

Mechanisms that Shape Social Media and their Impact on Society

Report on the State-of-the-Art in Research

Authors (alphabetical order):

Mikolaj Biesaga (University of Warsaw, Warsaw, Poland),
Guido Caldarelli (CNR, ISC Rome, Italy)
Salvatore Iaconesi (HER, Rome, Italy)
Fiorenza Lipparini (PlusValue, London, UK)
Marcin Napiorkowski (University of Warsaw, Warsaw, Poland)
Andrzej Nowak (University of Warsaw, Warsaw, Poland)
Oriana Persico (HER, Rome, Italy)
Szymon Talaga (University of Warsaw, Warsaw, Poland)
Erika Widegren (Re-Imagine Europa, Brussels, Belgium)
Karolina Ziembowicz (University of Warsaw, Warsaw, Poland)

This report is part of a project that has received funding by the European Union's Horizon 2020 (SMART 2017/0090) research and innovation programme.

The content of this report reflects only the authors' view. The contents of this report should not be interpreted as expressing a position of the European Commission.

Contents

ACKNOWLEDGEMENTS	4
EXECUTIVE SUMMARY	5
Introduction	5
Overview	5
Communication ecosystem is the central fabric of society	6
The Digital Revolution	6
Flow of information in the age of social media	7
Understanding how information spreads on social media	8
The role of narratives	9
CASE STUDY: Pilot experiment - reactions on debunking in the new media	10
Future research directions	11
Recommendations for practitioners and policymakers	12
Conclusions	14
DEFINITIONS	16
PART 1 – STATE-OF-THE-ART IN RESEARCH	17
Chapter 1 - INTRODUCTION	18
1.1 Introduction	18
1.2 Flow of Information in the Age of Social Media	19
1.3 The European Commission’s Actions to Combat Disinformation	20
1.4 Conclusions	21
Chapter 2 – QUANTITATIVE APPROACH	24
Understanding How Information Spreads on Social Media	24
2.1 Introduction	24
2.2 Science of Networks	24
2.3 Conclusions	31
Chapter 3 – QUALITATIVE APPROACHES	32
Narratives and Narratology	32
3.1 Introduction	32
3.2 Narratology	34
3.3 Conclusions	38
Chapter 4 – EMERGING THREATS AND SOLUTIONS	41
4.1 Introduction	41
4.2 Emerging Threats	41
4.3 Emerging Solutions	46
4.4 Computational Approaches	48
4.5 ORGANIC FACT CHECKING	50
4.6 MIXED APPROACHES	51
4.7 Conclusions	53
PART 2 – ORIGINAL RESEARCH	54
Chapter 5 – THE SMART PILOT OBSERVATORY	55

5.1	Introduction	55
5.2	The SMART Pilot Observatory	55
Chapter 6 –		66
COUNTERACTING THE SPREAD OF DISINFORMATION		66
PLATFORM TESTS		66
6.1	Introduction	66
6.2	Platform Tests	66
6.3	Computational Tests	68
	CASE STUDY 1: Climate change and the global economy	89
	CASE STUDY 2: Pilot experiment: Reactions on debunking in the new media	90
	CASE STUDY 3: Brexit: The perspective of EU, British and Russian Press	92
Chapter 7 – AN ARTISTIC APPROACH		135
Counteracting the Spread of Online Disinformation		135
7.1	Introduction	135
7.2	The Data Theatre Performance	135
7.3	To Fake or not To Fake	136
7.4	Feedback and Conclusions	140
PART 3 – CONCLUSIONS		143
Future Areas of Research and Policy Recommendations		143
Chapter 8 – CONCLUSIONS		144
8.1	Introduction	144
8.2	Areas for further research and consideration	145
8.3	Policy recommendations	148
8.4	Final Statements	152
References		153

ACKNOWLEDGEMENTS

We would like to express our deepest appreciation to all those who supported and helped us with ideas, information and suggestions for this report. This report is based on an eighteen-month effort to capture the most interesting and novel research work in this field and build original research to test the conclusions of this initial work. This report would not have been possible without the input and advice of numerous experts from different disciplines and stakeholders.

A special gratitude goes to the experts who agreed to be interviewed for this report (alphabetical order):

- Luca De Biase, Editor of Nova, Sole 24Ore
- Jens Degett, President of the European Union Science Journalist Association (EUSJA)
- Fosca Gianotti, Director of Research at CNR
- Christophe Leclercq, Founder of EURACTIV and Executive Chairman at Fondation EURACTIV
- Piero Molino, PhD, Research Scientist, Uber AI Labs
- Dino Pedreschi, Professor of Computer Science at the University of Pisa
- Guido Vetere, PhD, Professor of Artificial Intelligence at the Faculty of Applied Sciences and Technologies of the University of Guglielmo Marconi in Italy

We would also like to thank all the experts from the SoBigData++ Consortium and those who contributed to the report through the participation to the two workshops that were organised to brainstorm initial ideas for the report: Workshop on “How information spreads on social media and the possible impact on the European Elections” - *Wednesday 27th of June 2018 – European Parliament, Brussels*; Workshop on “Identity, Communication and Power: European Narratives and Insights from Recent Research” - *Tuesday 25th of September – European Parliament, Brussels*. In alphabetical order of their first name, a big thank you goes out to: Alberto Mittesteiner (Sky Group), Alice Andreuzzi (Catchy), Alice Pontiki (DG COMM – European Commission), Anne-Marijn Epker (Forum on European Culture), Arya Sofia Meranto (RAND Europe), Axelle Devaux (RAND Europe), Beatriz Marin Garcia (DG COMM – European Commission), Bernd Halling (Bayer), Celine Geissmann (Microsoft), Charles de Marcilly (European Political Strategy Centre), Christal Morehouse (Open Society Foundations), Dana Manescu (European Commission -DG COMM), Daniel Friedlander (Sky), David Chavalarias (Complex Systems Institute of Paris Ile-de-France), Didrik de Schaetzen (ALDE Party), Elisabeth Ardaillon-Poirier (European Political Strategy Centre of the European Commission), Elizabeth Crossick (RELX group), Federico Ottavio Reho (Martens Centre for European Studies), François Balate (European Youth Forum), Geraldine Barry (European Commission, Joint Research Centre), Gianni Riotta (Data Lab at Luiss University), Giovanni La Placa (DG Joint Research Centre (JRC) of the European Commission), Giuseppe Porcaro (Bruegel), Jan Michal (Representations in Member States at the European Commission), Jan M. Zajac (SoTrender), Jennifer Erin Kavanagh (RAND Corporation), Jeremy Bray (European Commission's Science Advice Mechanism (SAM) Unit at the Directorate General for Research and Innovation), Joe Lynam (BBC Correspondent – European Commission), Johannes Klumpers (Scientific Advice Mechanism Unit (SAM) in the European Commission), John Lisney (European Youth Forum), Julian Knott (Organisation for Economic Co-operation and Development (OECD)), Kacper Olczyk (Erasmus Student Network), Karolina Wozniak (Social Media Coordinator at the European Parliament), Konstantinos Kallergis (European Parliament's Spokesperson's Unit), MEP Lambert van Nistelrooij (Member of the European Parliament), Laura Smillie (JRC – European Commission), Luca Tacchetti (Catchy), Manon Brunel (Cabinet of Nicolas Hulot, French State Minister, Minister of ecological and inclusive transition), MEP Maria João Rodrigues (Member of the European Parliament, Vice-President of the S&D Group and President of FEPS – European Foundation of Progressive), Matthew C. MacWilliams (University of Massachusetts Amherst), Mikel Landabaso (DG COMM – European Commission), Naja Bentzen (External Policies Unit in the Members' Research Service, EPRS, European Parliament), Petros Fassoulas (European Movement International), Pierre Mejlak (European Commission's Directorate-General for Communication), Robert Madelin (Fipra International), Robert Soltyk (Directorate General for Communication of the European Commission), Sixtine Bouygues (Deputy Director-General for Communication at the European Commission), Sonja Afanasjeva (Young European Federalists (JEF Europe)), Stefan Moritz (European Entrepreneurs CEA-PME), Stefano Battiston (University of Zurich), Svetla Tanova-Encke (European Science-Media Hub, STOA – European Parliament), Szymon Talaga (University of Warsaw), Theo Karapiperis (STOA - European Parliamentary Research Service (EPRS)), Tomi Huhtanen (Wilfried Martens Centre for European Studies), Wolfgang Tillmans (Photographer and Artist).

A special thank you goes to the project officers from DG COMM, Paolo Cesarini, Ralph Dum, and Anni Hellman, for their continuous support and input during this period.

EXECUTIVE SUMMARY

Introduction

Just like a virus, an idea spreads across societies impacting individuals and societal structures. It mutates constantly, adapting rapidly to the changing circumstances, always ready to take part in a ruthless arms race between various forms and versions. Just like viruses, ideas also need a host to survive and multiply. They feed on globalisation and mobility, through suggestive narratives, metaphors and cognitive models ideas infect our society, our media, and finally – our brains, merging into an indistinguishable whole with our cultural and psychological DNA. With COVID19 currently raging in Europe it is all too clear that we must better understand the patterns of diffusion of both viruses and ideas. If the current pandemic teaches us anything, it is that effective communication is both the deadly strategy of the virus itself and our only hope of defeating it.

The information ecosystem of any society is central to the creation and maintaining both of power and counterpower, because societies organise across shared identities and beliefs. Society, like any other entity, is in constant evolution, developing in parallel with changing environmental transformations and adapting to the world around it.

Today we are living in the midst of a revolution in communication technologies that affect the way people feel, think, and behave. The arrival of digital technologies changed the world. Fundamental concepts such as *space* and *time* shifted meaning making it possible to communicate instantly across the globe and be in many places at once. At first, these powerful technologies were only seen as progress as industry after industry transformed. Recently however, this digital optimism has been gradually giving place to dystopian visions of total surveillance, growing inequality and massive unemployment. The current pandemic will certainly exacerbate the overwhelming sense of fear and inequality.

We are still far from having a full understanding of the dynamics through which information spreads across digital media. Social media have changed the way we access, produce and share information, transforming traditional mass-media into a system of intertwined horizontal communication networks. The imagined free, equal worldwide network, promised at the dawn of the internet and social media age, has become a software-based cybersphere, driven by economic and geopolitical interests. This fact has significantly altered how contents are conveyed and spread, exposing the media system to distortions and manipulations whose origin is often difficult to ascertain.

The aim of this report is to provide a systematic, empirical and theoretical analysis to better understand the mechanisms that shape social media and their impact on society.

The decisions we will take in the coming years will be crucial in defining how the digital media ecosystem develops and what the future of democracy will look like. Remembering the words of Winston Churchill: "We shape our buildings; thereafter they shape us", are we capable of creating a European model for the digital world in line with the European values of human dignity, freedom, democracy, equality, human rights and peace?

Overview

At the end of 2016, Oxford Dictionaries has selected "post-truth" as the international word of the year, after the contentious "Brexit" referendum and an equally divisive U.S. presidential election brought the impact of social media on society at the centre of the political debate. Digital online media, and social media specifically, became the villain in the story of democracy and governments across the globe started developing policies and actions plans to combat "disinformation".

This report wants to take a step back from focussing on specific symptoms of our changing media ecosystem and provide an interdisciplinary and comprehensive study on the stock of ongoing research on the mechanisms that shape social media to provide a more systemic and evidence-based approach to this challenge.

The report is divided into three parts. The first part provides a systematic, theoretical and empirical review of the state-of-the-art of ongoing research. From *Science of Networks*, research on *Algorithms* and *Cognitive Bias* (chapter 2) to *Computational Social Science* and *Narratology* (chapter 3) the report creates an overview of the most promising

research from computer science, neurosciences and cognitive sciences, artificial intelligence and network sciences, social psychology, philosophy, behavioural economics and law.

Chapter 4 focusses on the phenomena that stand-out as particularly threatening for democratic and societal processes: the rise of *computational propaganda* and *micro-targeting*, the rise of *deep-fakes* and the *weaponization of the debate* around ways to tackle online disinformation, with a particular focus on the role of bots and automated agents.

The second part reflects the original research done for this report to better understand some of the underlying questions surrounding the understanding of the mechanisms that shape social media and their impact on society. The research was performed by developing a pilot European Observatory and analysing how (dis)information spread around the European elections, with a particular focus on sentiment analysis and narratology.

The research performed demonstrates the important role of narratives in framing facts and information in a package that can be easily transmitted across society and how studying narratives can be the key to better understanding how ideas spread across social media and why certain ideas take hold over others. The use of this methodology in understanding the flow of information during the COVID19 outbreak could be key to finding a different way to inform society and to keep disinformation and misinformation at bay.

Lastly, the report aims to analyse the current situation, look at the most promising areas for future research and suggest some policy recommendation based on the workshops, interviews and collaborations with stakeholders.

Below you will find an overview of the main findings of the report.

Communication ecosystem is the central fabric of society

In his seminal book "Communication Power", Manuel Castells argues that "power does not reside in institutions, not even the state or large corporations. It is located in the networks that structure society." Underpinning the importance of understanding the communication fabric of a society to understand how it organises and develops.

Scholars such as Marshall McLuhan, Walter J. Ong or Jack Goody teach us that the history of humanity has always been a history of changing media. There have been three – some argue four – communication revolutions in history. The first big revolutionary technology was *writing*. Transition from orality to scripture enabled synchronising actions of thousands of people, giving birth to new empires, fuelling the agricultural revolution, but also severely disturbing the ancient way of life, as already Plato noted in his *Phaedrus*. It took a lot of time to learn the new definitions of "truth" and "lie", rebuild the networks of trust and adapt to this new world. The second revolutionary technology was *printing* resulting immediately in over a century of religious wars, and undermining in the feudal structures of knowledge and power. After a profound transformation the society has developed new regulations – such as intellectual property laws and defamation laws – to tame the technology. The very idea of the modern nation state has been a political answer to this revolution caused by the printing machine. Similarly, radio and television have shaped the mass propaganda of the 20th century. As we can see, all of these media-driven revolutions have been deeply ambivalent. Each of them has revolutionised communicational possibilities, nurturing science and education and bringing new forms of democracy and participation, but they have also provided new means for propaganda, surveillance and war.

What is evident is that the communication infrastructure is central to the spreading of ideas, creation of identity and community and in the distribution and organisation of power and counterpower on any society.

The Digital Revolution

Now we are in the midst of yet another revolutionary transformation of society resulting from the development of *digital* technologies. These technologies are still in their infancy, but their impact on society is already deep and undeniable. The public at large, the media and some thinkers have a tendency to see them as either black or white.

Evgeny Morozov (2012) in his *The Net Delusion* criticised "the google doctrine" – "the idea that Internet favours the oppressed rather than the oppressor" supported by a "naïve belief in the emancipatory nature of online

communication". His point of view was then a marginal one – a minority report against a predominant "cyber-utopianism" translating Fukuyama's notion of the "end of history" to the language of the new, digital era. Now, after changes in the real world and collective imagination epitomised by the Cambridge Analytica scandal and programmes such as "Black Mirror", Morozov's dark vision is widely accepted by politicians, experts and by the general audience. Most prominent pundits, futurologists and "historians of the future" generally agree that the internet will either destroy democracy or change it into something radically different, such as "dataism" [Harari 2015; Ferguson 2019].

Does this mean the digital revolution is the end of democracy and of democratic freedoms that have been painstakingly conquered in the last century? As any systemic innovation, a new communication model must be just considered a disruptive change, that must be understood and tamed before its positive aspects can be exploited.

A considerable body of evidence has demonstrated that digital media, and in particular social media, is the communication fabric of our lives, for work, for personal connection, for information, for entertainment, for public services, politics and religion [Castells, 2010]. A deeper understanding of the mechanisms and incentives that shape media and their impact on society could thus be crucial in assisting any reform efforts that focus on media legislation, privacy, intellectual property, regulation on new technologies like AI or Blockchain. Since democracy is intimately linked to the free flow of information among its citizens, the regulation of the environment in which it takes place is a fundamental pillar for guaranteeing democratic freedoms. It is not enough, however, to merely react to the ongoing change. **To protect European values and work for the better future, the legislators have to act proactively, not but by harnessing the innovations of yesterday, but by shaping the intellectual landscape of tomorrow. The aim should never be to stop or limit the innovation, but to nurture it and lead in a right direction.**

Flow of information in the age of social media

Since the appearance of Facebook and Twitter, the impact of social networks on the way people consume, produce and interact with information has grown exponentially. The rapid change that the media environment has undergone since the advent of social media can be seen in the quantity of information produced each day. The total amount of data in the world was 4.4 Zettabytes (ZB) in 2013. That is set to rise steeply to 44 ZB by 2020. According to a report from IBM Marketing Cloud [IBM Report 2017] 90% of the data in the world today has been created in the last two years alone, at 2.5 quintillion bytes of data a day.

This can also be seen in changing habits and behaviour of people - and particularly of younger generations - in the way they consume and interact with information. Every action today can be photographed, filmed and distributed to friends, family or beyond. The traditional border between "private" and "public" has been erased, enabling new practices of participation and protest. This dynamic of occupation of real and virtual space was crucial for El Movimiento 15-M, Occupy Wall Street, "twitter revolution" in Iran, multiple protests of the so called "Arab Spring", and Ukrainian Euromaidan. Paradoxically, as Nick Srnicek and Ales Williams [2015] note, the new media with their promise of global connectivity were used by protesters „against the abstraction and inhumanity of capitalism,... to bring politics down to the 'human scale' by emphasising temporal, spatial and conceptual immediacy." Digital revolution redefined political participation, but, as all of the mentioned protest movements teach us, new forms of public engagement are still far from being effective and huge mobilisation potential does not always translate to clear goals.

With the advent of blockchain the idea of direct, horizontal communication and "cutting out the middlemen" reached a new stage. Information is not produced in few pivotal nodes and consumed everywhere else anymore. Phenomena such as web 2.0, social media, crowdsourcing or civic investigative journalism mark the advent of the epoch of new "digital prosumers" [Ritzer and Jurgenson 2010], virtually eliminating the entry threshold and blurring the distinction between producers and consumers. Users do not only create the most valuable content; they also leave infinite petabytes of metadata that changed their status from a worthless by-product of communication to a priceless commodity. Since social media such as Facebook or Twitter, as they claim, are merely "platforms for communication", they do not oblige the rules taming the traditional media, and it is virtually impossible to hold them accountable for defamation, hate speech or fake news. The change affects the whole information ecosystem. To survive, the "traditional media" – print media as well as radio and TV – has to adapt and subordinate to the new financial and semiotic ecology. They have to publish more and faster than ever, compete for the limited attention of the users, manage their personas in the social media and flirt with the business since the money does not flow directly from the readers, but from the carefully profiled, data-driven advertising.

This dramatic shift in production, consumption and interaction with information is bound to have a deep impact not only on the media environment but also on most systems, whether looking at the micro-level of individuals - the way they think, see themselves and the world, interact with others and take decisions, including what to buy or whom they should vote for - or at the macro-level, in the way our democratic and political systems and economic models work.

But the media do not only give us information. They tell us stories. The flow of information and opinion on social media may be understood as a competition of different narratives. Indeed, shared narratives decide about who is to be treated as "us" and who is to be viewed as "them". A concrete example of this discrepancy comes from the media coverage of migration crises: most media describe with different narratives (and results) phenomena that are intrinsically connected by a cause-effect relationship, such as devastation in war zones and the arrival of refugees in countries like Greece, Spain and Italy. Our research shows similar incompatibilities in narrative framing of many other subjects: climate change, international relations, the image of the EU, etc. It suggests that the difference between manipulation and information is not anymore in the content, but much more in the form of the message.

There are a lot of controversies on how these media now shape social dynamics and the opinion on this has shifted over time, from overly optimistic to tragically pessimistic. Given the present and future importance of digital media, it is vital to go beyond these dialectics of fear and hope and build strong foundations for an evidence-based approach towards this changing media ecosystem. Without this knowledge, there will be no effective debunking of fake news, nor efficient communication with various audiences, especially in times of crisis.

Understanding how information spreads on social media

The emergence of the network society has created a whole field of research looking and trying to understand how information spreads on social media. Why do some things go viral? Why was Obama or Trump's social media campaign much more effective than their opponents? Why do people retire into the comfort of their own "echo-chambers" thus reaffirming their worldview creating an increasingly polarized world? Why has the information environment of social media developed in a way so different from that envisaged by its creators?

The impact of algorithms and microtargeting have radically changed the information we are exposed to. Humans tend to search for information consistent with their opinions and beliefs, a mechanism known as confirmation bias. The tendency is exploited by online platforms for information search and social networking and media, which employ recommendation algorithms to catalyse users' attention. As a side effect, the platform amplifies and reinforces individual bias, resulting in extreme polarization of opinions and filter bubbles at the social level, with dramatically negative consequences on the pluralistic public debate needed to nurture democracy. In addition, access to information is often maliciously biased by either commercially or politically motivated influence agents.

Research today demonstrates that it is impossible to describe with simple features the effects of social media in the development of our society as the complexity in having to take account of the actions of millions of individuals far extends our capacity. Social media are today an incredibly powerful instrument of news creation and distribution. The emergence and ubiquitous nature of issues like "fake news", "microtargeting" and "computational propaganda" or eco-chambers demonstrate the power that a diverse range of actors ascribes to social media. It is therefore of the utmost importance to understand the forces and the causes that generate this phenomenon that is seriously changing the present society. The good news is that social media also provide the data to study how communication patterns develop and spread. Every year brings new insights. We know more about the emotions (especially morally-driven outrage) that shape social media [Brady et al. 2017], and we understand better how the new information ecosystems distort the notion of objectivity and credibility (especially for younger generations) [Marchi 2012], finally we are also more aware how this informational architecture originally designed to hijack our attention for marketing purposes, may easily be abused by hostile agents to spread fear and misinformation, and interfere with democratic processes.

This report suggests some of the promising directions for further studies. Recent works developed by Dino Pedreschi and Fosca Giannotti have shown how algorithms and online platforms have a decisive effect on the development of opinions and consensus. Communication patterns on online platforms are changed by the incentive model and advertising-based rationale where attention is monetized, and the motive is not to "inform" users but to capture their attention as long as possible. The impact of this incentive model in spreading discord, undermining social cohesion and increasing levels of cyber bullying and hate speech can be found across the social media platforms. Recent

research has also proved that algorithms slow down pattern of consensus formation in society leaving confusion, fragmentation and online fights to linger for longer deeply undermining social cohesion and trust.

Human-centric AI, a body of research with a strong social dimension, may help us design novel platforms and mechanisms for the public access to news and information, focused on counterbalancing our built-in confirmation bias and transparently striving to expose people to assorted opinions, intelligently. It is possible to imagine mechanisms for helping individuals and communities become informed on controversial issues by offering multiple perspectives, connecting opposing views and conflicting arguments, and fostering critical thought. For example, a robot in a group conversation can highlight information that was unavailable to the group or suggest omitted sources of important information. Advances in person-machine interaction models based on explainable AI have the potential to reach novel cognitive trade-offs between our confirmation bias and our curiosity of novelty and diversity, making it possible for more sustainable and humanized information ecosystems to emerge.

This is, however, just a beginning of a very long scientific journey; Big Data analytics, AI and other tools can significantly support our understanding of this new information ecosystem and Europe should increase its efforts to do so.

The role of narratives

An increasingly influential stream of research demonstrates that our political cognition is always emotionally shaped. Narratives, playing a crucial role in this process, have become an important area of research and debate. Their importance has been particularly clear in recent years when entire societies have made choices that seem “rationally” counterintuitive.

As debates around migration, Brexit, climate change and vaccines become increasingly emotional policy-makers and scientists alike have had to accept that in the battle for hearts and minds of human beings, narrative will consistently outperform data in its ability to influence human thinking and motivate human action.

With a polarising society and highly-charged emotional debates on the table, it is ever more important to look at what can we learn from recent research in psychology, neuroscience, economics, sociology and behavioural science about how the minds work and how they are shaped.

Narratives are fundamentally social. They create interpretative communities bonded by a shared notion of reality. Only the things that fit the narrative schema are perceived as real.

It is through narratives, shared interpretations of the world, that people unite. Yet, narratives also have the power to divide and polarize. Not sharing the common narratives is one of the basic factors taken into accounts when delimiting borders between nations, ethnic groups, religions or social classes. “Us” and “them” are very much narrative structures.

The recent surge in disinformation, or fake news, draws from these basic principles. Not only by attempting to influence directly our definition of “us” and “them”, but also by authenticating certain messages by entangling them into systems of commonly shared values.

Two very interesting points were highlighted that are very relevant to the current discussions around the need for a new narrative in Europe.

First, research shows that narratives can be very persistent and extremely adaptive at the same time. Narratives both adapt to describe new realities and are pivotal tools of understanding of the changes occurring. This past decade, following the financial, economic, social and political crisis that have ravaged Europe, the call for a paradigm shift has become mainstream. At the same time, the old narrative, shaped by neoliberal economic theory, with its tenets that globalisation and unrestrained capitalism will be good for everyone has been proven wrong. With the vacuum that has been left in the wake of this change, numerous new (and less new) narratives have sprung up across the world trying to redefine our understanding of reality and the framework that should guide our future development.

Second, change of narratives is a vital part of every media revolution. The introduction of writing, printing press, radio and television challenged not only the repertoire of dominant narratives, but first and foremost the very form in which they were stored and transmitted. Scholars have demonstrated how new media facilitated narrative transformations that resulted in remarkable changes in politics, culture and society. Ancient empires were built not only on slave labour and military triumphs but first and foremost on the new stories, myths and laws distributed in the written form (Assmann, 2011); modern nation states are deeply rooted in the culture of print that provided new forms of political

debate but also the modern novel (Anderson, 2006) and the triumphs of the 20th century propaganda would have not been possible without the narratives transmitted in the mass media. Nowadays, in the dawn of the new, digital era we face another great change in the form and content of the dominant narratives fuelled by the new media, especially social media.

To verify empirically the importance of narratives in spreading and debunking fake news, we conducted a pilot experiment by publishing a blog post debunking anti-vaccination propaganda during the measles outbreak in Poland described below.

CASE STUDY: Pilot experiment - reactions on debunking in the new media

The availability and quality of information is crucial to effective functioning of society, especially in a time of crisis. A lot of false information regarding COVID19 is currently spreading in societies, especially via social media. One of the main strategies for fighting fake news is debunking, a strategy of confronting it with facts and accepted theories. This effort is not very effective, especially as it does not work with individuals that have strong belief in false information. A seminal study in social psychology (Festinger, Riecken & Schachter 1956), demonstrated that a religious group, predicting the end of the world at a specific date, more strongly believed in the prophecy, after their prophecy was falsified.

To understand the results of debunking fake news on social media, we conducted a pilot experiment in Poland focussed on debunking some of the anti-vaccine debates.

Conspiracy theories and pseudoscientific beliefs surrounding vaccinations provide a good example of dangerous phenomena occurring at the intersection of the social media, fake news and democracy. Research shows that Twitter, YouTube, Facebook and other influential social media provide a crucial boost for the international network of anti-vaccination activists (Smith and Graham 2019). Feeding on fear and chaos, spreading fake news and misinterpreting the data, the anti-vaccination movement is stronger than ever (Fernández-Luque and Bau 2015). Social media provide not only a new set of tools for spreading disinformation but also a great weapon against it. What is the balance between these two forces?

At the beginning of November 2018, a measles outbreak occurred in Pruszków – a city of around 60 000 inhabitants located at the suburbs of Warsaw. In just a few days over 20 people had been diagnosed with the disease that had been virtually eliminated in Poland. The situation became the main news story for several weeks. The anti-vaccination movement in Poland and its prominent leaders started disseminating their views on the situation extensively both via private social media and institutional channels.

To better understand this complex phenomenon, we conducted a pilot experiment with a limited range by posting online an article debunking main arguments of the anti-vaccination movement. All the comments were then collected and analysed qualitatively to identify dominant strategies of reaction on debunking. In the following 10 days over 45 000 viewers read the article and it received over 30 000 likes on Facebook and over 2 400 reactions on Wykop (roughly a Polish equivalent of Digg or Reddit), making it one of the most popular “findings” of the day on this popular portal.

Using qualitative methods, we analyzed 390 comments left under the text, on Facebook and Wykop to establish main strategies of reaction to debunking. Many comments were responses to other comments, enabling us to analyse a complex structure of discussion, where users sharing an anti-vaccination sentiment had to defend their views and engage in a debate not only with our article but also with other users representing an opposite view. “Polemic” structures were over 300% more frequent than “supportive” – anti-vaccination comments were posted mainly in a reaction to pro-vaccination ones and vice versa. Only a tiny fraction of comments referred directly to the commented article and only 2 of them attempted to discuss the presented arguments. This seems to be a very important and general pattern, that suggests that **in a situation of a threat to their worldview the supporters of the anti-vaccination theories do not react directly to the attack, but rather repeat their *credo* to consolidate individual and collective identity.**

The experiment demonstrates that the anti-vaccination reactions to debunking are strongly grounded in the liberal philosophy of “individual freedom” valued more than the “collective good” associated with communism (in Poland this definitely lends additional gravity to the argument). Pro-vaccination debunking is identified with “dehumanizing” ideology of science and progress, represented also by the mainstream media. Heuristics of availability are often used to point on *visible* and *near* negative consequences of vaccination for the individual (against the *elusive* and *distant* benefits for the community). Contrary to the research highlighting the importance of conformism in declared beliefs, being a member of an “anti-establishment” movement provides the users with a very attractive “rebellious” identity. Proofs of the harmful nature of the vaccines are derived very often from the everyday experience regarding not only medicine, but the world in general, especially economy, politics, and structures of power etc. These initial findings suggest that effective debunking must consider formal and social structures of the debunked arguments. Instead of focusing on proving the opponents are factually wrong, we should focus on understanding *why* they choose to accept these beliefs.

The original research, detailed in chapter 6, also explores the circulation of narratives in the information ecosystem of the digital revolution. By combining Natural Language Processing and more traditional methods of qualitative narratology we were able to work on a very large corpus of data (over 200 thousand articles). We used advanced algorithms to divide all the articles into main topics (based on frequency of co-existing vocabulary), and then we analysed qualitatively most representative examples from every topic to establish dominant narrative patterns and answer questions such as: “how is the climate change narrated in different media?”, “what metaphors of Europe are used in various narratives about immigration”, etc. Our results suggest a strong potential of reframing facts through different narratives.

But the “emotional narrative v. rational data” problem should not be viewed as an alternative. We are not forced to choose between emotions and data. On the contrary – the main challenge nowadays for scholars and policymakers is to learn how to wrap facts in stories to make them compelling, attractive and accessible for various audiences. **Common Europe is fundamentally a narrative project rooted in learning from the common past and imagining a shared future. It is obvious that the European project must adapt, actively searching for new forms of its collective mythology that work in the current ecosystem of the digital revolution.**

Future research directions

The report outlines key areas of future research that the analysis of the stock of research today points towards. The following is a list that attempts to highlight some of the most promising areas of research:

- 1) **Better understand the impact of social media on society and information flows and how they affect our cognitions, emotions and collective behaviour.** Social media and online platforms have a significant impact on the creation of opinion and identities in societies as well as in created (or destroying) social cohesion and trust. Better understanding these mechanisms and how the current incentive models affect society will be very important in developing a more systemic understanding of how our media ecosystem affects society.
- 2) **Promoting a system approach combining science of network and narratology.** A combination of complex network and narrative methods – also implying a multi-disciplinary approach bringing together hard sciences and social sciences - would allow to better understand the complex mechanisms that shape social media. A number of research questions could be explored, for instance: is adherence to specific narrative schema related to disinformation? To what extent and in which narrative communities? How is credibility of information established in social media, and to what extent narrative plausibility makes information trusted? What is the role of narratives in shaping the opinions of society, and how new media differ from traditional media in this respect? To what degree are narratives openly revealed vs. hidden and implied and how the use of social media and the interactions across different media are affecting this balance? What is the role of public and shadow narratives in influencing the public opinion? Can narrative approaches help to tackle echo-chambers phenomena and stir genuine debate on social media? To what degree individuals read material representing opposite political view and how to integrate completely divergent views, integrate opinions of opposing echo-chambers? While partial answers to these questions can be found in the following chapters, it is clear that a more systematic approach is required to really understand how information spread on social and traditional media, influencing our cognitive processes, understanding of reality and behaviours.
- 3) **Better understanding of long-term cultural approaches.** “How can we create a news ecosystem and culture that values and promotes truth?” [Lazer et al, 2017]. Long-term approaches aimed at changing the culture of data in our societies and economies, contributing to raise awareness about the wide range of phenomena linked to the circulation of disinformation on social and new media, are deeply needed. This broad area of research includes both the creation and evaluation of specific programmes and tools, designed and implemented by a broad range of stakeholders, including scientists, tech-platforms, civil society and media organisations as well as pedagogues, anthropologists, psychologists and artists. Common languages and strategies should be built to help collaboration among these stakeholders, and more research is needed to evaluate regulatory and non-regulatory scenarios and ethical and moral implications at stake. Sound impact evaluation systems should be designed in collaboration with the scientific community to understand the efficacy of different approaches to tackling disinformation at the micro, meso and macro level, so to allow for rapid adoption, replication, adaptation and scaling of effective measures. From our research for instance, the need for offline activities aimed at familiarising local communities with the very fact that they

are immersed in an info-system to which they constantly contribute data without even realising it emerged strongly. Artistic, performative and collective approaches to this awareness raising activity seem particularly encouraging and should be further supported.

- 4) **Better understanding of the working and possible implications of emerging threats.** More research is needed to understand the patterns followed by false information, particularly of visual and audio contents, and on closed networks such as Messenger, Snapchat or WhatsApp. Emerging phenomena such as deepfakes and, more in general, the evolution of bots in their interactions with cyborgs and trolls, is another area which would deserve attention, including in terms of finding reliable methods to derive representative samples of bots and humans on a given platform, also considering that most work so far has been limited to Twitter. Longitudinal analysis of complete datasets would allow a dynamic understanding of how pervasive systems of fake news provision are evolving over time, including in terms of impacts on target audiences (concerning for instance the impact of micro-targeting and computational propaganda on individuals' as well as groups' behaviours). This would also allow to timely understand the efficacy of countermeasures such as fact-checking and automatic debunking (whose impact seems limited looking at recent research).
- 5) **A European Approach to Artificial Intelligence.** Artificial intelligence (AI) is already a reality and it is expected to largely impact on economy, social behaviour and, eventually, political equilibrium. In this context, many actors are asked to address regulatory and ethical issues, assure robust quality assurance process and deliver real business value. No country can act in isolation or lead this process, as well as no single industry should control the technologies or dominate the market. The European Union is the best candidate to provide the requested critical mass, guarantee the transparency of the process and the respect of the ethical values. Indeed, the European scientific and industrial excellence can drive a fair competitiveness at global level. Science, industry and policy have the joint responsibility to make this happen, throughout an integrated ecosystem where funds, instruments, rules and measures should be coherent and effective.

Recommendations for practitioners and policymakers

The digital era has drastically changed the nature of the public space. New, powerful actors emerged, who control this space and who have access to the most effective tools. As shown in this report, although the study of the mechanisms that shape social media is in its infancy, this phenomenon will have a lasting impact on society. This report shows that there is a growing consensus among scholars that the current information ecosystem can be seen as a threat to social cohesion and our democratic way of life. Something that is today equally understood by policy makers across Europe, demonstrated in the importance that this has been given in the policy priorities outlined by the von der Leyen Commission. It seems today that the question is not about *IF* anymore but about *HOW*. The following is a list that attempts to highlight some of the most promising questions and concerns that practitioners and policymakers should keep in mind in the years to come:

1) **Organising and regulating the digital online media ecosystem.**

Seeing the importance of the information ecosystem for society, a different ecosystem would be essential to preserve and strengthen our collective intelligence. As demonstrated by a long history of research, a multitude benefiting from a healthy diversity of opinions and experience can intelligently respond to difficult questions; this capacity is however undermined the more polarized a society becomes.

As underlined by this report, the communication ecosystem plays a key role for every society and today power is played on the digital social media. This creates the need to regulate this new space in order to ensure that the very positive aspects of social media are delivered as opposed to this becoming a new medium that can be harnessed to manipulate societies to push specific interests. Examples of how this is done today are clear not only with the Facebook-Cambridge Analytica scandal but also during the Irish referendum.

It is important to remember that the traditional media landscape is one of the most highly regulated precisely to safeguard the information ecosystem and in the interest of society at large. Today there are different rules for traditional and new media and the choices we make today in how the new media environment should be regulated will shape how our political landscape will develop for the decades to come.

Several scholars have posited that in order to change the thinking on how to organise and regulate the information ecosystem, we should change the way we approach the question itself – change the narrative. It is undeniable that the communication infrastructure of society is pivotal in shaping society and that a pluralistic and transparent space, allowing for a multitude of opinions and perspectives to reach citizens is a cornerstone of our way of organising society. Whilst today the space where public is confronted with information is privately owned by foreign global conglomerates with their own interests and incentives, Europe should think about how the public space could again become a “public good”, independent from media, company and government interests with the role of preserving the social diversity essential for democracy.

2) **Proposing more transparent and democratic rules of access to data to guarantee the full protection of the right to privacy and enable the maximum use of scientific and educational possibilities that comes with big data.**

Current organisation of the media ecosystem creates a narrow and opaque oligopoly of the companies controlling the data. Limited access to data from private companies that today control Europe's data is very detrimental in being able to properly assess the mechanisms that shape social media and their impact on society. It also makes it very difficult to be able to assess the threats of “information wars” and other manipulative practices. The question also raises significant moral and ethical debates. Policy measures could also include new ways of looking at data ownership, something that GDPR has already introduced in the EU regulation arena: more than with specific technological solutions, societies and governments should come up with principles that inspire policies, and then evaluate the application of those policies against the guiding principles to make sure those are not betrayed by the implementation and that there are no clearly exploitable loopholes to circumvent the principles. It's an iterative process of trial, error and adjustment, but the starting point should be a humanistic declaration of rights of individuals with regards to the data they produce.

In this context, it is fundamental for public authorities at EU and national level, to help safeguard spaces where data and computation can be organised and performed as ever-evolving commons, reminding that a common is given not only by the data, but also by the relational ecosystem around the data and the fact that this ecosystem negotiates and shares practices and rules. This can only happen if we heavily invest in supporting people, small businesses and public administrations in becoming aware of the data they produce – and of the fact that sometimes they cannot be aware neither of the data they produce nor of the ways this is used – and if we are able to find easy, cost-effective and engaging ways to turn data into actionable information, which should then be followed by concrete actions. In fact, it seems difficult-to-impossible to imagine any solution to disinformation and to its effects on society without reflecting on the social architectures and processes that bring to these kinds of phenomena.

3) **Proposing alternatives to the current business model.**

Current advertisement-driven model of the new media leads to a constant Darwinian battle for attention and the survival of the angriest. The military, cold war like escalation of algorithms, BOTs, AIs and other computational cyberwarfare tools and techniques reflect the pitfalls of our current digital online media ecosystem: a world that is more aroused and polarised and unable to deal with the complexity of society. “Funding decentralized alternatives to the current media platforms might be far more effective than regulating them or trying to make arrangement with them. The decentralization of social platforms should be encouraged by active policies, which would also lead to overcoming monopolies. It is crucial to come up with a business model which can open a market that currently doesn't exist. This will probably happen, sooner or later, but if the Commission could accelerate this process by supporting and stimulating something that still seems to be far from being market-driven, this would be a key strategic move. To do that, we need all the wisdom we are capable of, this is really the best resource we can mobilise in Europe: our sense of liberty and sociality, our ability to integrate differences, our open research and critical thinking. The point is precisely to make decentralized social platforms a key ingredient of the European culture, to make them the object of a “social desire”. There is so much room for EU policy here: electric cars are incentivised, so why not social decentralised platforms?” [SMART Expert interviews].

4) **Develop the European approach to AI.**

By 2020 AI ability to counterfeit media will largely surpass those of AI to identify such media. As highlighted above, AI is already here and is already having an enormous impact on the economy, the media ecosystem and society. An European approach to AI is therefore urgently needed.

5) **Encourage inter sectoral collaborations and collective intelligence.**

Other types of solutions should be designed to be able to address the rising separation between citizens, the need to produce the conditions for solidarity, empathy and the formation of an inclusive, high quality relational ecosystem in society. These kinds of actions can be achieved by combining multiple disciplines and technologies. They should start with the conception and design of new education processes, professions and roles for citizens and professionals which are dedicated to "connecting the dots" in society, to obtain higher quality and more inclusiveness in relational ecosystems in cities and regions: using social and cultural actions, and also using data and computation in open ways, creating alternatives for the current "extractive" data and computation industries which bring data and computation out in the public sphere, where society can agree on rules, regulations and protocols in open, inclusive, readable, knowable ways.

Collaborations and interactions between science, technology, arts and design should take place to bring opportunities for social imagination in the public sphere, in which the results of scientific and technological innovation are not only showcased, but also act as the trigger for constructive, inclusive, collaborative imagination of future scenarios that are capable of bringing better relations, information, communication, knowledge and opportunities for exchange.

