes the beliefs that we already believe in: "consequently, Facebook users get a false impression of overall agreement with their own beliefs because of the lack of exposure to conflicting opinions. Beliefs provide peace of mind and satisfaction, which we do not want to lose" (Monteiro Borges and Rampazzo Gambarato 2019). The last method presented by Peirce is the scientific method: according to the latter, to be considered as truth, our belief must not be formed by an "individual opinion", but on the contrary by "something external that is not influenced by our own thoughts, namely, external events. A common belief based only on external events and shared by everyone is what Peirce calls the 'ultimate conclusion'" (Monteiro Borges and Rampazzo Gambarato 2019). It is impossible to reach this theorical opinion shared by everyone. In a way, the online participative encyclopedia Wikipedia could be considered like a first attempt to reach this common agreement (still with lots of mistakes, but which seem to be corrected by the "wiki"-community -another bias?- with the time being), with the help of some algorithms, but always with the human confirmation of the users' community.

2.3.4. Proximity and emotions favor the dissemination of fake news

As a consequence of these observations, Schkade et al. showed in 2007 that proximity (whether being geographic, social, etc…) may also contribute to the dissemination of fake news because users trust more (bias of confirmation) those who transferred the information to them (contacts, friends, family, colleagues, …), even if the information is false. This tendency may be amplified by the implementation of the new 2018 algorithm which prioritizes the publications of the "close" ones, leading to the creation of "opinion-ghettos" (Seargeant and Tagg 2019).

The literature review also showed the impact of (mostly negative) emotions on the dissemination of fake news: "the analysis explores the way that communication […] on Facebook apparently gives rise to recurrent examples of conflict, disagreement, or a sense of frustration with other interactants, which, we argue, is in part a result of the specific form of diversity which exists on the site. We refer to this as intradiversity and suggest that it results from the type of 'ego-centered' network" (Seargeant and Tagg 2019). As El-Sharawy showed, "Facebook favors emotional content that hits people whether or not it is true" (Bakir and McStay 2018). And this is one of the reasons why polarizing publications and fake news have more expositions and sharing on the platform. A recent quantitative study by Vosoughi, Roy, and Aral (2018) on spreading accurate and false information online demonstrate that false news spreads faster and reaches more people than actual news. They suggest that the degree of novelty and the emotional reactions of recipients may be the reasons behind this discrepancy. Hence, whether the content on Facebook is true or false seems less relevant than whether it reinforces our soothing, satisfactory, and comfortable fixed beliefs (Monteiro Borges and Rampazzo Gambarato 2019).

2.4. Conclusion

The literature review showed that the "social bubbles" are a risk for democracies and that Facebook's news feed ranking algorithm, for promoting personalized publications that are tailored to the predicted needs and wishes of each user, impacts the creation of these "opinion-ghettos". However, it would be exaggerated to blame only the platform's algorithm which can as well publish other points of view to its users to maximize its revenues. It seems useless to try to develop technological tactics to cheat the algorithm for a simple reason: the users are also responsible for the creation of these "social bubbles", given their behavior on the social

network: sensible to emotions, homophily, conformism, ipsedixitism, … As Daniel Geschke et al. (2018) notices, the trend to modern technology and social networks cannot be stopped and the platform will continue to invest in more and more sophisticated algorithms in order to deliver the most personalized experience to every user, to maximize their own revenues. "They increase individual attitudinal stability, and, thus, individual certainty and security. On a societal level, however, these processes are prone to increase attitudinal differences between opinion groups and individuals and to cut communication ties between them, leading to attitude clusters, societal fragmentation, and polarization […] So the digital world presents a genuine dilemma, where positive individual effects go along with negative societal effects" (Geschke, Lorenz, and Holtz 2018). Geschke et al. (2018) propose to work on the three levels of filters that were studied: (i) on an individual level, it seems important to increase the media knowledge of the users (how the filters work, which impacts, …); (ii) on a social level, they highlight the necessity to create new forms of alternative debates where the best consensus would be sought; (iii) they hope that a part of serendipity will soon be introduced again in the algorithms (Geschke, Lorenz, and Holtz 2018). As already developed in another article (Marret 2020), I advocate for the creation of a real digital ecosystem with the State (and its legislative control power), with the responsibility of the social networks platforms (which are interested in a long-term business model, and which should collaborate more with the searching teams), and, more importantly, with the conscient control of the citizens over the State actions and the platforms' engagements. The citizen is responsible for the quality of democracy, which reflects its own social organization and advancement. Social bubbles are inherent to the human psychological condition: in a way, the algorithms are just amplifying an already existing and persistent human tendency by reinforcing our individual beliefs (through personalized, oriented and exclusive news feeds and the possibility to get socially involved with huge groups of like-minded people).

3. The European regulation designed to protect the 2019 elections from disinformation and micro-targeting advertising

Abstract: In 2018, the Facebook/Cambridge Analytica case raised severe concerns about using personal data to manipulate the 2016 US presidential and the 2017 UK general elections. As a result, the European Union reacted fast to a general data protection regulation (GDPR). This study focuses on the GDPR mechanisms to avoid fake news and microtargeting effects on the electoral context. Preliminary results suggest that GDPR reduced social media's effects on the 2019 European Parliament elections by adopting a public and private co-regulation approach, which avoids counterproductive reinforcement of the auto-persuasion power of fake news. Finally, the article discusses the difficulties of applying the European data protection law in a new digital era to conclude that it is essential to continue developing other measures.

3.1. Introduction

The alleged unlawful processing of personal user data acquired from Facebook by the company Cambridge Analytica raised severe concerns about the impact of data protection infringements on the 2016 US presidential and the 2017 UK general elections. Potemkina observes that the European political leaders' and citizens' perceptions that social networks could also be harmful to democratic elections made the European Union (EU) rapidly react to it, adapting their legislation to avoid the new threat (Potemkina, 2019, p.3 e p.11). The European General Data Protection Regulation (GDPR), ePrivacy Directive, and Regulation project, amendments of key regulations on political parties financing, and framework of actors' responsibilities are among the most important decisions to sort out the potential problem.

We study how the 2018 European GDPR was designed to reduce the electoral risk of manipulation in the elections of the European Parliament in May 2019 by social media platforms. We follow Alemanno to focus on two potentially dangerous mechanisms of social media platforms affecting the results of elections; the viral proliferation of fake news (disinformation) and the unlawful use of citizens' data to target specific groups of strategic voters (micro-targeting and profiling) (Alemanno, 2018, p.1). Based on the literature and official reports review, we find three normative solutions to combat disinformation: co-regulation, legislative, and coercive approaches (Alemanno, 2018, p.1; Bode and Vraga, 2015, p. 619; Clayton et al., 2019, p.1; European Commission, 2018d, p.1; European Parliament 2019, L85 I/7; Haciyakupoglu et al., 2018, p.2; Lewandowsky et al., 2012, p.106; Mena, 2019, p.1; Nyhan and Reifler, 2010, p.303; Pennycook et al., 2019, p.1).

The EU chose the co-regulation system, which is understood as the best way to regulate, given its flexibility. Still, it needs to be continuously well supervised (detecting and swamping fake news with other sources of information). Furthermore, the legislative and coercive approaches tend to be counterproductive since they could reinforce the auto-persuasion power of fake news, as pointed by literature.

We also discuss the difficulties of enforcing the European regulation to a new digital economy in future elections since the "pluralist model of speech regulation" is effective only with the goodwill of publicly identified actors. In particular, it is necessary to focus on the new threat of the "satellite" digital campaigns organized by undefined actors and, therefore, apparently impossible to be regulated with the current EU model. Thus, it is essential to continue developing other measures to combat the disinformation phenomenon, such as fact-checking education at school and better collaboration between public authorities, the digital industry, and society.

Besides this introduction, section 2 contextualizes and discusses the different approaches to regulating fake news on social networks. Section 3 presents the chronology of the EU response and the challenging political microtargeting strategies used on those social networks. Section 4 discusses the EU legislative data protection framework in elections. Section 5 presents the EU regulatory approach for the 2019 elections and results. Section 6 discusses the limits of implementing regulatory approaches to a digital environment. Finally, the last section brings the main findings.

3.2. Social media, fake news, and existing regulatory responses

The technology evolution of digitalization has disrupted the media industry (Waldfogel, 2017, p.195). In this sense, social media platforms, where any speaker is virtually able to reach a vast audience in seconds, have posed a new communication pattern to individuals worldwide. Unlike the centralized communication of traditional media, social media offers a decentralized way of communication. Social media also provides the service of classifying individuals according to their preferences based on their data on the platform. Thus, a speaker can buy customized profiles and disseminate information uncheckable or committed with any moral code of conduct or another pattern of quality (Allcott; Gentzkow, 2017, p.211). Furthermore, social media allows and incentivizes massive dissemination of information - among them "fake news"- due to the internet information reproduction at low cost (virtually zero), the artificial intelligence (AI) agents that automate the work of human propagators (Haciyakupoglu et al., 2018, p.2), the anonymization of the authors and sources of financial funding of specific messages, and the creation of personalized posts ("dark ads") linked to the predicted behavior of each user by the algorithms and based on its collected personal data ("profiling" method).

The disinformation problem is not a new phenomenon; before the advent of the Internet, it was seen as propaganda in which the mass media had been a vehicle that both state and non-state actors exploited to push messages that distort the opinions and emotions of people largely for the promotion of particular political agenda or ideology (Posetti and Matthews, 2018, p.1-2). However, the huge capacity of social media to disseminate disinformation very fast to a large population has hugely escalated the problem.

A first difficulty in addressing the disinformation problem is to define "fake news". The European Commission has adopted "fake news" as ""verifiably false or misleading information that is created, presented, and disseminated for economic gain or to intentionally deceive the public, and may cause public harm"", ""disinformation does not include reporting errors, satire and parody, or clearly identified partisan news and commentary"" (European Commission, 2018, p.3-4). The BBC media group uses the definition "'false information deliberately circulated by hoax news sites to misinform, usually for political or commercial purposes.'" BBC also distinguishes it from false news (UK Parliament and BBC, 2017, p.1), while The Guardian suggests the definition of ""fictions deliberately fabricated and presented as non-fiction with intent to mislead recipients into treating fiction as fact or into doubting verifiable fact"' (UK Parliament and The Guardian, 2017, p.3). Allcott and Gentzkow define "fake news" as ""news articles that are intentionally and verifiably false, and could mislead readers"" (Allcott; Gentzkow, 2017, p.213). Nyhan, however, brings the concept of "misperceptions" when a speaker sincerely believes in some event and does not require his knowledge of a speaker's intent in making claims, which can rarely be established with certainty, and bring additional difficulties to any process of regulation (Nyhan; Reifler, 2010, p.303)

Even social media platforms have their definition of fake news. Since the 2020 US elections, they (Facebook, Twitter, among others) started to "regulate" the content published on their platforms. For instance, Facebook defines "fake news" as ""a catch-all to refer to

everything from news articles that are factually incorrect to opinion pieces, parodies and sarcasm, hoaxes, rumors, memes, online abuse, and factual misstatements by public figures that are reported in otherwise accurate news pieces"" (Weedon; Nuland; Stamos, 2017, p.4). Weedon et al. observe, however, that Facebook uses the term "catch-all" to minimize the scope of the disinformation concept (Weedon; Nuland; Stamos, 2017, p.4).

Once an adequate definition for "fake news" is established, it comes the need for how a regulatory system could deal with the disinformation. The problem is more challenging in an environment of digitalization with fast technological transformation, where regulatory apparatus always is delayed. We follow Koop and Lodge to think of regulation as an interdisciplinary concept but keep in mind the economic regulation in a broad sense to keep markets having good operation conditions for all individuals (Koop, Lodge, 2015, p.1-11)

Alemanno points three solutions to combat disinformation on the current social network platforms (Alemanno, 2018, p.2-4). The first is a '*state intervention*' since public authorities are expected to police the social media environment, where the main risk is creating a "Ministries of Truth" to "model" the citizens' point of view under an authoritarian power. The second solution is the '*accountability of the platforms for all the editorial contents* of what is published by third parties on their networks, which imposes penalties to entities that engage in content creation and circulation of "illegal content" through laws, regulations, or directives. The author classifies this approach as coercive since the legislator and the court can decide what is fake news or outsource this responsibility to social media. The author also points to the weakness of a platform "pay-as-you-go business model", reimbursed by "clicks", in playing the role of arbiter's truth. Both previous approaches are flawed by the delay in the judgment after fake news stories get denounced as potentially false since it lets time to make the story viral. Finally, the third approach consists of '*"swamping" the fake news*' with various news from other sources and different points of view on the user's page. The idea is to cancel the

"bubble filters" effect that encloses the user in a flow of information with similar points of view providing him/her with more context and alternative views.