This report shows that it will not be enough to support traditional tools for tackling the spread of disinformation to safeguard our democratic systems. The problem runs much deeper and there is a need for a more systemic approach in understanding the mechanisms that shape social media and their impact on society.

Conclusions

The present report aims to create scientific insight about the functioning of fast-evolving social networks weaved by social media and to assess which strategies can be deployed to ensure that the information exchanged is accurate and truthful, without limiting the freedom of speech that democratic institutions must provide. The essence of the proposed approach is in shifting the focus from the content of information to the very architecture of the informational networks. Actions like deploying counter narratives, debunking and fact checking can produce some results but they're basically unable to deal with psychological effects which derive from living in such a complex and problematic information ecosystem.

We believe that a multidisciplinary scientific perspective can be a valuable aid in creating a shared source of evidence on this topic. There is dire need for shared arguments on which to pivot, both to allow the debate to evolve and to take steps to counter the general crisis of truth and expertise the world is experiencing in recent years. In Europe, over 50% of people find information mainly on social media, an average which grows to 69% for young people aged under 29, and most people using social media to access information pay little or no attention to sources. Although the research data do not point to any silver bullet to solve the problem, we believe that the overall indications provided can be a valuable support for further studies that will undoubtedly be necessary.

In a seminal psychological experiment Paul H. Thibodeau and Lera Boroditsky demonstrated, that by merely enforcing a metaphor of "crime as a virus" researchers were able to change significantly the way subjects felt and thought about punishment and resocialization. Contrary to the ones working with the metaphor of "crime as a beast", the participants from the "virus" group were much more focused on social factors and networks of agency. Instead of focusing on individual responsibility, they rather viewed society as a complex system, believing that in order to fight crime effectively one has to "fix the economy, improve education, provide healthcare" etc.

One may say that, at least to some degree, we are all living inside Thibodeau and Boroditsky's experiment right now. The metaphor of a virus is omnipresent, and it certainly influences the way we think about all the aspects of social life. To fight the pandemic effectively, to recover from its devastating economic consequences, and to prepare better for future outbreaks, we need to reconsider radically the way we think about local communities and nation states, about common Europe and about the world.

In the "global risk society" it is not possible anymore to solve problems only by focusing narrowly on what is wrong. It is always the architecture of the whole complex ecosystem that shapes both the risks and most effective preventive actions.

We have created a world that is more connected and better informed than any time in history. With new means of expression and virtually infinite accessible knowledge it is also certainly more democratic. However, as the current pandemic reminded us so painfully, it is also more fragile than ever. The more complex the system grew, the more decentralised it became, the more susceptible it was for unfortunate events, cynical manipulation and ruthless, illicit actions.

In a year, the virus as a harsh reality will hopefully be gone. It should, however, remain present as a metaphor and keep guiding our steps towards a deeper understanding of the world we have created and continue to create. Because the one lesson the virus teaches us about the new, globalised world is: you cannot act effectively if you see things separately.

DEFINITIONS

Despite the huge popularity of the term “fake news”, disinformation and misinformation phenomena present on social networks are a broader and more diversified area. In this report, we will focus on disinformation as defined by the European Commission:

***Disinformation:** “verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and can cause public harm. Public harm comprises threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens’ health, the environment or security” [Commission 2018]*

However, it is important to notice that the term “fake news” is prevalent in literature, particularly in [Lazer et al, 2018] definition: “fabricated information that mimics news media content in form but not in organizational process or intent. Fake-news outlets, in turn, lack the news media’s editorial norms and processes for ensuring the accuracy and credibility of information. Fake news overlaps with other information disorders, such as misinformation (false or misleading information) and disinformation (false information that is purposely spread to deceive people)”.

We list below a brief set of definitions useful to describe the specific problems one can have to face when dealing with the circulation of disinformation on the internet.

Affective Polarization: Dislike of the “other side” [Lazer et al., 2018].

Astroturfing: the practice of masking the sponsors of a message or organization (e.g., political, advertising, religious or public relations) to make it appear as though it originates from and supported by a grassroots participant(s) [Source Wikipedia]. A more recent approach is that of the “lateral astroturfing” where the automated tweets firstly are posted by compromised accounts and they are deleted immediately after they are created [Elmas et al, 2019]

BOT, Social bot, Sybil account: computer algorithm that automatically produces content and interacts with humans on social media [Davis et al., 2016].

Computational propaganda: the use of algorithms, automation, and human curation to purposefully distribute misleading information over social media [Wooley & Howard 2018].

Echo chambers: the result of selecting a set of friends and information that adhere to our system of beliefs thereby forming polarized groups [Del Vicario et al., 2016].

False Amplifiers: methods, such as false news, disinformation, or networks of fake accounts aimed at manipulating public opinion [Weedon et al., 2017].

False News: News articles that purport to be factual, but which contain intentional misstatements of fact with the intention to arouse passions, attract viewership, or deceive [Weedon et al., 2017].

Homophily: when for example a user engagement in a discussion correlates with the number of friends having the same behaviour [Bessi et al., 2016].

Information operations: actions taken by organized actors (governments or non-state actors) to distort domestic or foreign political sentiment [Weedon et al., 2017].

Micro-targeting: is the technique of persuasion on electronic media, where each addressee receives a message specifically tailored to match his or her individual characteristics, usually personality traits, which are automatically predicted from digital footprints.

Misinformation: the inadvertent or unintentional spread of inaccurate information without malicious intent [Weedon et al., 2017].

Narrative: a complex (created of smaller units) systematic (following a more general pattern) representation that contains a change of state

PART I – STATE-OF-THE-ART IN RESEARCH

Chapter I - INTRODUCTION

I.1 Introduction

Since the appearance of Facebook and Twitter, the impact of social networks on the way people consume, produce and interact with information has grown exponentially. The rapid change that the media environment has undergone since the advent of social media can be seen in the quantity of information produced each day. The total amount of data in the world was 4.4 Zettabytes (ZB) in 2013. That is set to rise steeply to 44 ZB by 2020. According to a report from IBM Marketing Cloud [IBM Report 2017] 90% of the data in the world today has been created in the last two years alone, at 2.5 quintillion bytes of data a day.

This can also be seen in changing habits and behaviour of people - and particularly of younger generations - in the way they consume and interact with information. Every action today can be photographed, filmed and distributed to friends, family or beyond. This new era of mass self-communication [Castells 2010] that is self-generated in content, self-directed in emission and self-selected in reception reaches a potentially global audience through p2p networks and is the backbone of today's communication networks. This dramatic shift in production, consumption and interaction with information is bound to have a deep impact not only on the media environment but also on most systems, whether looking at the micro-level of individuals - the way they think, see themselves and the world, interact with others and take decisions, including what to buy or whom they should vote for - or at the macro-level, in the way our democratic and political systems and economic models work.

"The shift from traditional mass media to a system of horizontal communication networks organised around the Internet and wireless communication has introduced a multiplicity of communication patterns at the source of a fundamental cultural transformation, as virtuality becomes an essential dimension of our reality" says leading sociologist Manuel Castells in describing the basis for our new "networked society".

These enormous shifts are having huge ramifications in all fields and, in fact, more and more resources are being poured into this new medium to be able to better harness it and use it. In the political arena, candidates and parties are developing ever more sophisticated social media strategies to get elected - look for instance at the impact of big data analytics and microtargeting during the Brexit and 2016 US elections [The Guardian 2017, BBC 2017] -, while policy-makers are struggling to genuinely engage with their constituencies via these new channels. Indeed, as remarked by the Commission's White Paper on the future of Europe, "restoring trust, building consensus and creating a sense of belonging is harder in an era where information has never been so plentiful, so accessible, yet so difficult to grasp. The 24/7 nature of the news cycle is quicker and harder to keep up with and respond to than it ever has been before. More tweets are now sent every day than in a whole year ten years ago. And by 2018, around a third of the world's population will use social media networks".

Social media are becoming even more important also for companies that invest enormous amounts to boost their reputations and where - on a darker note - campaigns are launched to support the damaging of competitors' reputation. Businesses are also becoming increasingly good at using social media for their own commercial purposes: indeed, microtargeting practices in the marketing world are indeed increasingly sophisticated and adopted. And as proved by the first large-scale study on the effect of microtargeting carried out by Matz and al (in press) they are also extremely effective: indeed, the research, involving 3.7 million individuals, showed that persuasive messages that were matched to people's extraversion or openness-to-experience level resulted in up to 40% more clicks and up to 50% more purchases than their mismatching or un-personalized counterparts.

Although the era of social media is just in its infancy the dramatic impact it has had on the world has sparked the emergency of completely new fields of academic research. This part of the report will create an overview of the state-of-the-art in research on how (dis)information spreads on social media. For simplicity, we have divided this research into three main areas

1. Understanding how information spreads on social media (quantitative approaches);
2. Understanding the emergence of narratives and change of behaviour in the age of social media (qualitative approaches);

3. Emerging threats and solutions in tackling the spreading of disinformation.

Chapter 2 will analyse the first question and look at quantitative approaches to how information spreads on social media. Within this area we will be looking at different research fields from science of networks and research algorithms.

Social media do not only affect society by algorithms and changing patterns of behaviour. Research shows that the “content” of information is key in understanding how social media networks work and why certain information has more effect than other. This highlights the need for a more systematic and comprehensive approach that not only looks at the patterns of behaviour and spread of information, but that works closely with psychologists, political scientists, economists and other experts to redesign interdisciplinary and perhaps intersectoral ways to approach the understanding of this new realm. Chapter 3 will assess the qualitative approaches to understanding the emergence of narratives and change of behaviour in the age of social media.

Chapter 4 will address the questions relating to emerging threats and solutions to tackling the spread on online disinformation.

1.2 Flow of Information in the Age of Social Media

In the contemporary European societies, the flow of information is shaped in the interaction between the traditional and the social media. Social media play increasingly important role, and they gradually replace the traditional media. This happens first, because social media like Facebook, Twitter, and Instagram broaden their traditional role of the media for social interactions, becoming the major news sources. Second because the traditional media expand their online presence as they follow the trend and try to compete with the “new media”. Traditional newspaper and magazines gradually switch online and new media outlets are created with just online presence. This process undermines the structures of informational hegemony, that, since the advent of the printing press, have been secured by a very high entry threshold. Because of the infrastructural costs only the most powerful agents were allowed to communicate on a national or global scale, leading to many informational monopolies or oligopolies and a ecology of information where “the winner takes it all”.

Nowadays the flow of information is shaped by the interaction of the traditional and the social media. Even the most influential actors, such as nation states, celebrities or international broadcasting companies are forced to compete for the attention of the users on the same platforms. On YouTube or Facebook official press conferences of presidents coexist and compete with the messages from regular users, memes, fake news, spam and scam. Every time a transitional magazine publishes an article it may be shared and discussed on social media. On Facebook and Twitter the readers interact on the topics of the article by commenting on the content, reacting on it and sharing it with others. A new hegemonic structure emerges, since the interactivity is very selective, with just a few articles generating high interactivity, and the rest being virtually ignored. The interaction on social media defines which topics and articles become the focus of public attention. Social media also have a very important role in shaping evaluate different reaction the content and the opinions expressed in the article and thus in shaping attitudes of the readers. The traditional media is interested in publishing articles that engage readers. Interactivity becomes a decisive criterion for choosing topics. The topics of articles, that generate the highest interactivity on the side of the readers, are likely to be followed in the next publications. The flow of information in societies is thus defined but the positive feedback loop between the traditional and the social media. This mechanism is even stronger in the advertisement driven model enforced and championed by the dominant social, where no direct financial transfers are made between the readers and the media with attention and user data serving as a dominant currency. Potential to attract and engage readers and viewers decide, weather a topic becomes a widely reported news or remains in an obscure niche. At the same time social media often do not create their own content, but rather feed on the one provided by the more traditional outlets.

To understand the mechanisms that shape the flow of information in current societies it is critical to focus on the interaction between the traditional and the social media. This Report focusses on this interaction.

Humans are essentially narrative creatures. We have a very limited ability to receive, process, remember and communicate information that has not been packed into meaningful structures. Narratives provide the natural vehicle for the transmission of information in societies. In the everyday experience an incoming information is organized into stories in the process called narrative embedding. When trying to decide whether an item of information is true or false, individual rarely look for evidence, usually they check if the coming information fits already adopted narratives

using the criteria of narrative plausibility. This is one of the main reasons why debunking fake news by focusing merely on facts, data or numbers is not very effective. Fake news, as well as these true ones, are not collections of separate pieces of information but rather meaningful structures that are credible not because of single facts but because of a more subtle quality of "making sense", "sounding right" or "fitting in". To understand the flow of information in societies we thus need to concentrate on narratives, rather than on piecemeal items of information.

SMART thus focused on narratives and their interaction between the traditional and the social media. To achieve this goal we have combined advanced algorithms for quantitative analysis of large corpora of data with qualitative approach to identify the narrative patterns visible from the point of view of a reader.

In our original research, outlined in section 2 of this report, we have started with gathering a corpus of over 200 thousands of articles from various European, British and Russian media. Then we have used advanced Natural Language Processing tools to separate main topics. The method is based on sophisticated analysis of probabilities of co-occurrence of words in the article and it has been published and tested it gives results that make sense. Based on existing algorithms we have written a software that can analyze on a desktop server hundreds of thousands of articles and that combines different types of analysis.

It suggests strongly that one of the basic needs satisfied by the most popular articles is the need to attribute guilt. Abstract, distant, global processes are made understandable by giving them a human face or a corporate logo. This process may be interpreted as a narrative to restore the sense of agency.

1.3 The European Commission's Actions to Combat Disinformation



Fighting disinformation while defending media freedom and pluralism is fundamental to protect European democracy. The rapid development of digital technologies has not only changed the way that citizens consume news, but also transformed their ways of interacting. Online platforms connect citizens, enable them to create content and break down geographic and societal barriers.

Online platforms have become powerful information gateways, where the gatekeepers have financial interests in servicing the users with customised information. While this environment makes it easier for citizens to interact and express their political views, thus contributing to the healthy functioning of democratic societies, it also allows the rapid spread of harmful disinformation that seeks to disrupt democratic processes.

Evidence shows that foreign state actors are also increasingly deploying disinformation strategies to influence societal debates, create divisions and interfere in democratic decision-making.

The Commission defines disinformation 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. The aim of disinformation is to distract and divide, to plant seeds of doubt by distorting and falsifying facts, thus confusing people and weakening their faith in institutions and established political processes. It is our common duty to defend the core value of freedom of expression and to protect European citizens from disinformation. This calls for three-fold action – effective policy measures to regulate online content and service providers, innovative communication to build resilience against disinformation, and the continued defence of media freedom and pluralism.

Over the past five years, the EU has strengthened its efforts to tackle disinformation. Following March 2015 European Council conclusions, the East StratCom Task Force was set up in the European External Action Service (EEAS). Together with the Commission and EU delegations in non-EU countries, the Task Force communicates about Union policies in eastern neighbourhood countries, strengthening the media environment and supporting media freedom

and independence. It improves the EU's capacity to forecast, address and raise awareness about disinformation activities. The Commission has tackled disinformation from the legislative, security and communication perspective. In the last year, the Commission has underlined the importance of securing free and fair European elections. It has encouraged leading platforms to sign up to a code of practice against disinformation and put forward an action plan with proposals for a coordinated EU response to the challenge of disinformation. The platforms must now take their fair share of responsibility for ensuring free and unbiased speech in Europe.

The Action Plan focuses on four areas to build up EU and Member State capabilities and strengthen cooperation:

- 1) improving detection, analysis and exposure of disinformation;
- 2) building stronger cooperation and joint response through a new rapid alert system;
- 3) engaging with online platforms and industry, as per the code of practice signed by online platforms, leading social networks, advertisers and industry; and
- 4) raising awareness and empowering citizens through targeted campaigns and dedicated programmes promoting media literacy at European and national level.

Crucially, the fight against disinformation requires a constant outflow of fact-based messaging that helps citizens to distinguish reality from falsehood. The Commission responds to disinformation directly via its Spokesperson's Service, the Commission Representations in the Member States and on social media by drawing on the expertise of its Network against Disinformation, a group of Commission mythbusters, data analysis from social media and insights from behavioural science. The Commission provides factual and accurate information on its policies and political priorities as well as rebutting any disinformation that seeks to mislead European citizens as regards the EU.

In Member States, Commission Representations also play a part in mythbusting setting the record straight on the most persistent and commonplace falsehoods about the EU. In the rapidly evolving world of disinformation, institutions need to continue efforts to adapt very quickly, not only to changing political circumstances, but also to the changing technological landscape. The use of 'bots' (online robots that perform repetitive tasks based on algorithms) in communication is already a reality. In the near future, artificial intelligence will also be increasingly used to carry out communication activities. The EU Institutions and Member States need to continue efforts to adapt to and get ahead of this new reality. Education systems can play a part here: The Digital Education Action Plan could encourage more specialist training in artificial intelligence.

In order to continue fighting the increasing flow of disinformation, Member States and institutions need to team up to build on existing synergies and pool resources. Together, Member States and EU institutions need to make sure that the new rapid alert system is fully exploited. Going beyond information sharing, a real knowledge community with expertise on the different facets of disinformation is needed, drawing insights from research and *Décodeurs de l'Europe* - an initiative academia, fact-checkers, online by the Commission Representation platforms, technology experts and in Paris international partners. The Commission is ready to link up the expertise of its Network against Disinformation with the rapid alert system and provide a hub for EU policy related content to counter disinformation.

The Commission intends to facilitate the creation of a European multidisciplinary community to foster cooperation between all involved, in particular independent fact-checkers and academic researchers involved in the fight against disinformation. To this end, the Commission plans to establish a European Platform on Disinformation. The Platform will scale up collaboration between fact-checkers and academic researchers in order to ensure full coverage of the Union territory and facilitate the build-up and interconnection of relevant national organisations, including national disinformation centres.

1.4 Conclusions

The change in the flow of information towards a horizontal model, in which everyone can modify and disseminate news and opinions, certainly brings many negative aspects. These will certainly prevail over the positive aspects until the understanding of the new model allows us to adapt the rules to the many changes that are already present. Disinformation and the techniques and processes which are used to spread them are in constant evolution, and the definition of disinformation is itself problematic, as in most cases disinformation is the result of gradients of compositions of true and false facts. On top of that, the communities, technologies, networks, software platforms and online services used to spread these malicious communications are constantly evolving. Indeed, academics, organizations and companies that deal with these phenomena through their research, products and services agree that there is no single fits-all model, and that scenarios and battlegrounds are ever evolving. Significant efforts have been made both by the academic community as well as by media organisations, governments, civil society

organisations and citizens to respond to these challenges. Case studies include algorithms, fact-checking tools, efforts by the big new media giants such as Facebook and Google to address the issue. So far no silver bullet has proven effective but it will be important to look at the successes and failures of existing initiatives when trying to address the fundamental question of how to deal with the scale of the problem of fake news in today's media environment and its effect on our political and economic institutions. Lionel Barber of the Financial Times summed up the challenge (FT 16 September 2017) as follows:

"Fake news damages public trust in news media. Fake news undermines public confidence in our democratic discourse. Fake news exacerbates economic pressures facing quality news organisations. Finally, and perhaps most importantly, fake news highlights issues of responsibility and regulation in our fast-evolving media ecosystem."

Indeed, while fake news and propaganda always existed in the history of mankind [Soll, 2016; Mansky, 2018], the computer revolution that created the present connected world made the problem more and more pressing [Waldrop, 2017]. This happened because it is now possible for everybody to have access to a global audience and to be able to address political/economic/religious and all the other social issues with a simple access to the Internet.

As the number of people using social media as their main source of news increases steadily across the globe, the rapid spreading of disinformation online becomes a threat to our democratic societies, with rapidly decreasing levels of trust in governments, private organisations and traditional media. A recent Pew Research Center survey [Mitchell and Page, 2018] found that 63% of Americans do not trust the news coming from social media, even though an increasing majority of respondents uses social media to get the news on a regular basis (67% in 2017, up from 62% in 2016). Even more disturbingly, 64% of Americans say that fake news left them with a great deal of confusion about current events, and 23% also admit to passing on fake news stories to their social media contacts, either intentionally or unintentionally. Even though such data is somewhat questioned [Doshi et al. 2018], in Europe, over 50% of people find information mainly on social media, an average which grows to 69% for young people aged under 29. Interestingly, those who hold populist views are more likely to get news from social media in France, Italy, Spain and Germany, but not in Denmark, the Netherlands, Sweden or the UK. As for levels of trust, the difference in media trust between social media news consumers and non-consumers is small: no larger than 7 percentage points in all countries surveyed. Worryingly, most people using social media to access information pay little or no attention to sources there [Mitchell and all 2018]; even more worryingly, according to the The International Center for Journalists, while 71% of journalists use social media to find stories, only 11% use any social media verification tools [Dias, 2018].

In a seminal book from 2009 Cass Sunstein [Sunstein, 2009] analysed the problems for modern democracies in a world increasingly dependent on the connections established on the Internet, where violence and partisan Web logs become political parties and (as in the case of jihad) can even start and/or fight wars thereby contributing to spread violence in societies. Ten years after, [Bradshaw and Howard 2018], [US Office of the Director of National Intelligence 2017] and [Field et al 2018, UK Parliament 2018] found compelling evidence about nefarious activities implemented by foreign countries to influence democratic processes in Western countries. The Kremlin-backed Internet Research Agency (IRA) was particularly active in spreading online disinformation, with a recorded track of 1,102 Russian accounts tweeting with the hashtag #ReasonsToLeaveEU [Field et al 2018, UK Parliament 2018]. "An NBC report identified 2,752 Russian "troll" accounts that posted more than 200,000 tweets; these tweets earned 2.1 million retweets and 1.9 million likes. Twitter reported an even more expansive campaign that likely extended beyond the IRA, with 36,000 automated accounts posting 1.4 million tweets that earned 288 million views leading up to the election. On Facebook, Russian posts reached 126 million US Facebook accounts. On Instagram, which is wholly owned by Facebook, 170 Russian accounts created more than 120,000 pieces of content, which reached more than 20 million US accounts. The activities of the IRA were not limited to Facebook, Instagram, and Twitter: YouTube, Google+, Vine, Meetup, Pinterest, Tumblr, Gab, Medium, Reddit, and even PayPal, which helped sell its merchandise, were also targeted" [Gangware 2019]. The steadily increasing amount of online content as well as the technological platforms themselves are making it extremely difficult to counteract these campaigns: "for example, a UK Home Office press release noted that in 2017, ISIS used more than 400 unique platforms to distribute content. In the latter half of 2017 alone, ISIS used 145 new platforms for this purpose." [Gangware 2019]

While there is some agreement on ways to address this spectrum of issues – including fact-checking, greater transparency, automatic debunking and tailored education initiatives - in order to rebuild public trust in the media - it has also to be noticed that recently the Reuters Institute [Newman et al., 2017] reported that it may be hard to persuade people of facts that run counter to their own entrenched beliefs however clearly or transparently stated.

Greater scepticism towards the facts is the inevitable consequence of exposure to a wider range of perspectives. In itself this may not be a bad thing as long as it is supported by better source labelling, signals of quality, and improved news literacy [Nicodemo 2018]. Narrative theory can also provide valuable support in promoting a genuine debate on social media, while also contributing to identify and contain disinformation. Three elements here seem particularly crucial to understand the spread of fake news:

1. System approach: fake news functions in information ecosystems. These ecosystems consist of one or more senders (information source), media channel and recipients, but they are also constituted by shared values, symbolic codes, language, media-related habits and attitudes etc. Narratives are important elements of these ecosystems, providing an acceptable form even for information that contradicts attitudes or beliefs of a given recipient.

2. The principle of plausibility is of crucial importance for the spread of fake news. Both the authority/familiarity of the source and the congruence with the narratives existing in a given community decide if the news item is believed. This rule of "social credibility" seems to differentiate fake news from more traditional forms of propaganda. Fake news "have credibility not because direct evidence is available to support them, but because other people seem to believe them" [Sunstein, 2017]

3. Values: narratives are interwoven in systems of values of the groups to which they are addressed, which is also expressed in the use of symbolic codes specific of a given group. Fake news in many sources is presented together with true or unfalsifiable information (e.g. normative claims) that is commonly accepted by a community. For example, a Facebook profile may present historical trivia and memes about cherished heroes together with anti-migrant content. As remarked by [Lazer & Al. 2018] "concern over the problem is global. However, much remains unknown regarding the vulnerabilities of individuals, institutions, and society to manipulations by malicious actors". We know surprisingly little about the impact of fake news on people's political behaviours, and evidence about the effectiveness of countermeasures such as fact-checking, debunking and specific education programmes is mixed at best [Swire et al. 2017; Ecker et al. 2017; Jones 2017]. It is therefore essential to promote an interdisciplinary and systemic approach to the problem, calling tech-platforms to share with the scientific community data allowing them "to conduct a rigorous, ongoing audit of how the major platforms filter information." [Lazer & Al. 2018].

In the following chapters, we will examine the state of the art concerning research on fake-news and related phenomena, particularly from the point of view of network science and narratology, while also looking at recent tools and initiatives put in place by governments, practitioners and tech-platforms across the globe to prevent or counteract the spread of online disinformation. We will then focus on the original research developed by SMART Consortium partners through the Observatory, an online platform with both a centralized and decentralized component set-up with a view of observing online disinformation phenomena and experimenting with both computational, non-computational and mixed approaches to address their spread. Finally, we will conclude with some policy-recommendations on the way forward based on what we learnt in this journey also through constant interactions with experts and citizens alike.

Chapter 2 – QUANTITATIVE APPROACH

Understanding How Information Spreads on Social Media

2.1 Introduction

The emergence of the network society has created a whole field of research looking and trying to understand how information spreads on social media. Why do some things go viral? Why was Obama or Trump's social media campaign much more effective than their opponents? Why do people retire into the comfort of their own "echo-chambers" thus reaffirming their worldview creating an increasingly polarized world? Why has the information environment of social media developed in a way so different from that envisaged by its creators?

Research today demonstrates that it is impossible to describe with simple features the effects of social media in the development of our society as the complexity in having to take account of the actions of millions of individuals far extends our capacity. Social media are today an incredibly powerful instrument of news creation and distribution. The emergence and ubiquitous nature of issues like "fake news", "microtargeting" and "computational propaganda" or eco-chambers demonstrate the power that a diverse range of actors ascribes to social media. It is therefore of the utmost importance to understand the forces and the causes that generate this phenomenon that is seriously changing the present society [The Guardian 2017].

Humans tend to search for information consistent with their opinions and beliefs, a mechanism known as confirmation bias. The tendency is exploited by online platforms for information search and social networking and media, which employ recommendation algorithms to catalyze users' attention. As a side effect, the platform amplifies and reinforces individual bias, resulting in extreme polarization of opinions and filter bubbles at the social level, with dramatically negative consequences on the pluralistic public debate needed to nurture democracy. In addition, access to information is often maliciously biased by either commercially or politically motivated influence agents.

This chapter will look at the state-of-the art in research in trying to understand how disinformation spreads on social media from a science of networks and research algorithms perspective.

2.2 Science of Networks

Statistical approaches, complex systems and network theory are key instruments to describe the spreading of information on social media. Statistical physics in particular has proven very efficient in the detection of sudden change in the qualitative properties of the system (functional/un-functional). Here the role of the topological structure of the connections [Caldarelli 2007] between individuals in the present society has been investigated in a variety of papers [Quattrociocchi et al. 2014]. The key point is to detect and understand the role of superspreaders, i.e. the persons/sites who account for most part of the diffusion in the system.

As communication today provides similar patterns to any complex network, be it biological networks or computer networks, an obvious approach to better understand the spreading of information on social media is to consider the analogous diffusion patterns of disease in society. Yet this model has proven insufficient in explaining certain features of the spreading of information on social media as it ignores the different mechanisms of transfer for the disease with respect to the news. Indeed, for the diseases the role of superspreaders has been traditionally linked to measures of centrality [Lloyd-Smith et al. 2005, Reluga 2009]. Contrarily to expectations [Jackson et al 2011, Ghali et al. 2012], when considering social networks, traditional quantities as the PageRank or the degree centrality are not always successful in detecting the key-players. Much more successful approaches are related to the analysis of persistence of k-core structures [Pei et al. 2015, Gong et al 2016] across different social networks.

Although significant research has been done to understand the spreading of (dis)information on social media the complexity of the environment, the number of individuals, echo-chambers, make the system so highly complex that no simple model seems to be sufficient at this time. An important note to make is also the fast – and relatively unpredictable - pace at which people move from one platform to the next, changing patterns and preferred method of communication (text, photo, other) as technology and services progress. More interdisciplinary studies will be required in order to answer the fundamental questions relating to the working of social networks.

2.2.1 COMPLEX NETWORKS

Complex Networks are mathematical objects composed by vertices (i.e. people) connected by edges (i.e. their interaction). The mathematical properties of these structures are particularly easy to compute and correspond to specific features of the system, as for example communities.

One of the main advantages of the theory of networks is the possibility to describe in a quantitative way features that have been described so far only qualitatively [Caldarelli, 2007]. For this reason, complex network approaches are a powerful method to describe a variety of systems spanning across the digital and physical domains with a particular interest for social systems. Specifically, in social media they can mathematically describe their status and their dynamics [Agarwal et al., 2014] as the formation and evolution of communities [Pattabiraman et al., 2015]. Interestingly, the structures of various social networks interact and affect each other. For instance, news diffusion on the WWW or Twitter and Facebook cannot be studied in isolation or without reference to the backbone of the Internet [De Domenico et al., 2016]. The problem of diffusion of news becomes particularly difficult to describe since it happens on a multilayer network. Only recently scientists started to work on this perspective trying to link the diffusion in different layers of social media [De Domenico et al., 2013; Solé-Ribalta et al., 2015].

Spotting Communities

As mentioned above, the use of networks allows to quantify and visualize immediately communities of people behaving in a similar manner. There are various methods available to such purpose. Many researchers are now working to describe both the dynamics of the network of contacts and the dynamics on this network i.e. how the information travels along this structure [Weng et al. 2013]. The basic models for news propagation have been inspired by epidemics, in the sense that you must enter in contact with an idea from one of your neighbours (neighbours can be almost infinite in an hyperconnected world). You can define a threshold for the activation and study the cascade processes originated in this dynamic. A non-trivial generalization of these models to multilayer networks [De Domenico et al. 2016] tries to incorporate the interaction of different news platforms (Twitter, Facebook, Web sites) in the diffusion of news. Some authors described interaction between Twitter and Facebook by means of bipartite graphs of users and sources (from FB or Twitter) and analysed community structure in projected graphs [Del Vicario et al. 2018]. By using the platform "Hoaxy", [Shao et al., 2016, 2017] have been able to detect the typical spreaders of fake news from suspected sources. Hoaxy (<https://hoaxy.iuni.iu.edu/faq.html>) does no fact-checking itself, but visualizes the spread of fake news articles, hoaxes, rumours, conspiracy theories, satire, or even accurate reporting, and related fact checking online across social media. Temporal trends plot the cumulative number of Twitter shares over time. Diffusion networks display how claims spread from person to person.

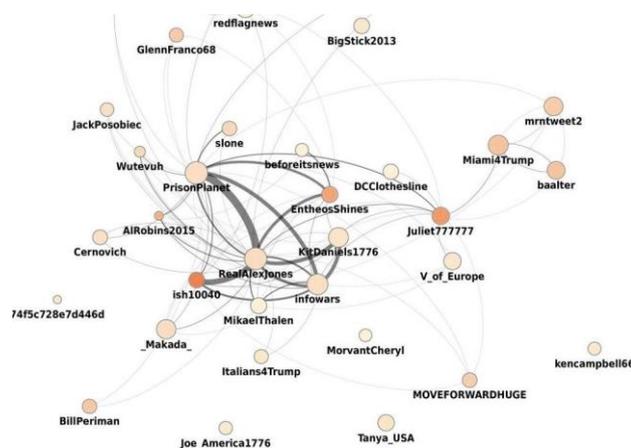


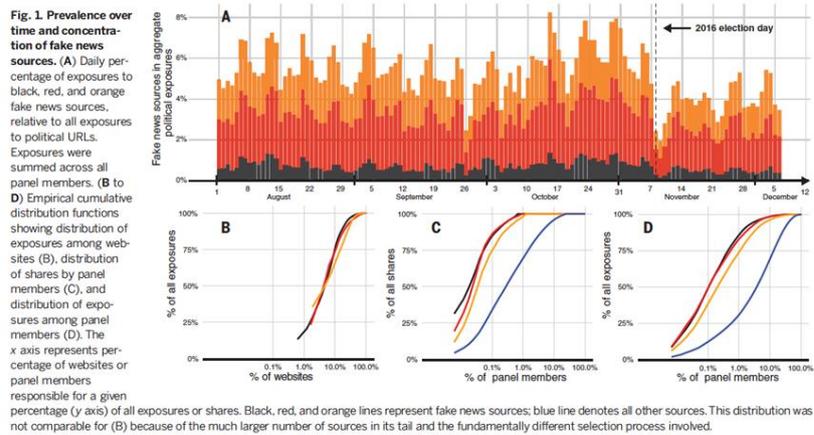
FIGURE 1 - RETWEET NETWORK OF THE STABLE MAIN CORE OF SPREADERS OF ARTICLES FROM LOW-CREDIBILITY SOURCES. FILTERING BY IN-DEGREE WAS APPLIED TO FOCUS ON THE 34 ACCOUNTS RETWEETING THE MOST OTHER ACCOUNTS IN THE CORE. NODE SIZE REPRESENTS OUT-DEGREE (NUMBER OF RETWEETERS) AND NODE COLOR REPRESENTS IN-DEGREE.

Most recently, [Grinbeg et al. 2019] made a very important study focusing on “three simple but largely unanswered questions: 1) How many stories from fake news sources did individuals see and share on social

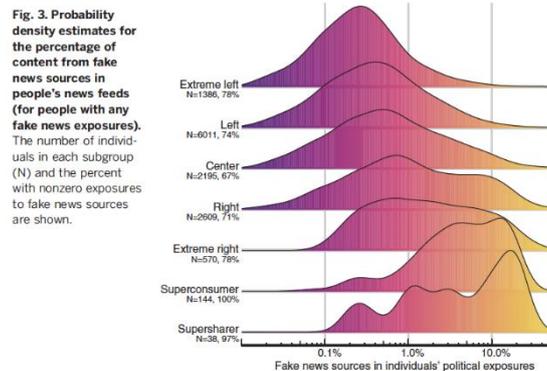
media? 2) What were the characteristics of those who engaged with these sources? 3) How did these individuals interact with the broader political news ecosystem?”. They examined exposure to and sharing of fake news by registered voters on Twitter and found that engagement with fake news sources was extremely concentrated. Only 1% of individuals accounted for 80% of fake news source exposures, and 0.1% accounted for nearly 80% of fake news sources shared. The most likely individuals to engage with fake news sources were found to be conservative leaning, older, and highly engaged with political news. A cluster of fake news sources shared overlapping audiences on the extreme right, but for people across the political spectrum, most political news exposure still came from mainstream media outlets.

To measure fake news more comprehensively, they labelled websites as “black” (if already identified by fact-checkers, journalists and academics as fake-news sources), “red” or “orange” via a manual annotation process of sites identified by Snopes.com as sources of questionable claims. Sites with a red label (e.g., Infowars.com) spread falsehoods that clearly reflected a flawed editorial process, and sites with an orange label represented cases where annotators were less certain that the falsehoods stemmed from a systematically flawed process.

There were 171 black, 64 red, and 65 orange fake news sources appearing at least once in the data.



Very importantly the authors also studied who shared most the fake news. Political affinity was also associated with the sharing of content from fake news sources. Among those who shared any political content on Twitter during the election, fewer than 5% of people on the left or in the center ever shared any fake news content, yet 11 and 21% of people on the right and extreme right did, respectively ($P < 0.0001$). A logistic regression model showed that the sharing of content from fake news sources (as a binary variable) was positively associated with tweeting about politics, exposure to fake news sources, and political affinity, although the disparity across the political spectrum was smaller than suggested by univariate statistics. Other factors such as age and low ratio of followers to followees were also positively associated with sharing fake news sources, but effect sizes were small.



2.2.2 COMPUTER SCIENCES APPROACHES

Research Algorithms and Social Media

Today, algorithms curate everyday online content by prioritizing, classifying, associating, and filtering information. In doing so, they exert power to shape the users' experience and even their perception of the world [Diakopoulos (2015)]. In the light of the issues raised above, another core area of research has been looking into the structure and biases of algorithms and their effect on social media.

Some of the most powerful influencers on elections today are the social media platforms and the algorithms they use to spread information [Guardian 2017]. Yet it is not possible for researchers to measure and effectively study these phenomena as the majority of data is not made available by the privately owned platforms (Twitter, Facebook, etc.). Philip Howard of the Oxford's Internet Institute highlighted "that there have been several democratic exercises in the last year that have gone off the rails because of large amounts of misinformation in the public sphere; "Brexit and its outcome, and the Trump election and its outcome, are what I think of as 'mistakes', in that there were such significant amounts of misinformation out in the public sphere." [The Guardian 2017]. Whilst over 60% of internet users in 2015 were completely unaware of the curated nature of their social media feeds [Eslami, M. et al. (2015)], these are in reality highly curated content where there is a clear algorithmic selection, based on a number of factors including payment history, popularity, how interactive use has been and actions of other. Researchers today agree that the impact of the algorithms of social media platforms should not be underestimated and that they have a high responsibility in the spreading of (mis)information.

The politics behind algorithms is becoming one of the most important questions: "who (and how) is setting the news agenda in the era of the algorithms?" leading to the question: Who Controls the Public Sphere in an Era of Algorithms? [Reed 2016]. Along these lines there is a growing concern about the rise of bots [Ferrara et al. 2016] as producers and amplifiers of fake news in systems like Twitter. DARPA for example organised a challenge to find the best algorithmic approaches to detect and measure bot influence from terrorists [Subrahmanian 2016]. There is also a strong debate if the use of algorithms increases or not [Messing et al. 2014] the presence of the so-called echo chambers [Bakshy et al. 2015, Bessi et al. 2016].

Algorithms and various types of Machine Learning, Deep Learning and Bayesian systems have been used by various groups worldwide to both spread and detect fake news.

[Bradshaw and Howard 2018] found evidence of organized disinformation campaigns in 48 countries in 2018, up from 28 in 2017, while a recent study authored by cybersecurity firm SafeGuard Cyber found that half the European population might have been reached by Kremlin backed disinformation campaigns about Europe between the 1st and the 10th of March 2019, propagated by 6,700 so-called "bad actors" including both troll and bot accounts. The proliferation of political bots and software allowing humans to communicate through multiple accounts at the same time have led to the creation of a broad array of detection algorithms. Among them, the most common are:

- a. **Classification algorithms** (supervised algorithms predicting categorical labels): Linear discriminant analysis; Quadratic discriminant analysis; Maximum entropy classifier; Decision trees; Naive Bayes classifier; Neural networks (multi-layer perceptrons); Support vector machines;
- b. **Clustering algorithms** (unsupervised algorithms predicting categorical labels): Deep learning methods; Hierarchical clustering (agglomerative or divisive); K-means clustering;
- c. **Ensemble learning algorithms** (supervised meta-algorithms for combining multiple learning algorithms together): Boosting (meta-algorithm); Bootstrap aggregating ("bagging"); Ensemble averaging; Mixture of experts;
- d. **General algorithms for predicting arbitrarily-structured (sets of) labels**: Bayesian networks; Markov random fields;
- e. **Multilinear subspace learning algorithms (predicting labels of multidimensional data using tensor representations)**: Multilinear principal component analysis (MPCA);

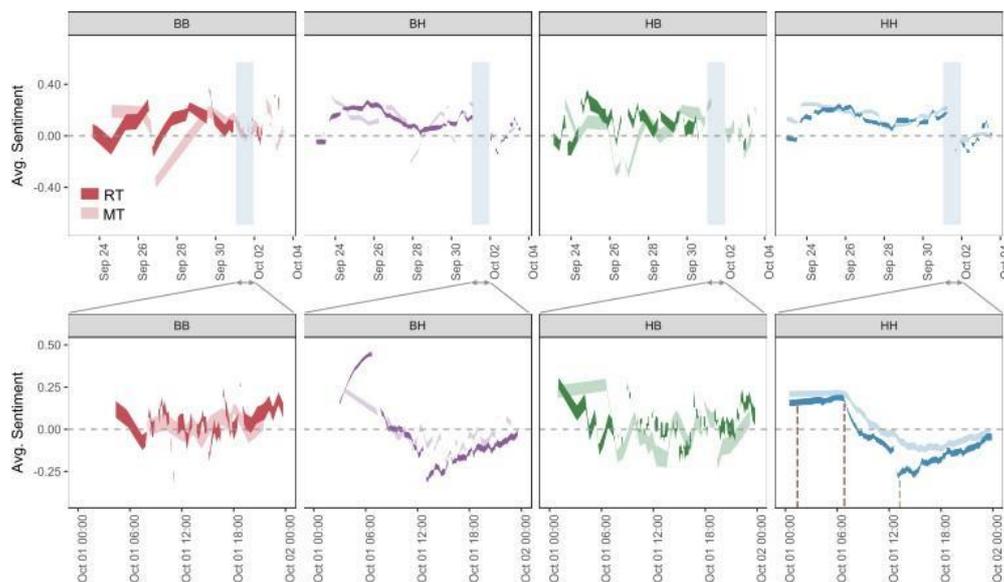
f. **Regression algorithms (predicting real-valued labels, supervised and unsupervised):** Gaussian process regression (kriging); Linear regression and extensions; Neural networks and Deep learning methods; Independent component analysis (ICA); Principal components analysis (PCA);

g. **Sequence labeling algorithms (predicting sequences of categorical labels):** Conditional random fields (CRFs); Hidden Markov models (HMMs); Maximum entropy Markov models (MEMMs); Recurrent neural networks.