Facebook has been experimenting with the third method voluntarily since 2017 with its product "Related Articles". The company's ambition is to provide ""easier access to additional perspectives and information, including articles by third-party fact-checkers"" (Facebook and Su, 2017, p.1). However, Bode and Vraga observe that this method still leaves open the deeper problem of algorithm fact-checkers quality (even when done through artificial intelligence): by whom and how will be evaluated news considered as fake (Bode; Vraga, 2015, p.631)? Mena and Clayton show that flagging false news on social media platforms may help the current efforts to combat sharing of deceiving information on social media (Mena, 2019, p.1) (Clayton et al., 2019, p.1). On the other hand, Lewandowsky et al. and Nyhan and Reifler observe that flagging fake news may reinforce false beliefs since labels as flagging may help people distinguish between true and false news stories, not knowing what is true or false (Lewandowsky et al., 2012, p.106) (Nyhan; Reifler, 2010, p.303). Pennycook et al. argue in the same direction, warning that labels may lead to a modest reduction in perceived accuracy of false news stories but also result in unlabeled false reports being seen as accurate (or "implied truth" effect) (Pennycook et al., 2019, p.1). Also, a maximization profits company regulated by itself represents the opposite incentives desired by regulatory institutions, since any good practice is not guaranteed if it reduces profits (Viscusi et al., 2005, p.358).

The European Union Commission adopted a co-regulatory approach, which combines state intervention and accountability of the platforms to deal with data protection. They avoided the *"swamping"* type of regulation given the controversy of its results shown in the literature.

3.3. Chronology, Regulation, and micro-targeting mechanism

We follow Potemkina to present a brief chronology of the EU legislative response to social media events (Potemkina, 2019, p.2.). On the 19th of March, 2018, the New York Times and The Observer published articles about the leak affair, or the Cambridge Analytica using private data of Facebook to sell access to profiles that could be politically influenced to people connect to Donald Trump previous the U.S. 2016 presidential election.

Facebook admitted two months later that (without admitting the information was sensitive) personal data of 87 million users had been transmitted to Cambridge Analytica through an application, including 2,7 million Europeans (official figures from Facebook, still contested by Cambridge Analytica). Thus, Cambridge Analytica helped the Republican Party send micro-targeted advertisements to "strategic" or specific profiles of American citizens to influence their votes. However, according to Levy, based on an experiment on Facebook users, social media algorithms did not change political preferences but may limit exposure to counter-attitudinal news and thus increase polarization (Levy, 2021, p.831). In this sense, the increased polarization can incentivize people to vote, remembering that the vote is not mandatory in the U.S.

Shortly after the leak, the European Commission (EC) protested and asked for a public hearing of the Facebook and Cambridge Analytica board members, including Mark Zuckerberg, and an investigation based on the information leak of the 2,7 million European users. The EC had kept close contact with the US Federal Trade Commission since disclosing data leaks. Mark Zuckerberg apologized in front of the US Congress on April 11, 2018. In the EU Parliament on May 23, 2018, Facebook agreed to collaborate for better supervision during the European elections of May 2019. On April 26, 2018, the EC published the official communication "Tackling online disinformation: a European Approach," which is the guideline of the EU

politics on the issue until now. On May 25, 2018, the general data protection regulation, GDPR, entered into force (it had signed on April 14, 2016). On June 25, 2018, the European Parliament adopted a resolution to force Facebook to comply immediately with the EU regulation, mainly on personal data (GDPR) and communication (e-Privacy Directive), and asked the European Union Agency for Cybersecurity (ENISA) to conduct an audit on the situation. On September 12, 2018, in his annual "State of the Union Speech" on the European Parliament, the President of the EUC Jean-Claude Juncker urged the European institutions to take actions for "fair elections" and announced his intention to undertake several measures to counter manipulation during the election campaign (Juncker, 2018, p.1). In September 2018, under the pressure of the EU Commission, Google, Facebook, Twitter, Mozilla, and several other private companies signed the "EU Code of Practice on Disinformation." The EC proposed a "security package" to the European Council in Salzburg on September 19 and 20, 2018, the "Commission guidance on applying Union data protection law in the electoral context." This package guides applying the GDPR in the electoral context and a list of legislative measures to adopt/amend the European laws to the new electoral context (European Commission 2018a). As an example, to amend the 2014 law regulation on funding of European political parties and foundations to allow financial sanctions in case of infringement to the new legislation about disinformation and personal data protection.

3.3.1. The reactive response of the EU to the microtargeting approach

Barbu (2014) defines microtargeting as a sampling process based on detailed segmentation of the target audience, mostly in online commercials, but it was first used during American election campaign lobbying. Tom Agan defines microtargeting as "a way to successfully create personalized messages or offers, correctly estimate of their impact (in regards to sub-grouping) and delivery directly to individuals" (Barbu 2014, p.44).

The engagement with the electorate is the basis of the democratic process. In the history of democracies, parties and candidates have always tailored electoral communication to groups of audiences, considering the specific interests of each of them. However, developments on micro-targeting of voters based on the unlawful processing of personal data, as revealed in the Cambridge Analytica case, have a different nature. It illustrates the challenges posed by modern technologies and the great importance of data protection in the electoral context. Data protection has become a key issue for individuals and for the functioning of democracies because it constitutes a serious threat to a fair, democratic electoral process and has the potential to undermine open debate, fairness, and transparency which are essential in a democracy (European Commission 2018c).

The innovative approach of social media on electoral marketing is deeply linked with the problematics of disinformation. Targeting the electoral audience was not born with the Internet advent, but novelty lies in its major amplification thanks to the power of internet high-speed networking (Haciyakupoglu et al. 2018, p.2). According to Allcott and Genztkow, the known false news stories that appeared in the three months before the 2016 American election, those favoring Donald Trump were shared a total of 30 million times on Facebook, while those favoring Clinton were shared 8 million times (Alcott; Gentzkow, 2017, p.212). Yet, people are much more likely to believe stories that favor their preferred candidate, especially if they have ideologically segregated social media networks.

Another challenge is the way the candidate/party manages to constitute its own database of voters' personal data: in the Cambridge Analytica case, the actors clearly used an illicit way to get those data. In particular, the data processor did not get a formal consent of the users for this specific processing and electoral finality (European Commission 2018c).

Profiling[2] can be used to micro-target individuals, namely to analyze personal data to identify the particular interests of a specific audience or individual to influence their actions. Micro-targeting may be used to offer a personalized message to an individual or audience using an online service, social media for instance. The combination of personalized micro-targeting fake news messages based on a profiling process of illicitly acquired personal data can be very harmful to the electoral process and to the credibility of the electoral results (European Commission 2018c; European Parliament 2016).

González observes that both sides in the 2012 and 2016 US presidential elections used personality profiling software. The author calls attention to products and apps - like IBM Watson, Crystal, and Apply Magic Sauce - to create profiles based on social media information and digital footprints and alternative ways to analyze personalities through social media profile photos and emotional analytics software (González, 2017, p.11). Condliffe minimizes the importance of profiling given the lack of evidence of "psychographics" to influence people's political behavior (Condliffe, 2017, p.1). However, Levy has strong evidence that microtargeting is able to increase polarization (Levy, 2021, p.831)

According to European Commission (2018a), microtargeting is not an illicit tool when not using illicit personal data, but the democratic process needs to be protected from the impact it produces on individuals, for example, when personalized messages affect the individual decision of voting. Micro-targeting can impact the decision and psychology of voters and rise of "strategic" (Alvarez and Nagler, 2000, p.57) and "tactic" votes (Dommett and Temple, 2018, p.195). For example, in the UK general elections of 2017, the authors highlighted the role of a

---

2 The EUC defines Profiling as a form of automated data processing used to analyze or predict aspects concerning for instance personal preferences, interests, economic situation, etc.

website, the "Swap my Vote"[3] that "uses social media to help pair voters who want to swap, each casting each other's preferred vote where it could count for more" (Dommett and Temple, 2018, p.195).

Figure 9. Screenshot of the website http://www.swapmyvote.uk for the 2017 UK general elections

The last elections of the European Parliament in May 2019 (the first European-scale ones after the entry in force of the GDPR on May 25, 2018), seem not to have been impacted significantly by political micro-targeted advertisements. However, it is still early to conclude whether the newly adopted European legislation (mainly the GDPR of 2018) and its supervised co-regulation approach based on partnerships with the major platforms (through the "EU Code of Practice on Disinformation" of September 2018) had a real and significant impact on this result. Compared to the 2016 US Presidential Elections tainted by the Cambridge Analytica affair, the EU Commission declared based on preliminary analysis that ""the regulation

---

[3] http://www.swapmyvote.uk, a site for connecting voters who wish to exchange their votes so that it has more weight in another constituency (because he considers that his constituency is electorally won or lost in advance)

contributed to exposing disinformation attempts and preserving the integrity of the elections while protecting freedom of expression. The highest turnout in the past twenty years (50.97 %) reflects the interest of the citizens for the Union and its importance for their lives"" (European Commission and High Representative of the Union for Foreign Affairs and Security Policy, 2019, p.9). On the other hand, the greater engagement of the population in the elections may also reflect a social media influence, since as posted by Levy social media does not affect opinions but polarization incentivizes the voters to vote in the elections for instance (Levy, 2021, p.831).

3.4. The EU legislative data protection framework in the context of elections

According to the European Commission (2018a), the data protection regime in place for the previous 20 years in the EU was not fully effective due to the fragmented application of the rules between the member states. The absence of formal cooperation between national data protection authorities and the limited enforcement powers of those authorities were the main cause. To solve these problems, GDPR harmonized key concepts - such as consent, empowerment of the users with the right to receive information about the processing of their data -, clarified the conditions under which personal data can be further shared with third parties - the introduction of rules on personal data breaches-, established a cooperation mechanism between the different national Data Protection Authorities (DPA) in cross-border cases, enforced their powers, and created the European Data Protection Board (EDPB). The EDPB groups, all national data protection authorities, and the European Data Protection Supervisor play a key role in the application of the GDPR by issuing common guidelines, recommendations, and best practices.

In case of infringement of European data protection rules, DPAs have the power to investigate and punish it. They also have the power to impose fines up to EUR 20 million or, in the case of a company, up to 4% of its worldwide turnover. ""In the electoral context, it is probable that the gravity of the infringement and the number of persons affected will be high. This might lead to the imposition of high-level fines, in particular considering the importance of the issue of citizens' trust for the democratic process"" (European Commission, 2018b, p.2)

The Directive on privacy and electronic communications - or "e-Privacy Directive" (Directive 2002/58/EC of the European Parliament and Council) - completes the Union data protection framework and is relevant in the electoral context as it rules on the electronic sending of unsolicited communications, including direct marketing. The e-Privacy Directive also lays down rules on the storing and accessing information stored, such as cookies that may be used to track a user's online behavior in smartphones and computers. The Commission's proposal for a Regulation on privacy and electronic communications, the "e-Privacy Regulation follows the same Principia and widens its scope to include internet-based electronic communication services" (European Commission 2018c).

3.4.1. Key obligations and rights of the various actors

According to the European Commission (2018a), the GDPR applies to all actors of electoral contexts - such as European and national political parties and political foundations, platforms, data analytics companies, and public authorities responsible for the electoral process – and they must process personal data (names and addresses) lawfully, reasonably, in a transparent manner, and use it only for specified purposes. Personal data in the electoral context includes special categories of personal data, the sensitive data, such as political opinions, trade union membership, ethnic origin, religious beliefs, and sexual orientations. Moreover, since

data analytics can infer sensitive information from sets of non-sensitive data, their processing also falls within the scope of the GDPR.

3.4.2. Data Controllers and Processors

European Commission (2018a) defines the data controller as ""the organization deciding (alone or with others) why and how the personal data is processed: the data processor processes personal data on behalf and under the instructions of the controller"", it may be the same organization as the data controller or an outsourced one. The ultimate liability lies with the data controller, in charge of taking measures appropriate to the risks and who should be able to demonstrate its compliance with the GDPR.

In the electoral context, several actors can be considered data controllers, such as political parties and foundations, individual candidates, and national electoral authorities. Platforms and data analytics companies can be (joint) controllers or processors depending on the degree of control over the processing.