A typical machine learning algorithm is the determination of the sentiment of a text. Indeed, sentiment and emotional analysis could, if combined with network approaches, be potentially useful in spotting fake news since, according to the US Department of State's Global Engagement Center, in disinformation some emotions i.e.– uncertainty, fear, and anger – are involved and they are the very characteristics that increase the likelihood a message will go viral [Gangware et al 2019].

Nowadays there are several instruments available on the market also in open source format. Amongst the commercial ones one of the most used is machinebox, used among others by Edell to detect fake news [Edell, 2018] (<https://machinebox.io/>). The almost alarming result of the experiment made in the "fakebox" project (<https://machinebox.io/cs/fakebox>) is that the only way to detect a fake-news is to look for the news that we can say are certainly true, that is only for factual statements [Edell, 2018].

A recent application can be found in [Stella et al., 2018], which investigates the trends in the retweet/mention of a tweet between the two classes of twitter users (bot and human). In principle, differences in this trend could allow to determine the nature of the user and disentangle automatic propaganda from the personal opinions of people. Details on the automatic recognitions of these trends through deep neural network methodology are presented in [Kudugunta and Ferrara, 2018].

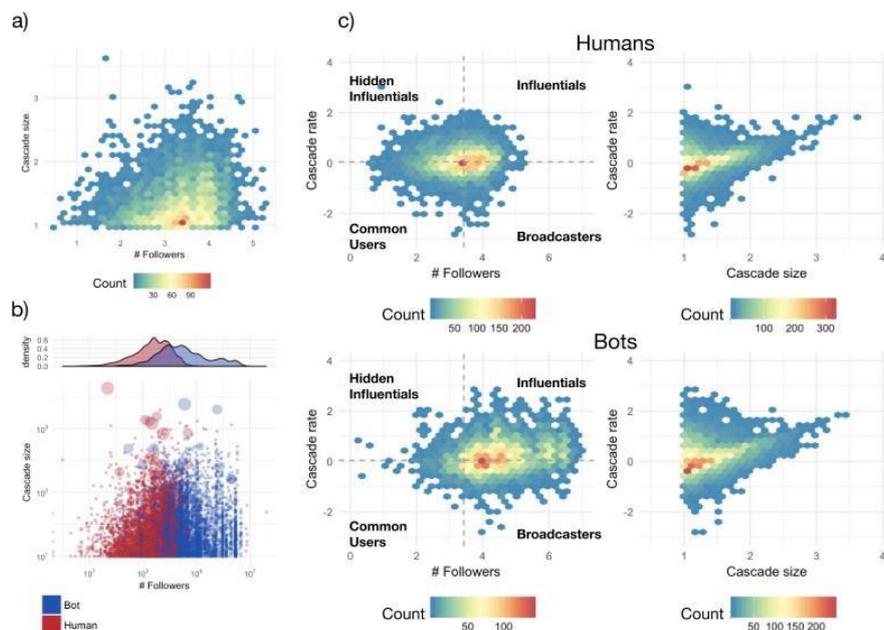


AVERAGE SENTIMENT SCORES FOR RETWEETS (FULL COLOR) AND MENTIONS (LIGHTER COLORS) OVER TIME FOR HUMAN-TO-HUMAN (HH), HUMAN-TO-BOT (HB), BOT-TO-HUMAN (BH), BOT-TO-BOT (BB). THE GREY BOX IN THE ABOVE PLOTS HIGHLIGHTS THE DAY BEFORE THE CATALONIA BALLOT. WHILE BOT- TO-BOT AND HUMAN-TO-BOT DISPLAY NO CLEAR TREND OVER TIME, HUMAN INTERACTIONS DISPLAY A POSITIVE PATTERN OF SENTIMENTS UNTIL SEPTEMBER 30, AFTER WHICH A DROP IN SENTIMENT UP TO NEGATIVE VALUES APPEARS IN HUMAN-TO-HUMAN AND BOT-TO-HUMAN INTERACTIONS.

Such interaction is of great importance to address also the distortions that may appear during electoral time. The role of bots in the recent European elections has been carefully studied, and while it seems almost unanimously [Artoni et al 2019], that they did not play a crucial role in the determination of the outcome, there is also some good evidence of the existence of a series of organized bots [Artoni et al 2019] ("We found evidence of connections between different disinformation outlets across Europe, U.S. and Russia, which often linked to each other and featured similar, even translated, articles in the period before the elections. Overall, the spread of disinformation on Twitter was confined in a limited community, strongly (and explicitly) related to the Italian conservative and far-right political environment, who had a limited impact on online discussions on the up-coming elections" as in the case of Italian elections of 2018 [Caldarelli et al. 2019]. The same issue of irrelevance has been claimed also by other authors who could

find in their data set the signs of the presence of organized network of bot [Cinelli et al. 2019]. For a recent review on these topics we suggest [Ferrara 2019].

An interesting case is that of the Italian elections for which some studies on Twitter predicting power have already been carried out [Caldarelli et al., 2014; Eom et al., 2015]. In a recently published paper, the basic mechanism of retweeting/mentioning is considered as the basis of the cascades of rumors. Such cascades are another quantity that allows to distinguish between humans and bots. In the figure below, we have a heat map of cascade size vs the size of initiators' social neighborhood (i.e., the number of followers). As expected, on average, the larger the number of followers the larger the cascade size, with very few exceptions. Figure 2 (b) shows the same data, with explicit information about user classification and the cascade diameter, ranging between 2 and 6. This figure shows a good separation between human and bot behavior. Deeper information cascades are generated mostly by humans with a high number of followers, with the remarkable exception of one, User01, who produced the largest cascade among humans and bots despite having less than 100 followers.



INFORMATION CASCADES DURING ITALIAN ELECTIONS. A) HEATMAP OF THE NUMBER OF USERS INITIATING INFORMATION CASCADES, AS A FUNCTION OF THE SIZE OF THEIR SOCIAL NEIGHBORHOOD (FOLLOWERS) AND THE SIZE OF THE GENERATED CASCADE; B) SCATTER PLOT OF THE SAME DATA, WITH POINTS ENCODING USERS' COLOR ENCODES BOT/HUMAN CLASSIFICATION AND SIZE ENCODES CASCADE'S DIAMETER; C) AS IN A) BUT CONSIDERING CASCADE RATE, DEFINED BY THE RATION BETWEEN CASCADE SIZE AND ITS DURATION, VS. NEIGHBORHOOD SIZE (LEFT PANELS) AND CASCADE SIZE (RIGHT PANELS), FOR HUMANS (TOP PANELS) AND BOTS (BOTTOM PANELS).

Most recently, researchers from MIT's Computer Science and Artificial Intelligence Lab (CSAIL) and the Qatar Computing Research Institute (QCRI) used machine learning to determine if a news source was accurate or politically biased. They took fact-checked data from Media Bias/Fact Check (MBFC) and fed it to a machine learning algorithm programmed to classify news sites the same way as MBFC by looking at common linguistic features across the source's stories, including sentiment, complexity, and structure. When given a new news outlet, the system was then 65% accurate at detecting whether it had a high, low or medium level of factuality, and roughly 70% accurate at detecting political orientation [Baly and Al 2018].

Large platform such as Twitter and Facebook are also investing heavily in developing their own detecting algorithms and modified their ranking algorithms to give less space to less reliable sources.

As remarked by [Gangware 2019], while computer sciences approaches have been successfully applied to detect and, to a certain extent, counteract fake news, we are still confronted with several challenges.

Firstly, sophisticated technologies that, until recently, were the apanage of a handful of large companies and research labs are becoming increasingly widespread. Bots are a case in point: while detecting spambots based on syntax, semantics, or network features is effective, detecting next generation political bots which instead of simply sharing or liking posts leverage on AI to produce adaptive contents based on user interactions is indeed a challenge. In addition, research on political bots has so far focused on "short periods of time, in a specific political context, and on Twitter. Questions remain about the average lifespan of a bot,

the transferability from one country or political context to another, and the efficacy of building algorithms to detect bots on platforms other than Twitter” [Gangware 2019]. When we look at photos and videos, while technology exists allowing to easily alter it, automated detection at scale is still difficult; at the same time synthetic media (manipulated or artificially-created video and audio content) are becoming increasingly easy to produce thanks to rapid advances AI and deep learning technology. As remarked by [Panetta 2017], by 2020 AI ability to generate counterfeit media will largely surpass those of AI to identify such media. To add problems, bots have been recognized as an evolutionary entity that changes shape and form in order to maintain its malicious feature of propagating disinformation. In a recent paper [Cresci 2019] authors claim that today’s bots are practically undistinguishable from legitimate accounts when doing a supervised bot analysis technique as demonstrated by a survey that could not spot bot presence [Cresci et al 2017].

Secondly, structural challenges exist. Even when algorithmic solutions allow to identify altered content, the huge and fast-growing mass of online content and platforms is at odds with the idea of thoroughly monitoring the information ecosystem. Even if this was the case, detection tools might prove useless unless they are widely adopted by users and platforms: “even if an effective detection method emerges, it will struggle to have broad impact unless the major content distribution platforms, including traditional and social media, adopt it as a screening or filtering mechanism” [Council on Foreign Relations Digital and Cyberspace Policy Program 2018].

Access to data is another important structural challenge: indeed, training the machine learning tools used in counter-disinformation efforts requires access to large datasets, which are currently mostly owned by private companies and hardly shared with researchers. Validated, empirical data on bots are also very difficult to find since their creators tend to remain anonymous. As remarked by computer scientist Piero Molino, “data access to researchers is certainly an issue. At the same time, companies that nowadays are the self-proclaimed keepers of such data are not held accountable to security standards (the situation just slightly improved after GDPR, but everyday there is a new report of millions of accounts’ information being exposed by those companies), so the current situation of having economically driven actors being the gatekeepers of people’s information is certainly suboptimal. On the one hand, research in sensitive fields like cybersecurity has so far shown how security by obfuscation rarely works, on the other hand, there are one million possible ways in which total freedom of access to people’s data can be used to damage them. **For this reason, more than specific technological solutions, societies and governments should come up with principles that inspire policies, and then evaluate the application of those policies against the guiding principles to make sure those are not betrayed by the implementation or there are no clearly exploitable loopholes to circumvent the principles.**” [SMART Expert interviews, 2019]. This is especially important since new methods and algorithms that allow to analyse massive amount of data seem to be rather effective in the determination

Concerns linked to privacy and compliance with regulations such as GDPR in Europe are also a potential barrier to accessing data. In this regard, a potential solution is the “privacy-by-design” approach suggested by GDPR. A recent example is the MIT’s OPAL project, developed around 3 principles:

- 1) Move the algorithm to the data: the idea is to move algorithm-execution on data at the location of the data repository, leaving full control on access to the repository’s owner and returning only aggregated data.
- 2) Use open algorithms: leave to everybody the possibility to study and vet algorithms to avoid privacy violations and any other issue stemming from the context of algorithms use.
- 3) Data is always in an encrypted state: data must be in an encrypted state while being transmitted and during computation.

While still in an early stage of experimentation, this approach might persuade tech-platforms to make their data available to researchers.

Finally, important gaps in our understanding of how AI works remain: “experts continue to lack an understanding of why complex AI products behave as they do. Neural networks underpin many of the tools people use every day. Yet their inner workings are sometimes beyond comprehension. Without understanding the AI that adversaries use to create and disseminate disinformation, methods to counter such techniques may be difficult to generate” [Gangware 2019]. Indeed, even when automated agents are created with the best intents their behavior can evolve in unpredictable ways, so, for instance, researchers with OpenAI decided to release only a small model of a machine learning system aimed at generating text based on brief writing prompts after realizing that the narratives produced by the fully trained model were basically impossible to recognize as non-organic. On the other hand, Microsoft’s “Tay” chatbot was

withdrawn after one day only on Twitter, as a few hours were enough for it to start mimicking the vitriolic racism and sexism it encountered on the platform.

As remarked by Prof. Guido Vetere "understanding the algorithms used to deliver certain contents to certain individuals is a huge challenge. As a community, we should be able to control these processes, but this is a very complex technical issue: probably, even those who work from within these systems do not know how to control something as deep and complex as the processes at the basis of platforms' behavior, especially if AI is involved. Coercion and regulations are not the right tools to overcome these issues. Instead, **governments – and the European Commission in particular – should take a proactive approach building policies and funding programmes to create and grow alternatives to the current centralized platforms, rather than trying to make arrangements with them**" [SMART Expert Interviews, 2019]. Along the same lines, Prof. Phil Howard from the Oxford Internet Institute observes that "in some circumstances, the ways coded automation interacts with or affects human users is unforeseeable — even by the software engineers who write such algorithms. In others, individuals and organisations work to build software that purposefully targets voters, activists, and political opponents. Politicized social bots are one version of potentially malicious automated programs; discriminatory algorithms are another. Understanding how technologies like these are used to spread propaganda, engage with citizens, and influence political outcomes is an urgent concern" [quoted by Jens Degett, President of the European Union Science Journalist Association in SMART Expert interviews, 2019].

According to Piero Molino, **technological approaches – which are “a posteriori” fixes, will hardly be sufficient to tackle online disinformation and – as we will see in section 8 - policies and incentives (which work on prevention, and therefore “a priori”) should also be put in place.** Indeed “we have at least three aspects of the current circumstances that, combined, led to the current situation. On the one hand you have the absolute low cost of content creation and diffusion, as creating it can either be done by really cheap labour together with semi-automated processes, and the diffusion is fully automatable. The second factor is the fact that it's almost impossible to make the people responsible for the generation of disinformation accountable for their actions, because of a lack in jurisdiction together with a lack in practical instruments of law enforcement on the internet. Already pinpointing who is responsible for the initial spreading is difficult (but not impossible, there are studies and techniques of network science virology that have been applied with great success to the task, but never left the academic lab so far and have not been adopted by law enforcement to the best of my knowledge), but even once the responsible people are discovered, there could be other problems in making them accountable (international laws, the exploitation of mass labour in countries where labour is cheap, lack of local laws in those countries, and so on). Finally, the third point is the general crisis of meaning and truth that the flooding of disinformation has given rise to (anti-vax, flat-earthers and similar groups) that would defend the freedom of spreading disinformation like they would defend freedom of speech, which highlights how thin the distinction really is and how, as a global society, we should probably start rethinking the meaning of freedom, truth, reality and opinion, if we want the current blur to be overcome” [SMART expert interview, 2019].

2.3 Conclusions

In conclusion, while computational approaches to tackling online disinformation have been successfully applied by both researchers and tech-platforms under certain circumstances (see chapter 4), according to most experts and based on literature review, addressing the spread of fake-news and its potentially negative impact on democratic processes and society will require a mix of automated and human driven interventions, from fact-checking and supervision of machine learning to policy action and educational campaigns.

Human-centric AI, a body of research with a strong social dimension, may help us design novel platforms and mechanisms for the public access to news and information, focused on counterbalancing our built-in confirmation bias and transparently striving to expose people to assorted opinions, intelligently. It is possible to imagine mechanisms for helping individuals and communities become informed on controversial issues by offering multiple perspectives, connecting opposing views and conflicting arguments, and fostering critical thought. For example, a robot in a group conversation can highlight information that was unavailable to the group or suggest omitted sources of important information. Advances in person-machine interaction models based on explainable AI have the potential to reach novel cognitive trade-offs between our confirmation bias and our curiosity of novelty and diversity, making it possible for more sustainable and humanized information ecosystems to emerge.

Chapter 3 – QUALITATIVE APPROACHES

Narratives and Narratology

3.1 Introduction

Narratives and narratology have in recent years become a key area of interest in understanding the spread of disinformation as an alternative or in conjunction with network science and the study of algorithms. Instead of focusing on the patterns of distribution, the starting point is the structure and content of the information and how it relates to the broader narrative ecosystem.

One of the most important findings of modern psychology is that the process of understanding is inherently constructive in nature [Bruner, 1991; Neisser, 1967], meaning that we understand through cognitive schemata, or mental models, which organize knowledge on important aspects of reality. Thus, the framework and codes used to organize our understanding of reality builds our view of reality and it is important to note that perceived reality is reality for individuals, groups and societies. In fact, social reality is constructed in interactions between individuals, and social interactions are necessary to sustain it [Wygotsky 1978, Berger & Luckmann 1966], meaning that social reality is often co-created by people in groups, in conversation.

It is this meaning, this shared view of reality that allows individuals to create shared goals and thus collective action needed for any society to function and develop [Fusaroli et al., 2014; Sebanz, Bekkering & Knoblich, 2006]. Thus, narratives are the basis of coordinating and mobilizing individuals and society.

In contrast to physical phenomena, properties of social reality depend more on the social process of construction than on objective properties of social objects and events. Social constructs are the results of human decisions, rather than objective reality. Stories or narrations are natural ways of giving meaning to situations. This process depends on the social context of individuals: people try to merge their own narratives into group narratives. Indeed, narrative schemas, through their repetitions in narratives, provide a common base for values and norms of a given culture [Gergen & Gergen 1973].

Personal narrations are developed following the templates provided by narrative schemas. Adopting a narrative is quite different from learning about a narrative. In contrast to narratives that individuals 'just know about', adopted narratives guide emotional reactions, reasoning, and action. Narratives may be adopted at different levels. At the most superficial level, narratives are just accepted as valid or rejected. At a deeper level, narratives are used to understand reality. At a very deep level of adoption, a narrative becomes a part of the personal identity of individuals. It is used to structure and give meaning to personal experiences and to guide decisions and actions [Singer 2004].

Narratives have the power of governing the life of individuals, social groups and nations. They organise the onslaught of facts and events into coherent, meaningful structures. Considered from this point of view, the current crisis of trust and expertise related to new media and fake news is not something new and unprecedented, but rather a new form of a very old challenge.

As humans, we understand the world through metaphors, images and stories. This is how we give meaning to reality. Telling stories is a "human universal" (Smith 2017) that has inspired anthropologists, scholars from humanities and social sciences, publicists, and self-proclaimed marketing gurus for decades. Narratives unite and divide. They are universal yet extremely varied. They seem almost eternal, not affected by the current of time, and yet they react so quickly to slightest changes in culture, environment and especially – in the available media.

Narratives unify us in one of the biggest imaginable communities – humanity itself. There is no culture that would not tell stories. Some scholars even tend to treat the repertoire of narratives of a given society as almost synonymous to its culture, highlighting bonds connecting mythological narratives with rituals, everyday practices, political systems or tools and media.

But narratives have also a tremendous dividing power. Not sharing a common narrative is one of the basic factors taken into account when delimiting borders between nations, ethnic groups, religions or social classes. "Us" and "them" are very much narrative structures.

At the level of a society, narratives play an important role in building the distinctive identity of a social group. They are culturally transmitted through storytelling and social interaction: through bedtime fairy tales in early childhood,

important readings, television programmes and, most recently, by TV series and social media exchanges. People communicate and co-create stories within their own reference groups (educational profile, social class, occupational group, etc.), narrative communities. They attach their personal narratives to the narratives of the groups they belong to. As a result of the inner cultural transmission, each narrative community keeps its own distinctive beliefs, norms and values. Shared narratives (both auto-descriptive and non-auto-descriptive) allow the establishment of inter-group connections. They answer fundamental questions presiding interactions and collaboration, such as who are we? What are our values and goals? What are our norms? How do we act? How do we relate to each other? Who are our friends? Who are our enemies? What is our history? What is our future? What other objects are important to us?

Phenomena related to the spread of disinformation on social media draw from these basic principles. Not only by attempting to influence directly our definition of “us” and “them”, but also by authenticating certain messages by entangling them into systems of commonly shared values. Even an information contradicting our worldview may become a part of our perceptive horizon if it is deeply rooted in a social imaginary we share, and forged with familiar images, symbols and metaphors. We tend to interpret reality by fitting events into existing narrative schemes [Schank & Abelson, 1995]. New, challenging, surprising events are described by existing narratives, following the rule of “imaginary analogies”. Narrative plausibility is thus a key factor responsible for accepting or rejecting certain ideas, news items or new information. Individuals rarely rely on fact checking to accept or reject new information or ideas. Rather they check if the narrative conveying the new information fits their already adopted narrative schema. Items that fit the narrative schema are adopted, because they are just another example of “what things are”, in contrast, information that does not fit the narratives is rejected, because “things like that do not happen”. It has been shown, for example, that juries in their deliberation and verdicts make decisions based on the plausibility of the narratives of the persecution and the defense, rather than following the instructions to consider each piece of evidence [Pennington & Hastie, 1992].

In Eder’s definition, “cognitive projects are embedded in a discursive universe beyond cognitive constructions. Therefore, the assumption is justified that there is something behind cognitive identity claims which determines their force (or weakness), their plausibility or their non-plausibility. This can be called the narrative plausibility of cognitive projects. Narrative plausibility is the result of an elementary social process. It results from stories that people tell each other, thus creating a space of narrative fidelity. Telling stories implies a social relationship and implies a space within which such stories circulate. This symbolic space is bordered by ‘shared stories’.” [Eder 2006: 257]. Indeed, narrative plausibility is key to drive collective action, and even the most conflictual and unfair situations require some narrative plausibility to be allowed.

Narratives are not static; they evolve and change. Social change and innovation are related to changes in common narratives: change in the interpretation of the current reality, change in the vision of the future and change in the memory of the past. This is because narratives are causally linked to action. By embedding their present context in a narrative structure and enacting their own assigned role in the resulting narrative, actors have a sense of the consequences of their action and a sense of what actions they need to perform to get somewhere. Individuals enact narratives, narratives are thus causally linked to behavior. Narratives shared in a social group represent platforms for group action.

This does not only apply to societal narratives, but also to organizational narratives, including in the private sector. So, for instance, [Lane and Maxwell 2005] have demonstrated that as a company radically changed its technology, its narrative changed to be congruent with the new vision. This change affected stories of how the company begun, what is shared by everybody who worked and/or work in the company and distinct from any other company, as well as the vision and goals for the future.

This also means that a change in narratives is intimately related to both societal and technological/commercial innovation and changes of narratives precede and cause change and innovation. According to [Lane and Maxwell 2005] innovations are always associated with ontological uncertainty. In reality, no-one really knows what the effects of an innovation will be. The uncertainty concerning the outcomes of an innovation can effectively hinder the innovation process. Actors use narratives to infer the likely consequences of an innovation. Narratives thus reduce the uncertainty and enable the introduction of innovations. Research in cognitive science suggests that narratives play a primary role in decision making under uncertainty [Tuckett & Nikolic, 2017].

Social and organizational changes involve several elements interlocked in feedback loops. Each important innovation induces a change in the relations between actors. Often the resistance against change is motivated by the opposition against a change in the relations between the actors involved (e.g. changes in power, centrality, access to information, connections to social networks etc.).

Basic narrative schemas seem incredibly persistent. We are able to understand and even enjoy stories that were told thousands of years ago or by people who are very different from us. Yet those same narratives that enable us to understand people distanced by centuries or thousands of kilometres and to empathize with them, push us to hate our neighbours, stigmatize them or even kill them. There is no war without a strong, shared narrative fuelling mutual hatred. But there can also be no lasting peace that is not backed by a plausible and attractive unifying narrative. The introduction of writing, printing press, radio and television challenged not only the repertoire of dominant narratives, but first and foremost the very form in which they were stored and transmitted. **Scholars have demonstrated how new media facilitated narrative transformations that resulted in remarkable changes in politics, culture and society.** Ancient empires were built not only on slave labour and military triumphs, but first and foremost on the new possibilities of knowledge transmission enabled by writing [Assmann, 2011]. Modern nation states are deeply rooted in the culture of print [Anderson, 2006] and the triumphs of the 20th century propaganda would have not been possible without the narratives transmitted by the mass media.

Narratives are especially important in understanding the information flow in society in the age of social media, especially in understanding the spread of fake news. **The flow of information and debates on social media may be to a large degree understood as a competition of narratives.** Individuals rarely relate news to facts, rather the congruence with adopted narratives decides about the decision to trust the news or to reject it as false. Narratives also decide which news items are attractive and become viral. Adopted narratives shape decisions, for example when voting in elections, and are the platforms for individual and group action, such as engagement in political activity or public demonstrations.

The most commonly adopted strategy for coping with disinformation and propaganda consists in debunking consecutive false claims or labelling certain sources of information as "unreliable". **But looking at fake news from the point of view of narratology highlights the importance of a systemic approach to tackling disinformation.** Only by examining carefully the informational ecosystem will it be possible to create a counteraction that will result not in reinforcing a false claim (e.g. by supporting a conspiracy-based worldview), but in changing attitudes and beliefs of the target audience.

Narratives are also the basis of group identities. Narratives exist in narrative communities and shared narratives decide about who is treated as "us". Narratives define who are our friends and who are our enemies. Social media facilitate the creation of echo-chambers; by facilitating choice of interaction partners they make it easy to communicate only with like-minded others, minimizing exposure to disconfirming information and diverging perspectives and opinions. This has a strongly polarizing effect on society, leading to increased fragmentation. Opinions become increasingly extreme, and common narratives that may be a basis for building consensus and common identity are becoming increasingly rare.

3.2 Narratology

Narratives are forms that our minds and cultures use to organise reality. In other words: "a narratory principle operates to provide meaning to the often-non-systematic encounters and interactions experienced in everyday life" [Sarbin, 1986]. For the purpose of this study we will define a narrative as a complex (created of smaller units) systematic (following a more general pattern) representation that contains a change of state. Below we will explore the five basic axis that define a narrative:

- a. change – narrative as a sequence of events
- b. state – narrative as a system of values
- c. representation – narrative as a sign
- d. systematic character – narrative systems
- e. complexity – basic components of a narrative

a. Change. Narrative as a sequence of events

The idea of change is at the very core of every narrative. There cannot be a narrative without a temporal axis. Change in a narrative may be analyzed as a sequence of consecutive states. A narrative pattern describes the logical succession of units in a story.

A narrative refers not only to a course of events, but also provides meaningful relations between them, in the form of cause and effect. Complex narratives (like novels, epic poems or journal articles) are often built of smaller sub-narratives. This means that narrative sequences can be hierarchically organized, nested in each other. A story may also be told non-chronologically (using for example flashbacks) or omitting some elements, especially because they are considered obvious to a model reader [Eco, 1984] and their detailed account would be redundant. This is precisely the difference between a story (a sequence of events as they took place) and a plot (a sequence of events as they are narrated). A plot implies an interpretative effort by a competent reader.

Fake news often uses this incompleteness of narratives to spread and improve their credibility by telling only selected parts of a bigger story and building on the fundament of an “encyclopaedia of a model reader”. This interpretative dependency of most narratives and their “openness” [Eco, 1984] makes them difficult to analyze, because it is necessary not only to take into account the narrated sequence of events, but also to reconstruct the untold elements of the story, often concerning values and attitudes internalized by the protagonists and model readers. This is an important factor that has to be taken into account in every attempt to analyze large corpuses of narratives, especially using semi-automatic technologies based on machine learning.

b. State. Narrative as a system of values

Narratives are a crucial medium for spreading and sustaining value systems in a given society. As Chamberlain and Thompson suggest [1998: 13]: “messages are not received transparently. Whether in words or images or both, they speak in forms, genres, which carry implicit meanings, expectations and associations; and at the same time their significance, meaning and impact are mediated and transformed by the values, predispositions and social world of the receivers themselves.”

The systems of values within narratives often take the form of symbolic codes typically deeply rooted in everyday experience. These codes provide a repertoire of symbols that serve as a reference point to describe changes of state [Jakobson, 2010]. They may refer to time (e.g. origin-end), space [Yi-Fu Tuan, 1977; Lakoff and Johnson, 2008; Mieletyński, 1981] (e.g. up-down, inside-outside), moral values (e.g. good-bad) or the experience of one’s body (e.g. illness-health, life-death). A narrative may thus, for example, tell a story of healing (change of state from illness to health) or invasion (movement from the outside towards the center of our world).

By exploiting a familiar repertoire of metaphors, symbols and images, fake news is able to boost their credibility. The codes are often interchangeable. For example, tribal identity may be represented through various animal figures or sexual encounters may be replaced by culinary procedures and taboos [Levi-Strauss, 1990; Leach 1989]. A single narrative often consists of many layers using different codes to highlight the positive or negative change of state. For example, cosmos and chaos may be represented by a horizontal spatial axis (the cosmos is inside the borders while chaos is lurking outside), a vertical spatial axis (the cosmos is above while chaos is the realm of the underground) and through a corporal experience (the cosmos is life and vitality while chaos is symbolized by illness and decay). According to [Mieletyński 1981], who adopted such an approach in his studies on myths, we only enter the domain of proper analysis when we attempt to understand the parallels between various codes that show how they are related to a certain system of values.

According to [Jakobson 2010] and [Levi-Strauss 1990] the temporal axis of a narrative and its dominant codes can be described as two dimensions of a narrative: diachronic and synchronic:

Diachronic	Synchronic
Explains the past	Explains the present/future
Necessary to read a narrative	Necessary to understand a narrative
Narrative sequence	Logical sequence

Two dimensions of a narrative (based on Mieletyński 1981:101)

Generating a narrative is always an act of projection of a synchronic (paradigmatic) axis into a diachronic (syntagmatic) one. [Greimas 1977] calls this mechanism "the narrativization of the taxonomy" in which a taxonomy (a paradigmatic system of values) serves as a basis for a narrative syntax. "One can establish a network of equivalences between the fundamental relations constitutive of the taxonomic model, and the projections of these same relations, or operations, bearing this time on already-established terms of this elementary morphology. The regulation of these operations would constitute syntax" [Greimas 1977: 27].

Consecutive repetitions of a single story through different symbolic codes increases credibility, narrative plausibility and creates a feeling of coherence: a mechanism which is often at the basis of online disinformation campaigns.

c. Representation. Narrative as a sign

As structures narratives are very abstract. But in real life a narrative is always a representation, which makes it a special form of a sign. Narratives may materialize as oral accounts, printed novels, movies or fake news. Every narrative is limited by the medium used to transmit it. To understand a success (or fail) of a certain narrative or to create successful counter-narratives it is necessary to take into account the specificity of a medium. Different levels of credibility are attributed to different media. We tend to trust written sources more than an oral account, and by default we are less skeptical towards visual sources than textual ones. Fake news is very often exploiting this naiveté in our perception of videos and photographs. Not only by manipulating the image (which is relatively easy using state-of-the-art digital techniques), but often just by manipulating the captions and creating a false context for a true image.

d. Systematic character. Narrative Systems

Narratives are understandable not only because they refer (via symbolic codes) to a system of values shared by a certain community, but first and foremost because they follow certain narrative schemas. It seems that there exist a very limited number of these schemas, defined by possible operations on the paradigmatic axis. It would be thus possible to juxtapose and compare various seemingly unrelated narratives by reducing them to a common narrative schema or even to create a periodic table of narratives that demonstrates relations between various narratives in a corpus.

The first scholar to adopt such a systematic view on narratives was Vladimir Propp. In his *Morphology of a Folktale* (2010) he demonstrated that all the folktales in a vast corpus he analyzed follow a very strict scenario that Propp broke into a series of functions. Seemingly different characters, objects or events may be perceived as equivalent if they play the same role in the general structure of a narrative. This approach inspired many scholars who adapted a similar perspective to various forms of narratives. For example, [Levi-Strauss 1990] applied Propp's finding to the analysis of myths; [Barthes 1975] to short stories; [Greimas 1987 and Genette 1972] created a general theory of narratology. The structuralist notion of narratives, represented to some degree by all the aforementioned scholars, draws also from De Saussure's [2011] idea of *langue* and *parole* in which a general system precedes particular realizations. In every society there exist not only a repertoire of stories but of basic structures that serve as elements or modules to create narratives. An interpretation is, thus, never focused on a single narrative but has to specify its place inside a narrative system. [Bal 2009] sums this up by saying that the aim of narratology is always a description not of a narrative but of a "narrative system". "On the basis of this description, we can then examine the variations that are possible when the narrative system is concretized into narrative texts." Such a systemic approach applied to online disinformation phenomena implies that we do not need to simply identify and debunk single false news, but whole narrative systems and informational ecosystems that provide narrative structures and symbolic codes. This is far from being simple, however, if we agree that every narrative may be described as a two-dimensional system consisting of a paradigmatic axis of values and a syntagmatic axis of change, basic narrative schemas may be reduced to a set of simple models representing various possible sequences of change expressed in different codes.

Each narrative may be described by a sequence of upward or downward changes in the narrative. The limited number of narrative schema stems from the fact that, if the number of the upward or downward changes is limited, the number of possible combinations is low. For the "no change" or stagnation condition, only three scenarios are possible: "low-low" (poverty, misfortune), "neutral-neutral" (normality, boredom) and "high-high" (prosperity, continuing success). For one change, the direction is either upward (success, growth) or downward (fall). With two changes more scenarios are possible.

Even though there are innumerable stories and narratives, research shows that most of the narratives follow just a few narrative schema [e.g Mc Adams et al., 2005]. For example [Reich 2005] defined 4 social narratives that are common for the American culture, but also for the European one. The Triumphant Individual Narrative describes a

situation where a poor individual achieves richness and glory by hard work and inventiveness. The Benevolent Community Narrative describes a group that roll up their sleeves and working together achieve a goal beneficial to the whole group. The Mob at the Gates Narrative describes a society that is too liberal and forgiving, which lead to weakness and therefore abuse at the hand of culturally inferior outsiders. The Rot At The Top Narrative refers to the malevolence of powerful elites that want to harm the rightful individual.

A basic set of narrative schemas is exemplified in the table below:

Schema	Discursive examples	Keywords	Sample symbolic codes		
			Spatial	Body integrity	Metaphor of home
Status quo	"Poland has always been catholic" "History repeats itself" "Europe doesn't want innovations and it doesn't need startups"	stagnation, stabilization, repetition,	Staying in place	Living	Home
Fall	The decline and fall of the European Union: is it time to rip it up and start again?	fall, decline	Climbing down or falling Moving backward Getting smaller Losing territories, being invaded	Disintegrating Getting ill	Ruin
Progress	Albania's European path: Making significant progress and ambitious for more	progress, getting better, success path, course	Climbing up or flying Moving forward Getting bigger Gaining territories, civilizing	Being more consistent Getting better, healing	Construction site
Trial	The EU overcame the legacy of two World Wars. Now we must overcome the legacy of the Cold War The EU overcame historical hostilities of states by voluntarily increasing their economic and social interdependencies ...	trial, redemption, penitence, learning from mistakes, rising from the dead, resurrection, healing, overcoming, obstacle, crisis, challenge	Hero's journey to the underworld and back again Coming back from exile Regainin g lost territories	Redemption Illness and recovery Death and resurrection	Coming home Return of the prodigal son Rebuilding

This very general structural typology of narratives happens to be surprisingly compatible with the attempts to categorize narratives rooted in the articles and social media expressions examined so far both as part of the literature review and looking at data collected online and revolving around Europe and the upcoming European elections (see chapter 6).

e. Complexity. Constitutive units of a narrative

Narratives are complex structures. To analyse (semi-automatically or manually), compare and categorize narratives it is necessary to parse them into chains of related events, actions or states. This goal may be accomplished by following and modifying classical narratological notions of basic units of narrative structures. Developing de Saussure's chain of phonemes – morphemes etc. [Benveniste 1971] defines discourse (very similar to narrative) as a linguistic entity bigger than a sentence. Similar ideas may be discovered in Roland Barthes "Introduction to the Analysis of Narrative." [1975], where a narrative is analysed as a sequence of "functions". Claude Levi-Strauss analyses myths similarly, using the term "mythem" to describe a basic meaningful unit of a mythical narrative.

Nowadays basic, undividable units of a narrative structure – atoms of the story – are often called narremes. They are isolated and studied by narratology, that helps to identify and to classify universal patterns present in narratives of different origin [Bonheim, 2000]. At the most basic level basic units of narratives may be reduced to transformations described in various semantic codes. Change in time constitutes the syntagmatic axis while the repertoire of values is represented on the paradigmatic axis. Following the horizontal (syntagmatic) axis of a narrative we perceive action (change of states), while following the vertical (paradigmatic) axis we perceive different overlapping layers of meaning expressed in various codes. [Levi-Strauss 1990] compares this two-dimensional structure to music scores in which vertical, narrative relations organized by metonymy are compared to melody, while metaphorical relations between various codes are similar to harmony.

In the table below, we provide an exemplary narrative iterating the same message by using three different codes: geopolitical, corporal and spatial (house related metaphors):

Geopolitical	"We don't see these people as Muslim refugees. We see them as Muslim invaders. For example, to arrive from Syria in Hungary, you have to cross four countries, all of which are not as rich as Germany, but stable. So they are not running for their lives there already." Viktor Orban for "Bild"
Body-related metaphors	"For us migration is not a solution but a problem ... not medicine but a poison, we don't need it and won't swallow it." Viktor Orban quoted by "The Guardian"
House-related metaphors	"No-one will tell us who we let into our own house" Viktor Orban, May 2015, when the European Commission proposed compulsory quotas to redistribute asylum seekers.

For Viktor Orban, and his model audience, Hungary may be described as a country, but also as a body or home. This creates an opportunity to switch between various codes to create a more suggestive and seemingly complex story, while really the arguments are just repeated, styled in different metaphors. The main problem with such narratives, as it has already been mentioned, is that they usually are not presented in their integrity. What we get is rather fragmentary and dispersed. Part of a narrative may be presented in one code (e.g verbal), another part in another one (e.g visual), while yet another may remain latent, unspoken or barely suggested. Moreover, while some narratives are dominant – produced and circulated in a top-down manner - others are rather bottom-up and can be harder to spot. Narratives are created on purpose and transmitted in a top down process by actors high in power and authority, as for example in the case of a government's policy or regulatory intervention. Such narratives represent "official narratives" and serve the purpose of controlling social dynamics. Every dominant narrative is usually accompanied by counter-narratives and/or "shadow narratives". Unofficial, shadow narratives are built in relation to the official ones, both in agreement or in opposition to them.

Official and shadow narratives are deeply intertwined in the case of new media, where comments and social media posts are usually fragmentary, meant to be perceived merely as links in much longer and complex narrative chains which in turn are dispersed in several articles or posts.

3.3 Conclusions

“The science of mind has lit up a vast landscape of unconscious thought – the 98 percent of thinking your brain does that you’re not aware of. Most of it matters for politics. The mind that we cannot see plays an enormous role in how our country is governed.”

George Lakoff

An increasingly influential stream of research demonstrates the integration of cognition and emotion in political decision-making. Political cognition is emotionally shaped. The role of narratives in shaping people’s minds has become an important area of research and debate, in particular in recent years when entire societies have made choices that seem “rationally” counterintuitive.

As debates around migration, Brexit, climate change and vaccines become increasingly emotional policy-makers and scientists alike have had to accept that in the battle for hearts and minds of human beings, narrative will consistently outperform data in its ability to influence human thinking and motivate human action.

With a polarising society and highly-charged emotional debates on the table, it is ever more important to look at what can we learn from recent research in psychology, neuroscience, economics, sociology and behavioural science about how the minds work and how they are shaped. Narratives exist within a narrative community where perceived reality IS reality. This is why a system approach is needed to understand the flow of information in modern societies. Narratives are crucial in this as the dynamics of social systems are based on meaning and narratives is how we communicate meaning.

It is through narratives, shared interpretations of the world, that people unite. Yet, narratives also have the power to divide and polarize. Not sharing the common narratives is one of the basic factors taken into accounts when delimiting borders between nations, ethnic groups, religions or social classes. “Us” and “them” are very much narrative structures.

The recent surge in disinformation, or fake news, draws from these basic principles. Not only by attempting to influence directly our definition of “us” and “them”, but also by authenticating certain messages by entangling them into systems of commonly shared values.

Two very interesting points were highlighted that are very relevant to the current discussions around the need for a new narrative in Europe:

- 1) Research shows that whilst narratives can be very persistent, when conditions change narratives change. Narratives need to be adapted to understanding new realities and are then pivotal in shaping the understanding of the changes occurring.
This past decade, following the financial, economic, social and political crisis that have ravaged Europe, the call for a paradigm shift has become mainstream. At the same time, the old narrative, shaped by neoliberal economic theory, with its tenets that globalisation and unrestrained capitalism will be good for everyone has been proven wrong. With the vacuum that has been left in the wake of this change, numerous new (and less new) narratives have sprung up across the world trying to redefine our understanding of reality and the framework that should guide our future development.
- 2) The persistence of narrative schemas seems also challenged by the rapid change of media. The introduction of writing, printing press, radio and television challenged not only the repertoire of dominant narratives, but first and foremost the very form in which they were stored and transmitted. Scholars have demonstrated how new media facilitated narrative transformations that resulted in remarkable changes in politics, culture and society. Ancient empires were built not only on slave labour and military triumphs but first and foremost on the new possibilities of knowledge transmission enabled by writing (Assmann, 2011); modern nation states are deeply rooted in the culture of print (Anderson, 2006) and the triumphs of the 20th century propaganda would have not been possible without the narratives transmitted in the mass media. Nowadays, in the dawn of the new, digital era we face another great change in the form and content of the dominant narratives fuelled by the new media, especially social media.

In a recent report published by the Joint Research Centre of the European Commission on “Understanding our political nature – how to put knowledge and reason at the heart of political decision-making” the link between disinformation, emotion, values, and identity on the way we think, talk and take decisions.

Chapter 4 – EMERGING THREATS AND SOLUTIONS

4.1 Introduction

This chapter of the research examines the various techniques that are used to distort the perception of information that is disseminated through social networks. As mentioned in previous chapters, misinformation is almost always the result of distortion, such as the citation of a sentence outside the context in which it was pronounced or the juxtaposition of verifiable data and facts with others that are not verifiable or of ambiguous interpretation. Current technologies allow those who have the intent to influence public opinion to reach increasing levels of realism, for example through the manipulation of images or the creation of deepfake videos that are indistinguishable from those that were really shot, thanks to the increasing effectiveness of Artificial Intelligence in deceiving both the human eye and the algorithms that should discover the deception. The private or even personalized transmission (for example through microtargeting) of the messages allows to increase their effectiveness further and makes it impossible to study their diffusion, especially when they pass from one user to another through private channels (as happens for example through Messenger, Telegram or WhatsApp).

4.2 Emerging Threats

In this complex environment three phenomena stand-out as particularly threatening for democratic and societal processes: the rise of computational propaganda and micro-targeting, the rise of deep-fakes and the weaponization of the debate around ways to tackle online disinformation, with a particular focus on the role of bots and automated agents.

4.2.1 COMPUTATIONAL PROPAGANDA AND MICROTARGETING

Computational propaganda has been defined as “the use of algorithms, automation, and human curation to purposefully distribute misleading information over social media [Wooley & Howard 2018, p. 3]. The conclusions of the Computational Propaganda Research Project at the Oxford Internet Institute, a project involving 12 researchers that worked with 65 experts and analyzed millions of posts in several countries, reveal the main trends in this form of online political influence. According to the report, as social media are one of the main platforms for disseminating news and political engagement (especially for young generations), they became one of the most important media platforms for political influence.

The importance of social media for political influence is due to a number of factors: the enormous popularity of Facebook as a platform for public life, widespread use of social media in many democratic countries as a means to share political news and information, especially during elections, and in authoritarian countries, to control and manipulate the population, especially during political crises. In countries with limited access to social media, these platforms provide a key communication channel among key players such as journalists, civil society leaders and political elites. In all the countries studied, civil society groups were targeted by online disinformation campaigns. The study found that social media bots have been broadly used to influence online discussions. In some countries, politically oriented content on social media is to a large degree controlled by governments and organized disinformation campaigns: for example, approximately 45% of Twitter activity in Russia is generated by automated accounts. Although the main goal of autocratic regimes is to control political orientation of their own citizens, they also try to influence population’s choices in other countries. As the authors state [Wooley & Howard 2018: 4], “Chinese influence campaigns target Taiwan, Russian influence is directed for example mainly at Ukraine, but also at Poland. (...) In democracies, individual users design and operate fake and highly automated social media accounts. Political candidates, campaigns and lobbyists rent larger networks of accounts for purpose-built campaigns while governments assign public resources to the creation, experimentation and use of such accounts”. The authors argue that the most effective computational propaganda uses a mix of humans and bots working together. Humans, often taking the position of trolls, create and curate the information, while bots spread it by reproducing or linking it, targeting individuals for specific influence attempts via increasingly sophisticated algorithms.

A large proportion of political discussion on social media platforms may originate from a small number of users. For example, in Poland a small number of right wing and nationalists accounts generates a relatively large proportion of political discussion on Twitter. This may involve ideologically motivated individuals but also paid propaganda workers. [Wooley & Howard 2018] also show that computational propaganda has been used heavily in Ukraine, in the Brazil 2014 presidential election, in the impeachment of the former president, and in the 2016 municipal elections in Rio de Janeiro, not to mention the US 2016 presidential elections and Brexit. The analysis reveals that "**Computational propaganda is now one of the most powerful tools against democracy.** Social media firms may not be creating this nasty content, but they are the platform for it" [Wooley & Howard 2018: 11].

Micro-targeting is probably the most advanced technique to implement computational propaganda campaigns. The approach captured the public attention with the controversy concerning violations of privacy and manipulation of public opinion during the Cambridge Analytica scandal, having been allegedly used during the 2016 US election in the campaign of one of the candidates [Grassegger & Krogerus 2017], in social media campaigns preceding Brexit [Howard & Kollanyi: 2016; Cadwalladr 2017] and in election campaigns in Brazil [Arnaudo, 2017]. Micro targeting was widely discussed by public media, leading to hearings in both the US Congress and the British Parliament, and resulting in changes in access permissions to private data collected by social media companies. The use of micro-targeting goes beyond politics, being used extensively in commercial marketing campaigns. While evidence on impact of political micro-targeting is scarce [Lazer et al 2018], research [Metz and al, 2017] has demonstrated that micro-targeting significantly influences users' online behaviours and purchasing decisions.

In summary, micro-targeting is the technique of persuasion on electronic media, where each addressee receives a message specifically tailored to match his or her individual characteristics, usually personality traits, which are automatically predicted from digital footprints. Micro-targeting allows the influencer to personalize persuasive messages in massive scale influence campaigns. The most striking feature of micro-targeting is that personality of the target of influence is not directly measured by questionnaires, interviews or rating, but rather is being assessed on the basis of easily available information - digital footprints – traces of activity of internet users, including Facebook Likes, location within a friendship network on Facebook, contacts on social media, links, posts, history of searches on search engines, and history of browsing. The technique is described in detail by [Kosinski et al 2016].

The scientific basis of micro-targeting has been developed in a research conducted using legally obtained data by Michał Kosiński and his collaborators who, differing from what was often claimed, never worked with Cambridge Analytica. Their research was published as a scientific work in peer reviewed journals. These findings were later replicated on Facebook with paid volunteers by another psychologist, Aleksander Kogan, whose work and data were used for Cambridge Analytica influence campaigns. **It has been proved that a broad range of psycho-demographic traits, including personality traits and intelligence can be measured with high accuracy on the basis of Facebook Likes** [Kosinski, Stillwell & Graepel 2012]. The authors have used a dataset of 58,000 volunteers, who agreed to share their data. The dataset concerned Facebook Likes, detailed demographic profiles (e.g age, gender, parental separation) the results of several psychometric tests as well as intelligence test, scales of happiness, drug use, and political orientation). The authors used statistic techniques to reduce high numbers of likes to smaller numbers of more abstract variables, and then applied multiple regression techniques to predict other variables on the basis of Facebook Likes. The results have shown that psychological and demographic characteristics can be predicted with high accuracy. For example, sexual orientation was predicted in 88% of cases, race (Caucasian or African-American) in 95% of cases and political orientation in 85% of cases. The accuracy of prediction of personality traits was close to original assessment by psychometric tests. Further research using similar methods [Youyou, Kosinski, & Stillwell, 2015] has shown that personality characteristics and answers to questionnaires (e.g. behavior) can be more accurately predicted on the basis of Facebook than by one's spouses or friends. The research has revealed that the accuracy of prediction grows with the number of digital footprints (e.g Facebook Likes) used as predictors.

Further work has shown that psycho-demographic characteristics can also be predicted on the basis of photographs of faces. In a series of studies [Wang, Kosinski 2017] over 3,000,000 images of over 70,000 individuals were collected from a public dating site. Information concerning sexual orientation of

each individual was also collected. From this dataset a set of approximately 4,000 gay and 4,000 heterosexual men and approximately 3,500 and 3,500 heterosexual women were created. A deep learning neural network pretrained for facial feature recognition was used to extract facial features from the faces. Logistic regression was used to predict sexual orientation from the faces. When the task was to distinguish on the basis of a single photograph which of the two faces represented a gay person and which a heterosexual one, the accuracy of the prediction was 81% for men and 71% for women, much higher accuracy than for human judges. The accuracy went up to 91% for men and 83% for women when the prediction was based on 5 images. As in the studies concerning digital footprints, the accuracy of prediction increased with the amount of information on which the prediction was based. The prediction was, however, much lower when the classifier tried to predict being a gay man (57%) and a gay women (58%) on the basis of a photograph, rather than to distinguish which of the two is gay and which one is heterosexual when simultaneously presented with two images of one gay and one heterosexual person, which was the procedure employed in the previously described studies.

There are a lot of indications suggesting that micro-targeting is highly effective. Part of its effectiveness is that it allows the source of communication to address the issues that are specifically relevant to the addressee of the message. It also allows the persuader to choose the style of arguments that are fit to the personality, values and beliefs of the target of the persuasion, the style that the target is most likely to resonate with. Micro-targeting can allow the sources of messages to tailor the messages to specific needs, motives and values of the targets of persuasion. In principle, it is also possible that the techniques of micro-targeting could be used to identify individuals who have adopted specific narratives and use the narrative schema underlying these narratives to formulate persuasive messages.

Micro-targeting can also use known information acquired in previous research about the factors that influence particular attitudes, voters' preference etc. Knowing for example that a particular candidate has more support among voters high on anxiety, one can identify likely supporters of the candidate as individuals high in neuroticism, and send them messages encouraging them to participate in the election and vote for the candidate who is portrayed as someone who offers safety and is capable of dealing with threats.

In the public debate on social media, the high effectiveness of micro-targeting has been discussed as one of the reasons why the outcomes of the 2016 US election and Brexit referendum had different outcomes than those predicted on the basis of surveys conducted before the votes [Cadwalladr, 2017; Grassegger & Krogerus 2017, Philip .& Kollanyi 2017, Risso 2018].

The effectiveness of micro-targeting was tested in three experiments by [Matz & Al 2017]. In these experiments an existing database (my.personality.org) which contained both personality measurements and digital footprints was used to establish which footprints (e.g. Facebook Likes) were predictive of a specific personality characteristic. Two separate messages were formulated, one formulated to fit preferences of those high on this characteristic, and one to fit the preferences of those who are low on this characteristic. For example the message advertising a beauty product for women addressed to highly extraverted individuals was "dance like no one's watching (but they totally are)" accompanied by a dancing woman in a crowd, and the message to the low extroversion individuals was "Beauty doesn't have to shout" and it was accompanied by a picture of lonely woman looking into a mirror. In the matching condition each message was sent on social media to those whose digital footprints were indicating a matching level of the personality characteristic, in non-matching condition each message was sent to individuals of opposite characteristic. The effectiveness of each message was assessed by measuring the number of clicks on the message and the resulting behaviour compliant with the persuasion attempt e.g. a purchase of the product or installing of a software.

The results of the three studies provide converging evidence for the effectiveness of psychological targeting in the context of real-life digital mass persuasion. Tailoring persuasive messages to the psychological characteristics of large groups of people was effective in influencing their actual behaviours and choices. Results show that persuasive appeals that were matched to people's extraversion or openness-to-experience level resulted in up to 40% more clicks and up to 50% more purchases than their mismatching or impersonalised counterparts. [Metz and Al 2017].

In a real-life micro-targeting campaign, the effectiveness of the influence would probably be significantly higher. In these experiments estimation of personality was based on a single like. In a real campaign full

history of digital footprints would provide much more accurate assessment of personality. Also, more than one personality characteristics could be used to design a personalized message. Moreover, the issues, in contrast to real life political issues, were not particularly involving on a personal level. One can thus expect that micro-targeting skilfully used in political influence campaigns would be much more effective.

These experiments taken together show that it is possible to target accurately large groups of people without measuring their personality explicitly by questionnaires. These experiments also show that large groups can be manipulated effectively by social media campaigns. These techniques may be used to influence voting choices of undecided voters and discourage voters who are likely to vote for the opposing party candidates from voting. The problem with micro-targeting from the perspective of democratic principles is that messages formulated by the same party or a candidate to different voters could not only be differently formulated but also have quite different, possibly conflicting meaning. In effect, when voting for a candidate, or a party, different voters would vote for quite different programmes or for a different vision of the same candidate. This inconsistency might be difficult to detect, because the message would be delivered to a specific individual, and not seen by others. This would lead to the paradoxical outcome of voters building a political programme instead of voting for it.

4.2.2 DEEPPAKES

A new frontier of disinformation is the use of deepfake, that is the production of fake video through computer synthesis of human images made with algorithms and artificial intelligence. The Deepfake phenomenon was first reported in December 2017 in the magazine Vice reporting about the creation of pornographic videos superimposing the images of other actresses. This was successively leading to its widespread reporting in other media. In the United Kingdom, producers of deepfake material can be prosecuted for harassment but there are calls to make deepfake a specific crime.

This new generation of fake-news has emerged due to recent development in the field of artificial intelligence in the use of deep learning algorithms in video information processing (Blitz, 2018). This technology allows even amateurs equipped with a desktop computer to swap someone's image and voice with a character in already existing videos by feeding real data about someone's face and voice into the fake image. It takes a number of pictures of the face of an individual, as well as samples of their voice, and the software powered by artificial intelligence can insert someone's face and voice in any video.

Potential for political misuse is immense: with a variety of experiments made with fake speeches of President Obama or President Putin .

It is possible, for example, to emulate someone's face and voice answering questions in real time, when in reality the answers are provided by someone totally different, even of different gender. Using this technique it may be possible, for example, to show an online news conference of a politician, where all the viewers can attest to authenticity of the conference, because they see the politician in real time reacting to authentic questions of journalists, but in fact all of this is technologically orchestrated, with the media experts of the opposing site fully controlling the answers. **Because first-hand visual experience is the subjective primary criteria of truth, deepfake technology strongly dilutes the border between fake and truth making it hard or nearly impossible to debunk fake news.** In [Galaway 2019] words: "images are easy to manipulate but more difficult to detect with the currently available image analysis and forensic tools [...]. While reverse image search tools exist, robust reverse video search tools are similarly needed to detect synthetic video content. Reverse image search can be used to identify the source of an image online by pointing to where else it has appeared, which helps people verify the origin of an image quickly. Google Reverse Image Search is one such tool. However, reverse video search tools are limited in their functionality, relying on thumbnails or dissected portions of videos".

4.2.3 A NEW ARMS RACE

As already mentioned in the introduction, in the wake of the scandal of Cambridge Analytica, public debate around micro-targeting and disinformation has acquired a military ring which is highly unusual in the public discourse around new technologies. The shift was already clear in [Cadwalladr, 2017] reconstruction of

what happened during the Brexit campaign: “this is not just a story about social psychology and data analytics. It has to be understood in terms of a military contractor using military strategies on a civilian population. Us. David Miller, a professor of sociology at Bath University and an authority in psyops and propaganda, says it is ‘an extraordinary scandal that this should be anywhere near a democracy. It should be clear to voters where information is coming from, and if it’s not transparent or open where it’s coming from, it raises the question of whether we are actually living in a democracy or not’. Most recently, the Commission’s communication on “Tackling online disinformation: a European Approach” has also focussed on the links between fake-news and security, highlighting that “the Commission and the European External Action Service will explore further options to develop strategic communications responses and other mechanisms, together with Member States, to build resilience as well as counter systematic disinformation campaigns and hybrid interference by foreign governments towards citizens and entities in the EU” and that “the Commission, in cooperation with European External Action Service, will report in June on the progress on bolstering capabilities to address hybrid threats, including cybersecurity, strategic communication and counter intelligence areas”.

This “military” approach is somewhat mirrored in both discussion and practice around detecting bots and automated agents. As pinpointed by [Lazer and al, 2018]: “Bot detection will always be a cat-and-mouse game in which a large, but unknown, number of humanlike bots may go undetected. Any success at detection, in turn, will inspire further countermeasures by bot producers”.

Clearly, it seems very difficult to strike a balance between protecting citizens’ right to privacy and to accessing reliable information and companies’ right to protect their IPs, while also avoiding the risks of censorship connected to regulatory intervention.

Today, both private platforms and governments tend to see data as an “extractive” industry: we pull data out of people for good (to provide them with more personalised, integrated and cost-effective services) or for worse (to sell them more products, to manipulate their opinion), but there are still very few examples of projects which look at data as something co-created with people and machines, including bots. This is also because, as technology becomes increasingly ubiquitous and frictionless, understanding the full spectrum of the data we generate at any time – as we walk across sensors and cameras, search the internet for products and services, make financial transactions using our cards or smartphones, get our heartbeats recorded, interact with smart grids and buildings - becomes increasingly difficult. Indeed, even for the most gifted technologists, understanding both what data we generate, who uses it, how – with what other data is recombined and analysed – and for what purposes, is nearly impossible. In this context, **it is fundamental for public authorities at EU and national level, to help safeguard spaces where data and computation can be organised and performed as ever-evolving commons, reminding that a common is given not only by the data, but also by the relational ecosystem around the data and the fact that this ecosystem negotiates and shares practices and rules.** This can only happen if we heavily invest in supporting people, small businesses and public administrations in becoming aware of the data they produce – and of the fact that sometimes they cannot be aware neither of the data they produce nor of the ways this is used – and if we are able to find easy, cost-effective and engaging ways to turn data into actionable information, which should then be followed by concrete actions. In fact, it seems difficult-to-impossible to imagine any solution to disinformation and to its effects on society without reflecting on the social architectures and processes that bring to these kinds of phenomena.

Disinformation would be difficult to achieve in a cohesive, empathic, educated society. As a matter of fact, most disinformation phenomena are based on separation within society. Algorithms, artificial intelligences and, in general, computational techniques “classify”, meaning that they subdivide the population they are addressing (whether it is consumers, electors, users of a service, etc) and their interactions into classes, in order to be able to personalize the interactions, actions and communications. This is true whether we’re talking about marketing, political communication, product placement or other possible cases.

This classification does not correspond to a sense of belonging and of solidarity: unlike in other times in human history, being grouped by institutions (in this case, computational institutions) into classes which are exposed to different conditions, does not bring people in these classes together. Most of the time this fact is not even knowable and understandable by people, who are separated without even knowing it [Tufekci, 2014].

The military, cold-war like escalation of algorithms, BOTs, AIs and other computational cyberwarfare tools and techniques falls into this scenario and draws stamina from this scenario: in a world that is progressively more angry, aroused and polarised escalation brings results that are communicable to the public, but which are not able to deal with the complexity of society. At the same time, the advancements bring on further advancements from the other factions. The result is a technical and methodological progression of those same mechanisms which are at the core of the problem, and a society which is progressively more divided and extreme, and in which people's rights and liberties are progressively eroded. [Iaconesi, 2017]

For this, solutions should be designed which are able to tackle this scenario not only with the technical approaches which see the development of tools, techniques and methods able to progressively confront with the rising standards and achievements of disinformation operators. These solutions should go beyond the idea of implementing tools and services, and would assume the form of actions which take place in the public sphere, and that combine technologies, tools and engagement strategies in order to 1) promote a culture and awareness of data and computation, 2) enact processes which stimulate social imagination and 3) adopt artistic and design/creative approaches to trigger solidarity, empathy, social cohesion.

Other types of solutions should be designed to be able to address the rising separation between citizens, the need to produce the conditions for solidarity, empathy and the formation of an inclusive, high quality relational ecosystems in society.

These kinds of actions can be achieved by combining multiple disciplines and technologies. They should start with the conception and design of new education processes, professions and roles for citizens and professionals which are dedicated to "connecting the dots" in society, to obtain higher quality and more inclusiveness in relational ecosystems in cities and regions: using social and cultural actions, and also using data and computation in open ways, creating alternatives for the current "extractive" data and computation industries which bring data and computation out in the public sphere, where society can agree on rules, regulations and protocols in open, inclusive, readable, knowable ways.

Collaborations and interactions between sciences, technology, arts and design should take place to bring opportunities for social imagination in the public sphere, in which the results of scientific and technological innovation are not only showcased, but also act as the trigger for constructive, inclusive, collaborative imagination of future scenarios that are capable of bringing better relations, information, communication, knowledge and opportunities for exchange.

4.3 Emerging Solutions

Necessarily the debunking procedures and software are different for different use cases (i.e. limiting the circulation of disinformation on WhatsApp is particularly complicated [Daniel Funke, 2018] and identifying deepfakes is nearly impossible, while debunking on Twitter is relatively easy). In any case, there is the urgency of providing answers to this problem, and various stakeholders are nowadays creating lists of software and websites devoted to debunking fake news [Bishwas, 2018]. In the words of an expert of this research area (D. Lazer) **"There needs to be some regular auditing of what the platforms are doing and how much this information is spreading, because there is a collective interest in the quality of the information ecosystem that we all live in"** [Nicodemo, 2018].

Attempts to limit the spreading of disinformation on social media can be classified in two main categories: **"(i) those aimed at empowering individuals to evaluate the fake news they encounter, and (ii) structural changes aimed at preventing exposure of individuals to fake news in the first instance"** [Lazer et al, 2018]. In the first category we mainly find fact-checking – both organic and automated - and targeted education programmes, while the second category includes software developed to debunk fake news and targeted policy actions.

Concerning **fact-checking**, solutions span from web-platforms run by volunteers and/or professional organisations who evaluate factual claims of news reports (for instance Snopes and Fullfact or dedicated sections of traditional

media such as the Washington Post), to providing contextual information on content inserted by intermediaries, such as those used by Facebook.

As highlighted by [Lazer and all 2018] though, there is no agreement on the efficacy of fact-checking, and paradoxical effects having also been demonstrated. Indeed, as mentioned in the section on narratives, people tend not to question the credibility of information which comply with broadly adopted beliefs and narratives. Selective exposure (i.e. only finding information which is aligned with one's pre-existing beliefs), confirmation bias and desirability bias are all well-known phenomena which while magnified on social media are hardly influenced by fact checking of a given fake news story. Moreover, given that people tend to remember information or how they feel about it, but not the context in which information is provided, while also being more likely to accept familiar information as true, repeating false information – even just to disprove it – may increase an individual's likelihood of accepting it as true. Of course, research proving the opposite also exists [Swire et al 2018; Ecker et al 2018]. For instance, one study of 7,200 participants found that counterfactual information can change partisan opinions when the information is presented with strong evidence [Kim 2018].

Attitudinal Inoculation, has also been proven effective, that is similar to how a vaccine builds resistance to a virus: the idea is to warn people that they may be exposed to information that challenges their beliefs, before presenting a weakened example of the (dis)information and refuting it. This strategy can better inform, and even immunize, participants to similar misleading arguments in the future [van der Linden 2017].

Concerning **educational programmes** aimed at developing critical skills and the ability to discern reliable sources on social media, while initiatives in this area keep multiplying [Jones 2017], there is still no scientific evaluation available proving the efficacy of this educational interventions, and particularly their long-term impact. As suggested by [Lazer and al 2018] it will be important to establish a sound impact evaluation framework, including longitudinal studies.

Concerning **algorithmic solutions**, a broad array of initiatives have been implemented both by independent organisations and by tech-platforms [Weedon et al 2017; Crowell 2017] trying either to identify and prevent bots from disseminating information or by trying to include quality measures in the algorithms presiding the ranking of contents, however, once again, information on how these solutions work or on their impact is very scarce.

Policies and regulations aimed at counteracting the circulation of disinformation are also being piloted across the world. On the 9th of May 2019, a bill was passed by Singapore's parliament requiring online media platforms to amend or remove contents considered false by the government, with penalties for perpetrators including prison terms of up to 10 years or fines up to S\$1m (\$735,000). A mix of regulatory action and economic (dis)incentives could also be imagined, as highlighted by computer scientist Piero Molino: "what I imagine is a policy that disincentives the spread of disinformation by increasing the cost of access to the means of distribution, which would make running a gigantic viral campaign on social media like the ones that are happening before elections in European countries really expensive to do. At the same time the policy would have to make both the sources and the perpetrators of disinformation accountable for their actions, imposing strong disincentives on the creation of such content (high fines, imprisonment and any other form of law enforcement). This, combined with technologies allowing to identify disinformation spreaders and increased accountability for the social media platforms that are used as vehicles for disinformation (by allowing, for instance, rights to law enforcement to tear down entire discussion threads without having to pass through the usual Facebook/twitter process) could help mitigate the issues" [SMART Expert Interviews]. The risk is of course giving too much power to governments in setting what is true and what isn't.

Policy measures could also include **new ways of looking at data ownership**, something that GDPR has already introduced in the EU regulation arena: "more than with specific technological solutions, societies and governments should come up with principles that inspire policies, and then evaluate the application of those policies against the guiding principles to make sure those are not betrayed by the implementation and that there are no clearly exploitable loopholes to circumvent the principles. It's an iterative process of trial, error and adjustment, but the starting point should be a humanistic declaration of rights of individuals with regards to the data they produce. And once those rights are defined, for instance stating that all the data one produces directly or indirectly belongs to them, which seems a pretty reasonable basic right that currently is always disattended, then there could be policies in place like the right to sell or lend one's data and the right to decide how, who, if and when the data will be accessed and used. A policy like this would transfer the problem from "Should facebook / twitter / google share their data?" to the single individuals deciding for themselves. But this is just an example, and there is no consensus yet on the principles to begin with, so there's a lot of work to be done" [Piero Molino, SMART Expert

Interviews]. Of course, such an approach would radically affect tech-platforms business models, facilitating the rise of decentralized platforms and making mass disinformation campaigns extremely more expensive.

As highlighted by prof. Guido Vetere, funding decentralized alternatives to the current media platforms might be far more effective than regulating them or trying to make arrangement with them: “the decentralization of social platforms should be encouraged by active policies, which would also lead to overcoming monopolies. The idea is that we should support a next wave of better platforms, rather than try (or pretend) to control existing ones. A possible path starts from supporting the specification of functionalities, models and schemas (i.e. an Application Programming Interface) by calling academia, public research, developers, communities of practice on collaborative, open platforms such as GitHub. There’s no need for huge investments for that, just some seed money to support the organisation and have a handful of right people involved. Then we give the right amount of resources to a lot of small nodes willing to implement the network based on that specification. It is reasonable to assume that each node could be financed with at most one thousand euros to start up. If we consider that an EU Flagship project sizes up to 1 Billion, a ‘Decentralized Flagship’ could support thousands of independent nodes in each Member State. There are already attempts to go in this direction out there, such as Diaspora and Mastodon, but these are small and isolated initiatives. The W3C is at work on the problem of decentralized identifiers, and Tim Berners-Lee is striving to restore the original Web architecture: all these efforts should be encouraged and sustained. If the Commission launched a decentralized Flagship for social networking, this could really make a difference. Of course, a decentralised system is certainly more complex and less efficient than a centralised one: the way Facebook works on user inputs (‘like’ or whatever reaction) relies on the big architectural advantage of having all the data accessible by means of a single logical endpoint. Centralised systems are easier to create and manage. And critical mass is important too: there are over 2 bn people already on Facebook, which is really a heavy legacy. Moreover, this mass is a huge revenue generating asset, since if a company (say a European one) wants to know to whom they could advertise their products, the only option they have now is to buy profiling data from Facebook or Google. People have not a clue of what their data is worth. So, for sure, when we talk of decentralization, we are not talking about an easy endeavour. At the same time, this is not impossible. Tech platforms are subject to fast-changing and disruptive trajectories: if we could set-up a decentralised platform which had some desirable feature (primarily, grant user privacy) I think we could make it happen. For instance, if I dealt with my provider’s node – instead of a centralized global platform - I could reliably use a customized algorithm to get my timeline as I like and not as Facebook wants it. This way, I could have a totally personalised and transparent service. Decentralization and pluralism are crucial here: should I end up not trusting my provider anymore, I could easily switch to another one, without data or connection loss. Many people would be more than happy to pay, say, 5 euros per month to access a social network which, besides being secure and respectful of privacy, could also offer the possibility to control its behaviour and to smoothly escape if something goes wrong. At the end, we need to make decentralisation “sexy” by improving both users’ safety and experience. **It is crucial to come up with a business model which can open a market that currently doesn’t exist.** I believe this will happen anyway, sooner or later, but if the Commission could accelerate this process by supporting and stimulating something that still seems to be far from being market-driven, this would be a key strategic move. To do that, we need all the wisdom we are capable of, this is really the best resource we can mobilise in Europe: our sense of liberty and sociality, our ability to integrate differences, our open research and critical thinking. The point is precisely to make decentralized social platforms a key ingredient of the European culture, to make them the object of a “social desire”. There is so much room for EU policy here: electric cars are incentivised, so why not social decentralised platforms?” [SMART Expert interviews].

A selection of fact-checking sites and platforms as well as of mixed solutions aimed at countering the spread of disinformation on social media is presented below, while specific education programmes have not been considered in this paper.

4.4 Computational Approaches

BOTOMETER

This website ([https://botometer.iuni.iu.edu/!](https://botometer.iuni.iu.edu/)) checks the activity of a Twitter account and gives it a score based on how likely the account is to be a bot. Higher scores are more bot-like. Botometer is a machine learning algorithm trained to classify an account as bot or human based on tens of thousands of labelled

STORYZY

This company (<https://storyzy.com/>) uses algorithms of artificial intelligence (i.e. Natural Language Processing) "A social media user can use Storyzy's Quote Verifier to validate a quote by pasting the text from an online news in the box and hitting check. With a database of over 15 million quotes, it extracts reported speech from the online textual news material, along with other metadata. The text is converted to data using NLP and verified for authenticity" [Bishwas, 2018].

TWITTERTRAILS

TWITTERTRAILS (<http://twittertrails.com/>) is an investigative and exploratory tool, allowing to analyse the origin and spread of a story on Twitter. The issue of the validity of a story is attached to the credibility of the source originating it. It therefore provides information on the sources more than on the story itself. Furthermore, it provides visualizations to trace not only the origin, but the spread of a story. In October 2018 they have already examined 587 stories. For each story, the algorithm used measures how widely the story spread and how skeptical users were about its validity. By measuring crowd behavior, the TwitterTrails algorithm allows journalists to investigate claims and determine whether they appear to be true or false [Metaxas et al., 2015].

USERFEEDS

This Polish company (<https://userfeeds.io/>) uses a protocol for establishing information relevance in crypto-economic networks. In particular it forms a network on the basis of ranking algorithms that verify the quality of the news. With respect to standard protocols of transmissions as TCP/IP, SSH HTTP they aim at designing a transmission that takes into account the reliability of the parts involved in the transmission. The Userfeeds protocol uses one message type - a 'proof-of-evaluation' - that can be processed by both consensus and personalization algorithms. The owner of the company also describes an economics of news based on these tokens and argue about the economic value of the relevance of news produced in the net. It creates a content network that depends on ranking algorithms that use digital tokens such as Bitcoins to verify to the quality of the item.

4.5 ORGANIC FACT CHECKING

BUTAC

The blog (<http://www.butac.it/>) BUTAC (from Italian "bufale un tanto al chilo", i.e. "hoaxes by the pound") is based on volunteers looking for info and debunking one by one memes and news present on WWW, Twitter and Facebook).

FULLFACT

Fullfact is an independent fact-checking charity operating in the UK. Further to fact-check Prime Minister's Questions and BBC Question Time and publish a round-up of the main claims that have been made every week, the organization monitor news on primary citizens' concerns according to the IPSOS MORI Issues Index. In addition, Fullfact pushes for corrections where necessary, and work with government departments and research institutions to improve the quality and communication of information at source, including by developing and making available factchecking toolkits to give people the tools they need to make up their own minds.

SNOPES

The Snopes.com web site, established in 1994 to research urban legends, is today the oldest and most renewed fact-checking website in the world, widely regarded by journalists, folklorists, and laypersons alike to debunk false information and relying on a team of 16 people who are precluded from donating to, or participating in, political campaigns, political party activities, or political advocacy organizations.

4.6 MIXED APPROACHES

CHECK4SPAM

The company (<https://check4spam.com/>) has been founded by engineers working at IBM, is based on human fact-checking and analysis of web traffic. "Some of the categories found on their website include promotions, internet rumors, and jobs among others. Interestingly, the founders have gone on record to say that most of the material they verify is political in nature." [Bishwas, 2018]

COMPASS

The idea of assigning one score to the news sources is at the basis of the search engine Compass (compass.catchy.ai). Users themselves vote for the plausibility of what they found on the Internet, the algorithm ranks both users and sources creating an archive of information about the score that both received from the "wisdom of crowds". Such score is used to determine plausibility of future news and plausibility of the score assigned by determined users. Framing this information should help people to find orientation whenever there is an overwhelming amount of information.

CREDBANK

This is an annotated dataset of tweets ranked according to their credibility. Data set accompanying the original paper [Mitra and Gilbert, 2015] are available from the web site <https://github.com/compsocial/CREDBANK-data>.

FACEBOOK FACT-CHECKING PROGRAMME

Facebook has been one of the favorite playgrounds for both Russian and Chinese propaganda efforts. Facebook is indeed vulnerable to the spread of disinformation and influence through both covert and overt strategies. Bots and trolls have created fake accounts spreading fake text, image, and video content through posts, pages, and paid advertising. At the same time, state-backed media companies often spread overt influence through legitimate accounts and pages that use paid advertising to disseminate their messages. As a reaction to the overwhelming evidence that disinformation was shared widely over Facebook during the 2016 US presidential campaign, Facebook announced the rollout of several features to help combat disinformation. First, the company made it easier to flag false news to platform administrators by allowing users to click on the news post in question and select from pre-set reporting options. Facebook also began enlisting the help of third-party fact-checkers to review reports from the community. After reviewing, these third-party fact-checkers provide a rating on the trustworthiness of an article. News deemed inaccurate automatically appears lower on a user's newsfeed and displayed with a "disputed" label to warn users who may read or share the story on Facebook.

Facebook additionally began testing changes to its newsfeed rankings, weighing its algorithm against articles that present disinformation warning signs.

Concurrently, Facebook began reducing the financial incentives for false news. That included preventing fake news sites from “spoofing” the domains of real ones. A malicious actor can no longer imitate a legitimate news site’s domain in an attempt to deceive readers.

Facebook’s actions were not limited to the immediate wake of the 2016 race. In 2018, they took additional steps to further restrict disinformation. Following Facebook’s debut of the “disputed” label a year prior, the company found that many of the users who read the disputed information and warning label actually became more inclined to share the news. In response, Facebook changed its approach and began shrinking the link in the newsfeed and including a number of related articles that debunked the news in question. The platform also increased its advertising controls, announcing that all political and issue-specific ads and advertisers would be vetted. Advertisers now need to have their identity and location confirmed by Facebook before receiving authorization to run ads on their platform. Furthermore, all political and issue-based advertisements are clearly labeled with “political ad” in the top corner, along with information about who funded the ad.

By November 2018, Facebook had also dramatically scaled up its fact-checking operation. It had expanded its network of third-party fact-checking providers to 23 countries. The company also began to implement machine learning tools to help these fact-checkers. The tools can spot warning signs in articles and help their human counterparts prioritize their efforts.

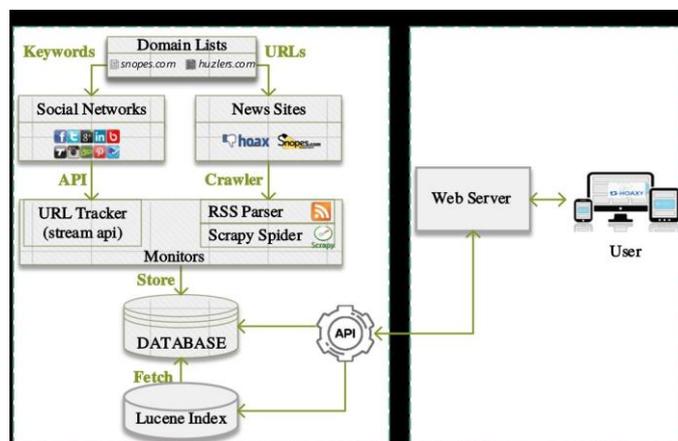
An independent study [Allcott et al 2018] confirmed that, while Facebook made indeed an effort reducing by over 50% the number of its users' interactions with fake-news websites between 2016 and July 2018, in July 2018 Facebook engagements with fake news sites still averaged roughly 70 million per month.

FACTMATA

This startup (<https://factmata.com/>) began as a fact-checking site, but AI methods are currently being developed. At the moment, the aim of the platform is (from their web site) “building a system to provide more context on information and perspectives on news, so that ultimately we can provide the public with a better informed opinion on the world. Our platform will assist journalists, fact checkers and anyone in the public to detect, verify and fact check media information in close to real time”.

HOAXY

This is an academic-based platform (<https://hoaxy.iuni.iu.edu/>) interested in the scientific problem of the recognition of fake news. A careful description of this platform can be found in [Shao et al., 2018], but the basic functioning of it is to analyse and categorize the news present on the WWW by assigning a flag of credibility to the sources thanks to the activity of third-party journalistic and fact-checking organizations. A list of credible/not credible organization is also listed in public repositories [Shao et al., 2017].



HOAXY SYSTEM ARCHITECTURE FROM [SHAO ET AL., 2018].

4.7 Conclusions

Just as with any systemic change, the revolution that has affected communication technologies in the last decade has highlighted the limits of the rules governing the exchange of information, which are now inadequate because they have been designed for an outdated system. The result of the delay in adopting a new model. Any delay in adapting the legislative framework to the new model constitutes a danger for democracy. Computational propaganda and micro-targeting, the rise of deep-fakes and the weaponization of the debate around ways to tackle online disinformation are all emerging threats.

Whilst there are a number of different solutions being developed and tested, so far, there is no silver bullet. When it comes to measuring the effectiveness of countermeasures such as fact-checking, most scholars seem to agree on their ineffectiveness: both [Chavalarias, 2018 and Guess et al; 2018] show that debunked news very rarely reach fake-news consumers. [Lazer and al, 2018] highlighted the lack of evidence on the efficacy of specific education and training programmes on new media, and the debate around if and how to regulate tech-platforms seems also scarcely supported by scientific evidence [Lazer and al, 2018].

Technological approaches – which are “a posteriori” fixes - will hardly be sufficient to tackle online disinformation and policies and incentives (which work on prevention, and therefore “a priori”) should also be put in place. It will be important to develop a wholistic and systemic approach to tackling the spread of disinformation on social media and develop a healthier media ecosystem for the 21st century.

PART 2 – ORIGINAL RESEARCH

Chapter 5 – THE SMART PILOT OBSERVATORY

5.1 Introduction

Based on the conclusions of Part 1 of the Report, a Pilot Observatory was built to test some of the claims and research models studied in this section. In particular, the consortium wanted to combine quantitative and qualitative analysis to see which tools and systems could work to create a more comprehensive understanding of the spread of disinformation on social media, the impact on society and identity.

In the light of the timing of the research period, it was decided that the study should focus on European narratives around the European elections. Narratives thus have the capacity to govern individuals, groups and nations and pave the path for the development of society. Narratives evolve as societies and conditions change.