Companies based outside the EU also have to comply with the GDPR when their processing activities relate both to the offering of goods and services to residents and monitoring their behavior in the EU. This is the case of companies outside the EU, contracted by European companies to process the personal data of European electors. If the data processor is not a European organization, it needs to have a representative inside the EU, officially registered by a national DPA (European Commission 2018c).

3.4.3. Special conditions for "sensitive data"

Election data controllers or data processors can only process personal data - including those obtained from social media – under the GDPR principle of lawfulness. Election data can be processed only based on a limited number of relevant grounds, under the individual's consent, or under the performance of a task in the public interest (with some limitations regarding this second point). In addition, storing information, or gaining access to information already stored in the terminal equipment (computer, smartphone, through cookies, for instance), must comply with the e-Privacy Directive requirements, which means that the users must give their consent again.

Public authorities involved in the electoral context have the right to process personal data (lists of electors, for instance, containing name, surname, electoral number, physical address, etc.) to comply with legal obligations (the organization of the election, for example). However, political parties and foundations may do so only if authorized by the law of a Member State and only for advertising in the electoral context.

The processing of sensitive data, like the electoral profile of the individuals, is strictly prohibited by the GDPR, except for the political parties' members. According to the "purpose limitation" principle, data processing should be specified at the time of collection and can only be further processed for a compatible purpose. Data collected by brokers or social media platforms for commercial purposes cannot be further processed in the electoral context. Political parties and foundations are responsible for ensuring that the data they receive from a third party has been obtained "lawfully". In this case, the company should ask again for the user's consent.

3.4.4. Transparency

According to the European Commission (2018a), the Cambridge Analytica case illustrated the importance of fighting opacity and adequately informing the individuals about what they are contracting. They frequently have asymmetric information, not knowing who processes and for which purposes are their personal data used. According to the GDPR, the data controller must inform the individuals every time it intends to collect personal data, and at each stage of the processing, with the following information; the identity of the controller, purposes of the processing, recipients of personal data, source of the data when not collected directly from the person, the existence of automated decision-making, and any further information necessary to ensure fair and transparent processing. Also, this information has to be provided in a ""concise, transparent, intelligible, and easily accessible form, using clear and plain language"" (European Commission, 2018a, p.6). However, the feasibility of providing information to individuals at "each stage of processing" is still an open question.

3.4.5. Profiling, automated decision-making, and micro-targeting

The EC defines "profiling" as a way to automatize data processing to analyze or predict aspects concerning personal preferences, interests, economic situations, etc. Companies can use profiling to micro-target individuals. Profiling analyzes personal data to identify particular interests of individuals aiming to influence their actions (European Commission 2018a). This can be made through personalized advertisements on web browsers or applications through the use of "cookies". In the electoral context, Cambridge Analytica mined the data collected through social media users to create voters' profiles. Then employing artificial intelligence to classify and identify voters who can be more easily influenced to affect the outcomes of the elections. Thus, the GDPR obliges all data controllers – political parties or data analysts working for them - to inform the individuals of their consequences when they use such

techniques. It also provides that ""individuals have the right not to be subject to decisions based solely on the automated processing of their data"" (European Commission, 2018a, p.7).

3.4.6. Security, accuracy, and impact assessment

The GDPR created a framework of supervision of the data controllers. Thus, the latter must notify any personal data breach in their system to their national DPA within 72 hours at the latest. They also must inform the individuals affected by a data breach without delay when the personal data breach is a risk to their rights and freedoms. In addition, the GDPR states that organizations and actors involved in the electoral process are fully responsible for the accuracy of personal data when collected and compiled from various sources. Finally, the GDPR requires the data controller to carry out a "data protection impact assessment" before using any data process which could risk individuals' rights and freedom; this is the case when a data controller uses the "profiling" methods or when it processes sensitive data on a large scale (European Commission, 2018a, p.8).

3.4.7. Rights of individuals

The GDPR recognizes individual rights to access and request the deletion of their personal data to protect voters. Individuals also have the right to have corrected their incorrect, inaccurate, or incomplete data if the processing is based on withdrawn consent, or the data is no longer needed, or the processing is unlawful. The GDPR also recognizes the right to object to processing even if the organization argues it is based on the "legitimate interest" or the "public interest" grounds. Thus, the right is not subject to decisions based solely on the automated processing of the personal data that means, the "profiling" methods, but includes the right to lodge a complaint to a supervisory authority and the right to a judicial remedy (European Commission, 2018a, p.9).

All these obligations and rights, derived from the GDPR and e-Privacy Directive, are intended to protect the individuals' personal data, particularly their sensitive ones.

3.5. The EU approach for the 2019 elections: supervised co-regulation with the "Code of Practice on Disinformation", and legislative threats

The EU published in October 2018 the "EU Code of Practice on Disinformation" as a practice to avoid problems in the elections of the European Parliament of May 2019 (European Commission 2018b). The code exemplifies the approach to tackle the disinformation problem based on a "partnership" with the major digital media companies - Facebook, Google, Twitter, Microsoft, Mozilla, among others. The signatories commit to scrutinizing advertisement placements to reduce revenues of the purveyors of disinformation, be transparent with the origin of the political advertisements, guarantee the integrity of information, and empower consumers and the community.

The code also provides key indicators of performance (KPIs). The signatories committed to delivering an annual account of their work to counter disinformation in the form of a "publicly available report reviewable by a third party." Furthermore, the signature of the code was followed by an assessment period of 12 months, during which the signatories met regularly to analyze its progress, implementation, and functioning.

Online platforms and trade associations representing the advertising sector have submitted a baseline report in January 2019, setting out the inventory of the measures taken to comply with their commitments under the code. Between January and May 2019, the EC monitored the implementation of the commitments by Facebook, Google and Twitter with particular pertinence to the integrity of the European Parliament elections in May 2019. In

particular, the EC asked the three platforms (signatories to the Code of Practice) to report monthly the actions undertaken to improve the scrutiny of ad placements, ensure transparency of political and issue-based advertising, and tackle fake accounts and malicious use bots. The EU published the companies reports and its assessment (European Commission and High Representative of the Union for Foreign Affairs and Security Policy, 2019, p.4) on its website, keeping them for five months.

By May 2019, the EC concluded that the major platforms made significant progress on political advertising transparency, their services' integrity, and scrutiny of ad placements. However, the EC urged the platforms to improve their cooperation with fact-checkers in all member states and empower users to detect disinformation better. Platforms should also make additional datasets available to the research community (in particular, the EC claimed official access to the algorithms for academic purposes).

However, already in August 2018, Facebook launched a new public report, Ad Library Report, which ""lets people see how many political and issues ads were run in a given country – as well as aggregated advertiser spend and top searched keywords in the Ad Library"" (Facebook, 2018). One additional initiative of Facebook needs to be highlighted: the so-called "EU war room" in Dublin. Facebook built a team of 40 specialists working in an operation center to counter digital threats that would undermine the European Parliament elections (Scott, 2019, p.1). The team counted with coders, digital engineers and specialists in all of the EU's 24 official languages, and it was split along national boundaries: the digital monitoring was not limited to disinformation but also to illegal contents, including hate speeches (Hinds, 2019, p.7).

3.5.1. Effects of EU Code on 2019 parliament elections

In an open letter to the European press in February 2019, Mariya Gabriel (European Commissioner for the Economy and the Digital Society) and Sir Julian King (European Commissioner for the Union of Security) reminded us that if the results of elections prove to be insufficient, they could propose other measures, including regulatory ones (Gabriel and King, 2019, p.1). But many observers doubted the determination of the EU, in particular, because of the strong lobbying work done by those major platforms in Brussels. According to the data protection NGO Corporate Europe Observatory, ““*while 2,25 million euros get Facebook rank 19 on the list of corporation's biggest EU lobby budget, it ranks 4th among the corporations with the most lobby meetings at the Commission*”” in 2017 (Corporate Europe Observatory, 2018, p.2). In March 2019, the UK national newspaper The Guardian, based on the leak of internal Facebook documents, revealed a ““*secretive global lobbying operation targeting hundreds of legislators and regulators in an attempt to procure influence across the world*””. The document includes details of how Facebook had lobbied European politicians to strategically head off “overly restrictive” legislation. The document includes extraordinary claims that ““*the Irish prime minister said his country could exercise significant influence as president of the EU, promoting Facebook's interests even though technically it was supposed to remain neutral*”” (Cadwalladr and Campbell, 2019, p.1).

Ahead of the EU elections, the NGO Avaaz conducted a Europe-wide investigation on disinformation on Facebook for the period February to May 2019 in France, Germany, Italy, Poland, Spain, and the United Kingdom. The inaugural investigation was published on May 22, 2019, just one day before the first countries voted. Avaaz reported almost ““*700 suspect pages and groups to Facebook, which was followed by over 35 million people and generated over 76 million “interactions” (comments, likes, shares) over the period. Facebook had taken down 132 of the pages and groups reported, together the pages taken down reached 762 million*

estimated views."" Interestingly, the pages removed had more than twice the number of followers compared to the main European far right parties combined (Avaaz, 2019, p.5A).

Between November 2018 and March 2019, SafeGuard Cyber, a private company that develops platforms to detect threats in digital channels, analyzed almost 3,5 million posts on Twitter, Facebook, Instagram, and YouTube to evaluate Russian misinformation campaigns. The report focused on the period of 1-10 March 2019 and on "bad actors" (bots, trolls, and hybrids which are humans using software). To determine misinformation contents, they used a tool that aggregates the data from 155 fact-checking sites (such as "Politifact", "EU vs Disinfo") in 53 different languages and a database containing over 500.000 known troll and bot accounts. The main findings were the following: (i) misinformation agents worked within clear narrative categories; (ii) the message was suited for an European Audience (iii) the tendency was to amplify already existing content, rather than creating new content, underlining already existing societal and political tensions (for instance, the most used categories of narrative by Russian misinformation were Brexit in the UK, the "yellow vest" movement and the low popularity of President Macron in France, irregularities about EU funds, and supporting Euroscepticism), (iv) content was often related to hashtags that could have been picked by bots automatically and shared rapidly (like 2,3 posts per second). Real users could also be used to amplify through hashtags; (v) as the narrative exploited existing tensions, some states with lower Eurodeputies representation were bombarded by bad actors' messages. One example is the Netherlands (with over 3% of Eurodeputies allocation in the European Parliament) which received 10% of Russian bad actors, due to the tension around the rise of the Party of Freedom; (vi) analyzing Twitter accounts, they found that 12% of the accounts following Jean-Claude Juncker's official Twitter profile were probably "bad actors". Mr Otavio Freire, the co-founder of SafeGuard Cyber, affirmed to The Guardian that "our report reinforces the need for a new approach to security, as today's bad actors are not at all hindered by the cybersecurity tactics of yesterday" (Boffey,

2019, p.2). In conclusion, the report had shown the existence of Russian interference during the European Parliament elections. This was the major fear of the European Union (SafeGuard Cyber, 2019).

According to the EC, the European anti-fake news strategy seems to have been relatively successful during the last campaign for the European Parliament of May 2019 (European Commission and High Representative of the Union for Foreign Affairs and Security Policy, 2019, p.9-10): so far, no significant case of massive-scale disinformation has been publicly opened. On the contrary, the partnership of the European Commission with the main social media platforms resulted in better transparency of the political advertising sources both for the users and the regulator. This has been recognized by independent actors and media as well. A study by the Oxford Internet Institute found that less than 4% of news sources shared on Twitter ahead of the European elections were disinformation content, while mainstream professional news outlets received 34% of shares. ““*According to FactCheckEU, there was less disinformation than expected in the run up to the European elections and it did not dominate the conversation as it did around the past elections in Brazil, the United Kingdom, France or the United State*”” (European Commission and High Representative of the Union for Foreign Affairs and Security Policy, 2019, p.9). This innovative approach is based on a partnership, i.e., a supervised co-regulation of the platforms, which run the risk of consistent fines at the European and national scales. This partnership is a shared responsibility of all relevant actors: EU institutions, member states, private sector/online platforms, fact-checkers, civil society, and researchers: a sort of "ecosystem" to fight political fake news messages. However, the EC stated that a lot of progress still needed to be done by the platforms to ensure the sincerity of the electoral results (innovation never stops) and urged the platforms to give academic researchers access to one of their best-kept industrial secrets: the algorithm.