All evidence from recent years points to the fact that we are currently living through a period of change that will affect the narratives of the future and the shared understanding of reality that will shape the current decades based on two key characteristics:

1. The changing media environment
2. The changing conditions of society

European Narratives

Narratives have the power to define era, bring together people and ensure the accomplishment of complex and arduous goals. The European project has been defined as the most remarkable political project in history and narratives have played a significant role in making this possible. Maria João Rodrigues set out to give an overview of why narratives are so important and how they provide meaning and a sense of direction. Narratives are both cognitive and emotional. Based on her experience with the construction of the EU she gave an overview of the seven key stages of European narratives:

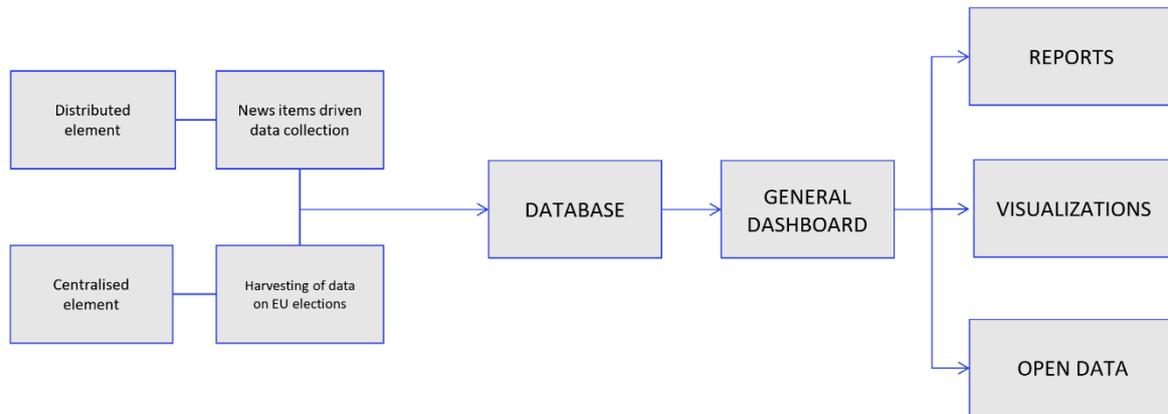
- Peace: The original narrative of the European project was the “never again” argument following the devastation and horror that the two world wars of the first half of the twentieth century had brought upon the continent.
- EU vs USA and Japan: In the 1980's the narrative shifted to one based on the competitive challenge versus the United States and Japan
- Internal Cohesion and Solidarity (Delors): With the fall of the Berlin wall and the ramifications of this the focus shifted to internal cohesion and solidarity
- Globalisation/Digital: At the dawn of the new century, the narrative shifted again to addressing globalisation and the digital transformation
- China/ BRICS: With the arrival of new super-sized countries (by 2050 the three most populated countries in the world will be in South-East Asia: China, India and Indonesia) challenging the world view and the role of Europe in the world the European narrative shifted towards one of Europe's role in this new world. Harmonious society and adaptation of narratives to other cultures became topics of discussion.
- Crisis: Then came the crisis and with this austerity and increasing focus on sovereignty. In this milieu the purpose of Europe in protecting citizens was challenged. There was a wide-spread sense of loss of control.
- Today: What will the post-crisis narrative be? How do we regain control? What are the biggest challenges that will define the coming ten years? Climate change? Who is in control of the interest and the information ecosystem? Trade agreements? Migration?

5.2 The SMART Pilot Observatory

Based on the State of the Art and on the SMART Consortium partners' technical know-how and tools, a platform was developed allowing to harvest, analyse and visualize contents from the main social networks (Twitter, Facebook, Instagram) and using a variety of techniques such as natural language processing, sentiment and emotional analysis and network analysis. The platform, which also features a distributed component in the form of a plug-in allowing users to evaluate news for their credibility, is conceived as a tool to support researchers, journalists, policy makers and citizens in understanding how information and disinformation on topics of public general interest

spread over the internet. Given time and budget constraints, the platform was set-up in July 2018 to collect and analyse data relating to European elections and, more in general, European narratives. A second focus on Italian elections was added in early 2019.

The platform architecture is as follows:



The centralized element consists in a data collection, analysis and visualization system. The data collection process was initiated in July 2018, and currently consists in 3 data capture processes: EU Elections, EU Narratives, Migration and the Italian Elections. For EU Elections, Twitter and Instagram posts were captured featuring:

- The hashtags #Euelections2019, #Euelections, #EuropeanElections
- The keywords (using boolean operators) "Europe AND Elections", "Future AND Europe"

For EU Narratives Twitter and Instagram posts featuring the following keywords that indicate temporal perspective (using boolean operators) were captured:

- "stagnation AND Europe"
- "stabilization AND Europe"
- "repetition AND Europe"
- "fall AND Europe"
- "decline AND Europe"
- "progress AND Europe"
- "getting better" AND Europe"
- "success path" AND Europe"
- "course AND Europe"
- "trial AND Europe"
- "redemption AND Europe"
- "penitence AND Europe"
- "learning from mistakes" AND Europe"
- "rising from the dead" AND Europe"
- "resurrection AND Europe"
- "healing AND Europe"

- "overcoming AND Europe"
- "obstacle AND Europe"
- "crisis AND Europe"
- "challenge AND Europe"
- "hubris AND Europe"
- "falling out of grace" AND Europe"
- "wasting chances" AND Europe"
- "fail AND Europe"
- "disappoint AND Europe"

A specific case of study related to Italy made a further investigation on a topic likely to trigger the decisions of voters in the country, that is to say the reactions of people to the news on migrants crossing the Mediterranean. The list of hashtags analysed is in the table below:

Hashtag Translation (meaning) in English

immigrati (Immigrants)

migranti (Migrants)

ong (Ngo)

scafisti (boat drivers/human smugglers)

seawatch (a ngo operating in the Mediterranean Sea)

barconi (barges/boats)

clandestini (illegal immigrants)

Guardia Costiera Libica (Lybian coast guard)

naufragio (Shipwreck)

sbarco (Disembarkation)

The data coming from the Observatory is completely anonymised and/or aggregated, so that it is possible for other researchers to safely visualize, download and use it from the project's website.

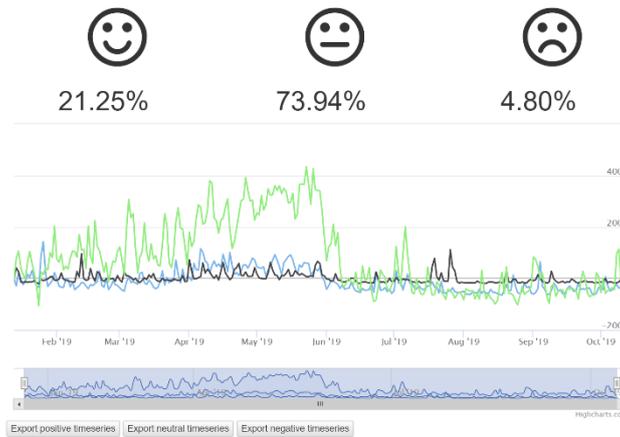
The data includes social media posts and web articles dealing with the afore-mentioned topics, as well as the social interactions which they triggered (shares, appreciations, comments, mentions) and thus define a dynamic network of relations which evolve through time.

In September 2019 about 1 million expressions had been collected, from 383,129 users.

Expressions from selected websites have also been collected by Sotrender to prepare a special focus on narratives (see chapter 8), and the focus on Poland was added as a result of a first round of analysis of the data performed in early 2019. Concerning languages, the data capture system focuses on a selection of fundamental languages (English, French, German, Italian, Spanish) on which a full analysis is provided using Natural Language Processing. At the same time, a "capture all" approach only uses Discourse Analysis and Statistical/Structural analysis, but applied to 54 languages, to intercept more emergent phenomena, for example detecting shadow narratives.

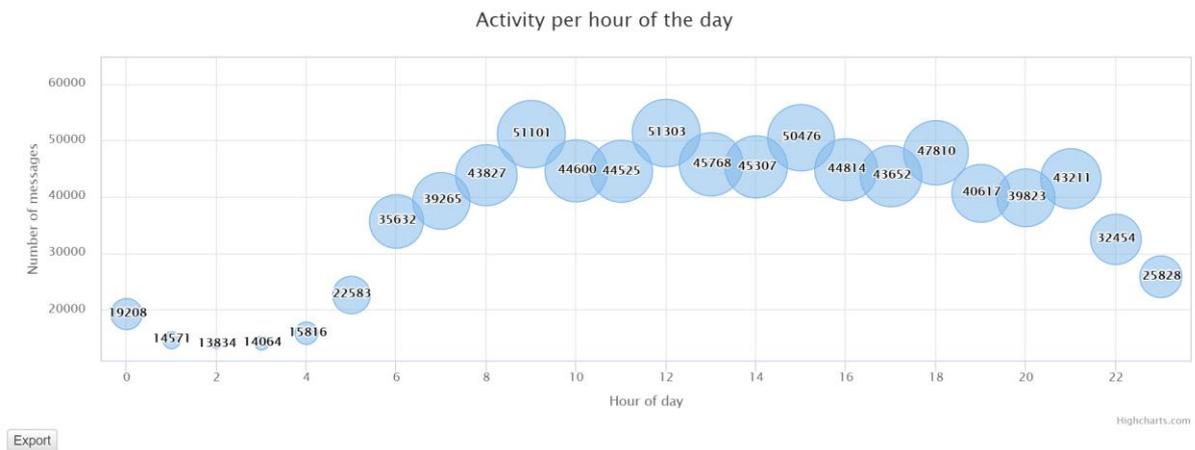
Users can visualize data according to different parameters and look at the totality of data or at monthly, weekly and daily samples. A simple dashboard allows to have an overview of what's going on in terms of sentiment, emotions, activity per hour of the day, topics, topics relations, top-users and users relations.

The collected data is distributed across time according to the sentiment relating to it as for following image:

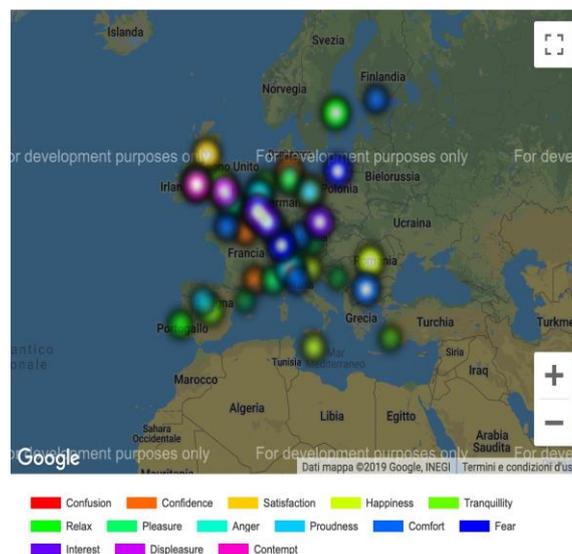


In the image total contents collected are showed in green, blue shows positive sentiment and black shows negative sentiment, allowing us to identify peaks in these dimensions. The drop of attention after the EU elections is clearly showed.

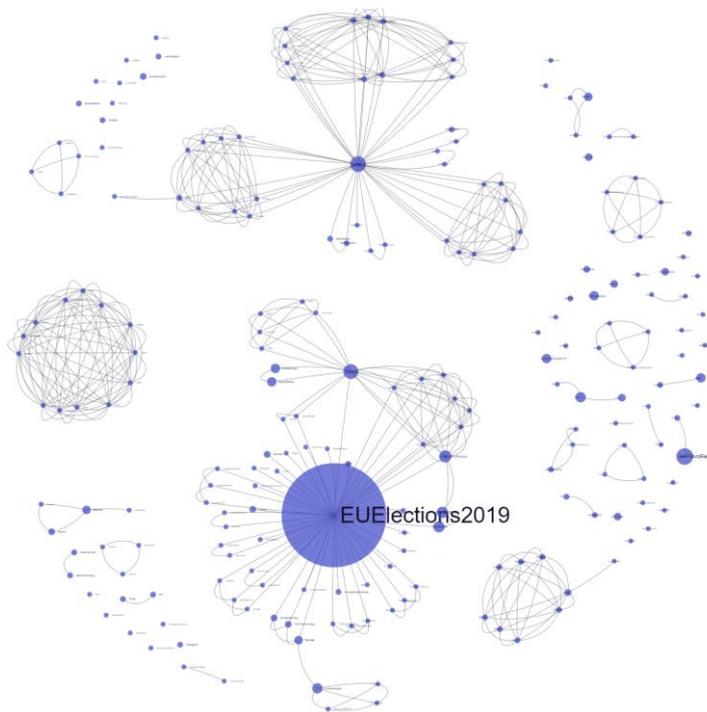
The following image shows when contents are posted during the day:



It is also possible to see where contents are produced, and the emotional picks linked to them:



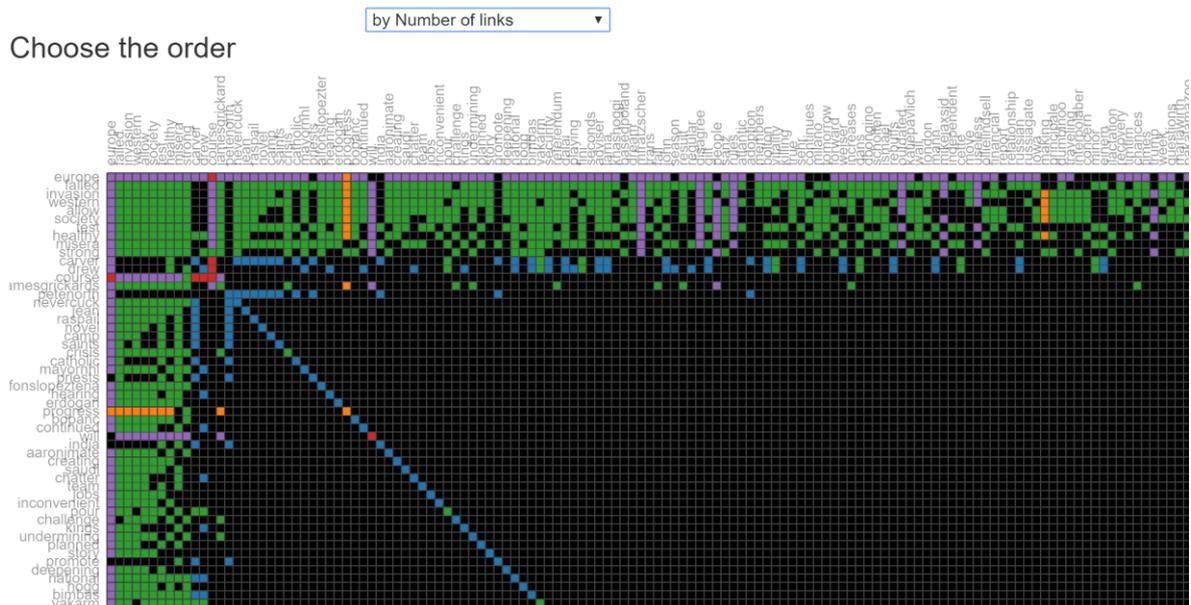
While the relational networks among users based on topic treated is showcased below:



It shows that if people talk about Topic A, they also talk about Topic B, and it is useful to understand what topics are associated, so that you know, for example, that Election is associated to Europe, but also to Migration.

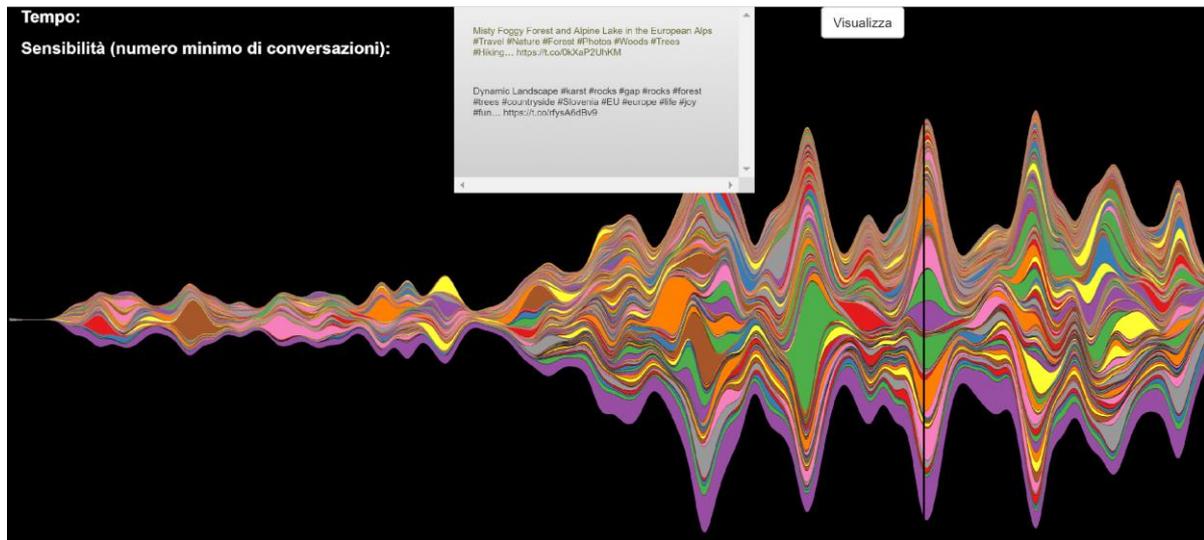
More sophisticated analysis and visualizations can be observed, including different kinds of correlations (by words, by number of links and by emotion) or communities developing around key topics. Word searches can also be performed to retrieve posts about a certain topic.

Co-Relations

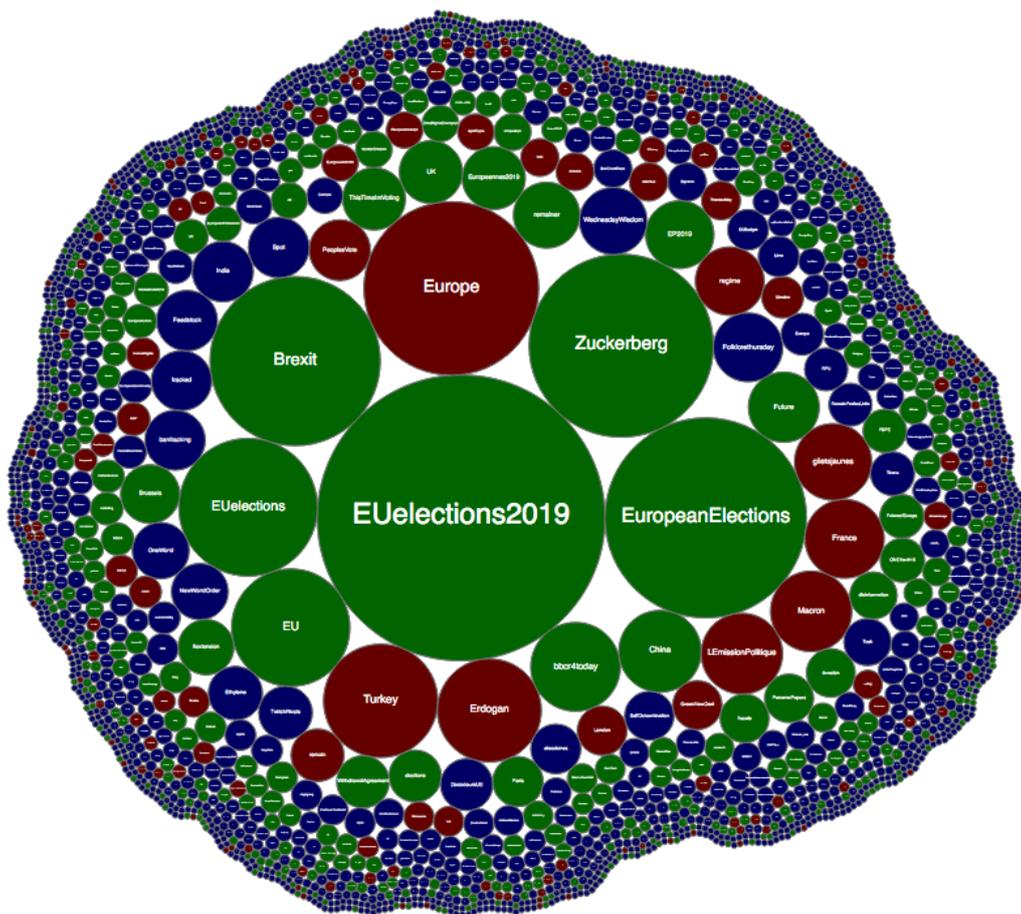


Of course, all the data is available in aggregated form in Json format so to allow more new analysis and visualizations which combine data and analysis techniques in different ways. So for instance, we have created a visualization allowing to follow different topics on the timeline, observing when they appear or leave the public discussion, how dominant they are, and to what kind of posts they refer to. This type of diagram allows exploring when and how topics enter the scene, allowing us to gain better understandings about temporal relations. When users hover over

one of the colours (each of which corresponds to a topic), the other ones fade out, allowing them to understand when the topic entered and left the scene of public communication:



The topic navigation tool is an instrument which allows to follow the evolution of the topics of discussions, as it varies across time. A simple "From date / To date" selection dialog allows selecting date ranges of interest, to understand and visualize what were the most discussed topics:



In the visualization, colours are associated to sentiment, allowing to have a vision of what the scenario is.

All these visualizations have their "export data" button, which allows to download the data being visualized, to enable people (for example researchers and journalists) to use it in their work.

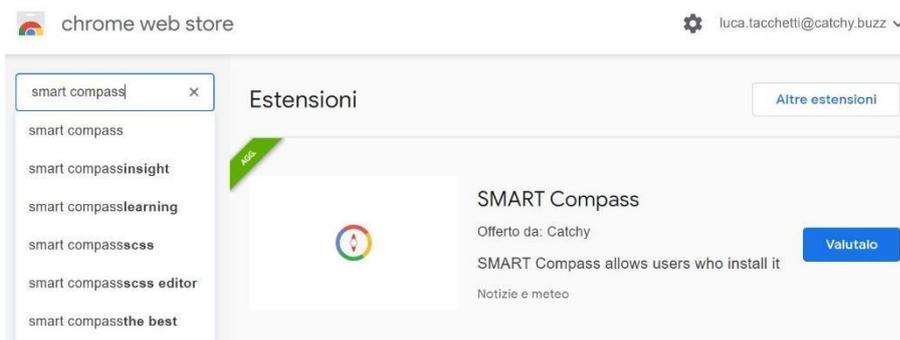
For example, we have been systematically using the data export feature to:

- Create date and topic-based archives of data, useful for studying the evolution across time and topics.
- Create test data extractions, to experiment techniques studied in the state of the art (for example to recognize disinformation phenomena, bots, or to measure the impacts of certain narratives etc, as per the state of the art documentation of the SMART project): results are described in chapter 8.
- "Freeze" the data which regards interesting phenomena, to be able to reference it in the resource.

On top of this, the availability of the Observatory has allowed us to very easily create specific focuses on important phenomena as they arose: for example, on April the 4th we turned on a focus on Poland.

As the datasets grow, we were able to perform original research observing narrative communities around Europe and the upcoming elections, and to better understand how both information and disinformation spread on social media and on a set of relevant websites. In this regard, the University of Warsaw with the support of SoTrender, collected almost 200,000 articles from European and British news outlets, with a view of understanding which main topics selected sources covered and how the news was presented. In other words, the goal was to discover whether the news follows a positive or the negative narrative. Topic analysis was used to identify 20 main topics, which were further broken down into 10 more specific topics characterizing the debate around Europe. A generative statistical model based on the distribution of words probability in each document was used: texts with similar distribution fell into the same topic. Afterward, SoTrender computed the sentiment of each document, and its evolution over time (see chapter 8 for further information).

Concerning the distributed element of the platform, and with the goal of enriching the contents of the Compass platform and allowing experts and journalists to focus on specific news so as to start the fact-checking process, in October 2019 Catchy created a plugin for Compass, which can be downloaded from [Google's chrome web store](#).

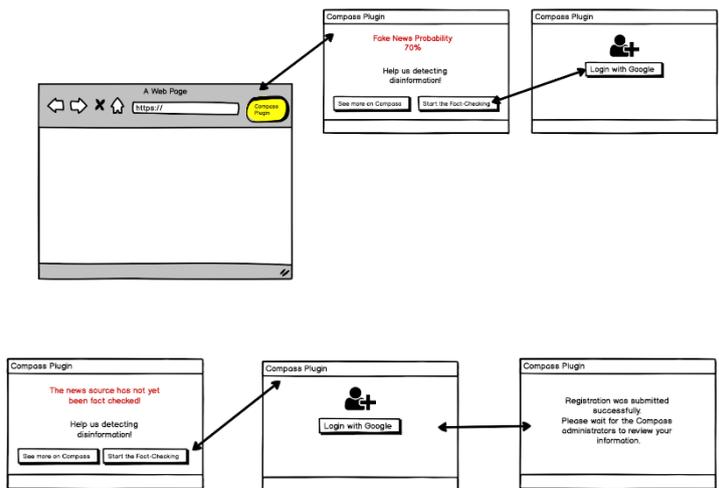
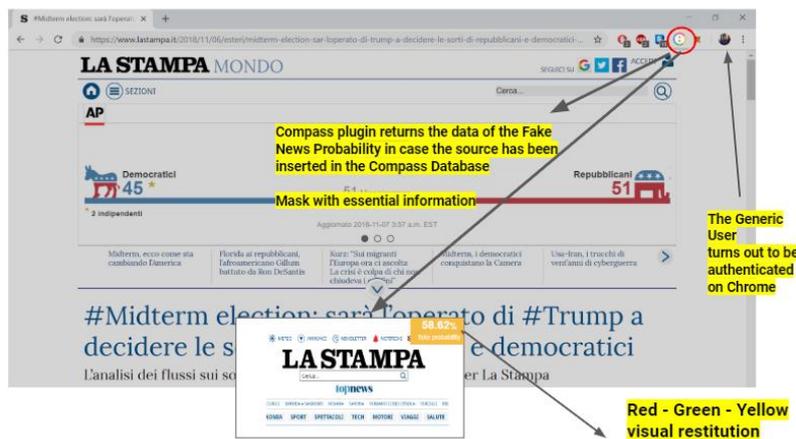


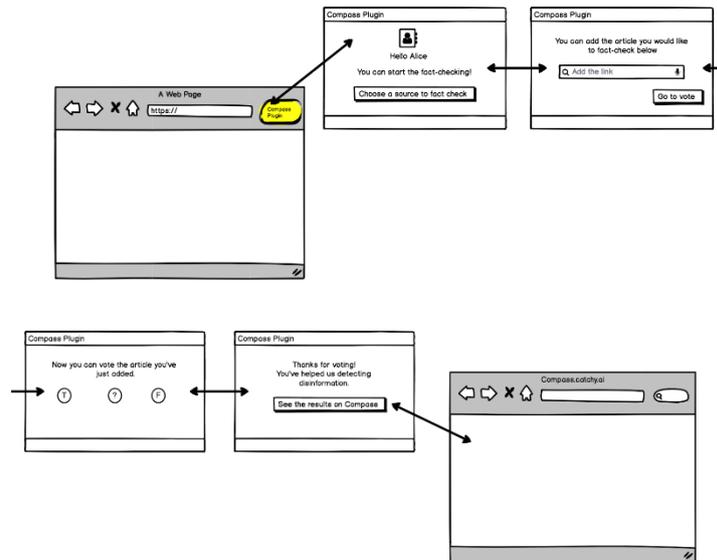
It is an extension of Google Chrome that allows users who install it to choose any article they are reading and that they do not consider reliable, to insert it as a source of news in the compass platform, to vote for it and to allow other experts to express their own preference in terms of reliability. The structure of the plugin is the result of an analysis of user experience design and information architecture. The information architecture was modelled building mock-ups and wireframes to describe the distribution of the contents. An agile approach was used, firstly writing and modelling user stories, or the main features of the plugin. Listed below are the user stories considered as macro features of the plugin.

Cod	User Stories
01	I as a Generic User can download the "Compass" plugin on Chrome.
02	I as a Generic User (who has already installed the plugin) can browse an information site and click on the plugin to know the fake news probability of the source I am browsing if it is included in the Compass db.
03	I as a Generic User (who has already installed the plugin) can decide to authenticate myself on the Compass plugin as a Researcher
04	I as a Generic User (who has signed in) have to wait for the authorization by the Compass administrators to become a Researcher

05	I as Researcher (logged in) can decide to insert the source of the article I am reading in the db compass to start fact checking
06	I as a Researcher (logged in) can vote for the reliability of the news source

At this point we have moved on to the design of low and high level wireframes.

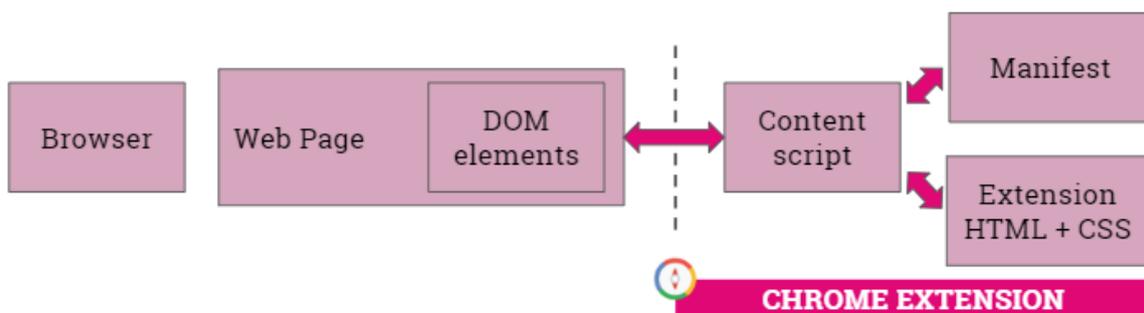




Finally, the Chrome extension Compass was developed. The Chrome extension is written in Javascript, with a UI defined in HTML and CSS. The basic properties of the extension, the configurations and the way it interacts with the pages served by the browser are managed within a file called manifest.json.

The extension deals with:

- recover the id of the user logged in with a Google account, performing the Oauth with the provider;
- extract the source of the page you are viewing;
- interface through Ajax calls to the Compass API to recover the source's authenticity;
- send user votes again via Ajax.

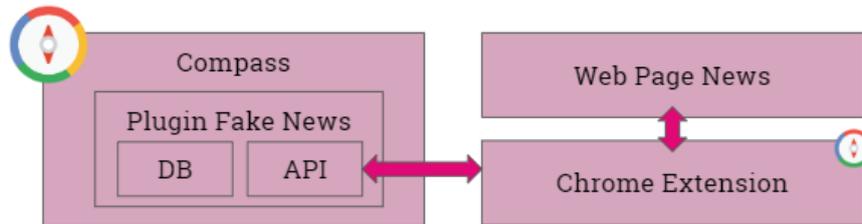


On the Compass side, the Wordpress plugin has been extended to manage the evaluation of the sources, adding APIs that can be contacted by the Chrome extension.

The APIs receive the following parameters as input and log in to the plugin tables in the Wordpress database:

- url of the page containing the news
- User ID
- Source
- Channel
- date of insertion of the news on the system

On the platform side, the Chrome extension is managed as an additional channel that joins social media and RSS.



The extension operation flow is summarized in the following steps:

- On the news page, the user selects the Compass extension (hereinafter CE)
- If the user is already logged in to a Google service, CE will retrieve the Google user ID token from the browser and show him his name and avatar
- If the user is not logged in, CE proposes to log in
- CE obtains the source from the url and sends a request to Compass
- If the source already exists on the Compass database, the platform responds with the percentage of truthfulness of the source, which is shown on CE
- If the source does not exist, CE reports it
- Selecting "start the fact checking", CE shows the user the URL of the news and a voting mask. Selecting one of the values and then pressing the "send vote" button, CE sends the data to Compass via API. Compass saves the data on the db and the source is now available for consultation.

Currently, the Platform does not include data from the Observatory; while this is not strictly necessary, in the concept we described the fact that, as the news items are placed under observation (for example fact checking using Catchy's Compass platform), capturing processes could start on the observatory to assemble the data which shows how the suspicious news is linked/shared/commented, to what relational ecosystems these actions correspond, and similar actions which would provide context to the analysis; the Human Ecosystems platform which is at the foundation of the Observatory can host such data capture processes and, thus, it is possible to have the plugin trigger them through API calls, and then to visualize the results.

The current overall platform constitutes a working prototype that was evolved across several different design/development iterations, to be able to explore the widest possible array of existing and innovative solutions for providing researchers and policy makers with standard sets of tools that can be used to confront with disinformation, so that different researches can be coherently and consistently compared and replicated. Given the objectives, the project team chose to dedicate additional resources to test and assemble a wide variety of tools coming from the most promising research outputs – mentioned in the state of the art –, rather than to spend time and resources on costly integrations between HER's and Catchy's platforms, which are used as a baseline. For example, in its current state, if some disinformation phenomenon is unveiled using either HER's or Catchy's platforms, or through the variety of tools installed and configured on the project's servers, it is still possible to activate the other base platforms and tools, with the only current limitation that it has to be done manually (for example by turning on the monitoring for a certain URL on either Catchy's or HER's base platforms, with a few clicks). Therefore, the current solution constitutes a working prototype whose functionalities can be further automated when transforming it into a market-ready product.

In the next session, we will demonstrate how we used the Observatory to perform some original research on how disinformation spreads online.

Chapter 6 – COUNTERACTING THE SPREAD OF DISINFORMATION PLATFORM TESTS

6.1 Introduction

Our objective with SMART was to create a toolkit supporting researchers and practitioners to easily monitor how information spreads on the web and social networks, allowing to add further instruments, so that the toolkit can constantly be updated, maintained and used.

To test the toolkit, we performed a series of tests according to two main modalities:

- Tests based on platforms, using their interfaces
- Tests based on computation, preparing some datasets to be processed by an algorithm-driven technique, to try to understand a determined communication phenomenon

The Human Ecosystems software was used as our main infrastructure to collect public content from the web and from social media sites, and to perform basic initial analysis of this data, which in turn allows to make it easily available for further analysis and visualization.

More in detail, for the purposes of the experiment described below, we used Human Ecosystems to collect data from:

- public Twitter profiles of around 200 news outlets, public figures (like Donald Trump, for example) and the interactions that take place around these (for example comments and shares)
- the web presence of a total of 100 European news outlets in English, Italian, French, German, Polish, Greek, Spanish, Swedish and Dutch

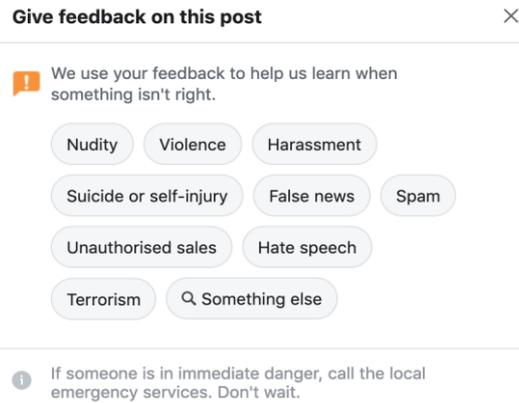
6.2 Platform Tests

Based on (Gangware, C. & Nemr, 2019) we ran a series of platform experiments to check how Facebook is performing with tackling the spread of online disinformation. The platform interactions described in the paper can be seen in the images below. Below each image the tests that have been performed and their results are described.

According to the authors:

“first, the company (Facebook) made it easier to flag false news to platform administrators by allowing users to click on the news post in question and select from pre-set reporting options”.





Test performed:

- Take a problematic news item on Facebook, report it, wait for feedback (eg: notification, news taken down, other)
- Perform 10 tests on 10 different problematic news items from 10 different user accounts

Result:

- No feedback received
- The news items remain unmodified and public
- Facebook establishes a trustworthiness score from 0 to 1 for users who report¹

“Concurrently, Facebook began reducing the financial incentives for false news. That included preventing fake news sites from “spoofing” the domains of real ones. A malicious actor can no longer imitate a legitimate news site’s domain in an attempt to deceive readers.”



Test performed:

¹ See <https://www.washingtonpost.com/technology/2018/08/21/facebook-is-rating-trustworthiness-its-users-scale-zero-one/>

- Search for FB pages and groups of 30 well-known fake-news operators which use spoofed names of well-known news operators (such as in the picture above, where “Il Fatto Quotidiano” becomes “Il Fatto Quotidaino” and “Il Giornale” becomes “Il Giomale”).
- Measure the impacts of their posts.

Results:

- 24 out of the 30 selected Fake News pages still have FB pages and groups.
- The most recent 20 posts of the 24 remaining operators (that is 20x24=480 posts) have had a total of 122,513 likes, 13,769 comments and 68,522 shares.

“The platform also increased its advertising controls, announcing that all political and issue-specific ads and advertisers would be vetted. Advertisers now need to have their identity and location confirmed by Facebook before receiving authorization to run ads on their platform. Furthermore, all political and issue-based advertisements are clearly labeled with “political ad” in the top corner, along with information about who funded the ad.”

The screenshot displays the Facebook Ad Library interface for the page 'Partito Democratico'. At the top, the page's profile information is shown, including its name, handle (@partitodemocratico), and 302,779 likes. Two 'Page transparency' boxes are overlaid on the page image. The first box indicates the page was created on 27 Apr 2009 and its name has been changed once. The second box shows that the page has spent a total of €133,742 on ads related to social issues, elections, or politics between March 2019 and June 2019. Below the page information, a list of ads is displayed, all of which are marked as 'Inactive' and are categorized as 'About social issues, elections or politics'. The ads were launched in May 2019.

Test performed:

- Select 30 pages of political organizations (eg: parties)
- Browse their ad history
- Ask a sample of 50 people from the same country of the party 1) whether they have seen the ad or not, 2) if the label “political ad” was shown, 3) if they saw the ad high up or down low in their Facebook timelines

Results:

- Out of the 50 people surveyed in each country, an average of 8 for each country remembered seeing the ads, and confirmed that they displayed the label (highest 12 in France; lowest 1 in Italy)
- Of the remaining people, an average of 12 for each country declared that they saw at least 1 ad from a political organization which was not labeled as such (although they did not remember which page was publishing it, and seen that Facebook does not allow to visualize past ads, it was not possible to verify this information) (highest 22 in Germany; lowest 6 in UK).

6.3 Computational Tests

This is the second type of tests we performed using the Observatory to gather data on EU Elections from the web and social networks, before adapting it to the format required by the authors whose tool was being tested. As previously described, data included

- posts and interactions (for example comments and shares), across public Twitter profiles of around 200 news outlets and public figures (like Donald Trump, for example).
- the web presence of a total of 100 European news outlets in English, Italian, French, German, Polish, Greek, Spanish, Swedish and Dutch

From these, we aggregated those contents which were dealing with EU Elections, as identified by hashtags like #EUElections2019, #EuropeanElections, and other declinations of these principal ones, in all the above-mentioned languages (for example: #EUElecciones2019 for Spanish).

The resulting data comes under the form of a text with metadata including:

- **The text:** the textual content of the article/post/comment/..., in which all references to subjects which are not the initial public subjects of the collection are replaced with generated, anonymous, unique IDs
- **The source:** the emitter of the content (article/post/comment/...):
 - if it is a public subject (for example profiles on social media of public personalities or organizations; or the indication of a news outlet in the case of a web article), the source is quoted as for the actual profile if it is not, it is an anonymized and assigned an auto-generated unique ID
 - **The date and time** when the contents was published
- **The connections with the reactions** (retweets, shares, comments)

The base data can be downloaded in anonymized, aggregated form. It consists of a total of around 1 million messages, articles and reactions, from the sources detailed above. The reactions to the conversations (likes, shares, comments) define 2,553,289 relations. The computational tests below have been performed onto this relational graph and its metadata.

6.3.1 THE SPREAD OF LOW-CREDIBILITY CONTENT BY SOCIAL BOTS

To better understand the role of bots in spreading news online we tested the assumptions of (Shao & Al., 2018).

Talking about the pattern of activity of the bots that contribute to spreading low-credibility content, the authors show how *"likely bots are more prevalent in the first few seconds after an article is first published on Twitter than at later times. We conjecture that this early intervention exposes many users to low-credibility articles, increasing the chances than an article goes "viral"."* Moreover, a *"strategy often used by bots is to mention influential users in tweets that link to low-credibility content"*.

By adding up the implications of both observations, (Shao & Al, 2018), designed a system to identify bots according to how different social media user profiles adhere to these patterns and behaviors (i.e., by acting in the early seconds from the publication of low credibility content, and by mentioning influential users in the messages).

From a SMART perspective, the objective was also to connect the software used for the project to collect data and information from the web and social media sites, to the software used by the authors of the referenced article, to set-up a process which could be easily repeated in other occasions and with other dataset.

Tests performed:

- Installed Hoaxy backend
- Installed the Botometer API
- Integrated data from Human Ecosystem SMART instance (exported and formatted using Hoaxy data structure, then imported into the Hoaxy database)
- Configured Hoaxy to take into account the top 20 posts (the posts which obtained higher numbers of likes, shares and comments) from each account
- Ran bot detection on the accounts that shared tweets about the 30 most frequent topics in our data captures

Results:

The results are shown on the following table:

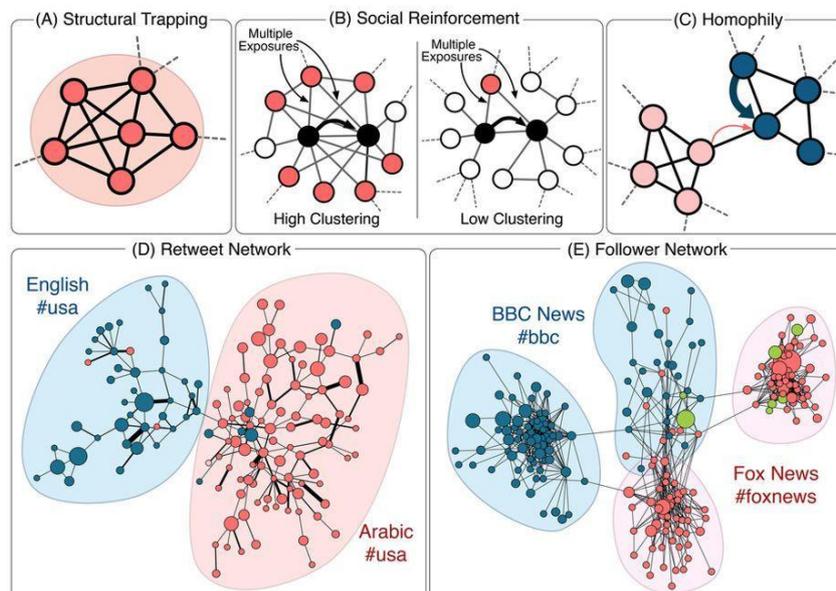
Topic	Number of identified bots (n>4)
EUelections2019	17
Brexit	11
Europe	37
disinformation	31
Bridgend	1
Ford	22
EU	2
DeniedMyVote	26
future	14

The table refers to Shao & Al's 2018 paper technique for identifying BOTs using the prescribed technical infrastructure and configured with a credible level of reliability ("n>4" as indicated in the referenced paper). Even this simple extraction on the most common topics shows how BOTs have colonised a wide array of types of online conversation, ranging from specific discussion on Elections, to ones on the "future", to emergent activist actions and more.

As with all the following tests, this also shows the practicality and usefulness of having a streamlined toolkit which can be used to monitor online phenomena: once the phenomena is identified (in this case: EU Elections 2019) it is easy to run the data capture and, when done, run the tests in sequence to browse and compare results.

6.3.2 VIRALITY PREDICTION AND COMMUNITY STRUCTURE IN SOCIAL NETWORKS

The second test was run based on (Weng & al., 2013) to detect communities' structures and promptly identify the conditions in which disinformation spread on Twitter. More specifically, we aimed at detecting the following characteristics described by (Weng & al., 2013):



(A) **Structural trapping**: dense communities with few outgoing links naturally trap information flow.

(B) **Social reinforcement**: people who have adopted a meme (black nodes) trigger multiple exposures to others (red nodes). In the presence of high clustering, any additional adoption is likely to produce more multiple exposures than in the case of low clustering, inducing cascades of additional adoptions.

(C) **Homophily**: people in the same community (same color nodes) are more likely to be similar and to adopt the same ideas.

(D) **Retweet Network**: diffusion structure based on retweets among Twitter users sharing the hashtag #USA. Blue nodes represent English users and red nodes are Arabic users. Node size and link weight are proportional to retweet activity.

(E) **Follower Network**: Community structure among Twitter users sharing the hashtags #BBC and #FoxNews. Blue nodes represent #BBC users, red nodes are #FoxNews users, and users who have used both hashtags are green. Node size is proportional to usage (tweet) activity, links represent mutual following relations. (From Weng et al. 2013)

To uncover the network structure of our SMART sample of data we installed the Igraph package for Python (<https://igraph.org/python/>) and used the available algorithms for community detection, in particular

- [Optimal Modularity](#) [Good et al 2010]
- Edge Betweenness [Girvan et al 2002]
- Fast Greedy (<https://skerritt.blog/greedy-algorithms/>)
- Info Map [Rosvall et al 2008]
- Fast Greedy (2004)
- Walktrap (2005)
- Eigenvectors (2006)
- Spinglass (2006)
- Label Propagation (2007)
- Multi-level (2008)

Once the community structure is given in parallel with [Weng et al 2013] we considered four baseline models to validate the spreading observed:

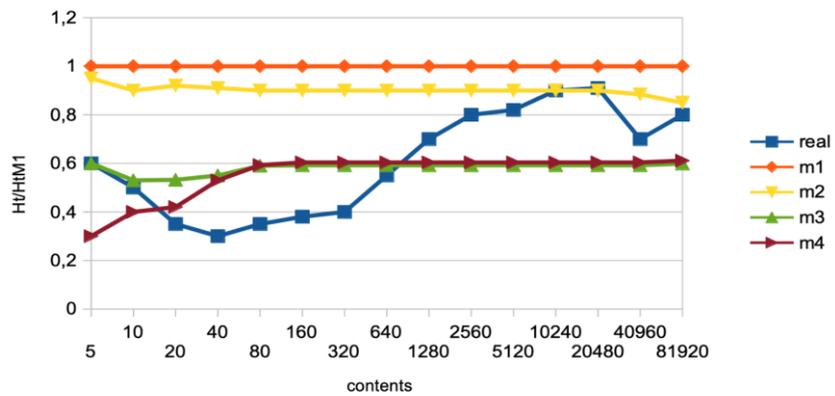
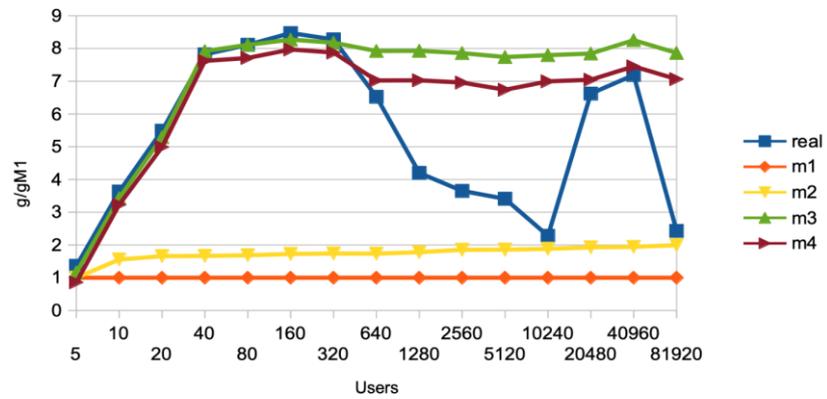
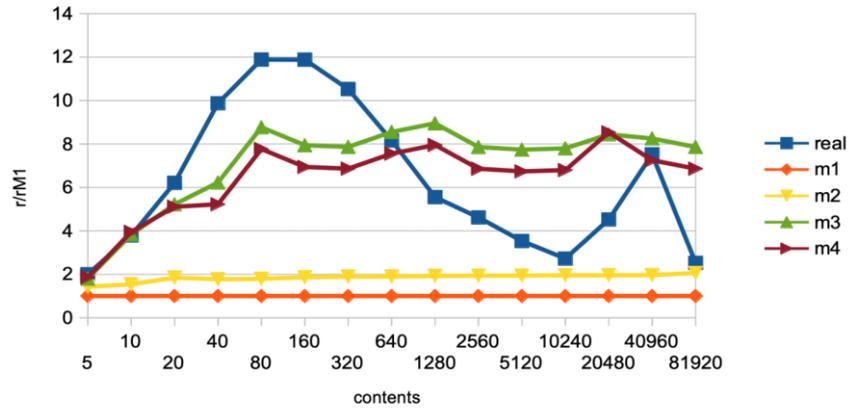
- **M1** *The random sampling model* assumes equal adoption probability for everyone, ignoring network topology and all activity.
- **M2** *The simple cascade model* simulates the spreading of simple contagions.
- **M3** *The social reinforcement model* employs a simple social reinforcement mechanism in addition to considering the network structure.
- **M4** *The homophily model* where users prefer to adopt the same ideas that are adopted by others in the same community

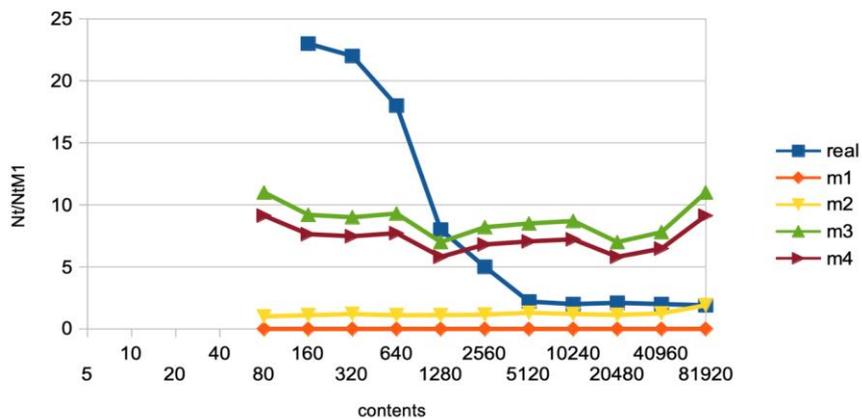
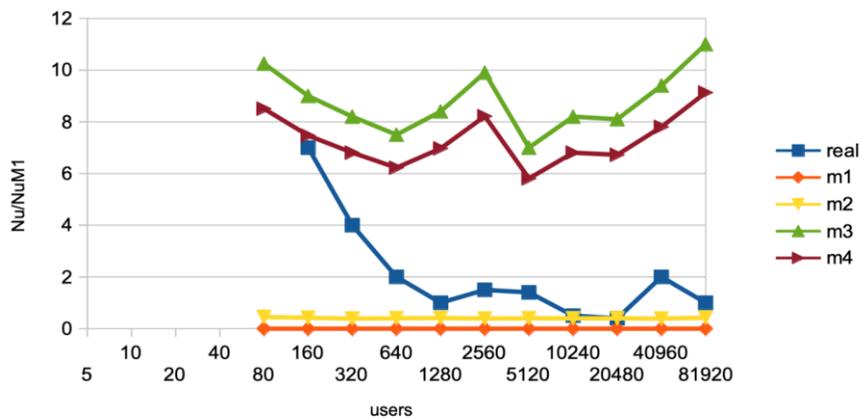
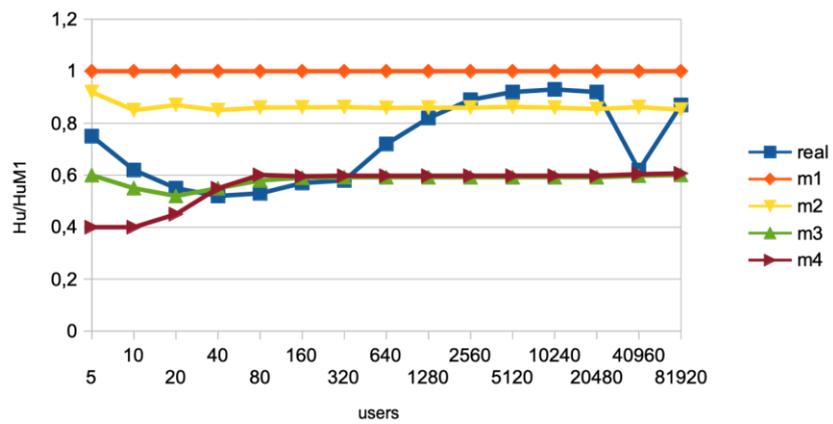
Accordingly we simulated propagation of content (contagion), using the 4 models (M1, M2, M3, M4) and compared it with actual propagation (eg: take 10 articles that are particularly viral and the sequence of the hops across the network/communities, and compare with what happens in the models). We focused on new memes emerging in the communication, defined as those with fewer than 20 tweets during the previous month. Finally, as per the method described in the reference paper (Weng, Menczer, Ahn, 2013) we calculated:

- a. the concentration of a meme h based on the proportions of contents in each community
- b. the usage-dominant community $c^t(h)$, as the community generating most content with h
- c. the usage dominance of h , $r(h)$, as the proportion of contents produced in the dominant community $c^t(h)$ out of the total number of contents $T(h)$ containing the meme
- d. the *usage entropy* $H^t(h)$ based on how tweets containing h are distributed across different communities

- e. the relative usage dominance and entropy, calculated using M_1 as baseline
- f. the *adoption dominance* $g(h)$ of h , as the proportion of the $U(h)$ adopters in the community with most adopters
- g. the *adoption entropy* $H^U(h)$, computed based on how adopters of h are allocated across communities

Results are as follows:





In all these diagrams, it is important noticing which lines have levels which fall below the other ones. This is the key to reading them, as it allows to understand how real-life news propagation of meme/viral news items go above or below certain specific referenced models.

For example, looking at the curves for the adoption dominance $g(h)$ (calculated as the proportion of the number of users who adopt meme h in the community with the most adopters of the meme): if the real life news propagation curve falls below both the social reinforcement (M3) and homophily (M4) curves, for a high number of contents, it means that content h spreads easily, and tends to be a simple contagion and,

thus, it is more probable that it will go viral (because it doesn't need social reinforcement or the homophily, it just spreads faster and more easily).

Or, for example, if we study the usage entropy H_t for content h , measured on how tweets containing h are distributed across different communities, if the real life curve for a high number of contents goes above curves M3 and M4, it means that content h overcomes community boundaries more easily than the theoretical curves defined through social reinforcement and homophily based models and, thus, it is more likely to go viral.

Even given the differences due to the different times, contexts, numbers and types of users, contents and networks involved, it is possible to see, by comparing these results with the ones in the referenced paper, that the architecture and shapes of the phenomena are the same: viral content requires as little reinforcement as the simple cascade model, while non-viral ones need at least as M3 and M4 (this can be seen by noting how the different curves fall above or below one another).

We thus achieved our objective: to create a streamlined toolkit which can be used to calculate these parameters, which can in turn be used to train machine learning processes enabling us to predict what is the probability of contents to go viral on the base of their time of propagation, the number and types of users involved, their layout across communities.

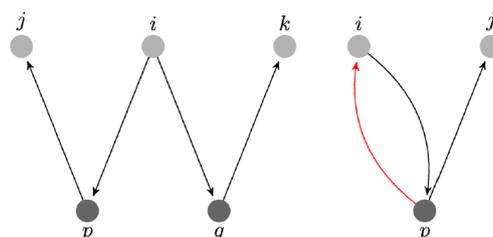
The toolkit can be used by any researcher: just like the consortium partners started a data capture on a communication phenomenon (EU Elections) and then were able to export the data in one command and execute a standard set of tests (the ones selected from the State of the Art) and browse/study the results, another researcher could do the same studying a different phenomenon. This would allow for a number of benefits to be widely available:

- ease of use and quickness: the data capture could be configured minutes after the phenomenon is identified, and after the data capture the tests could be run immediately
- standardised tests: a community of researchers could maintain the selection of tests (we now have chosen the most relevant ones, as emerged from the State of the Art), to have a standardised set of tests that would make the results credible and comparable
- transparency and accountability: the tests are selected from the most important and relevant publications in the research domain, and use software which is fully inspectable (open source), which means that they are supported by the entire research community and that they are completely transparent and accountable

This describes a model in which the international community of researchers in this area agree on the sets of tests to be performed, and these tests are made available on easy to use, standardised testing facilities that are completely shared, transparent and accountable, so that the results of these analyses can be shared, compared, and used as evidence.

6.3.3 ELECTIONS AND MIGRANTS IN ITALY

We rely on the existing literature, and in particular on the analysis of $\sim 10^6$ tweets exchanged during the last Italian elections held on March 4, 2018 [Becatti et al 2019]. First, we employed an undirected representation of the network of retweets by distinguishing certified from non-certified users. Then we identified groups of verified users by their interaction with the opposite layer, following the recipe of [Gualdi et al., 2016; Saracco et al., 2017]: if two verified users are retweeted more than expected by the non-verified ones, they are likely to be related.



Then, we analysed the community organization of the resulting network and measured the polarisation of unverified users according to this division: as observed in other studies. As in [Bovet and Makse, 2018], we simply projected the bipartite and directed network of tweets and retweets onto the users layer: a directed edge between i and j in the projected graph indicates that j has retweeted i 's posts at least one time.

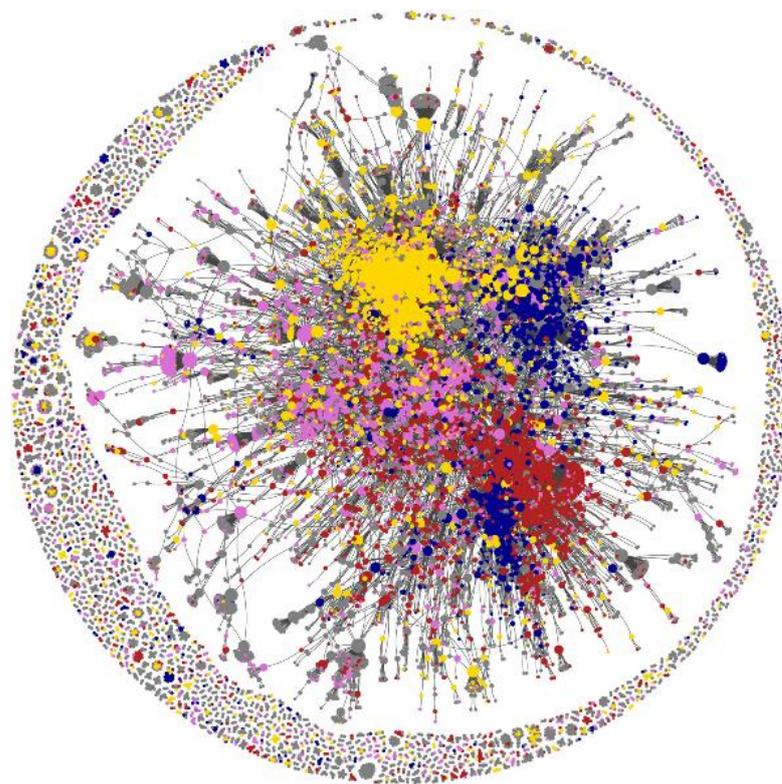
By means of the Python module tweepy we have used the Twitter Search API to download a sample of all tweets posted from January 28 to March 19, 2018.

We then used these data on Italian Elections to create a map of the Political parties alliances. For this reason, we expected this procedure to be mostly applied to those people considered of public interest, such as actors, politicians, newspapers, TV channels, radio channels etc.

In order to obtain groups of verified users based on their activity, we projected the information contained in this bipartite network on the verified users layer. The result of the classical projection methods is a weighted monopartite network: two users are connected if they share at least one common neighbour on the opposite layer and the edge between them is weighted by the number of non-verified users who have simultaneously retweeted their posts. Given the projected and validated network of retweets, we performed a reshuffled community detection procedure, i.e., the Louvain algorithm (Blondel et al., 2008) ran several times with a rearranged nodes' ordering. The partition with the highest modularity was then selected: by doing so we overcame the original algorithm's order dependence (Fortunato, 2010)

Given the division in political alliances, we started analysing the topological characteristics of the subgraphs made by each group. In order to facilitate the understanding, we represent each community with a colour reminiscent of the colour used during the elections to represent the coalitions.

As a first step, we have split the sample of Italian-speaking users into two groups, the verified and non-verified users. This variable is directly available from the API and characterises those users that have requested to be authenticated by the system.

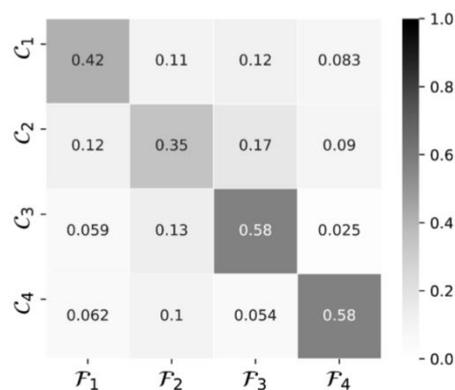


Nodes' colour identifies the user's community while nodes' dimension is proportional to their out-degree in the validated graph, i.e., the number of significant spreaders they have. Each plot focuses on the structure of the subgraphs of the directed network generated by each community. Nodes' dimension is again directly proportional to their out-degrees in the subgraph: the larger the node, the higher the number of times that user has been retweeted by the other accounts. Nodes' colour is instead related to whether the account has been verified or not: blue for verified users, orange for non-verified ones.

Therefore, the distinction between the two types is actually made by the users themselves, since in principle any account can request to be verified: when it does, Twitter guarantees that the account is authentic. After this distinction, we build the bipartite network of retweets between verified and non-verified users: an edge between two users indicates that one has retweeted the other's content at least once during the available time period. In order to obtain groups of verified users based on their activity, we project the information contained in this bipartite network on the verified users layer. The result of the classical projection methods is a weighted monopartite network: two users are connected if they share at least one common neighbour on the opposite layer and the edge between them is weighted by the number of non-verified users who have simultaneously retweeted their posts.

Given the projected and validated network of retweets, we have performed a reshuffled community detection procedure, i.e., the Louvain algorithm [Blondel et al., 2008] runs several times with a rearranged nodes' ordering and the partition with the highest modularity is selected: by doing so we overcome the original algorithm's order dependence [Fortunato, 2010]. As a result, we found four communities (yellow Movimento 5 Stelle, Purple mostly journalists, Red for Partito Democratico, Blue for the right-wing parties (Former Popolo della Libertà)) in the graph of users shown here in different colors.

The main result is that most of the retweets of users come from the same community as shown in the diagram on the right. The first community corresponds to M5S and shows a strongly connected block of



(mostly non-verified) nodes that retweet among themselves and with the verified accounts of the community, the most central of which are the Twitter accounts of the newspaper Il Fatto Quotidiano and its journalists Marco Travaglio, Peter Gomez and Antonio Padellaro. In the other communities, journalists and newspapers do not form such a strong core. It is interesting to note that the M5S political leader Luigi Di Maio does not belong to this central community but is located in a small community outside this large component. The second community has central nodes that are the verified accounts of newspapers (see for example La Repubblica or Il Corriere della Sera) and information channels (such as Sky TG24, Tg La7, Agenzia Ansa or Rainews). Some politicians such as Pietro Grasso or Giuseppe Civati are present in this group,

together with the political parties of Rifondazione and an extreme-left party (whose account is not verified). These politicians represent the most extreme, left-leaning orientation who have not encountered a commonality of interests and supporters with the accounts in the community of Partito Democratico and we indeed notice that they belong to different communities. In the third community we identify a central block of mostly non-verified users. The most retweeted figures are Matteo Renzi and the account of Partito Democratico, as shown by their high values of out-degrees. The remaining verified nodes are mostly well-known characters in the political scenario (such as Maria Elena Boschi and Carlo Calenda), as well as newspapers (see for example Il Foglio or IlSole24Ore). Among the non-verified users we have the accounts of the Partito Democratico political parties related to the areas of Milan and Rome. In the fourth community we found two quite separate clusters; one of them is centred on the accounts of people belonging to Lega Nord, such as Matteo Salvini, Claudio Borghi and the party Lega-Salvini Premier. On the other side there are the accounts of Forza Italia and Gruppo FI Camera and some of its exponents like Silvio Berlusconi or Renato Brunetta. The two verified nodes of Giorgia Meloni and Fratelli d'Italia (the political party she is leading) receive retweets from both sides, nevertheless being closer to the Lega pole. Another popular node is a neo-nazi party (which account is not verified), that has its own circle of retweeters and share some interactions with the subgroup of Lega Nord.

Users with the highest out-degree identify the most popular ones, whereas highest out-degree posts are the most "viral". We identify significant content spreaders with a procedure that allows to statistically validate the connections that cannot be explained by users' tweeting activity and posts' virality, using an entropy-based null model as benchmark.

The analysis of the directed network of validated retweets reveals signals of the alliances formed after the elections, highlighting commonalities of interests before the event of the national elections. In order to detect the source of possible fake news, we started by analysing the structure of the Twitter community and the presence of unverified users within them that can be the signature of bot presence.

We display the political coalitions (i.e., M5S, left-leaning and right-leaning alliances), respectively, in yellow, red and blue, while we use purple for the information channels community most of the non-verified users have an extremely unbalanced distribution of their interactions with the members of the other alliances, since they mostly retweet content shared by people from their own community rather than from different one.

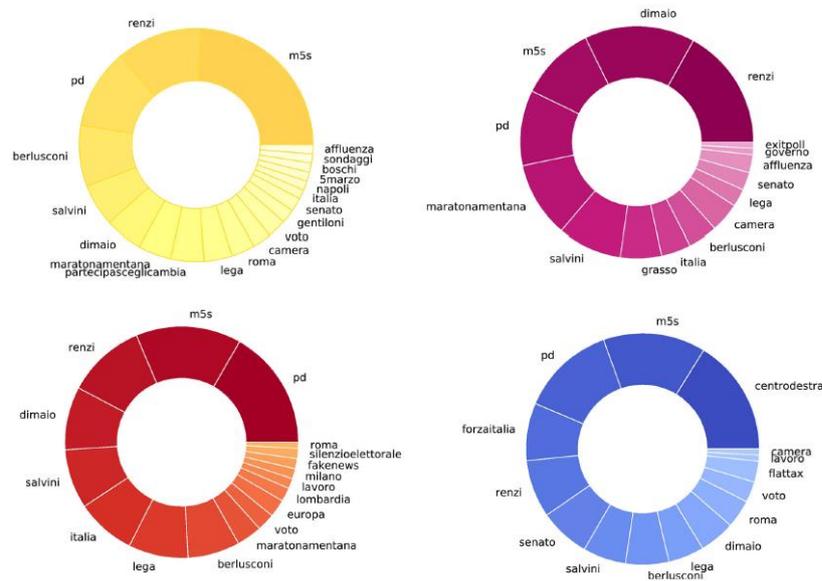


Fig. 1 Hashtags used by the four network communities

Here we show the proportion of hashtag (on top of the ones indicated for the crawl) that have been used by the members of each of the four communities.

By construction, the behaviour of non-certified users is exploited to understand the division of the political sphere into clusters. An analysis of the obtained communities shows that the most central accounts are mostly newspapers, journalists and information channels in general, each of them belonging to one of the previously listed clusters. Looking at the hashtags included in the posts published by the verified users in each group, we see that the most used keywords are referred to the party itself and its members, followed by keywords related to political competitors.

However, when also non-verified users are included in the analysis, the behaviour changes and the most popular keyword becomes the name of the major competitor. Given the obtained division in political alliances, we studied the behaviour of the remaining non-verified users towards these groups. More specifically, we observed the fraction of retweets directed towards each alliance: we observe a strongly polarised behaviour, since the majority of the uncertified accounts in the bipartite network of retweets mostly interact with one community only. In order to strengthen this result, we also performed a different analysis comparing the distribution of polarisation values observed for users with the same number of interactions. Even in this case, the distribution is skewed towards the higher numbers, indicating a focus towards the same group of users. At this point we focussed our attention on the identification of significant news spreaders and collected a new set of data on migrants.

Migrants coming to Italy

To identify the role of bots in this environment we merged the application of the lightweight classifier for bot detection used also above with the analysis of complex networks via entropy-based null-models [Caldarelli et al 2019]. Once cleaned the system from the random noise via the application of the null-model, we studied the effects of social bots in retweeting a significant amount of messages on Twitter, without entering in the highly sensitive matter of the veridicity of the messages exchanged. We applied this analysis to a tweet corpus concerned with the Italian political propaganda about migration in the Mediterranean Sea from Africa to Italy through Libyan ports. We selected tweets according to the presence of specific keywords and analysed the network of messages and the related accounts across a one-month period. We measured that the most effective hubs in the considered network have a number of bots-followers higher than average.

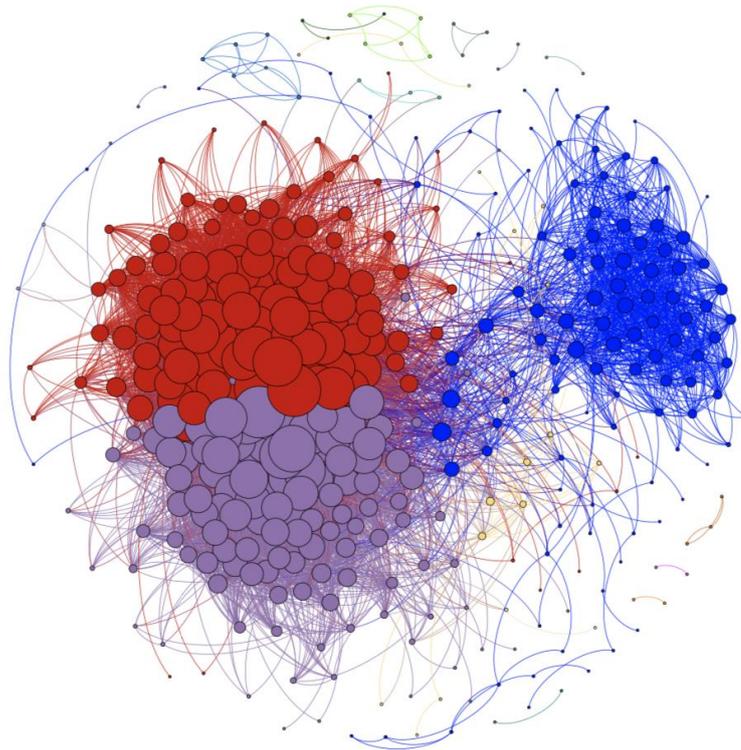
We collected 1,082,029 tweets, posted by 127,275 unique account IDs, over a period of one month (from January 2019, 23rd to February 2019, 22nd). By relying on the technique introduced in [Cresci et al. 2015], all the accounts have been classified either as genuine or as bots. This classification led to 117,879 genuine accounts and 9,396 social bots. All collected tweets were stored in Elasticsearch for fast and efficient retrieval. Following the methodology presented in (Bovet and Makse, 2018), we constructed the directed network of retweets among users and selected the names with the highest outdegree and in-degree. In order to statistically validate our findings, we constructed the bipartite and directed network of tweets and retweets introduced in Sections 'Bipartite directed network of tweeting/retweeting activity', 'Bipartite Directed Configuration Model and directed validated projection' and performed the validation procedure described therein. The outcome of this analysis is a monopartite, directed network of users, in which an edge from i to j indicates that the latter retweeted contents posted by the former a significantly high number of times. The visualisation of this validated network helps to understand the actual composition of each coalition, as well as the possible interconnections between them. For instance, we observed that part of the connections between one community and another one happens between verified and unverified users, where, in most cases, the latter retweeted some posts from the former. However, we also see connections involving newspapers and information channels belonging to different coalitions, confirming again their essential role and centrality in spreading news on the social networks.

Based on these conclusions we realized a similar analysis on a large corpus of Twitter data, generated by collecting tweets about migration, and focusing on the case of Libya-Italy flows (Caldarelli et al 2019). For the data collection operations, we developed a crawler based on Twitter public Filter API, which provides real-time tweet delivery, filtered according to specified keywords. We selected a set of keywords compatible with recent chronicles. The table below lists the selected keywords. The filtering procedure was not case-sensitive. The keywords have been selected because they are commonly used in Italy when talking and writing about immigration flows from North Africa to the Italian coasts and about the dispute about the holder of jurisdiction for handling emergencies, involving European countries and NGO's.

KEYWORDS

immigranti	Immigrates
migranti	Migrants
ONG	NGO
scafisti	boat drivers as humans smugglers
Seawatch	Seawatch
barconi	barges/boats
clandestini	illegal immigrants
guardia costiera libica	Lybian Coast Guards
naufragio	ship wreck
sbarco	Disembarkation

By comparing the real system with a null-model, we can highlight all the contributions that cannot be explained only by the fixed constraints. The filtering returns a directed network in which the arrows go from the authors to the retweets and it reduces the number of nodes to 14,883 users and of links to 34,302

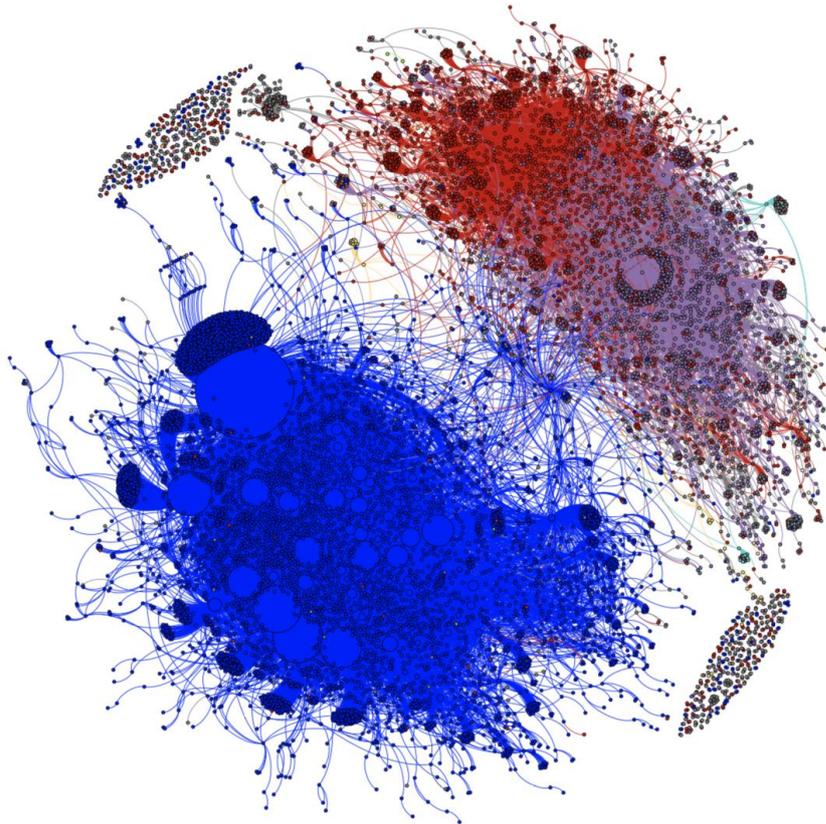


The network resulting from the projection procedure. In blue, accounts tied to the Italian government, to the right wing, to Movimento 5 Stelle and the official account of the newspaper 'Il Fatto Quotidiano' with its journalists. In tomato red, the accounts of the Italian Democratic Party (PD) and its representatives, as well as some representatives of smaller parties on the left of PD. The purple group includes several NGO's, politicians on the left of PD, different online and offline newspapers. In orange, some official accounts related to the Catholic Church. Finally, in turquoise we found smaller groups related to the Malta government (including the Prime Minister Joseph Muscat and some of his ministers), and in green we even found a soccer commentators community.

With respect to the previous study on the election campaign, here we find several differences. While the M5S group and their supporters were distinguishable from the right wings during the election campaign, in the propaganda regarding the Mediterranean migration this is not the case. Similarly, the left wing representatives, both inside and outside the Democratic Party, are much closer than during the election campaign. Smaller communities are linked to Malta prime Minister Joseph Muscat and part of his ministers involved in the discussion for the aid of migrants and castaways. As mentioned in the previous sections, the validated links go from the authors to the retweets. In this sense, the effectiveness of an author can be derived by its ability to reach a high number of most relevant nodes: this principle is finely implemented in the Hubs-Authorities algorithm. Remarkably, we observe a non-zero overlap among the bots in the list of the validated followers of human users. To the best of our knowledge, this is the first time that such a phenomenon is detected. The presence of bot squads, retweeting the messages of two or more strong hubs, increases the visibility of their tweets.

We detect two main groups of such accounts, the other being composed by a maximum of 2 common bots. The first one includes 22 genuine accounts (9 of which are in the top 10 hubs), sharing 22 bots. In this set, some users share a relatively high fraction of bots: there is the case of a right-wing account sharing all his automated followers with both Meloni and Salvini. All the accounts that are mostly effective in delivering their messages (i.e., the Hubs) refer to the blue area the figure below, where we can find representatives of the government in charge and the right wing. The first account referring to a community different from the blue one is the official account of the newscast 'TgLa7', at position 176th in the hub ranking.

The contribution of bots to the visibility of the various accounts shows that the fraction of bots that significantly retweet the content of two right wing political leaders (Mr. Salvini and Ms. Meloni) is greater than the incidence of bots in the validated network. Interestingly enough, other hubs show a smaller presence of bots among their followers, even if their hub score is not that different from the two political leaders.



The directed validated network. Nodes have been coloured according to the partition in Figure 3. The dimension of each node is proportional to its hub score: the biggest node (in blue) is the account of Matteo Salvini

6.3.4 NARRATIVES AND ONLINE DISINFORMATION

The goal of our research was to understand how narratives function in the age of social media. To achieve this goal we have conducted a pilot study in which we analyzed articles in EU English-language magazines, selected British, and Russian magazines published between January 2017 and March 2019. Overall, approximately 200 thousand articles were included in our analysis. We also analyzed reactions to these articles on social media. We use the interactivity index on social media as one of our main measures - so we can exactly tell what the impact of each article on the activity on social media is. An Internet research company - SOTRENDER -, collected social media reactions.

This research offers a comprehensive view of the narratives in the recent public discourse in Europe in the age of social media. To analyze the articles for narratives, we combined massive automated machine analysis of the text with human interpretation – by reading the most representative articles and performing semiotic analysis of narratives. We have combined existing methods in automated text analysis - e.g. topic analysis, entity analysis, sentiment analysis, sentiment distributions -, in a special package that highlights the narrative content of the articles. Our analysis focuses on the dominant narratives in the EU press; however, we have also compared the perspectives of British and Russian media outlets on selected issues. The topics of narratives that are of special interest for EU have been analyzed in more detail.

This research has a pilot character. It shows, however, the potential of a large-scale, systematic analysis of the contents of EU press and social media reactions to the articles, for evidence-based decisions. The results of such analyses can be used by policy makers, social organizations, business and the media, significantly enhancing the knowledge of narratives in Europe and its presence in the press and social media. To realize the full potential of the proposed method, the research needs to be further developed and more systematically implemented.

Introduction

Narratives enable us to understand the world and to share this understanding, since they are not only capable of combining objects and actions into meaningful structures, but also to transmit these structures through various media from oral speech, through writing and print, to television and new media. We may thus say that narratives fulfill two crucial goals: they connect an individual to the reality by providing the sense of agency, and they bond individuals together by establishing a narrative community that share similar myths and stories.

From this point of view, the European Union has been a narrative project from the very beginning. Focused on overcoming the difficult past and setting ambitious goals for the future, the very idea of European unification is narrative in both the aforementioned meanings – it provides a certain interpretation of facts (economy, politics, environment, trade etc.), while also creating community bonds through shared stories.

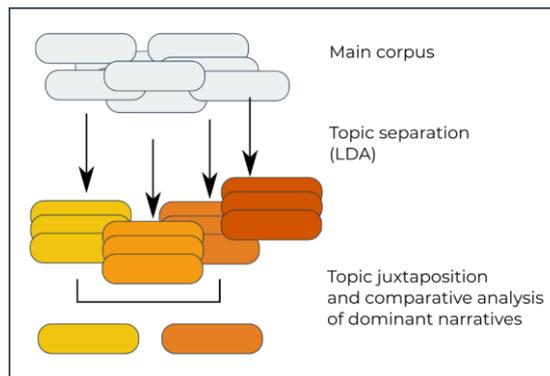
For centuries the process of creating and distributing narratives has been radically asymmetrical. Privileged groups of rulers, priests, experts etc. maintained an “interpretative monopoly”, secured by high thresholds imposed by traditional media. It required a lot of resources to build a temple, erect a monument, print a book or broadcast a TV channel. Last decade has completely eroded this model. The shift towards a much more symmetrical “narrative market” certainly may be viewed as an opportunity. It plays a pivotal part in the process of democratization, creating a participant civic society and emancipating minorities. At the same time, a world where narratives are no longer provided and controlled by privileged minorities open new opportunities for propaganda, disinformation and manipulation. Many studies show how the rapid development of social media resulted in the “death of expertise”, dawn of the “age of polarization” or “age of the post-truth” and of the seemingly unstoppable invasion of fake news.

We suggest that the same processes that caused these disturbances may be used to overcome them. Not by reversing the process and re-establishing the control of the elites over the narratives, but by using AI, algorithms and machine learning to create a more human-centered market of narratives in which values and common goals are more important than purely market-driven incentives.

In our research we have used advanced data-mining and Natural Language Processing tools and supplemented them with qualitative analysis of selected texts to explore the subtle balance between the top-down production of narratives by selected European news outlets and their bottom-up reception in social media.

Analyzing a large corpus of 200,000 articles we were able to derive a coarse-grained picture of the semantics of public discourse and to divide the whole corpus into 30 topics. Furthermore, we wanted to go deeper and create a scheme for extracting more narrative-like patterns, such as leading actors and relations between them. We believe this non-standard approach combining three distinct types of data analysis (semantics, content sentiment and social media reactions) by itself can provide important insights, both in terms of confirmation of existing hypotheses based on either existing literature or common sense, as well as in terms of detection of new surprising patterns of correlations. However, we were also able to devise a method for detecting actors and relations between actors present in textual data based on modern Natural Language Processing (NLP) techniques such as Named Entity Recognition (NER) and Part-of-Speech (POS) tagging. This is still the first step on a long road to fully automated narrative analysis at scale, but as we will see, it can be already useful and provide nontrivial insights.

After dividing the corpus into 30 topics and selecting the most representative and most popular articles for each topic, we have performed a semiotic analysis of selected articles to establish whether there exists a dominant narrative pattern associated with each topic. To establish a correlation between the topics (automatically generated) and the narrative patterns (qualitative analysis) we have juxtaposed pairs of topics similar in terms of subjects (for example: migration, economy, environment) and compared dominant narrative patterns in each topic. This analysis revealed that seemingly similar articles may differ in terms of narratives and that these “narrative reframing” may result in significantly different reception in social media.



We also prepared a more detailed case study of a corpus of articles belonging to the topics of climate change, environment and economy. This case study, taking into account also Russian-funded and British media outlets, has also suggested the vital importance of narrative framing in the process of news distribution and reception.