3.6. The limits of the EU bureaucratic approach in a digital environment

According to Balkin (2018), the view of free expression that characterized much of the twentieth century is no longer adequate to protect free expression nowadays. This is because while the last century featured a dualist model of speech regulation with only two well-identified basic kinds of players - territorial governments and private speakers – the new century is pluralist, with multiple players. The author uses a triangle picture to illustrate the multiple actors and their interactions (ways of regulating speech); the first corner includes the nation-states and the EU, and the second sets privately owned internet-infrastructure companies, including social media companies, search engines, broadband providers, and electronic payment systems, and the third displays different kinds of speakers, legacy media, civil society organizations, hackers, and trolls. Figure 9 illustrates this "pluralist model of speech regulation" (Balkin, 2018, p.2014).
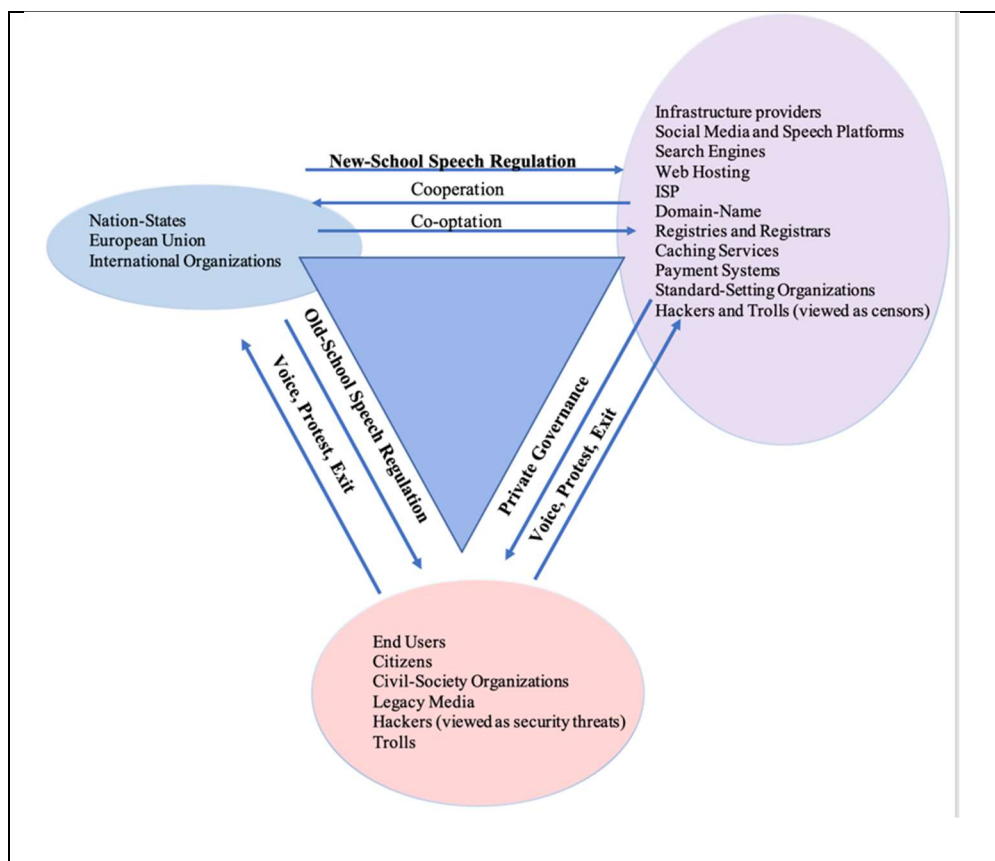

Figure 10. The pluralist model of Speech Regulation. Source: Reproduced from Balkin (2018)

According to Balkin, the "practical ability to speak in the digital world emerges from the struggle for power between these various forces, with "old-school", "new-school", and private regulation directed at speakers, and both nation-states and civil-society organizations pressuring infrastructure owners to regulate speech" (Balkin, 2018, p.2014). However, the author points that that this configuration creates three problems: ""*first, nation-states try to pressure digital companies through new-school speech regulation, creating problems of collateral censorship and digital prior restraint. Second, social media companies create complex systems of private governance and private bureaucracy that govern end users arbitrarily and without due process and transparency. Third, end users are vulnerable to digital surveillance and manipulation*""" (Balkin, 2018, p.2011).

The XXth century model of free speech regulation is no longer adapted to the new digital environment: yet, despite all the positive impacts of the EU approach on the electoral context within a digital environment. It seems that this approach is only adapted for a dualist model in transition (toward a triangle model), where the actors are still very well identified (to be supervised, and fined if necessary), and, finally, of good willingness (see the concept of co-regulation that we already discussed in this article). But this European approach does not seem able to regulate the action of robots, trolls, hackers, and others coming from outside the European Union. One thing is to make the data controllers judicially responsible for every piece of information which is published, processed, and for any breach in their enormous databases of personal data (through a quite heavy bureaucratic process which, until now, seems to have been relatively successful); another one is to make sure that, even with this process, the final choices of the individuals will not be altered at the moment of voting because they were exposed to huge volume of disinformation just before closing the electoral process. And nowadays, the conjunction of AI, big data, robots, bots, trolls, and others are able to pass through the regulation, even more, when they come from outside the European Union.

3.6.1. The current threat of campaigns organized by authorities outside the EU: extraterritorial legal application and robots

Due to the very nature of the information, which is reproduced with a virtual null cost (Shapiro and Varian, 1999, p.3-7), and the Internet being a worldwide web, there are no boundaries of national states for information dissemination. Thus, the action of one state in regulating information dissemination may fail if it is not coordinated with other states and actors since information spillovers to all countries. In addition, a state can become very powerful if it is able to disseminate controlled information direct to specific groups.

""to date, most proposed legislation against fake news does not directly address the issue of extraterritorial application. However, some proposed bills do have extraterritorial implications. Germany's Network Enforcement Act mandated the establishment of a local point of contact for transnational technology companies to cooperate with local law enforcement authorities on takedown requests. The proposed Honest Ads Act, although framed generally in terms of protecting US domestic order, targets the role of foreign nationals and seeks to prevent contributions, expenditures, and disbursements for electioneering communications… in the form of online advertising"" Haciyakupoglu et al., 2018, p.6)

But once again, this would be only efficient in the case of well-defined actors coming from outside the EU. In the case of Russian interference during the last elections, it has been impossible to identify the responsible organizations behind it. For instance,

""considerable evidence exists demonstrating that entities affiliated with and acting at the direction of the Russian Federation have sought to influence the direction and outcome of a series of major elections in Western democracies. These incidents include notably the referendum over whether the United Kingdom should exit the European Union and

episodes during the 2016 and 2017 U.S. and French presidential elections"" (Gvosdev, 2019, p.321).

How can the European approach limit this "foreign" interference when no organization is identified as the responsible behind this spread of fake news?

New dynamics brought about by technological advancements are a concern for governments that want to use their laws to fight fake news. Ministers of Justice in three German states, for example, have proposed anti-botnet legislation to reduce the impact of automated social media accounts in disseminating fake news.

""*Jenna Abrams, a popular Twitter account that attracted up to 70,000 followers through its support for US President Donald J. Trump and advocacy of far-right views, for example, is believed to have been run by the Russian propaganda machine to discredit the Democrats. The role of automated accounts in influencing elections was raised during the US Senate hearings as well*"" (Haciyakupoglu et al., 2018, p.6).

The GDPR does not specifically address the worldwide character of information dissemination by bots and makes again the social media platforms solely responsible (they would have to delete fake and robot accounts). But, in fact, the only possible retaliation measure would have been to condemn the social media platforms that published such fake news, sent in high volume and velocity by robots. As discussed, ""when fake news stories do get denounced as potentially fake, or the interim judge is ready to take action, it is already too late and the story has gone viral"" (Alemanno, 2018, p.4).

In addition, regulating companies is more similar to usual regulatory practices in the real world (in opposition to the virtual world), making regulatory activities easier to implement given the history of the regulation (Viscuse et al., 2005, p.362). Also, the social media platforms are worldwide monopolist or oligopolist companies.

Finally, one can consider that the EC itself acknowledged that it was a lost battle (with the protection tools as of today), investing massive amounts of financial resources in education, communication, counter-information (through the website https://euvsdisinfo.eu for instance), and forming the EU East StratCom Taskforce in 2015 to counter Russia's disinformation campaigns (Haciyakupoglu et al., 2018, p.10).

3.6.2. The new threat of "satellite campaigns" organized by undefined actors within the EU

Dommett and Temple investigating the 2017 general elections in the UK (to decide which government would implement the voted Brexit), focused on a new development in the online electoral campaigns: the increased visibility of digital infrastructure offered by non-party organizations to encourage voting and campaigning. For instance, the authors note that "innovations such as Momentum's "My Nearest Marginal" app[4], fundraising sites such as "CrowdPac", and campaigning hubs like the "Progressive Alliance" or "Campaign Together" were seen to empower and connect individuals to contribute to electoral campaigns via non-traditions routes" (Dommett and Temple, 2018, p. 194). All these organizations are located in the UK, close to the Labour Party, and don't have contractual or juridical links with the Party. "It suggests that, in addition to Whiteley and Seyd's categories of the central party campaign, centrally coordinated local campaigns, and purely locally directed campaigns, we can also identify campaigns originating beyond party structures and control: those termed here 'satellite' campaign" (Dommett and Temple, 2018, p.194).

---

4 My Nearest Marginal is an application used in the 2017 UK elections by the Momentum movement (a British political organisation described as a grassroots movement supportive of the Labour Party) to direct activists flooding into strategic "swing" constituencies (Rees, 2017, p.2).

One of the benefits of satellite campaigns is the potential for innovation: as organizations less restricted by legal requirements and responsibilities (they are not considered as political parties or foundations in the scope of the GDPR for instance), these bodies have the space to innovate and trial new tools that traditional parties may be wary of promoting: the example of the UK site "Swap my vote", already discussed here, is significant.

The main issue of the rise of this "citizen-initiated campaigning", in the electoral context, is that they are not considered as official political parties, but, in fact, they do act like political parties.

The GDPR seems to be efficient in a digital environment where every actor plays fully identified, and with a certain transparency. When it deals with robots, botnets, hackers, or even "citizen-initiated" campaigns, i.e., when it is impossible (or very complicated, or very time-consuming) to identify the responsible organization behind the campaign, the European approach as of 2019 seems to be useless.

3.7. Conclusion

Disinformation and illicit profiling strategies become a national security issue when they undermine the foundations of the nation-state. In this regard, fake news could serve as a tool for disinformation campaigns at a massive scale: the intentional dissemination of false information for influencing opinions or policies of the receiving audience. It is currently too early to assess definitively the impacts of the European legislative initiatives against fake news and illicit profiling strategies due to problems pointed by Haciyakupoglu et al (2018), such as issues on the definition of fake news, the global dimension of cyberspace *vis-a-vis* the territorial boundaries of legislation, challenges in identifying the actual perpetrator of fake news, and the sophistication of disinformation campaigns. In this sense, the authors call attention to the urgent

necessity to "reconcile" online regulations with the offline regime to avoid problems like pornography or infant pornography which is prohibited in some countries, but publicly accessible on the internet, and the controversial use of VPN.

The European Union has launched a "pack" of legislative measures and co-regulation system with the social media platforms to combat the disinformation and illicit profiling phenomena in the electoral context, which seems to be, until now, the most "avant-gardist" and efficient approach, in the current state of the art and according to the literature. However, progress and efforts still need to be made by all actors to follow the race for innovation launched by the protagonists of disinformation. In spite of all possible efforts, technology is always a head start over legislation.

This is why, in spite of these necessary (but not sufficient) legislative measures, it seems urgent to continue developing other types of measures, in the short, and long terms. In the immediate term, it is important to continue developing the fact-checking efforts, the counter fake news website and communication, the fake news flagging directly on the social media platforms (potentially with the help of algorithms, artificial intelligence, and machine learning). On the long term, it seems important to develop measures to promote the ability of (social) media decryption in the education of children at school (it will be helpful for adults too), support new social practices against fake news such as individual responsibility before sharing messages (checking and authenticating the sources and author, reading the information extensively), and to clearly define the responsibilities of the technology companies (Haciyakupoglu et al., 2018, p.3).

It is important to consider the creation of a new ecosystem to fight the fake news in the electoral context counting with: a State able to define new rules of conduct for the electoral campaigns, which protects the data of its citizens by law, and which imposes a minimum of transparency rules to the social media platforms, technology companies aware of their social

and democratic responsibilities, which auto-regulate themselves (under the State supervision), and which collaborate with the civil society sharing useful research data, and a civil society (users -individuals and organizations-, scientific research) which fully assumes its role as a controller of both the State and the private technology companies.

Considering all those elements, the approach of the European Commission to protect the sincerity of the results of the election for the European Parliament in May 2019 seems to have brought some interesting progress in the vast debate of personal data protection through a reactive answer to the Cambridge Analytica affair, a regional collaboration of all the national DPAs, an incentive to the NGOs and journalists consortiums to develop fact-checking platforms, and an innovative partnership based on co-regulation with the main social media platforms. In addition, it is important to highlight the growing necessity of more transparent governance at the head of the social media companies, especially in their policy of posts validation criteria, and, to be more exhaustive, in the construction of their algorithms.

Observation: an extract of this article has been accepted for presentation at the conference "Artificial Intelligence: democracy and social impacts" organized in 2021 by Humanity Seminars – Center for Artificial Intelligence (C4I) – Humanides USP and is to be published as one of the chapters of the book based on the seminar topic and/or in a dossier of the journal "Revista de Estudos Avançados" (USP)

(http://hseminar.webhostusp.sti.usp.br/2021/?page_id=1262&lang=pt)

4. Investigating potential effects of the new Facebook algorithm on the 2019 French Yellow Vests movement

Abstract: The global covid-2019 pandemic has accelerated the digital transformation of political campaigns and social movements. Thus, it is important to know the challenges this transformation is posing to societies. We study whether introducing a new version of Facebook's news feed algorithm in January of 2018 has facilitated the Yellow Vests' social manifestation on France's streets from November of 2018 to the end of the year 2019. The new algorithm reinforces and prioritizes connections among people with similar preferences or characteristics for some specific issues (the Facebook "groups"). Building on the literature, which documents strong evidence of connections between the social media and the movement, we explain how this algorithm played a key role in transforming digital anger into a radical social movement in the streets. Also, we investigate how the lack of governance on the social networks could lead to violent real-life manifestations, resulting in negative net gains in the long run and a high potential to destabilize democracies.

4.1. Introduction

We investigate whether the introduction of a new version of Facebook's news feed algorithm in January of 2018 has facilitated the Yellow Vests manifestation on France's streets from November of 2018 to the end of the year 2019. Based on the literature, we discuss the possible mechanisms that could connect the two events. Thus, we employ a case study to address some potential effects that the new social media have in changing societies. Results suggest a significant relationship between the events.

More specifically, to evaluate whether the new 2018 Facebook algorithm facilitated the emergence of the Yellow Vests movement we use the concept of "filter bubble" effect described by Pariser (2011). Filter Bubble seems to be particularly applicable to the Yellow Vests movement through various mechanisms that the new 2018 Facebook algorithm catalyzed since it is connected to a relative homogeneity of the Yellow Vests movement's actors, or people younger and less educated than the mean of the French population.

Finally, we conclude by discussing the benefits and costs of the movement for French society.

Besides this short introduction, section 2 presents the context of the movement and the change in the Facebook algorithm; section 3 investigates the mechanisms linking both events; section 4 discusses the benefits and costs of the movement for French society. Finally, section 5 summarizes our main conclusions.

4.2. Background

The French protest started after the government tried to pass laws to increase gasoline taxes, and to decrease the roads speed limits in November 2018. According to Boyer et al. (2020), the movement officially emerged in France on the 17th of November 2018 with a massive blockade of public roads over half of the country's municipalities, and rapidly became a full-blown protest against the government and all the representations of the "elite" of the society. The following protests occurred almost every Saturday in the main cities and principally in the capital Paris until around November 2019 with little protesters.

Dormagen and Pion (2021) estimate that around three million French people (over a total of 67 million people) have participated to the protests, which is comparable to the May 1968, the biggest post-WWII French social movement. Beyond material damage caused on public

patrimony, the cost in human lives was very high: 11 dead and 4439 injured people (including police forces), around 12000 arrests, 3100 condemnations, and 400 prison sentences.

Boyer et al. (2020) point out that the movement was highly decentralized regarding its origins outside traditional parties and labor unions, and its mobilization was done by means of social media and the Internet. Brancaccio (2020) reinforces that the moment of "digital gestation" between October 12 and 16, 2018, was "crucial" as the on-line petition received hundreds of thousands of signatures and the two principal Facebook groups ("virtual *agoras*") were created: "the element of the *commoning* of anger in virtual space served as a premise for the desire to meet physically, which happened somewhat later in the occupation of the roundabouts".

On the 11th of January 2018, the Head of Facebook's news feed Adam Mosseri announced a very important update of the Edgerank algorithm, which is responsible for filtering and prioritizing the publications on each user's wall. According to Mosseri (2018), the algorithm determines the frequency of a post appear in the News Feed of the individuals, based on the number of reactions, comments, and shares. This update prioritizes posts that spark conversations and meaningful interactions between people by predicting which posts people want to interact with friends (MOSSERI, 2018).

While Zuckerberg (2018) argues the change aimed to improve the quality of the time spent on Facebook with more meaningful social interactions and less content of business, brands, and media, Ladrova (2018) explains that the change affects the profits of Facebook since the administrators of company pages have to resort to paid advertising to get their posts available for the user ("IDIMT-2018", 2018).

Rogers and Niederer (2020) point to the three-pronged strategy that would favor meaningful connections (family and friends), the trusted sources (user-surveyed media), and the local news in the news feed over more far-flung businesses, brands, and media. One of the

major impacts of this update is the new prevalence of the Facebook Groups (users who share common interests and subscribed to a specific group) among the Facebook Pages (of public people, brands, or traditional media press) feed of the users.

The concept of "filter bubbles" has been well documented (and sometimes partially questioned) since the invention of the expression (PARISER, 2011). According to Fu et al. (2016), the self-interest and communal incentive could drive Facebook user's content-sharing intention, but their effects depended on the content types. Further, the effects of self-interest incentives were found only among the users who focus on their close friends (bonding-focus), but not among those who focus on distant friends (bridging-focus).

Oeldorf-Hirsch and Sundar (2014) point out the role of the "engagement" - any active action someone takes on a Facebook page: like, comment, share, location, tag - as crucial for Facebook to analyze and predict each user's profile, and to offer more interactions with users who share the same interests. The authors underline the importance of the acquaintances' implication - known as virtual "friends" on Facebook - for the process of assimilation and sharing of received information. Their results show that user involvement in the news content depends on the social affordances of the site, particularly those that allow for audience customization and those that drive network feedback. Asking the network's opinions and targeting specific friends led to greater involvement in the news content. In addition, discussion through comments led to a greater sense of influence and greater participation in sharing the news story. These findings highlight the importance of encouraging individuals to act as sources of information in their networks to drive engagement in current events in the changing news landscape.

A study realized by the Belgian marketing company Soprism[5] and the French national statistics institute INSEE[6] (Figure 10) in 2020 allows us to confirm that the French Facebook users are much younger than the general population. Still, the youngest generation (under 24 years old) is under-represented - more kind to use other platforms such as Instagram, Snapchat, or TikTok. Almost 50% of the French users of the social network are between 25 and 44 y.o., twice more than the general population.
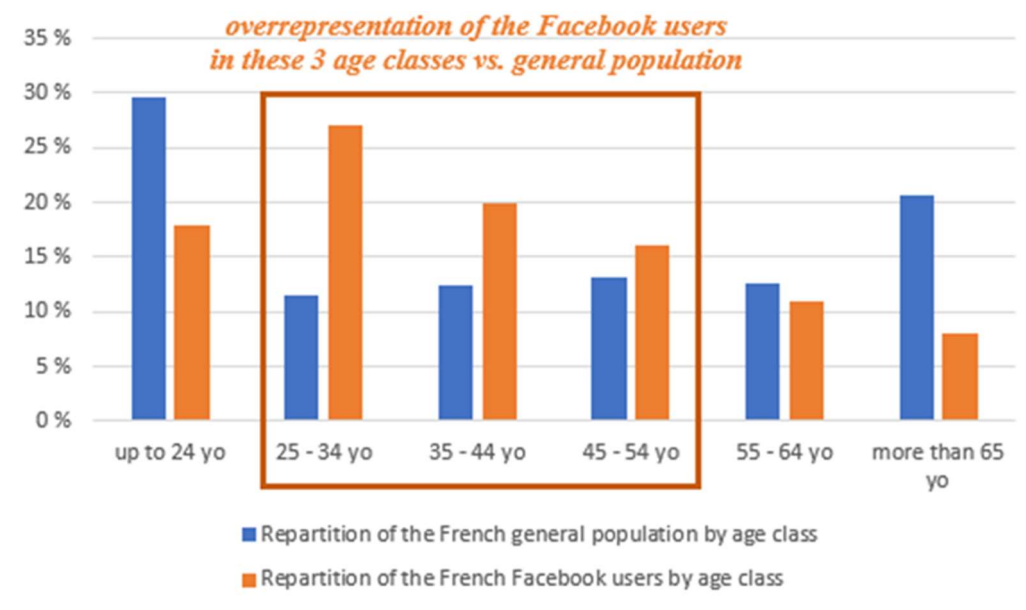

Figure 11. Comparison of the repartition of the French Facebook users vs. the repartition of the French general population by age classes (2020). Source: INSEE and Soprism (2020)

This 2020 Soprism study also informs that only 26% of the French Facebook users have a diploma compared to a proportion of 38% in the general population. Moreover, according to a 2019 study made by the French Ministry of Education[7] which represents a very qualitative significant gap as the French academic structure shows that the youngest generations are much

---

5 https://www.lsa-conso.fr/qui-sont-les-utilisateurs-de-facebook-et-d-instagram-en-france-chiffres,351919

6 https://www.insee.fr/fr/statistiques/2381474#figure1_radio2

7 https://publication.enseignementsup-recherche.gouv.fr/eesr/FR/T666/le_niveau_d_etudes_de_la_population_et_des_jeunes/#:~:text=La%20France%20est%20le%2022e,%25)%20(graphique%2023.01

more "academically educated" than the older ones (Figure 11). Therefore, it seems possible to conclude that the typical French user of Facebook is much younger and much less academically educated than the general population, which also confirms the homogeneity of people in a filter bubble as posted by Pariser (2011).

	25-34 y.o		35-44 y.o.		45-54 y.o.		55-64 y.o.	in %
	Women	Men	Women	Men	Women	Men	Women	Men
Up to baccalaureate diploma	72,6	66,2	70,0	62,4	54,3	47,0	40,1	35,6
A higher diploma than baccalaureate	52,2	43,5	49,7	41,9	36,9	32,0	25,0	23,8

Figure 12. Highest diploma by age and sex - Source: 2019 INSEE[8]

Enjolras et al. (2013) studied the impact of "social media and mobilization to offline demonstrations" using Norwegian data. The result indicates that the typical participant mobilized via social media comes from lower socioeconomic status and is younger than those mobilized by other mobilizations. In addition, the authors evidence the role of social media and its capacity to recruit new and different segments of the population as an alternative to "mainstream media and well-established political organizations and civil society."

It is essential to underline the relative political homogeneity of the actors of the Yellow Vests movement: Algan et al. (2019) demonstrate that they bring people together whose life satisfaction rates are meager, regardless of their agreement on how to respond. Most of them are former voters of Marine Le Pen (far-right on the French political spectrum), Jean-Luc Mélenchon (far left), or abstainers (in this order). They share a more radical critique of state and government than either of these electorates while having more median positions on moral issues such as tolerance for minorities. Analysis of the geography of the roundabouts confirms the original character of this movement; the North-East and the South-West are the strong points of the mobilization, i.e., two regions where Marine Le Pen and Jean-Luc Mélenchon did their

---

8 https://www.insee.fr/fr/statistiques/4797586?sommaire=4928952&q=%C3%A9ducation

best scores in the presidential election of 2017 (ALGAN et al., 2019). Boyer et al. (2020) confirm that in France, women, youth people, workers of lower socio-professional categories, and the voters of populist parties are over-represented amongst Facebook users.

4.3. Sociography of a social mobilization

A study realized by the Institut Montaigne and Elabe in March 2019 (ELABE and INSTITUT MONTAIGNE, 2019) tried to define the main characteristics of the people which considered themselves as members of the yellow vests' movement:

-gender: 53% are men and 47% are women (this presence of women is a rather rare phenomenon in major social movements in France).