Importantly, this is a work-in-progress. The main goal of this project was to devise a preliminary version of a systematic data-mining framework able to extract narrative-like patterns from textual data and general discourse analysis on a big scale. Obtained results are promising, but the method certainly requires further development and the corpus of articles should be expanded to other news outlets, especially non-English, to create a more representative picture of public discourse in Europe.

Narratives: a mixed approach

Narratives make facts meaningful, satisfying the primal human drive of understanding reality. Narratives order events into a story, giving the sense of agency to certain actors, while rendering others as passive. Through this structuring power, narratives are tightly connected to the collective understanding of agency, responsibility, and social norms.

The same facts may be reported through various narratives. This enables changing the attitude, sentiment, and sense of agency or responsibility without altering the facts themselves. Narratives may be, thus, perceived as a pivotal tool of framing, that often happens unconsciously both for the authors and readers.

In our research, we apply a narratological understanding of narratives in which a narrative is a representation that contains a change of state. This notion of change was a crucial factor allowing us to determine whether a given article is a narrative or not, and in some cases, even to compare levels of narrativity roughly measured by the importance of changing states in a given text. The level of narrativity in a given corpus and its dominant narrative structure may serve as important indicators of some important trends organizing the collective imagination.

Who is rendered as an agent? Certain individuals, nation-states, European institutions, or maybe Europe itself, personified as a single, active entity? Who is perceived as responsible for the necessary change or keeping the status quo? Is a change imagined as possible, or is the world presented as a static place? Are human actors capable of making significant change or is the change restricted to the blind forces of nature or fate?

We have addressed all these questions by supplementing the automatic quantitative analysis of the corpus with qualitative semiotic analysis in the form of a limited pilot study. Latent Dirichlet Allocation² (LDA) divides the material basing on the vocabulary used. The sentiment analysis and subject-verb-object analysis also use the outer surface of the language (semantics) rather than deeper levels (narrative structure). To compensate for that, after dividing the corpus into 30 topics using LDA, we have compared a group of most representative articles in every topic to determine the level of narrativity and dominant narrative structures.

We have then compared topics that were semantically close (i.e., related to similar or same issues) to determine whether the differences in used vocabulary detected by LDA correlate with different narrative structures. For each article in the TOP 10, we have analyzed the dominant active agents (subjects of

²

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.

actions), actions and goals, dominant narrative structures, and level of narrativity, as well as space/field where the actions took place. This enabled us to detect significantly different framings related to the same topics, such as migrants or EU politics.

This comparative analysis revealed what seems to be significant patterns, however, it is very important to highlight that these are just initial results that certainly require further study and establishing a more precise method of comparing narrative structures.

In the following sections we will describe the methodology of the narrative analysis work package of the SMART project. We start with the general overview of the goals of the empirical part of the work package. Next, we describe the data gathering process and briefly present the technical computer infrastructure we set up and used. Finally, we provide a description of statistical and machine learning methods and related computational software as well as non-standard and custom metrics we developed for a quantitative description of the data.

Sources

In sum, we have collected approximately 200 thousand articles from EU, British, and Russian press. Most of our analysis present data from a much smaller sample (slightly over 32 thousand) of articles published in European magazines. Table 1. presents detailed information about both corpora.

Table 1. Description of the publishers

Source	Origin	Articles	Facebook	
			Likes	Followers
The Parliament Magazine	UE	887	4502	4848
EUobserver	UE	3663	141859	142034
Euronews English	UE	19224	2040663	2083956
EURACTIV	UE	4472	40179	42370
POLITICO Europe	UE	4212	143028	149377
Russia Today	Russia	28987	5510636	5600911
Sputnik News	Russia	42927	1252461	1269249
Daily Mirror	UK	41553	3114909	3073220
The Guardian	UK	35875	8160962	8072817
Daily Express	UE	21898	1334230	1349291

The articles have been collected with web-crawlers (bots for scraping data from the internet) and stored in a document database (MongoDB) together with additional metadata, such as language tags and sentiment values.

We limited our data collection processes to articles promoted by the publishers on Facebook. This allowed us – thanks to our business partner SoTrender – to combine the information on the reactions of Facebook users. Social media data was provided to us by SoTrender, which extracted it via the official Facebook API.

Narrative analysis

Our approach to narrative analysis was based on NLP techniques, namely Named Entity Recognition (NER) and Part-of-Speech (POS) tagging. For that, we used a full-featured, state-of-the-art NLP library for Python named *Spacy* (<https://spacy.io/>). It is widely used in research and industry and is well tested, thus we could rest easy that our core linguistic annotation is of high quality.

Based on the output provided by *Spacy* we designed a method for extracting subject-verb-object (SVO) triplets from articles. In general, SVO triplets are sequences of tokens corresponding to a subject (active agent), verb (action), and object (passive agent). The method was based on syntactic analysis of sentences using POS tags. Furthermore, we limit our data to SVOs that could be reliably identified – based on entity recognition – with known actors such as politicians, nation-states, corporations, social/political groups, and well-defined events (i.e., Brexit). Furthermore, in some analyses, we used vector space-based representation of the semantics of verbs to determine semantically similar SVOs. The vector space representation we used was the one provided by *Spacy*.

Sentiment

The first variable computed after the collection of articles was the sentiment. We built our sentiment metric on top of the widely used sentiment model for English language texts called *VADER*.³ It is a parsimonious rule-based model tailored specifically for web-based data such as articles, blogs, and other social media posts. For every text, it outputs rates for negative, positive and neutral sentiment as well as a compound score ranging from -1 (most negative) to 1 (most positive). We wanted to have one synthetic measure capable of expressing both negative and positive scores, so it seemed natural to use the compound score. However, we found that, although it detects the general sentiment class (negative vs. positive) well, it also tends to approximate maximum or minimum value for long texts regardless of the actual emotional intensity. To solve this, we defined a corrected sentiment index as:

$$\textit{Sentiment} = \textit{Compound} \times (1 - \textit{Neutral})$$

This is a simple metric that weights the compound score by the actual rate of non-neutral sentiment.

Topic modeling

The next part of our data-processing pipeline was topic modeling. The goal of this stage was to detect a number (depending on a particular analysis) of topics and assign each document to a topic that was the most representative. For this task, we used the Latent Dirichlet Allocation (LDA). It is a hierarchical probabilistic model that is efficient at scale and widely popular. It has been proven to work very well in many practical settings⁴. We used a full-featured, industrial-strength implementation of it provided by the Gensim library for Python programming language (<https://radimrehurek.com/gensim/>).

Facebook reactions

As was mentioned before, data on Facebook reactions on articles was provided by SoTrender, which extracted it using the official Facebook API designed for such purposes. In most cases, in further analyses, we used the so-called Interactivity Index (INI), which is a synthetic metric that captures the intensity of user interactions with a given piece of content. It is widely used in social media analytics and is defined as follows:

$$\textit{INI} = \#Likes + 4 \times \#Comments + 16 \times \#Shares$$

However, in empirical data (at least in our case) INI tends to have highly right-skewed distributions similar to the lognormal distribution. Hence, to normalize data we used transformed INI:

$$\textit{INI}_T = \log(\textit{INI} + 1)$$

Narratological analysis

The top 20 most relevant articles for each topic were then analyzed qualitatively for dominant narrative patterns using basic narratological tools derived from the work of Wladimir Propp, Roland Barthes and Algirdas Greimas⁵. The first step of the analysis included identifying dominant actors, objects and actions. Results from previous NLP analysis were taken into account as a basis of the manual analysis. Then the narrative pattern, understood as change of values in time, was analyzed to establish whether there is a dominant temporal pattern for the whole topic and if the paired similar topics are different in terms of dominant narratives.

Results

In this section, we present summaries of selected topics from 30 topics detected with Latent Dirichlet Allocation (LDA) (1 topic is likely to be an ambiguous fit and is flagged as BAD FIT). Topics are numbered starting from 0. In this report, we present only the list of topics and a more detailed analysis of selected topics which seemed very relevant for the European debate.

We have juxtaposed some of the topics that were similar in terms of subject in pairs to perform a comparative analysis of narrative patterns.

³

Hutto, C. J., & Gilbert, E. (2014, May). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. 10.

⁴

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

⁵

Propp, Vladimir. *Morphology of the Folktale*. Vol. 9. University of Texas Press, 2010. Barthes, Roland, and Lionel Duisit. "An introduction to the structural analysis of narrative." *New literary history* 6.2 (1975): 237-272.

Since the new European Commission has declared the fight against climate change its main goal, we have also supplemented the analysis with an additional detailed case study presenting a detailed topic analysis and narrative analysis of the climate change debate in the selected European, British and Russian media. We have also looked at how European British and Russian media portray Brexit.

30 main topics discussed in EU magazines

- Topic 0: Energy and Environment (N = 889)
- Topic 1: Elections and Politics (N = 1411)
- Topic 2: Art, Entertainment, and Pop Culture (N = 2161)
- Topic 3: EU Regulations (N = 367)
- Topic 4: Trump and American Politics (N = 1405)
- Topic 5: Russian Influence (N = 474)
- Topic 6: Migrations (N = 1028)
- Topic 7: Political Economy (N = 781)
- Topic 8: Attacks and Terror (N = 2456)
- Topic 9: Health and Environment (N = 496)
- Topic 10: Science (N = 423)
- Topic 11: Sport (N = 888)
- Topic 12: British Politics (N = 656)
- Topic 13: Nuclear Politics (N = 1920)
- Topic 14: Public Health and Technology (N = 639)
- Topic 15: Refugees and Global South (N = 796)
- Topic 16: Middle East Affairs (N = 220)
- Topic 17: Industry, Trade, and Tariffs (N = 639)
- Topic 18: Italy and France (N = 415)
- Topic 19: Poland, Romania, Hungary, and EU (N = 832)
- Topic 20: EU Politics (N = 1164)
- Topic 21: Financial Scandals and Corruption (N = 250)
- Topic 22: Privacy Policy and Scandals (N = 485)
- Topic 23: Catalan Crisis (N = 566)
- Topic 24: Not Interpretable (N = 544)
- Topic 25: Disasters (N = 1358)
- Topic 26: Balkans (N = 295)
- Topic 27: Brussels Politics (N = 636)
- Topic 28: Brexit (N = 1569)
- Topic 29: Sexual Harassment and Gender (N = 556)

Qualitative analysis of dominant narrative patterns

Below we present a comparative analysis of dominant narrative patterns. These analyses involve a mix of automated analysis of text and qualitative human processing. Automated text analysis has indicated topics and the most typical articles for each topic. We have read the most typical articles for each topic and used the narrative interpretation approach to identify the narrative schema in the selected topics. Here we present different, coexisting narrative schema identified by the automated analysis of text on similar topics. They can be understood as different ways of framing or thinking about similar topics.

Topic 6 (migrations) v. Topic 15 (refugees, global south)

Dominant narrative structure in Topic #6: an EU country is deciding whether to open its borders for the incoming refugees or not.

The narrative has a very strong spatial and temporal dimension. Key action of the narrative (decision on opening the borders) is linked with the approaching group of migrants, usually pictured in a boat. The approaching boat serves as a ticking clock that marks the flow of time.

This narrative may be compared to two images, widely distributed in contemporary European culture. (1) It resembles a conventional image of an action movie hero deciding whether to cut the blue or the red cord to disarm a bomb. A ticking clock is making his decision not only crucial for survival but also more and more urgent. (2) It also derives (most likely unintentionally) from the Trojan horse myth. Opening the gates for the approaching boat may be easily linked to the tragic tale about the end of Troy, subverting the positive message intended by the communicators.

Dominant narrative structure in Topic #15: situation in a distant land is getting worse.

This narrative also has a clear spatial and temporal dimension. Narrative space is created by presenting an image of "a land far, far away". The history happens "elsewhere", "far from here", "in a different reality". The temporal dimension is clearly a narrative of decline⁶. It may be presented as a vector pointing right and downwards. The more we move from the good mythical past ("the golden age"), the worse the situation gets.

This narrative drives from two main images. (1) The decline and fall – the narrative best illustrated by Edgar Allan Poe's *Fall of the House of Usher*. However, it is the fall of a closed, distant land, not a decline that touches us directly, opposite to the equally popular narrative of *our decline*, illustrated by Edward Gibbon's *History of the Decline and Fall of the Roman Empire* or Oswald Spengler's still popular *Decline of the West*. (2) Biblical history of plagues: famine, hunger, war and diseases – images driven from the *Apocalypse* or the Old Testament organize the social imaginaries of the western reader even in the secularized societies⁷. This narrative is particularly dangerous when connected with the just-world fallacy. It is much easier to accept the tragedies, especially the distant ones, by framing them as a result of a sin/guilt⁸.

Main structural differences between the two sets of articles:

1. A difference between a view from here (Europe) and a view from there (war- and poverty-stroke countries).
2. Articles in the first group (#6) are significantly more narrative, they always describe an event or a change (eg. growing statistics). It is always clear why we should read/write right now: these are urgent and precise matters. There are strong actors (states, NGOs, less often migrants) that perform visible actions. The results are usually clear.
3. Articles in the second group (#15) describe static states. Actors are seldom personalized.
4. It may be a hint that the media do not cover properly the causal (and time/spatial) relations between here and there and the migrant crisis is presented as two separate issues. It significantly reduces the explanatory power of narratives.
5. The analyzed sample provides an interesting insight: the media coverage of a topic may be „complete" but clearly separated into two different topics/sections. Indeed, reasons and roots of a crisis and its results – together with action taken by European countries and institutions – are rarely described in the same article.

Table 1

	Topic #6	Topic #15
Active agent	EU member states	no active agents or NGOs, strong passivity in narratives
Goal	resolve internal conflicts, refugees are not subjects, but objects of conflicts	Refugees are the victims of wars, poverty and natural catastrophes

⁶

Zerubavel, Eviatar. Time maps: Collective memory and the social shape of the past. University of Chicago Press, 2012.

⁷

See: Taylor, Charles. Modern social imaginaries. Duke University Press, 2004.

⁸

Lerner, Melvin J. "The belief in a just world." The Belief in a just World. Springer, Boston, MA, 1980. 9-30.

Dominant narrative structures	relating current events, conflicts	static view of the world
Dominant field of action	Europe	outside Europe

Topic 3 (EU regulations) v. Topic 27 (Brussels politics)

Dominant narrative structure in Topic #3: the industry, driven by greed, is ruining the order of the world and threatening the safety of consumers. Regulators need to act to reestablish order.

This narrative resembles a classic folktale in which the order of the world is destabilized by an irresponsible action often connected with breaking some rules⁹. Words such as “harmony”, “rules”, “regulations” are thus confronted with terms such as “risk”, “danger”, “threat”. It is worth to note that consumers are rendered as passive. In the classic folktale structure, they occupy the place of an enchanted/enslaved princess, while the regulators play the active part of the knight facing the chaotic dragon of the industry. This links the structure of the narrative with the structure of power that is represented by the very form in which the story is told. The same structure has been identified in many tales, such as medieval romances, where the distribution of active and passive roles served the purpose of legitimizing the social order¹⁰. Europe is presented here as a cosmos of rules (as opposed to chaos) – a safe kingdom where the rule of law is successfully reestablished. This is a positive image of the EU, even if a particular narrative deals with some obstacles or troubles. The rules are always established to protect the weak.

Dominant narrative structure in Topic #27: a politician manipulates the rules to obtain more power.

This narrative also focuses on the rules, but this time the rules are not reestablished in order to conquer the chaos. They are rather a chaos themselves, a ladder used by the clever politicians to climb higher in the Brussels structure. This is a very common narrative in popular culture nowadays. It dominates in the famous series *House of Cards* and *Game of Thrones*, where the quotation “chaos is a ladder” is used directly by one of the characters. Politics in this narrative is only about getting more power¹¹. Europe is presented here as an arena of constant fight between the politicians. It is a very negative image, where the rules are only set to serve the strong and powerful.

Main structural differences between the two sets of articles:

1. 8 out of 10 articles in the TOP 10 of Topic #3 refer to the case of the legislation process around the nutrition and ingredients labelling of alcoholic beverages. They all adopt a basic structure highlighting the conflict between industry and consumers. Legislators *should, need to be or have to* protect consumers. Therefore, the dominant narrative structure of the analyzed texts is not descriptive but imperative/normative. They often refer to plans (often unsuccessful) or duties using modal verbs. Particular politicians and even parties are either not mentioned or situated at the margin of the narrative. The EU as a whole or its agencies (European Commission) are the active agents in the narrative.
2. This may be perceived as an interesting example of the “normative power of Europe”¹² that is usually mentioned in the context of geopolitics and international relations. Interestingly, the analyzed corpus proves clearly that the term could also apply to the single market and related regulations. If we abstracted from the dominant case (regulation of alcohol labeling), compensating for the imperfections of the applied model of automatized generation of topics, what we would see would be a universal structure of European legislation narrating itself to the wider public, explaining its goals, purposes, methods and ambitions. This may be perceived as building the “normative power” by wielding a narrative about how this power should be executed, how to do it effectively and even by narrating mistakes and failures. **An organization that is able to speculate credibly on its power to regulate the world, is also able to create an image of this world where it**

⁹

Propp, Vladimir. *Morphology of the Folktale*. University of Texas Press, 2010.

¹⁰

Dugas, Don-John. "The Legitimization of Royal Power in Chaucer's "Man of Law's Tale"." *Modern Philology* 95.1 (1997): 27-43; Ingham, Patricia Clare. *Sovereign Fantasies: Arthurian Romance and the Making of Britain*. University of Pennsylvania Press, 2015.

¹¹

Pilipets, Elena, and Rainer Winter. "House of Cards–House of Power: political narratives and the cult of serial sociopaths in narrative politics in American quality dramas in the digital age." *Politics and Politicians in Contemporary US Television*. Routledge, 2016. 113-126;

¹²

See: Manners, I. (2002). Normative Power Europe: A Contradiction in Terms? *JCMS: Journal of Common Market Studies*, 40(2), 235–258.

is situated in the position of normative power. The voice and strong normative structure of the analyzed articles is therefore much more important than the content. The European Union and its institutions are operating on the level of axiology. In traditional mythologies this was the narrative position available exclusively for gods and heroes.

3. The dominant event in **topic #27** is a conflict around the 'lead candidate' system and the internal procedures of the EU Parliament. The active agents in the narratives are either politicians or political parties. Their goal in the narratives is always a seat or a function – becoming a Commissioner, Commission President or EU Parliament President. The EU is presented as an arena of permanent conflict between parties, fractions and ambitious politicians.
4. Performed actions are often verbal – X attacks, accuses, threatens or approves Y, sometimes they are connected to the legislative process (X votes for Y) but the legislative issues are always directed 'inwards', not 'outwards' i.e. they regulate the political life in Brussels not the everyday life of the citizens.

Table 2:

	Topic #4	Topic #28
Active agent	EU as a whole / EU institutions	politicians or political party
Goal	protect the citizens	win some seats / functions / positions
Dominant narrative structures	normative/imperative (<i>what we have to/should do</i>)	relating to current events
Dominant field of action	everyday life of citizens	political life in Brussels

Topic 10 (science) v. Topic 14 (public health & technology)

Dominant narrative structure in Topic #10: Humanity has accomplished something remarkable.

This is a relatively simple narrative describing an adventure – fulfilling a goal, reaching a destination, achieving a success. A task has been completed. This narrative is often inscribed in a wider meta narrative structure of progress, which is often connected with the Enlightenment. The history of humanity is rendered here as a great march towards a better future.

Dominant narrative structure in Topic #14: Europe is facing a challenge created by a new technology. This challenge may only be overcome by adapting.

This narrative is much more complicated. The subject (usually Europe, sometimes humanity as a whole) faces a challenge created by a new technology that seriously impact the current socio-economical reality. This narrative may be associated with something as personal as e-cigarettes or as broad as vaccines, the future of work or cancer research. The technology creates a challenge, sometimes by reframing an existing problem and opening new possibilities, posing a threat to the existing social order. The answer is always more technology, but also smarter regulations. This narrative is also strongly entangled in the Enlightenment meta narrative of progress. However, it also uses some classical motives from the heroic myths.

Main structural differences between the two sets of articles:

All 10 articles in TOP 10 of topic #10 cover space-related topics. 9 of 10 narrate current progresses in space missions. Active agent is usually a space agency (NASA) or a satellite. Europe is often mentioned as a key partner in aerospace industry. What is striking, however, is that in this topic there are no mentions of links between science and everyday life. Research is presented as a purely scientific endeavor not changing the everyday life of European citizens. Articles in topic #14 cover instead the relations between science/technology and the everyday life and wellbeing of European citizens.

This may be an interesting insight for policymakers and communicators: science/technology are perceived and narrated twofold: as an abstract endeavor or as a life-changing mechanism.

Table 4:

	Topic #10	Topic #14
Active agent	Humanity	Europe

Goal	expand, develop	overcome an obstacle
Dominant narrative structures	progress meta narrative	narrative of challenge
Dominant field of action	abstract space	everyday life of citizens

CASE STUDY 1: Climate change and the global economy

To supplement the information about dominant narratives and topic structures in the field of ecology and climate policy we have performed an additional Latent Dirichlet Allocation (LDA) narrowing the full corpus of over 200,000 articles to a more coherent sub-corpus containing texts related to Economy, Environment, and Energy. It contained 9,514 articles (EU sources – 3,975, Russian sources – 4,408, and UK sources 1,131). On this smaller corpus, we run another LDA to divide it into 10 topics:

Topics identified in the subcorpus

- Topic 0: Social Issues (N=1273)
- Topic 1: Cars (N=468)
- Topic 2: EU (N=1073)
- Topic 3: Nord Stream 2 (N=488)
- Topic 4: Climate Change and Energy (N=920)
- Topic 5: Economy and market (N=1063)
- Topic 6: Food and Agriculture (N=825)
- Topic 7: Oil market and sanctions (N=1131)
- Topic 8: Russia, Infrastructure, and Gas (N=282)
- Topic 9: Trump's Trade War (N=1020)

Topic-source distribution

Comparative analysis of the topics shows clearly that most of them are dominated either by the EU/British media or Russian media. The topics covered by various sources differ and so does the dominant vocabulary. It is especially worth to note how actively Russian media are covering the topic of international politics (especially sanctions, trade wars and infrastructure) and Nordstream 2.

From the point of view of topic-source distribution, agriculture seems especially interesting (Topic 6), since 8 out of 20 most fitting articles are from The Guardian and 10 out of 20 are from Sputnik and RT. This equal distribution may suggest important similarities between these media outlets in terms of narratives about food and agriculture. In this case it signifies mostly critic of GMO and catastrophic visions of bee-extinction, pesticide overuse, cancer risk etc. These similarities would certainly require further research.

Interactivity analysis

Despite significant differences between interactivity patterns for various sources, a certain pattern seems to emerge, showing a group of subjects that provoke most interactions:

1. global inequality
2. anti-Americanism (including anti-Trump messages)
3. anti-corporate content

The analysis strongly suggests that one of the basic needs satisfied by the most popular articles **is the need to attribute guilt** for the economic disturbances and the climate catastrophe. Abstract, distant, global processes are made understandable by giving them a human face or a corporate logo. This process may be interpreted as **a narrative to restore the sense of agency**.

Global politics and ecology

Our pilot study suggests that the view on the economy-climate presented in RT and Sputnik is much more spatial than temporal (narrative). A significant majority of most typical (topic-fitting) and most popular (interactivity) articles are constructed **in the form of a map**, not a story. The paragraph above is representative for this peculiar way of explaining the world:

"The UK's shock move opens up a wider question about energy security on the continent, especially in light of American attempts to prevent the development of Nordstream 2 – a pipeline connecting Russian gas to the German market. Berlin regards the new pipeline as vital to its interests, and Moscow sees it as a way to prevent a hostile Ukraine from interfering in its energy exports to the European Union."

In this very short fragment 6 geopolitical agents are mentioned: US, UK, EU, Russia, Germany and Ukraine. They are all connected by the network of interactions. Verbs such as "to open", "to connect" may be

interpreted as lines drawn on the map, while “attempts to prevent development” or “hostile [...] interfering” are obstacles that disrupt the flow of resources. This very flow of resources, described in terms of connectivity, but also economic interests and even friendship or mutual respect, is presented as the highest value.

The United States are often described as the enemy of connectivity “force-feeding Europe its liquefied natural gas, which is three times more expensive than buying it from Russia”, or “illegal sanctions”, while Russian gas saves Britain while it “turns to sanctioned Russian energy to avoid big freeze”.

In Russian media outlets climate and the economy are not depicted as a process (a catastrophe to avoid) but as a game. This enables creating a picture that is “revealing” (it promises to show the “hidden” structure of the world) and at the same time avoiding any mentions of climate change.

CASE STUDY 2: Pilot experiment: Reactions on debunking in the new media

To understand the results of debunking fake news on social media, we conducted an experiment by observing social media reactions to debunking anti-vaccine false beliefs.

The availability and quality of information is crucial to effective functioning of society, especially in a time of crisis. A lot of false information regarding Coronavirus is spreading in societies, especially through social media. One of the main strategies for fighting fake news is debunking, a strategy of confronting it with facts and accepted theories. This effort is not very effective, especially as it does not work with individuals that have strong belief in false information. A seminal study in social psychology (Festinger, Riecken & Schachter 1956), has demonstrated that a religious group, predicting the end of the world at a specific date, more strongly believed in the prophecy, after their prophecy was falsified.

Conspiracy theories and pseudoscientific beliefs surrounding vaccinations provide a good example of dangerous phenomena occurring at the intersection on the social media, fake news and democracy. Research shows that Twitter, YouTube, Facebook and other influential social media provide a crucial boost for the international network of anti-vaccination activists (Smith and Graham 2019). Feeding on fear and chaos, spreading fake news and misinterpreting the data, the anti-vaccination movement is stronger than ever (Fernández-Luque and Bau 2015). Social media provide not only a new set of tools for spreading disinformation but also a great weapon against it. What is the balance between these two forces?

To understand better this complex phenomenon, we conducted a pilot experiment with a limited range by posting online an article debunking main arguments of the anti-vaccination movement. All the comments were then collected and analysed qualitatively to identify dominant strategies of reaction on debunking.

At the beginning of November 2018, a measles outbreak occurred in Pruszków – a city of ca. 60 000 inhabitants located at the suburbs of Warsaw. In just a few days over 20 people had been diagnosed with the disease, that had been virtually eliminated in Poland. The situation became the main news story of the first half of November. The anti-vaccination movement in Poland and its prominent leaders also communicated their views on the situation extensively both via private social media and institutional channels. We collected the reactions of the anti-vaccination movement to the outbreak and confronted their main arguments in a blog post. The text polemized with the claims of the anti-vaccination movement, providing data and links to reliable sources.

In the following 10 days over 45 000 viewers read the article and it received over 30 000 likes on Facebook and over 2 400 reactions on Wykop (roughly a Polish equivalent of Digg or Reddit), making it one of the most popular “findings” of the day on this popular portal.

Using qualitative methods, we analyzed 390 comments left under the text, on Facebook and Wykop to establish main strategies of reaction to debunking. Many comments were responses to other comments, enabling us to analyse a complex structure of discussion, where users sharing an anti-vaccination sentiment had to defend their views and engage in a debate not only with our article but also with other users representing an opposite view. “Polemic” structures were over 300% more frequent than “supportive” – anti-vaccination comments were posted mainly in a reaction to pro-vaccination ones and vice versa. Only a tiny fraction of comments referred directly to the commented article. Only 6 comments referred directly to the commented article and only 2 of them attempted to discuss the presented arguments. This seems to be a very important and general pattern, that suggests that **in a situation of a threat to their worldview the supporters of the anti-vaccination theories do not react directly to the attack, but rather repeat their *credo* to consolidate individual and collective identity**. 8 main strategies of coping with the “attack” were extracted as a result of qualitative analysis. Some of them have distinguishable “sub-strategies”:

Strategy	Number of comments (from the corpus of 148 ant-vaccination comments)*
sub-strategy	
Freedom is the highest value	9
Critique of the official medicine and dehumanization of patients	5
Vaccines are more dangerous than the disease	17
Children are vaccinated to early	4
The hidden truth	36
"I'm proud not to vaccinate"	8
Follow the money	10
Attacking the pro-vaccination movement	25
Ad personam arguments	21
Accusing of manipulating the data or requesting more data	14
Accusing of ridiculing the anti-vaccination movement	6
Pointing the inconsequences of the pro-vaccination argumentation	5
Quoting sources	33
Vaccines are ineffective	15
Other means are more effective (eg. hygiene)	3
Symmetry	21
*Numbers do not sum up to 148 because some comments use more than one strategy, some of them do not use any distinguishable strategy.	

The experiment demonstrates that the anti-vaccination reactions to debunking are strongly grounded in the liberal philosophy of “individual freedom” valued more than the “collective good” associated with communism (in Poland this definitely lends additional gravity to the argument). Pro-vaccination debunking is identified with “dehumanizing” ideology of science and progress, represented also by the mainstream media. Heuristics of availability are often used to point on *visible* and *near* negative consequences of vaccination for the individual (against the *elusive* and *distant* benefits for the community). Contrary to the research highlighting the importance of conformism in declared beliefs, being a member of an “anti-establishment” movement provides the users with a very attractive “rebellious” identity. Proofs of the harmful nature of the vaccines are derived very often from the everyday experience regarding not only medicine, but the world in general, especially economy, politics, and structures of power etc. These initial findings suggest that effective debunking must take into account formal and social structure of the debunked arguments. Instead of focusing on proving the opponents wrong, we should focus on understanding *why* they choose to accept these particular beliefs.

CASE STUDY 3: Brexit: The perspective of EU, British and Russian Press

In this section we compare the perspectives on Brexit presented by European magazines, including the results of the analysis of British and Russian Press.

EU perspective

Table 5: 15 the most typical articles for Topic 28

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
POLITICO Europe	Close but no deal: UK, EU scramble to salvage Brexit talks	-0.0780	6	3	0	54	0.999
EUobserver	May on whistle-stop EU tour to seek new backstop pledges	0.1714	3	1	0	19	0.994
POLITICO Europe	Donald Tusk: Post-Brexit deal should be 'as special as possible'	0.1808	8	2	2	48	0.995
POLITICO Europe	London's Brexit time bomb is about to blow	0.2025	152	28	16	664	0.993
POLITICO Europe	Donald Tusk: UK Brexit plan 'will not work'	0.1482	65	5	46	329	0.997
POLITICO Europe	EU leaders to affirm Brexit deal 'not open for renegotiation'	0.1740	20	5	0	100	0.982
The Parliament Magazine	Hans-Olaf Henkel: EU trying to punish UK in Brexit talks	-0.0849	0	0	0	0	0.951
POLITICO Europe	Theresa May: UK and EU 'not far apart' in Brexit talks	0.0754	20	1	14	92	0.995
POLITICO Europe	Philip Hammond: UK will enforce hard border in Ireland if there is no Brexit deal	0.1700	47	9	24	287	0.992
POLITICO Europe	Michel Barnier: EU ready to 'improve' Irish border proposal	0.1706	10	0	11	54	0.915
POLITICO Europe	Only Ireland matters in final Brexit exam	0.1303	21	4	1	89	0.991
POLITICO Europe	May's Northern Ireland backers denounce 'broken' Brexit promise	0.1036	4	0	1	8	0.990

Euronews English	UK eyes Canada-style trade deal with EU	0.1616	58	10	4	238	0.996
POLITICO Europe	UK and EU move to 'continuous' Brexit talks	0.1227	33	1	10	89	0.996
The Parliament Magazine	Theresa May meets with Roberto Gualtieri on Brexit	0.1170	0	0	0	0	0.992

Table 5: 10 the least typical articles for Topic 28

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
	UK approaching its 'darkest hour'	-0.0739	11	2	11	87	0.272
	Nine European countries to formalise EU defence force plan	0.0634	23	2	3	67	0.272
	No known cases of deported 'Windrush' residents says Britain	-0.1630	26	7	5	158	0.272
	May and Macron to sign Calais border treaty	0.1533	57	8	4	201	0.273
	France and Germany back Dutch on human rights sanctions	-0.1144	28	6	4	140	0.274
	"Hostages" to the Brexit talks: fears and frustrations grow in Gibraltar	-0.0488	34	6	62	378	0.275
	UK could return to blue passports	0.1327	104	14	64	584	0.276
	The reality behind the €7 'Brexit bombshell visa'	-0.1464	28	2	6	84	0.276
	British royals in Paris on Brexit 'charm offensive'	-0.1171	62	6	18	230	0.2766
	Ex-UK PM Cameron: Brexit not going as badly as we thought	-0.0568	10	2	2	50	0.2768

The highest interactivity is elicited by the articles: *London's Brexit time bomb is about to blow* but also *UK could return to blue passports* and *Donald Tusk: UK Brexit plan 'will not work'*.

Daily Express	Brexit negotiator Verhofstadt says transitional deal must be SAME terms as EU membership	0.0777	216	38	244	1800	0.997
Daily Express	UK needs 'a MIRACLE!' Juncker issues brutal Brexit slapdown on progress of talks	0.1967	204	40	221	1728	0.997
Daily Express	'Who will finally have the courage?' Tusk rips into UK Brexit position as PM's deal fails	0.2230	354	71	341	2854	0.997
Daily Express	David Davis will NOT be in Brussels for start of fifth round of Brexit talks	0.1047	176	23	132	1072	0.996
The Guardian	Divisions exposed over £66bn Brexit divorce bill	-0.0419	152	45	157	1500	0.996
Daily Express	'I can't tell you what it means!' Merkel's farcical admission over EU's negotiating tactic	0.1218	130	28	85	918	0.996
Daily Express	REVEALED: Merkel and Macron toughen EU's Brexit stance HOURS before May's Brussels dinner	0.1079	177	63	236	2129	0.996

Table 7: 10 the least typical articles for Topic on Brexit in British press

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
The Guardian	Bad Brexit deal would be disaster for NHS, says Jeremy Hunt	-0.0552	632	229	704	7112	0.3575
Daily Express	THERESA MAY: We WILL leave the European Union	0.2280	282	42	238	1906	0.3583
Daily Express	Brexit <u>news</u> : The entire future of Britain is at stake and THIS is how many MPs are bothered	0.1840	175	18	173	1155	0.3585
The Guardian	Following Theresa May's great escape, Brexiters plan their final battle for Britain	0.1890	227	33	214	1611	0.3587
Daily Express	Brexit SHOCK: Farage, Rees-Mogg AGREE Donald Tusk's deal plan - but will May play ball?	0.2049	185	46	106	1345	0.3600
Daily Express	British Ambassador ON THE SPOT over UK-Canada trade – 'Do you take us for GRANTED?'	0.1430	79	10	77	547	0.3601

RT	Good Friday Agreement: Has Brexit put Northern Ireland's peace at risk?	-0.1149	174	23	69	818	0.998
RT	How Brexit debacle is uncovering and unraveling the dis-United Kingdom	0.1984	87	19	20	471	0.998
RT	Explainer: Why is Northern Ireland such a big deal in Theresa May's Brexit negotiations?	0.1824	46	10	15	266	0.998
Sputnik	'PM May Wants the Best of Both Worlds, Impossible Compromise to Make' - Analyst	0.1655	23	0	6	47	0.998
Sputnik	PM May: Scottish Economy Lags Behind Rest of UK Due to One-Party System	0.1855	22	2	1	58	0.998
Sputnik	Backstop is Far From Only Problem May is Facing - Journalist	-0.1846	11	0	0	11	0.997
Sputnik	UK 'Act of Supreme Selfishness' Means Ireland Will Not Back Down on Soft Border	-0.1949	36	4	11	144	0.997
Sputnik	UK Lawmakers Overwhelmingly Vote Down May's Brexit Deal	0.1964	44	10	5	224	0.997
Sputnik	No 'No Deal' Vow, No 2 Brexit Ref, No Fee for EU Citizens: No Massive May Change	-0.1335	13	2	1	49	0.997
Sputnik	Scotland's Leader Will Seek Authority for New Independence Vote	0.1827	82	9	31	350	0.997
Sputnik	According to Leaked Report, Brexit Will Significantly Impact the Irish Economy	-0.1420	0	2	12	108	0.997
Sputnik	Brexit One Step Closer, Despite Legal and Scottish Challenges	0.1084	63	8	3	203	0.997
Sputnik	Between Three Rocks: Tough Spot for UK PM Over Belfast, Dublin, Brussels Demands	0.1396	11	1	1	31	0.997
Sputnik	EU Officials' Pensions & Greek Hotels: What Will Brexit Divorce Bill Cover?	0.1055	54	9	1	202	0.997

Table 9: 10 the worst fitting articles in topic on Brexit in Russian press

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
--------	-------	-----------	-------	--------	----------	-----	-----

RT	Blasphemy case against Stephen Fry dropped due to lack of 'outraged people'	-0.051	1095	142	238	4319	0.535
RT	Apple missed deadline for multi-billion euro tax payment to Ireland	0.113	632	186	108	4040	0.536
Sputnik	Supreme Court Brexit Appeal - 'Phony, Theatrical War': Gina Miller Exclusive	0.123	63	10	3	235	0.536
RT	Crossing the Brexit red line: Ministers threaten to quit after May's war cabinet flop - reports	0.075	222	45	64	1198	0.536
Sputnik	Prof on Brexit Division: 'Economic Catastrophe' vs 'Mlns of Muslims on the Door'	0.158	9	3	3	69	0.537
RT	EU leaders endorse contentious Brexit deal at emergency summit	0.175	169	28	41	781	0.537
RT	'People are entitled to change their mind': Tony Blair confirms he is fighting to cancel Brexit	0.034	1091	193	664	6835	0.538
RT	Operation 'No deal'? UK army braces for Brexit chaos as May faces fierce rejection of divorce plan	0.146	261	67	84	1669	0.539
Sputnik	UK PM May Will Fight Any Attempt to Oust Her as Leader - Spokesman	-0.165	44	3	9	128	0.539
Sputnik	Amber Rudd Resigns as UK Home Secretary Amid Migrant/Windrush Scandal	0.175	43	2	2	83	0.539

EURACTIV	Carbon prices seen hitting €55 in 2030, hastening 'major' coal-to-gas switch	0.0739	0	0	0	0	0.896
EURACTIV	No decarbonisation without electrification	0.1293	5	1	0	21	0.876
The Parliament Magazine	Low carbon future: Alternative technologies have crucial role to play	0.1020	3	0	0	3	0.872
EURACTIV	EU slows down in race for renewables, energy efficiency	0.1995	3	3	0	51	0.862
The Parliament Magazine	Biofuels have an important role to play in the energy union	0.2598	0	0	0	0	0.859
Euronews English	Which European nations lead the way with renewable energy? Euronews Answers	0.1569	149	23	21	601	0.855
EURACTIV	Study dampens industry projections for 'renewable gas' in 2050	0.0852	2	0	0	2	0.855
EURACTIV	Equinor: Oil majors have 'muscle' to join renewables race	0.1479	1	1	0	17	0.846
EURACTIV	UK-backed report identifies 'viable business models' for carbon capture technology	0.1505	1	0	0	1	0.841
EURACTIV	Future of gas: Decarbonise or go bust	0.1058	0	0	0	0	0.839
POLITICO Europe	UN report lays out radical path to combat global warming	0.1736	9	2	2	49	0.837

Table 11: 10 the least typical articles in Topic 0

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
EURACTIV	Corporate reporting group launches two-year transparency project on green finance	0.1668	0	1	0	16	0.271
EURACTIV	Mayor of Astana announces big ambitions	0.1500	0	0	0	0	0.271
Euronews English	How Brussels is failing to punish EU states over their deadly air pollution	0.1655	31	2	1	67	0.279
Euronews English	European CO2 shortage threatens World Cup beer supplies	-0.1121	62	12	15	314	0.271
EURACTIV	EU closer to 'genuine Energy Union' as MEPs support gas supply solidarity	0.2056	15	3	0	63	0.270