-age: working people are at the heart of the movement: 50–64-year-old 30% (+5 points compared to their weight in the French population), 35–49-year-old 28% (+3) and 25–34-year-old 16% (+1).

-professional situation: it is a movement of the working classes, the yellow vests' members are mainly workers (26% are workers, +9 compared to the weight of workers in French society), employees (21%, +4) and retired people with a low income - (17%, =), who live in rural municipalities (28%, +5) and small towns with 2,000 to 20,000 inhabitants (21%, +3), but also in towns with more than 100,000 inhabitants (26%, -3).

-education: 50% of yellow vests' members have a level of diploma lower than the baccalaureate, 28% have a level of diploma equal to the baccalaureate, and 22% have a diploma of higher education, these proportions are respectively 54%, 16% and 30% in the general population in 2017.

-economic situation: the economic situation of the yellow vests' members is significantly worse than the average French population, their standard of living is €1,486 per month compared to €1,777 per person on average for the sample of French people interviewed.

-feeling of social downgrading: they are less numerous than the French average to declare themselves happy (64% against 73%), 65% of yellow vests' members have difficulties to financially survive until the next payroll (+17 points compared to the French average), purchasing power is by far the subject that concerns them the most (50%, +10), ahead of pensions (30%, +4), employment (28%, +3) and immigration (27 %, +4).

4.4.Mechanisms connecting FB and Yellow Vests Movements

4.4.1. *A facilitation of the digital interconnections*

In 1973, the American sociologist Mark Granovetter developed his theory of the "Strength of weak ties": a network is made of strong and weak tied between its participants/users. The strength of the ties is constructed by time passed together, emotional intensity, intimacy, and reciprocity of the links between different agents. Strong ties are developed with close friends, whereas weak ties are produced with superficial acquaintances. The originality and apparent contradiction of this theory (at the base of any algorithm of actual social media) lie in the fact that "weak ties are more likely to link members of different small groups than strong ones, which tend to be concentrated within particular groups" (GRANOVETTER, 1973). The ties between indistinct and unknown Facebook users of various groups - which share common characteristics like political opinions, demographic or geographic proximities, history of navigation and "likes" on the social network, etc. - can be stronger than the links with real friends/family members since they are genuinely diverse and allow to enter other "social networks" than those constituted by the strong ties.

Egebark and Ekström (2018) have studied the impact of the "threshold effect" of the Facebook groups on the engagement of "outside" users: "one Like from a single stranger had no impact. However, increasing the size of the influencing group doubled the probability that

subjects expressed positive support. Friendship ties were also decisive. People were, on average, four times more likely to press the Like button if a friend, rather than a stranger". They conclude showing that "both group size and social proximity matters when opinions are shaped" and that "once the number of recommendations reaches a certain threshold there will be a 'snowball effect".

4.4.2. The use of an unfiltered anger speech on the network that promotes the feeling of group belonging

As Boyer et al. (2020a) notice "Facebook was used as means to organize protests and share demands, but as conflicts with the police intensified, main topics of interest were progressively shifted towards police violence and government critiques". Furthermore, its "textual analysis documents a situation where mobilization on Facebook becomes more radicalized over time, and tends to depart from its original goals, as anger grows among the remaining participants" (BOYER et al., 2020b): an anger aimed at all the representants of the "establishment" (as government, unions, and capitalistic elite) who are considered as responsible for the their social despair; some attacks against minorities are more sporadically observed -racism, xenophobia and homophobia-The conversations on Facebook groups (posts publications, posts commentaries, or messenger chat) seem to act like a "car interior" where the participants feel comfortable to speak without any filter or moderation, and then protected and encouraged to radicalize their own speech by the group effect[9]. Godefroid (2020) emphasizes the notion of "implication": a personal commitment of the speaker in the debate, and which can trigger in him a whole spectrum of (sometimes obsessing) emotions. According to this study,

---

9 https://www.affordance.info/mon_weblog/2018/11/gilets-jaunes-facebook-bleu.html

most of these platforms are precisely configured to promote the abundance of content of a controversial nature. The implementation of the Facebook's new algorithm in January 2018 which promotes the users' "engagement" precisely acts like a catalyst of polemics as the components of the controversy correspond perfectly to what favors comments: the controversy arouses the reactions, animates passions which are sometimes violent and, by its polarizing character, leads to Internet users to argue over the given question through long discussions, sometimes interminable because they confront irreconcilable points of view. A publication that arouses controversial exchanges increases its visibility and consequently increases its exchanges, and so on (GODEFROID, 2020). A study by Hazard Owen (2019) shows a significantly increase engagement (reactions, comments, likes, shares) of Facebook users and in particular a sharp increase in the reaction "angry" under journalistic publications[10].

4.4.3. *A strong geographic correlation between the on-line and off-line mobilizations thanks to a reduction of the information asymmetries*

Boyer et al. (2020) demonstrate a strong correlation between the number of members of local Facebook's yellow vests groups, the number of publication and the number of blockades during the first mobilization of November 17th, 2018 all over the country (3000 blockade points according to the police affairs). These very strong correlations testify to the close link between the coordination of the movement via the Facebook groups and the physical action points; this digital coordination allowed an efficient collective action thanks to a reduction of the information asymmetries (BOYER et al., 2020a).

---

10 https://www.niemanlab.org/2019/03/one-year-in-Facebooks-bigalgorithm-change-has-spurred-an-angry-fox-news-dominated-and-very-engaged-news-feed/

Another study by Enikolopov et al. (2020) found that "a 10% increase in VK -*a Russian equivalent of Facebook*- penetration increased the probability of a protest by 4.6% and the number of protesters by 19%" and that "social media induced protest activity by reducing the costs of coordination rather than by spreading information critical of the government" (ENIKOLOPOV; MAKARIN; PETROVA, 2020).

However, following authors like Levy (2001), we argue that the effects of the algorithm on individuals using social networks is principally behavioral: it affects our actions (our participation to debates or elections, our anger against some groups for instance) but not so much our preferences (political preferences in our case). According to Levy (2001), "exposure to counter-attitudinal news decreases negative attitudes towards the opposing political party", and he finds "no evidence that the political leaning of news outlets affect political opinions" (LEVY, 2021). In other words, it seems that social networks do not influence our opinions, but they strongly impact our attitudes and our actions (both on- and off-line).

4.5. Discussion: a mobilization that was worth the cost of it?

In December 2018, French President Macron announced that his government would release 10.3 billion euros to support fiscal and social measures[11]: (i) cancellation of the fuel tax (which was the starting point of the dispute), (ii) the minimum wage of the French is increased by 100 euros (around 6% direct increase), (iii) the amount of the social contribution for the most modest retirees is frozen, (iv) employers can pay their workers up to 1000 euros in bonus without being taxed (without a great success). This movement made it possible to organize in January 2019, a "great national debate" (according to polls, one in two French people believe it has enabled citizens to debate public policy). But these measures are not felt in the daily life of

---

[11] https://www.rtbf.be/info/monde/europe/detail_france-qu-ont-obtenu-les-gilets-jaunes-en-un-an?id=10367007

yellow vests; most of them only saw it as a communication operation and feeling of frustration remains. This is why the mobilization (on- and off-line) remains very active during the first semester of the year 2019, with even more radicalized manifestations every Saturday in Paris.

The human toll of the manifestations is particularly heavy[12]: during the first 12 months of the movement, 11 people died (mainly in traffic accidents). According to the French Interior Ministry[13], 2495 protesters and 1944 police or firemen officers were injured. Due to the uncommon and extensive use of flash balls arms by the police officers (more than 13900 shots), 24 of these wounded protesters lost an eye and five had their hands torn off. According to the French Ministry of Justice, the police carried out a total of 12,107 arrests, which resulted in 10,718 police custody. Nearly 2,000 sentences were handed down (40% of which were in prison) and as many cases closed without follow-up. According to the French Interior Ministry, 313 judicial inquiries "for suspicions of police violence" were opened by the General Inspectorate of Police (IGPN), the "police of the police", after negative reactions of a part of the civil society. The material balance is also heavy with shops closed, cars burned, damage to street furniture, looting, blockages every Saturday in Paris during almost 12 months ... The French Ministry of Economy estimated at the end of March 2019 that the economic cost of the demonstrations could "go as far as" at 0.2 point of GDP", or 4.5 billion euros, spread over 2018 and 2019 (a parliamentary report, published in mid-July 2019, estimates for its part that the movement of yellow vests cost the French economy 0.1 point of growth in the last quarter of 2018). According to a senatorial report at the end of June 2019, the insurance companies had paid 217 million euros in compensation to repair the material damage linked to the acts of vandalism committed on the sidelines of the marches.

---

[12] https://www.cnews.fr/france/2019-11-15/mobilisation-blessures-arrestations-un-de-gilets-jaunes-en-chiffres-840398#:~:text=11%20morts%20et%20plus%20de,Int%C3%A9rieur%20arr%C3%AAt%C3%A9s%20au%204%20octobre.

[13] https://www.cnews.fr/france/2019-11-15/mobilisation-blessures-arrestations-un-de-gilets-jaunes-en-chiffres-840398

As Zeynep Tufekci explains[14], the main limitations of modern social movements initiated online are inherent to their digital origins: lack of leadership, contradictory debate within the group, and then a flagrant lack of long-term social consensus about the revendications. The government had no responsibility to negotiate and just reacted to the street pressure. Thanks to the power of the algorithm, the speed of the movement formation has a price to pay; the gains are exclusively short-term. According to Boyer et al. (2020), the movement also had a counter-intuitive effect in the European elections of May 2019, when the vote share for the government party increased. The authors observe that "exposing a life zone to a blockade increases its vote share for LREM by 0.4 percentage points (a 1.7% increase relative to the national average). This effect is reminiscent of other historical events, such as the elections following May 1968 in France". In fact, "blockades did push more LREM supporters into voting for LREM, rather than more LREM opponents into abstaining from voting" (BOYER et al., 2020b).

This social crisis, however legitimate it may be, has highlighted some flagrant weaknesses of the French republican model (low political representativeness among others), without making it possible to respond to the numerous demands of the demonstrators, and at an enormous material and human cost for the society. This movement, initiated on social networks, seems to have benefited from the critical magnifying glass effect that these communication channels allow (groups, filter bubbles): for several months, the movement of yellow vests, although poorly organized, imposed its media and therefore political agenda on the government, and for some weeks (at the start of the movement) some influent commentators described the situation as a kind of Coup d'Etat which could have brought down the elected government and the republican institutions. This impact of Facebook on the public agenda is also identified by a Spanish study realized in 2018: "we find that when Facebook is a relevant news referral,

[14] https://www.ted.com/talks/zeynep_tufekci_online_social_change_easy_to_organize_hard_to_win?language=fr

people are less likely to mention the top MIPs -most important problems- for a representative sample of the Spanish population" (CARDENAL; GALAIS; MAJÓ-VÁZQUEZ, 2018). However, the movement quickly ran out of steam, due to a growing divergence in its demands within its members first, and a growing gap with the aspirations of the rest of society (to the great difference of the liberal demonstrations of May 1968). Almost 3 years after, the legacy of the movement is very weak in terms of social changes, and in comparison with the human and material costs incurred to obtain these benefits. The movement participated to greater polarization of the French society until now, and a kind of new conservatism attitude of the French politicians (afraid to implement new reforms).

4.6. Conclusion

According to the literature, implementing the new 2018 Facebook algorithm had a powerful impact on the digital connection of citizens who would not have had the occasion to meet otherwise through the platform's "groups." These connections strongly fostered the emergence of the French Yellow Vests social movement in the streets. The mechanism worked by incentivizing users to build a group where they feel ideologically comfortable. There is evidence that the influencing effect did not change political preferences but changed attitudes. For societies, it's essential to ask whether this movement, which appeared thanks to social media and imposed its political agenda for months, had a representative legitimacy and if the mobilization's results were worth the costs. The literature suggests that the French elected government had no control of the movement that potentially could create opportunistic behaviors to earn some advantages by violence rather than by democratic elections. Also, the movement gains for protesters were short-run, not persisting in time. This scenery raises concerns on the representativity of democracies since the movement probably would not have emerged if its members had a better feeling of political representation. Moreover, democracies

are suffering a lack of social media regulation: the only current form of "regulation" is a kind of private self-regulation by the platforms themselves, under their own "politics". It seems essential for societies to be able to punish digital actions which would have been punished in real life and have better control over the algorithms used by the social media: regulation based on three pillars (i) broader transparency of the tech companies over their algorithm and politics, (ii) more vital legislation by the elected government to regulate them, and (iii) a civil society which plays its role of controller of this ecosystem.

5. Final general considerations

While they have allowed democratization of access to information, new media technologies have been accused of reducing the specter and quality of political debate.

On the one hand, the literature review showed that the "social bubbles" are a risk for democracies and that Facebook's news feed ranking algorithm, for promoting personalized publications that are tailored to the predicted needs and wishes of each user, impacts the creation of these "opinion-ghettos". However, the users are also responsible for the creation of these "social bubbles", given their behavior on the social network: sensible to emotions, homophily, conformism, ipsedixitism. Geschke et al. (2018) propose to work on the three levels of filters that were studied: (i) on an individual level, it seems important to increase the media knowledge of the users (how the filters work, which impacts, …); (ii) on a social level, they highlight the necessity to create new forms of alternative debates where the best consensus would be sought; (iii) they hope that a part of serendipity will soon be introduced again in the algorithms (Geschke, Lorenz, and Holtz 2018).

On the other hand, disinformation and illicit profiling strategies become a national security issue as they undermine the legitimacy of the electoral processes in our democracies. In this regard, fake news could serve as a tool for disinformation campaigns on a massive scale: the intentional dissemination of false information for influencing the opinions or policies of the receiving audience. Faced these new challenges, the European Union has launched a "pack" of legislative measures and a co-regulation system with the social media platforms to combat the disinformation and illicit profiling phenomena in the electoral context, which seems to be, until now, the most "avant-gardist" and efficient approach, in the current state of the art and according to the literature. However, progress and efforts still need to be made by all actors to

follow the race for innovation launched by the protagonists of disinformation. In spite of all possible efforts, technology is always a head start over legislation.

The irruption of the French Yellow Vests social movement in 2018 has largely been facilitated by the implementation of a new algorithm by Facebook which encouraged the information exchanged between users sharing common points, who would not have had the occasion to meet otherwise through these "Facebook groups". This major social event is a good illustration of the impacts in the real life of the phenomenon described in the two first articles: these connections strongly fostered the emergence of the French Yellow Vests social movement in the streets. The mechanism worked by incentivizing users to build a group where they feel ideologically comfortable. There is evidence that the influencing effect did not change political preferences but changed attitudes. For societies, it is essential to ask whether this movement, which appeared thanks to social media and imposed its political agenda for months, had a representative legitimacy and if the mobilization's results were worth the costs. The literature suggests that the French elected government had no control of the movement that potentially could create opportunistic behaviors to earn some advantages by violence rather than by democratic elections. Also, the movement gains for protesters were short-run, not persisting in time. This scenery raises concerns about the representativity of democracies since the movement probably would not have emerged if its members had a better feeling of political representation. Moreover, democracies are suffering a lack of social media regulation: the only current form of "regulation" is a kind of private self-regulation by the platforms themselves, under their own "politics".

This is why, in spite of necessary (but not sufficient) legislative measures, it seems urgent to continue developing alternative types of measures, in the short, and long terms. It seems essential for societies to be able to punish digital actions which would have been punished in real life and have better control over the algorithms used by the social media: regulation based

on three pillars (i) broader transparency of the tech companies over their algorithm and politics, (ii) more vital legislation by the elected government to regulate them, and (iii) a civil society which plays its role of controller of this ecosystem.

This study highlighted the limits of the European legislation which once had been a model exported to other countries over the world (LGPD in Brazil for instance). The European institutions prepared a new legislative pack of measures to adapt to these new challenges. On July 5, 2022, the European Parliament adopted the final versions of two new "avant-gardist" texts: the Digital Markets Act (DMA) and the Digital Services Act (DSA), which should come into force in 2023 and whose general principles are to strengthen competition law in the face of the GAFAM "monopoly", and to condemn in digital life everything that is reprehensible in real life. The European institutions introduce the concept of systemic/essential platforms to identify the digital platforms they will directly supervise and regulate (the smaller or national platforms staying under the supervision of the member-States): (i) a strong economic position (at least 7.5 billion euros of turnover achieved in the European Economic Area or a market capitalization/market value of at least 75 billion euros with activity in at least three States members), (ii) the control of an "essential platform service" (search engine, social network, messaging, online marketplace, etc.) used by at least 45 million Europeans per month and at least 10,000 professionals per year in the Union; (iii) exceeding these thresholds during the previous three years ("solid and durable" nature of its position on the market). The DMA legislation must limit the many advantages through which these systemic/essential platforms can maintain market dominance. Faced with their sometimes-unfair practices, the text aims to impose a certain number of obligations ex-ante: today, the fines sanctioning breaches of competition law often come late, which does not encourage companies to change their deep behavior. These essential platforms will no longer have the right to promote their own other

services; the main messaging services will have to be interoperable with all their competitors; finally, personalized micro-targeting advertising will require explicit consent from the user.

The DSA legislation seeks to limit the dissemination of illegal content (incitement to hatred or violence, harassment, child pornography, the apology for terrorism, etc.) and the sale of illegal products online applying a strong general principle: to condemn in digital life everything that is reprehensible in real life. While the DSA does not call into question the limited liability of platforms with regard to the illegal content and products they host (the notion of "passive host"), they must on the other hand offer a tool allowing users to report them: once this report has been made, they will then have to remove these contents and products or quickly deactivate access to them. Platforms have an obligation to cooperate with "trusted flaggers" of the civil society. These are bodies, associations or individuals labeled within each State by virtue of their expertise and who will have their notifications processed as a priority. The DSA also prohibits targeting people with online advertisements based on their religion, sexual preferences, health information, or political beliefs. Targeted advertising is also prohibited vis-à-vis minors. While the European institutions did not achieve to force the platforms to publish their algorithms, the latter must explain the operation of their recommendation systems, which reinforce the visibility of certain content for a user according to his personal interests (very large platforms are also required to offer users an alternative recommendation system that is not based on their profiling).

To give the European Union the means to enforce all of these rules, the European Commission could recruit more than 200 people and create a high-level European center on the transparency of algorithms.[15]

---

[15] https://www.touteleurope.eu/societe/numerique-que-sont-le-dma-et-le-dsa-les-reglements-europeens-qui-veulent-reguler-internet/

These initiatives seem to be the first step toward the three-pillar ecosystem that this study recommends. It is likely that this new stage of European regulation of the Internet in its economic area will serve as a model for other countries in the coming years.

Observation:

- An extract of the 2nd paper "*The European regulation designed to protect the 2019 elections from disinformation and micro-targeting advertising*"

    - was published as a pre-print on the NUPRI website - USP's International Relations Research Center (http://nupri.prp.usp.br/nupri-working-paper-04/),

    - was presented at the 8th Meeting of ABRI – Brazilian Association of International Relations

      (https://www.encontro2021.abri.org.br/trabalho/view?ID_TRABALHO=5718)

    - -is in press to be published in the journal Artificial Intelligence: democracy and impacts 2021

      (http://hseminar.webhostusp.sti.usp.br/2021/?page_id=1262&lang=pt),

- An extract of the 3rd paper "*Investigating potential effects of the new Facebook algorithm on the 2019 French Yellow Vests movement*" was presented at the 8th Meeting of ABRI – Brazilian Association of International Relations (section 4)

   (https://www.encontro2021.abri.org.br/trabalho/view?ID_TRABALHO=5718)

6. References

- Alemanno, Alberto. 2018. "How to Counter Fake News? A Taxonomy of Anti-Fake News Approaches." "Tackling Online Disinformation.Pdf." https://www.cambridge.org/core/services/aop-cambridge-core/content/view/EA53D30745F601834218DDD7DB90950A/S1867299X18000120a.pdf/div-class-title-how-to-counter-fake-news-a-taxonomy-of-anti-fake-news-approaches-div.pdf (March 30, 2020). *European Journal of Risk Regulation* 9(1): p.1–5.

- Algan, Y. et al. Qui sont les Gilets jaunes et leurs soutiens ? Observatoire du Bien-être, v. 2019– 03, 14 fev. 2019.

- Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31(2): p. 211 e 36.

- Alvarez, R. Michael, and Jonathan Nagler. 2000. "A New Approach for Modelling Strategic Voting in Multiparty Elections." *British Journal of Political Science* 30(1): p. 57.

- Avaaz. 2019. *Far Right Networks of Deception. Avaaz Investigation Uncovers Flood of Disinformation, Triggering Shutdown of Facebook Pages with over 500 Million Views Ahead of EU Elections.* https://secure.avaaz.org/campaign/en/disinfo_network_report/ (October 10, 2019). p. 5A.

- Balkin, Jack M. 2018. "Free Speech Is a Triangle." *Columbia Law Review* 118: 47.p. 2011-2014.

- Bakir, Vian, and Andrew McStay. 2018. "Fake News and The Economy of Emotions: Problems, Causes, Solutions." *Digital Journalism* 6(2): 154–75.

- Bakshy, E., S. Messing, and L. A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348(6239): 1130–32.

- Barbu, Oana. 2014. "Advertising, Microtargeting and Social Media." *Procedia - Social and Behavioral Sciences* 163: p. 44.

- Beam, Michael A., Myiah J. Hutchens, and Jay D. Hmielowski. 2018. "Facebook News and (de)Polarization: Reinforcing Spirals in the 2016 US Election." *Information, Communication & Society* 21(7): 940–58.

- Bode, Leticia, and Emily K. Vraga. 2015. "In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media: In Related News." *Journal of Communication* 65(4): p. 619 e 631.

- Boffey, Daniel. 2019. "241m Europeans 'May Have Received Russian Linked

Disinformation.'" *The Guardian*: 3.p. 2.

- BOYER, P. C. et al. Les déterminants de la mobilisation des Gilets jaunes. Revue économique, v. 71, n. 1, p. 109, 2020a.

- BOYER, P. C. et al. The Gilets Jaunes : Offline and Online. p. 50, maio 2020b.

- BRANCACCIO, F. Introduction. South Atlantic Quarterly, v. 119, n. 4, p. 846–855, 1 out. 2020.

- Bruns, Axel. 2019a. "Filter Bubble." *Internet Policy Review* 8(4). https://policyreview.info/node/1426 (August 4, 2020).

- ———. 2019b. "It's Not the Technology, Stupid: How the 'Echo Chamber' and 'Filter Bubble' Metaphors Have Failed Us." In Madrid, Spain, 12.

- BOUTYLINE A.; WILLER R.; The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks; Political Psychology, Vol. 38, No. 3, 2017; doi: 10.1111/pops.12337; 2017

- CARDENAL, A. S.; GALAIS, C.; MAJÓ-VÁZQUEZ, S. Is Facebook Eroding the Public Agenda? Evidence From Survey and Web-Tracking Data. International Journal of Public Opinion Research, 15 dez. 2018.

- Cadwalladr, Carole, and Duncan Campbell. 2019. "Revealed: Facebook's Global Lobbying against Data Privacy Laws." *The Guardian*. https://www.theguardian.com/technology/2019/mar/02/facebook-global-lobbying-campaign-against-data-privacy-laws-investment (October 10, 2019).p. 1.

- "CELEX_52018PC0636R(01)_EN_TXT.Pdf."

- "CELEX_52019JC0012_EN_TXT.Pdf."

- Chitra, Uthsav, and Christopher Musco. 2019. "Understanding Filter Bubbles and Polarization in Social Networks." *arXiv:1906.08772 [physics]*. http://arxiv.org/abs/1906.08772 (August 3, 2020).

- Clayton, Katherine et al. 2019. "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media." *Political Behavior*. http://link.springer.com/10.1007/s11109-019-09533-0 (March 30, 2020).p. 1.

- Clement, J. 2020. "Digital Users Worldwide 2020." *Statista*. https://www.statista.com/statistics/617136/digital-population-worldwide/#:~:text=How%20many%20people%20use%20the,in%20terms%20of%20internet%20users. (August 3, 2020).

- Condliffe, Jamie. 2017. "The Right-Wing Propaganda Machine May Not Be as Smart as You Think.".p. 1.

- Cooper, Paige. 2020. "Algorithme Facebook : comment fonctionne-t-il en 2020 ? Comment l'utiliser à votre avantage ?" : 16.

- Corporate Europe Observatory. 2018. *Post-Scandal Facebook: Will the EU Stop Treating the Tech Giant as a Trusted Partner?* https://corporateeurope.org/en/power-lobbies/2018/05/post-scandal-facebook-will-eu-stop-treating-tech-giant-trusted-partner (October 10, 2019).p. 2.

- Dingler, Tilman, Ashris Choudhury, and Vassilis Kostakos. 2018. "Biased Bots: Conversational Agents to Overcome Polarization." In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp '18*, Singapore, Singapore. http://dl.acm.org/citation.cfm?doid=3267305.3274189 (August 4, 2020).

- Dommett, Katharine, and Luke Temple. 2018. "Digital Campaigning: The Rise of Facebook and Satellite Campaigns." *Parliamentary Affairs* 71(suppl_1): p. 194 e 195.

- DORMAGEN, J.-Y.; PION, G. « Gilets jaunes », combien de divisions ? - Une tentative de comptage. Le Monde Diplomatique, fev. 2021.

- EGEBARK, J.; EKSTRÖM, M. Liking what others "Like": using Facebook to identify determinants of conformity. Experimental Economics, v. 21, n. 4, p. 793–814, dez. 2018.

- ELABE and INSTITUT MONTAIGNE Les Gilets jaunes: la partie émergée de la crise sociale française ?, https://www.institutmontaigne.org/analyses/les-gilets-jaunes-la-partie-emergee-de-la-crise-sociale-francaise mar. 2019.

- ENIKOLOPOV, R.; MAKARIN, A.; PETROVA, M. Social Media and Protest Participation: Evidence From Russia. Econometrica, v. 88, n. 4, p. 1479–1514, 2020.

- ENJOLRAS, B.; STEEN-JOHNSEN, K.; WOLLEBÆK, D. Social media and mobilization to offline demonstrations: Transcending participatory divides? New Media & Society, v. 15, n. 6, p. 890–908, set. 2013.

- European Commission. 2017. *Proposal for a Regulation of the European Parliament and of the Council Concerning the Respect for Private Life and the Protection of Personal Data in Electronic Communications and Repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications).* https://about.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/ (April 5, 2020).
- ———. 2018a. A contribution from the European Commission to the Leaders' meeting in Salzburg on 19-20 September 2018. Free and Fair elections *Commission Guidance on the Application of Union Data Protection Law in the Electoral Context. Free and Fair Elections.*
- ———. 2018b. *EU Code of Practice on Disinformation.*
- ———. 2018c. *Proposal for a Regulation of the European Parliament and of the Council Amending Regulation (EU, Euratom) No 1141/2014 as Regards a Verification Procedure Related to Infringements of Rules on the Protection of Personal Data in the Context of Elections to the European Parliament.*

- ———. 2018d. Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions *Tackling Online Disinformation: A European Approach*.

- European Commission, and High Representative of the Union for Foreign Affairs and Security Policy. 2019. Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions *Report on the Implementation of the Action Plan Against Disinformation*.

- European Parliament. 2009. Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services (European Parliament) *Directive 2002/58/EC Concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector and Regulation (EC) No 2006/2004 on Cooperation between National Authorities Responsible for the Enforcement of Consumer Protection Laws*.
- ———. 2014. *Regulation (EU, Euratom) No 1141/2014 of the European Parliament and of the Council of 22 October 2014 on the Statute and Funding of European Political Parties and European Political Foundations*.
- ———. 2016. GDPR *Regulation of the European Parliament and of the Council - of 27 April 2016 - on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*.
- ———. 2018. *Regulation (EU, Euratom) 2018/673 of the European Parliament and of the Council - of 3 May 2018 - Amending Regulation (EU, Euratom) No 1141/2014 on the Statute and Funding of European Political Parties and European Political Foundations*.
- ———. 2019. *Regulation (EU, Euratom) 2019/493 of the European Parliament and of the Council - of 25 March 2019 - Amending Regulation (EU, Euratom) No 1141/2014 as Regards a Verification Procedure Related to Infringements of Rules on the Protection of Personal Data in the Context of Elections to the European Parliament*.

- Facebook. 2018. "Facebook Ad Library." https://www.facebook.com/ads/library/ (April 3, 2020).

- "Facebook, Inc. Class A Common Stock (FB) Financials _ Nasdaq." 2019. *Nasdaq Financial Statements*. https://www.nasdaq.com/market-activity/stocks/fb/financials (August 3, 2020).

- Facebook, Sara Su. 2017. "New Test With Related Articles." : 5.

- France 24. 2018. "Ce Que Contient La Loi Française Contre Les 'Fake News.'" https://www.france24.com/fr/20181121-loi-contre-fake-news-definitivement-adoptee-macron-election (April 5, 2020).

- FU, P.-W.; WU, C.-C.; CHO, Y.-J. What makes users share content on facebook? Compatibility among psychological incentive, social capital focus, and content type. Computers in Human Behavior, v. 67, p. 23–32, out. 2016.

- Gabriel, Mariya, and Julian King. 2019. "Désinformation : Facebook, Twitter et Google doivent mieux faire." : p.1.

- Geli, Charles. 2018. "Byung-Chul Han: 'Hoje o indivíduo se explora e acredita que isso é realização.'" https://brasil.elpais.com/brasil/2018/02/07/cultura/1517989873_086219.html?id_externo_rsoc=FB_CC.

- Geschke, Daniel, Jan Lorenz, and Peter Holtz. 2018. "The Triple-Filter Bubble: Using Agent-Based Modelling to Test a Meta-Theoretical Framework for the Emergence of Filter Bubbles and Echo Chambers." *British Journal of Social Psychology* 58(1): 129–49.

- GODEFROID, P. Polémique et mouvements sociaux. Étude de cas sur la crise des gilets jaunes. p. 116, 2020.

- González, Roberto J. 2017. "Hacking the Citizenry?: Personality Profiling, 'Big Data' and the Election of Donald Trump." *Anthropology Today* 33(3): p. 11.

- GRANOVETTER, M. S. The Strength of Weak Ties. p. 22, 1973.

- Haciyakupoglu, Gulizar, Jennifer Yang Hui, V S Suguna, and Dymples Leong. 2018. "Countering Fake News." *Nanyang Technological University* S. Rajaratnam School of International Studies: 24. p. 2, 3, 6 e 10.
  -
- Herlocker, Jonathan L., Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. "Evaluating Collaborative Filtering Recommender Systems." *ACM Transactions on Information Systems (TOIS)* 22(1): 5–53.

- Hinds, Shari. 2019. "The European Union Approach to Disinformation and Misinformation The Case of the 2019 European Parliament Elections." University of Strasbourg European Master's Degree in Human Rights and Democratisation. p. 7.

- John, Nicholas A., and Shira Dvir-Gvirsman. 2015. "'I Don't Like You Any More': Facebook Unfriending by Israelis During the Israel-Gaza Conflict of 2014: Facebook Unfriending in Israel-Gaza Conflict 2014." *Journal of Communication* 65(6): 953–74.

- Juncker, Jean-Claude. 2018. *State of the Union 2018. The Hour of European Sovereignty.*

- Koop, Christel and Lodge, Martin (2015) What is regulation? An interdisciplinary concept analysis. Regulation and Governance. ISSN 1748-5983. p. 1-11

- Levy, Ro'ee. 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." American Economic Review, 111 (3): 831-70. DOI: 10.1257/aer.20191777. p. 831

- Lewandowsky, Stephan et al. 2012. "Misinformation and Its Correction: Continued

Influence and Successful Debiasing." *Psychological Science in the Public Interest* 13(3): p. 106.

- •

- Marret, Christophe. 2020. "The Impact of Social Media on Elections." *NUPRI-USP Working Papers*

- Mcclurg, Scott D. 2003. "Social Networks and Political Participation: The Role of Social Interaction in Explaining Political Participation." *POLITICAL RESEARCH QUARTERLY*: 16.

- McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27(1): 415–44.

- Mena, Paul. 2019. "Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook." *Policy & Internet*: poi3.214. p. 1.

- Milan, S, and C Agosti. 2019. "Personalisation Algorithms and Elections: Breaking Free of the Filter Bubble | Internet Policy Review." : 8.

- Monteiro Borges, Priscila Monteiro, and Renira Rampazzo Gambarato. 2019. "The Role of Beliefs and Behavior on Facebook: A Semiotic Approach to Algorithms, Fake News, and Transmedia Journalism." *International Journal of Communication* 13: 16.

- MOSSERI, A. Bringing People Closer Together. p. 9, 11 jan. 2018. https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/ (August 1, 2020).

- "Networks_Report_Update_Page_July_2019 Avaaz.Pdf."

- Nyhan, Brendan, and Jason Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* (32): p. 303

- OELDORF-HIRSCH, A.; SUNDAR, S. S. Posting, commenting, and tagging: Effects of sharing news stories on Facebook. Computers in Human Behavior, v. 44, p. 240–249, 2014.

- Owen, Laura Hazard. 2019. "One Year in, Facebook's Big Algorithm Change Has Spurred an Angry, Fox News-Dominated — and Very Engaged! — News Feed." *Nieman Foundation at Harvard*: 9.

- Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. London, UK: Penguin.

- Pennycook, Gordon, Adam Bear, Evan T. Collins, and David G. Rand. 2019. "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings." *Management Science*:

mnsc.2019.3478. p. 1.

- Posetti, Julie, and Alice Matthews. 2018. *A Short Guide to History of Fake News and Disinformation*. International Center for Journalists. p. 1-2

- Potemkina, Olga. 2019. "Regulation of Social Networks in the Upcoming Elections to the European Parliament." *Contemporary Europe, Institute of Europe of Russian Academy Sciences* 2(88): p. 2, 3 e 11.

- Rastegarpanah, Bashir, Krishna P. Gummadi, and Mark Crovella. 2019. "Fighting Fire with Fire: Using Antidote Data to Improve Polarization and Fairness of Recommender Systems." In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, Melbourne VIC Australia: ACM, 231–39. https://dl.acm.org/doi/10.1145/3289600.3291002 (August 4, 2020).

- Rees, Emma. 2017. "What Made the Difference for Labour? Ordinary People Knocking on Doors." *The Guardian*. https://www.theguardian.com/commentisfree/2017/jun/12/labour-knocking-on-doors-jeremy-corbyn-momentum (May 4, 2020). p. 2.

- ROGERS, R.; NIEDERER, S. (EDS.). The Politics of Social Media Manipulation. Nieuwe Prinsengracht 89   1018 VR   Amsterdam   Nederland: Amsterdam University Press, 2020.

- SafeGuard Cyber. 2019. *Contactless Actions against the Enemy*. SafeGuard Cyber. https://www.safeguardcyber.com (October 10, 2019).

- Scott, Mark. 2019. "Inside Facebook's European Election War Room." *Politico EU*. https://www.politico.eu/article/facebook-european-election-war-room-dublin-political-advertising-misinformation-mark-zuckerberg/ (October 10, 2019). p. 1.

- Seargeant, Philip, and Caroline Tagg. 2019. "Social Media and the Future of Open Debate: A User-Oriented Approach to Facebook's Filter Bubble Conundrum." *Discourse, Context & Media* 27: 41–48.

- Shapiro, C. and H. Varian. 1999. *Information Rule*. Harvard Business School Press. p. 3-7

- Sunstein, Cass. 2001. *Echo Chambers: Bush v. Gore, Impeachment, and Beyond*. Princeton, N.J.: Princeton University Press.

- UK Parliament, and BBC. 2017. *Written Evidence Submitted by the BBC*. UK Parliament. data.parliament.uk/WrittenEvidence/CommitteeEvidence.svc/EvidenceDocument/Culture, Media and Sport/Fake News/written/48758.html (October 10, 2019).

- UK Parliament, and The Guardian. 2017. *Written Evidence Submitted by Guardian News & Media*. UK Parliament.

data.parliament.uk/WrittenEvidence/CommitteeEvidence.svc/EvidenceDocument/Culture, Media and Sport/Fake News/written/48259.html (October 10, 2019).

- Viscusi WK, Harrington JE, and Vernon JM. 2005. Economics of Regulation and Antitrust. 4th edition. The MIT Press. p. 358.

- Weedon, Jen, William Nuland, and Alex Stamos. 2017. *Information Operations and Facebook*. . Corporate. p. 4.

- "World Population Clock_ 7.8 Billion People (2020) - Worldometer." 2020. *World Population Clock*. https://www.worldometers.info/world-population/ (August 3, 2020).

- Zuckerberg, Mark. 2018. "Bringing People Closer Together." https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/ (August 1, 2020).