Figure 5 a, b Sentiment analysis and interactivity for topic 0

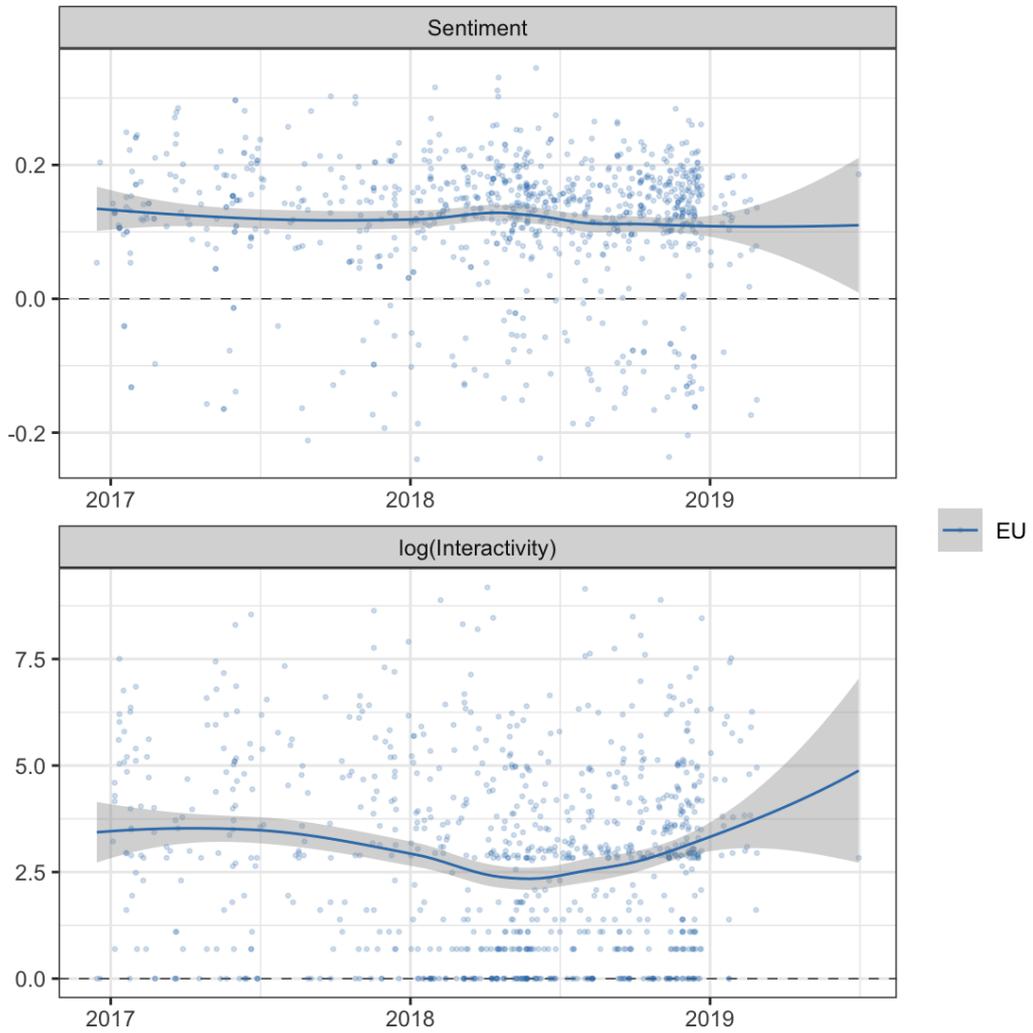
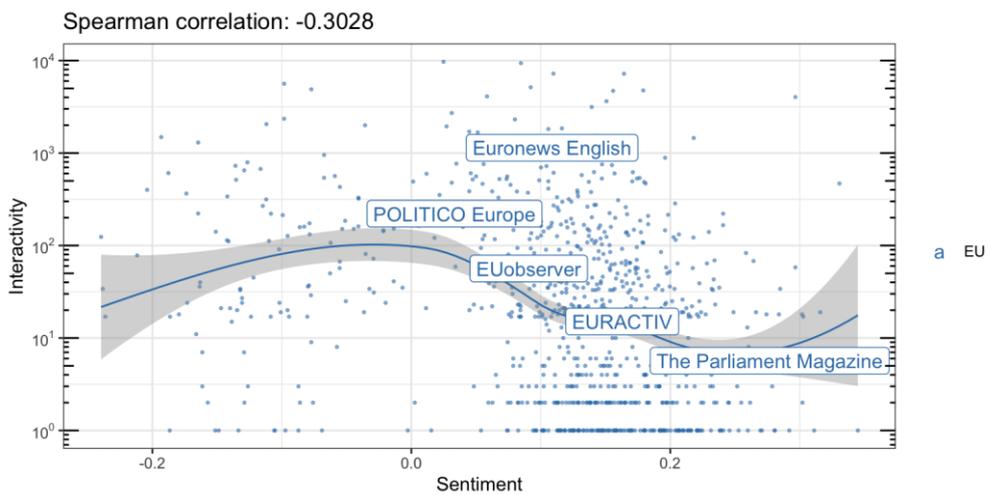


Figure 6 Relation between sentiment and interactivity in topic 0



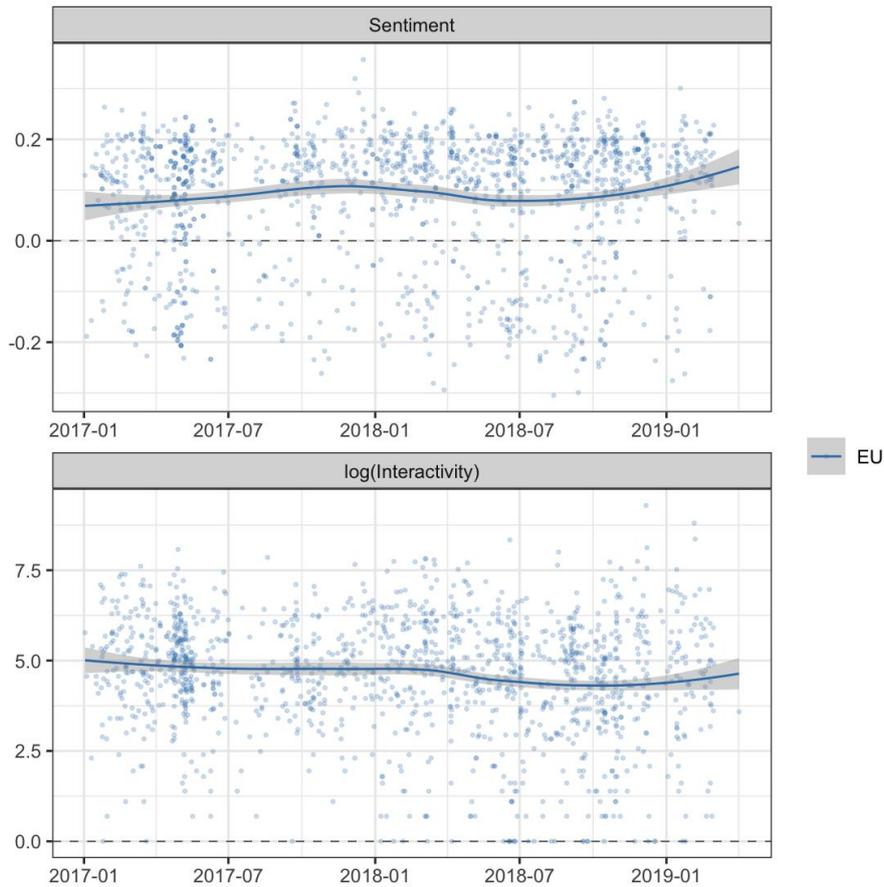
The negative correlation between sentiment and interactivity means that articles in energy and climate crisis topic with negative sentiment tend to attract more reader attention and generate stronger readers' reaction.

Topic 1. National election and politics

The topic concerns national elections and politics. Because the collection of data was terminated in March 2019, the main body of the corpus of texts concerns national elections. Most central to the topic are German elections. In subject-verb-objects triplets, however, the competition between Macron and LePen dominates the triplets.

Table 12: 15 the most typical articles for Topic 1

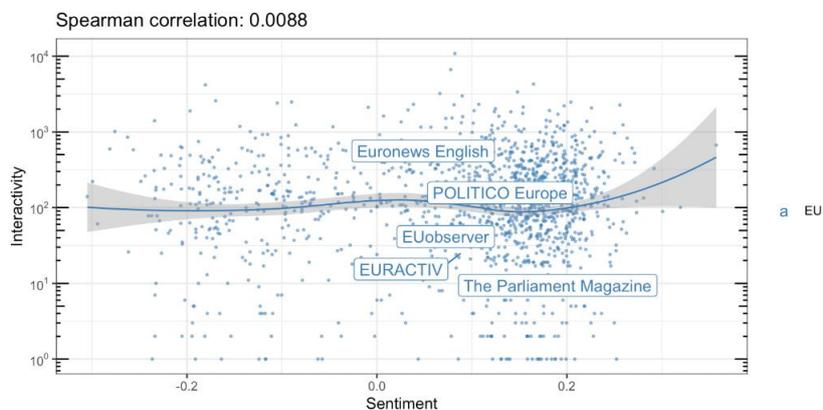
Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
POLITICO Europe	Merkel's twilight heralds German turmoil	0.2181	26	2	5	78	0.998
POLITICO Europe	How Merkel's successor could change the political landscape (or not)	0.1998	23	3	5	91	0.998
POLITICO Europe	A weakened Merkel still gets her way	0.1607	73	8	9	237	0.998
POLITICO Europe	Annegret Kramp-Karrenbauer elected to succeed Merkel as CDU leader	0.1816	357	87	40	1909	0.998
POLITICO Europe	Merkel's coalition lives to fight another day — just	0.1748	13	4	3	89	0.998
POLITICO Europe	Angela Merkel's Bavarian hangover	-0.2049	6	1	3	34	0.998
POLITICO Europe	The long, painful end of Angela Merkel	-0.0326	22	4	2	94	0.997
POLITICO Europe	Sweden braces for political uncertainty as far right makes gains	0.1696	32	9	2	184	0.997
POLITICO Europe	Markus Söder elected chief of Merkel's Bavarian allies	0.2182	4	0	0	4	0.996
POLITICO Europe	German election Merkel to step down as CDU leader	0.1781	132	62	10	1164	0.995
Euronews English	Merkel's Bavarian allies suffer historic setback in election	-0.0425	63	13	4	287	0.994
POLITICO Europe	Swedish Centre Party leader to head new government talks	0.1471	30	2	3	74	0.993
Euronews English	German voters 'undecided' hours before polls open	0.1840	73	12	119	741	0.993
Euronews English	Germany: Merkel's CDU victorious in Saarland	0.1392	108	12	11	344	0.992
POLITICO Europe	Angela's endgame: How Merkel's gamble will play out	0.1211	8	3	1	60	0.992



As we can observe, the majority of articles on politics has positive sentiment and the average sentiment gets more positive over time. At the same time the interactivity goes down, so EU citizens are getting less involved with politics.

The Figure 9 below shows the relationship between sentiment and interactivity. In this topic, as the correlation that is 0.00 indicates that there is no relation between the sentiment and interactivity, what means that positive and negative articles involve the readers in a similar fashion. It is interesting, because for most of topics the relation between the sentiment and interactivity is negative, what means, that articles with negative sentiment attract more attention and provoke more intense reaction of the readers. In the Figure below there seems to be is an increase of interactivity at very high positivity, but this could be an artefact due to a very small number of articles in this topic that are very positive in sentiment.

Figure 9. The relationship between sentiment and interactivity



Topic 6. Migrations

Table 14: 10 the most fitted articles in Topic 6

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
Euronews English	Malta agrees to let rescue ship dock before sending migrants elsewhere	-0.0745	132	12	48	516	0.981
EURACTIV	Greece starts bailing out Germany on refugees	0.0822	2	0	0	2	0.867
EURACTIV	Frontex: Arrivals of refugees to EU by sea two-thirds lower in 2016	-0.0718	1	0	0	1	0.866
POLITICO Europe	How the UN migration pact got trolled	0.0917	25	14	14	305	0.854
EURACTIV	Frontex: Africans made two thirds of illegal arrivals to EU in 2017	-0.1231	1	2	0	33	0.854
Euronews English	Italy demands sanctions on Malta following another row over migrant boats between the two countries	0.0812	133	13	61	585	0.852
Euronews English	Cyprus struggles with increased influx of asylum seekers	-0.0490	53	8	4	197	0.828
POLITICO Europe	Morocco rules out building EU offshore asylum centers	0.0155	25	2	14	113	0.818
Euronews English	Germany 'made most attempts' at sending asylum seekers to other EU states	0.1150	134	28	57	810	0.813
Euronews English	European Commission says NGO migrant rescue vessels must obey the law	0.1203	122	10	44	458	0.812
Euronews English	EU eyes more money to process refugees in Africa	0.1060	34	0	10	74	0.810
POLITICO Europe	Frontex chief: Increase in migrants reaching EU via Spain	0.0686	19	13	2	235	0.806
EURACTIV	Migrant deaths in the Mediterranean decreased in 2018, UN agency says	0.0022	2	1	0	18	0.802
Euronews English	Nearly a thousand migrants returned to Libya by coastguard	-0.2087	193	16	33	581	0.799
Euronews English	Little EU solidarity on migration	0.0801	13	0	1	17	0.790

Table 15: 10 the least fitted articles in Topic 6

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
Euronews English	On the ground at the Bosnian-Croatian border where migrant tensions are rising	-0.1102	184	35	77	1052	0.279

Figure 11 a,b. Sentiment and interactivity for the Topic 6

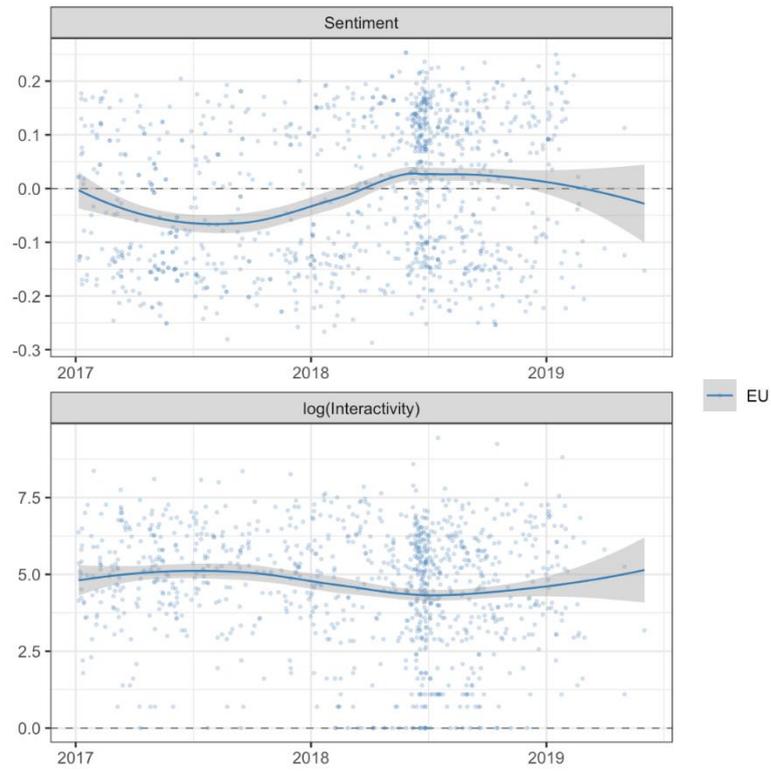
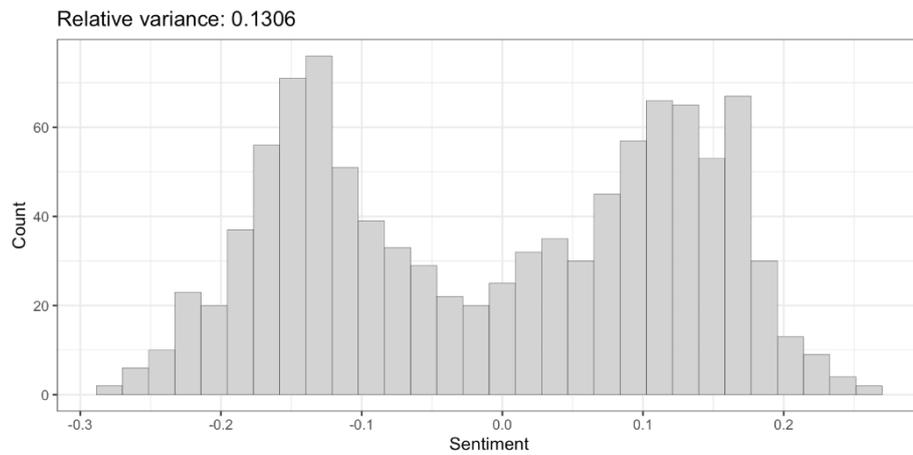


Figure 12. Distribution of sentiment in the Topic 6.



As we can see in figure 14, the distribution of the sentiment in topic 6, migrations is clearly bimodal, with one group of articles with positive sentiment and the other with negative sentiment.

Topic 15. Refugees and global south

Table 16: 15 the best fitting articles for topic 15

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
POLITICO Europe	Young heroes with invisible scars	-0.2174	2	4	0	66	0.898
Euronews English	Somaliland on the brink of famine, warn NGOs	0.0131	157	23	8	557	0.865
Euronews English	Targeting education for refugee children in Kenya	-0.1599	35	6	0	131	0.864

Euronews English	The twin evils of famine and war kill children in South Sudan	-0.2814	163	32	6	699	0.839
Euronews English	The battle to provide healthcare in Central African Republic	-0.1893	3769	269	39	8229	0.813
POLITICO Europe	Looking for Havel's legacy	0.1960	31	7	1	147	0.778
Euronews English	Bangladesh: Rohingya migrant crisis	-0.1482	17	3	10	105	0.777
Euronews English	NGOs scale up humanitarian aid for Rohingya refugees	-0.1448	20	5	3	112	0.776
Euronews English	Alarm raised about cholera epidemic in Yemen	-0.1729	8	2	0	40	0.776
Euronews English	Malian refugees in Mauritania: education in an emergency context	-0.0821	6155	610	83	16247	0.762
Euronews English	Mauritania: a 360° view from the Mbera refugee camp	0.0869	4	1	0	20	0.737
Euronews English	Thousands of DRC children at immediate risk of starvation, UN warns	-0.2177	34	13	10	282	0.733
Euronews English	Displaced in Ethiopia:	-0.1391	11	1	0	27	0.719
Euronews English	One of the 'worst droughts in living memory' - two years without rain in Baidoa, Somalia is affecting millions	-0.1874	58	9	2	210	0.716
Euronews English	Denmark to force 'ghetto' children to learn about democracy and Christmas	0.0582	525	51	46	1525	0.716

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
	Venezuela's government in denial as doctors warn of return to medical 'stone age'	-0.1945	44	12	1	240	0.27
	Merkel calls for stricter immigration rules and faster deportation	-0.2996	339	39	90	1323	0.27
	Toilets can be a matter of life and death. This city takes that seriously.	0.1601	39	7	5	171	0.27
	Which EU country is most popular for Europe's undergraduates?	0.0923	40	5	12	168	0.27
	Balancing work and children	0.1390	9	3	2	65	0.27

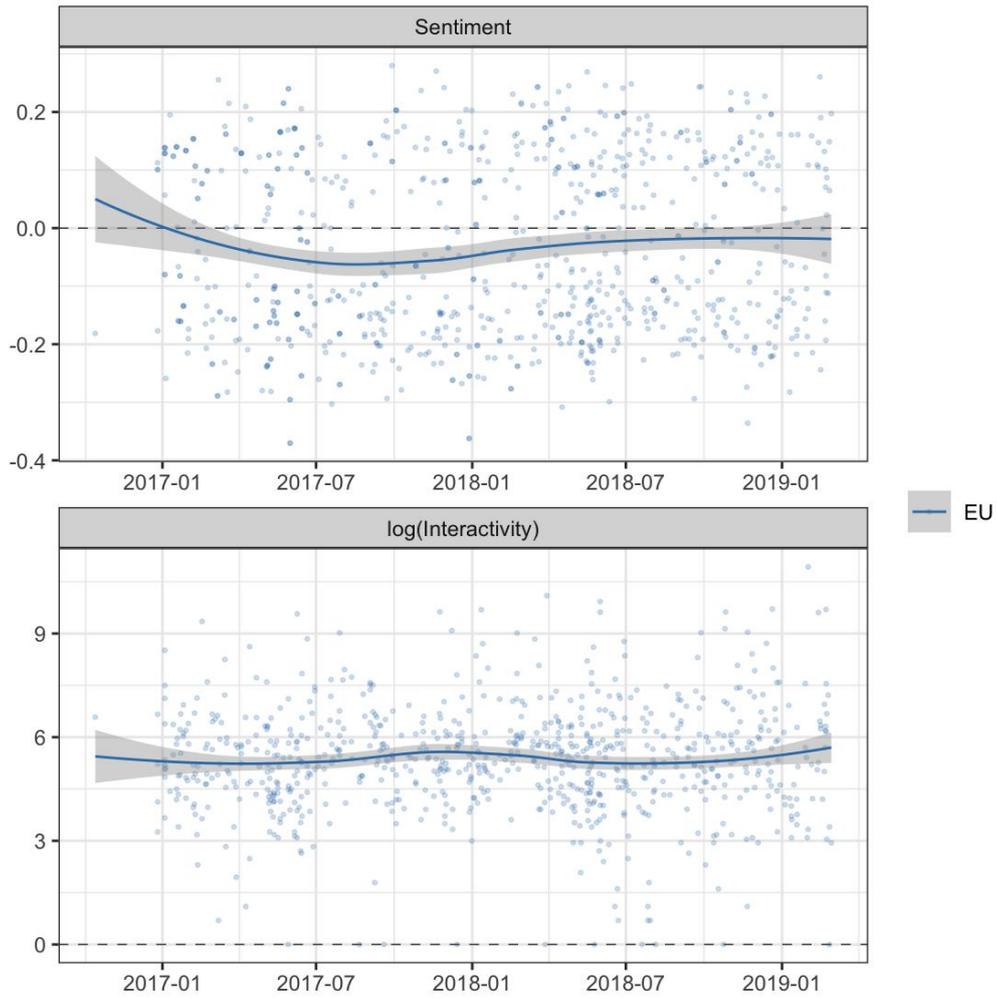
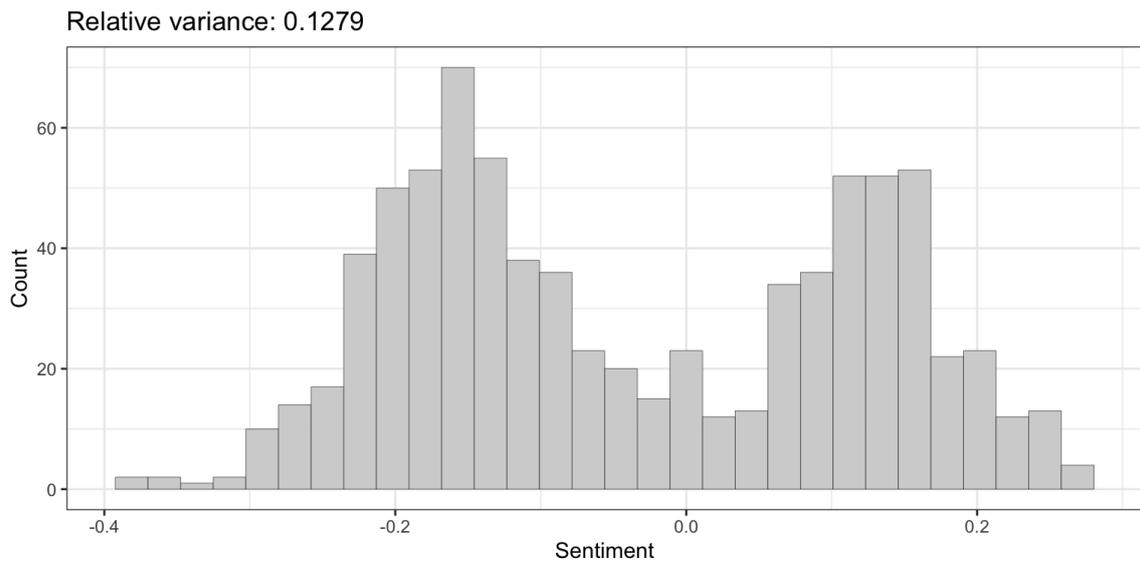


Figure 18. Distribution of sentiment in articles for topic 16



As we can see in Figure 18, the distribution of sentiment for topic 16 is bimodal, with the prevalence of negative sentiment.

Topic 3. EU regulations

Table 17: 15 the best fitting articles in topic 3

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
The Parliament Magazine	Alcoholic beverage labels: Time for harmony	0.1478	1	0	0	1	0.992
POLITICO Europe	World's cartoonists on this week's events	0.0000	75	19	12	427	0.986
The Parliament Magazine	Will PNR be a humiliating déjà vu for EU policymakers?	-0.1726	0	0	0	0	0.935
The Parliament Magazine	EU Commission report on alcohol labelling is yet another delaying tactic	0.0945	0	0	0	0	0.929
The Parliament Magazine	Alcohol labelling rules must address key specificities	-0.0575	0	0	0	0	0.922
The Parliament Magazine	What will new voluntary EU alcohol labelling rules look like?	0.1128	0	0	0	0	0.914
The Parliament Magazine	Alcohol labelling should be about what's best for consumers - not industry	0.1287	0	0	0	0	0.908
EURACTIV	EU health chief not satisfied with industry's alcohol labelling proposal	0.0580	1	0	1	5	0.873
The Parliament Magazine	Current file on alcohol labelling is only tinkering around the edges	-0.1400	0	0	0	0	0.863
The Parliament Magazine	EU must not adopt a 'one-size-fits-all' approach to alcohol labelling	0.1374	0	0	0	0	0.809
The Parliament Magazine	Alcohol labelling: Is the Commission putting business interests before EU citizens' health?	0.1926	1	0	0	1	0.848
The Parliament Magazine	New EU firearms directive is a balanced proposal	-0.2618	3	0	4	19	0.839
EURACTIV	EU Commission seeks 'quick fix' solution on alcohol labelling proposals	0.1012	0	0	0	0	0.796
The Parliament Magazine	ePrivacy regulation is a good first step	0.1455	0	0	0	0	0.791
The Parliament Magazine	The time for ePrivacy is now	0.1675	1	0	0	1	0.732
EURACTIV	Wine industry wants its products' labelling in the next CAP	0.0941	0	0	0	0	0.7728

Table 18: 10 the worst fitting articles in topic 3.

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
--------	-------	-----------	-------	--------	----------	-----	-----

Figure 17 a,b. Change in time of sentiment and interactivity for topic 3

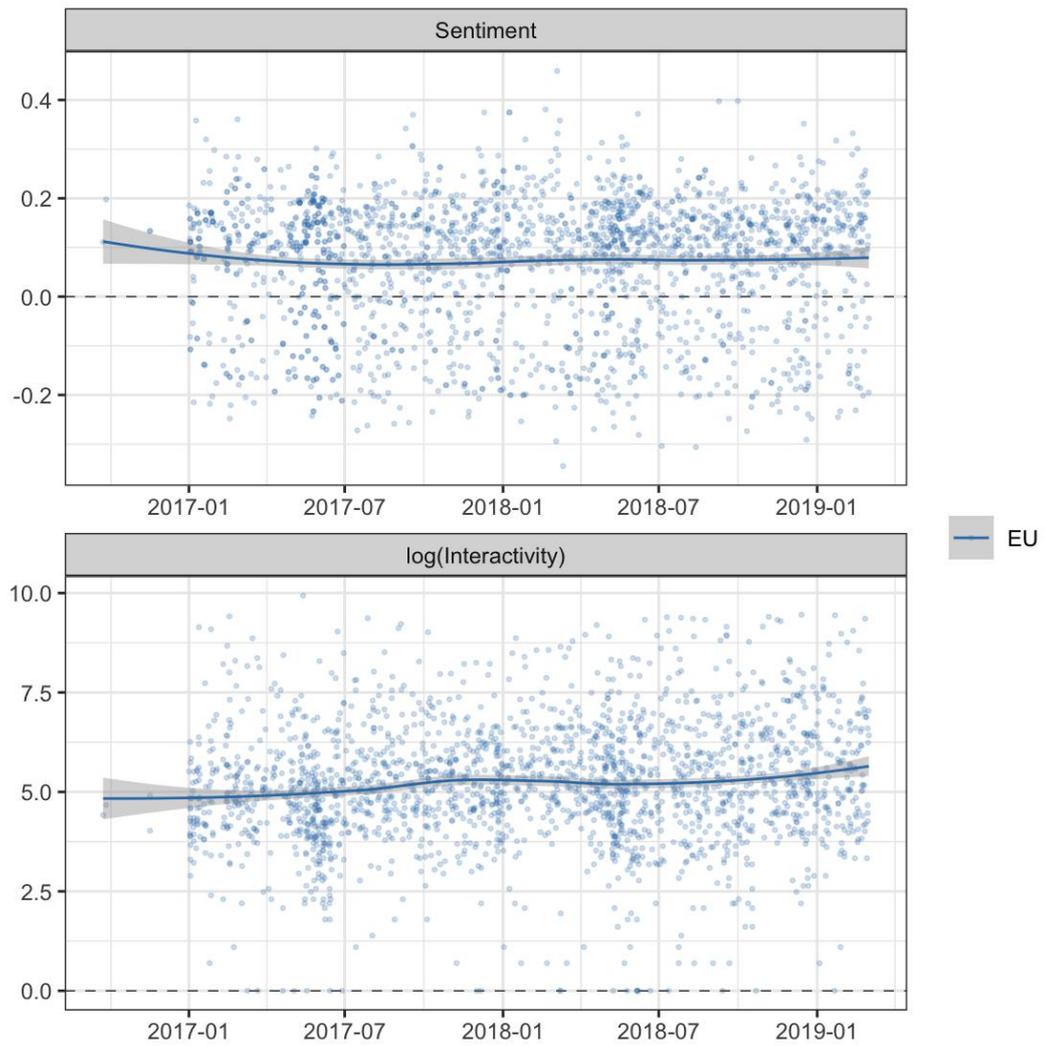
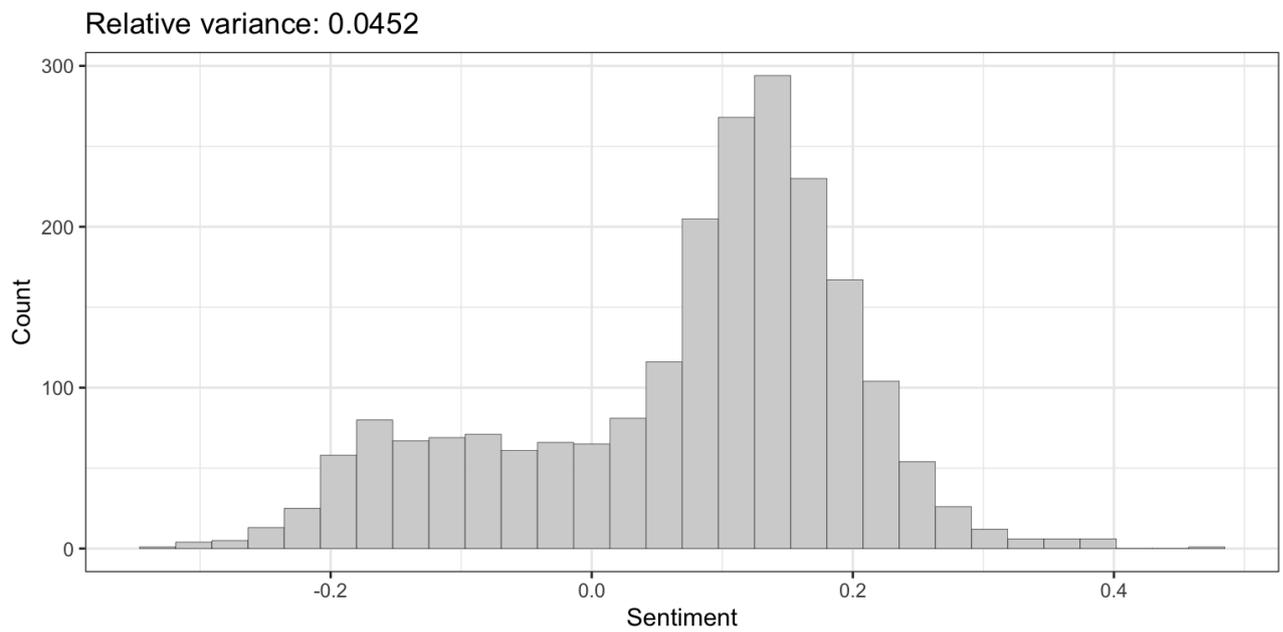


Figure 18. Distribution of sentiment in articles for topic 3



As we can see in Figure 22 the dominant in articles on this topic, sentiment is positive

Topic 7. Political economy

Table 19: 15 the best fitting articles in topic 7

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
POLITICO Europe	Why are France's Yellow Jackets so angry?	-0.2785	32	12	4	240	0.896
Euronews English	Watch: Why are tensions running high in Venezuela?	-0.0567	16	3	8	96	0.887
Euronews English	Eurozone unemployment falls to post-financial crisis low	-0.2216	8	0	0	8	0.842
POLITICO Europe	The Turkey-Venezuela mutual admiration society	-0.2232	11	4	3	87	0.817
Euronews English	Central Bank steps in as Turkish lira tumbles	-0.1557	41	13	17	317	0.797
Euronews English	Eurozone economy starts 2017 robustly	0.1149	20	2	0	52	0.795
POLITICO Europe	French PM admits 'mistakes' with Yellow Jackets	-0.0859	6	0	2	14	0.793
Euronews English	UK and US growth forecasts down	0.1390	33	6	10	169	0.779
Euronews English	Moody's keeps Portuguese rating stable, upbeat on recovery	0.1667	55	4	2	127	0.779
Euronews English	Turkish economy faces crisis ahead of election	-0.0131	83	9	31	351	0.774
Euronews English	Eurozone inflation heads up	0.0971	11	1	1	31	0.770
Euronews English	Maduro rejects Europe ultimatum	0.1121	33	4	9	133	0.745
Euronews English	Spanish retail sales slip as inflation rises	0.0824	8	1	0	24	0.742
Euronews English	Eurozone unemployment holds steady	-0.0373	34	4	1	102	0.732
POLITICO Europe	Yellow Jackets rattle Macron's foot soldiers	-0.0910	2	0	0	2	0.732

Table 20: 10 the worst fitting articles in topic 7

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
Euronews English	Talking tactics - eurozone shake-up plan	-0.0625	15	2	2	55	0.273
Euronews English	Gilets jaunes protests to have 'important impact' on French economy: Bruno Le Maire	-0.2512	35	4	11	143	0.273
Euronews English	Thousands protest in Madrid demanding higher pension increase	0.0427	97	16	13	405	0.274

Figure 20: Sentiment analysis and interactivity for topic 7

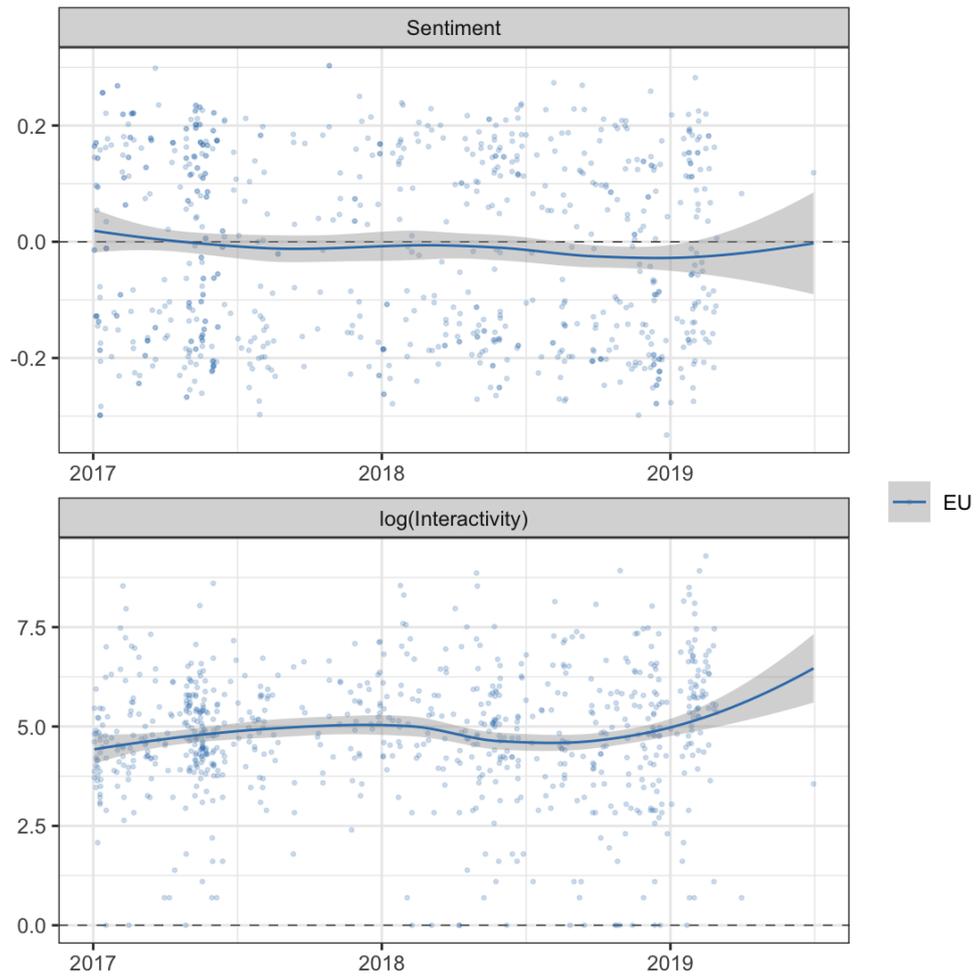
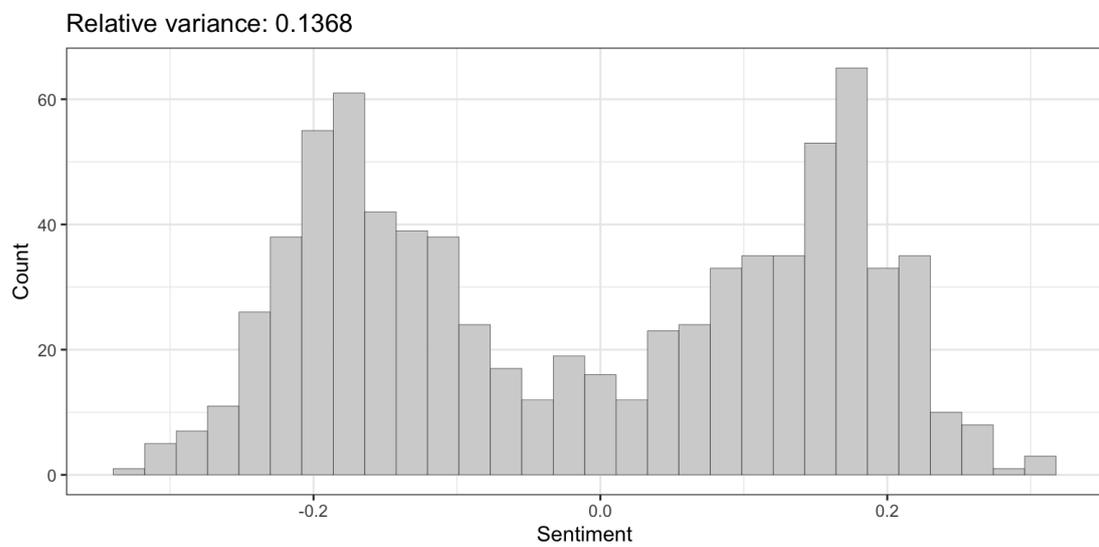


Figure 21: Distribution of sentiment in articles in topic 7



As we can see, the distribution of sentiment is bimodal, with approximately equal number of articles characterized by positive and negative sentiment.

Topic 10. Science

Table 21: 15 the best fitting articles in topic 10

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
Euronews English	Cassini-Huygens's extraordinary journey around Saturn	0.1040	20	2	0	52	0.998
Euronews English	NASA launches mission to explore below Mars' surface	0.0638	53	7	1	169	0.893
Euronews English	NASA's Mars mission hopes to find what's inside the red planet	0.0370	39	5	1	123	0.873
Euronews English	NASA launches ice satellite	0.0028	75	19	9	415	0.851
Euronews English	BepiColombo sets off to solve the mysteries of Mercury	0.0787	2824	350	42	8592	0.838
Euronews English	Planning a holiday to Mars? Don't forget to check the weather!	0.0498	48	8	3	188	0.824
Euronews English	End of space mission for Cassini	0.1238	56	1	1	76	0.815
Euronews English	Facing the furnace: BepiColombo mission to visit Mercury	0.0665	20	4	0	84	0.814
Euronews English	Aeolus: Forecast is good for first ever wind-measuring satellite	0.0723	2261	133	14	4445	0.810
Euronews English	NASA's InSight lander: What's next?	-0.0066	54	7	5	186	0.809
Euronews English	Voyager probe 40th anniversary	0.0657	82	12	6	298	0.805
Euronews English	Here comes the sun: NASA's journey to the sun starts next year	0.1014	68	14	8	324	0.796
Euronews English	Surfing scientists and algae hunters use Sentinel-3 to study coastline	0.0803	1759	34	10	2343	0.793
Euronews English	Europe's space mission to Mercury begins	0.0583	85	12	4	293	0.792
Euronews English	Is this the best solar eclipse photo bomb?	0.0763	96	10	4	272	0.783
Euronews English	Life beyond earth? Try Saturn's moon Enceladus	0.059043 6	73	8	3	213	0.774

Table 22: 10 articles worst fitting articles in topic 10

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
Euronews English	Wind energy takes a toll on birds, but now there's help	0.1129	32	10	2	200	0.276

Figure 23 a,b . Sentiment analysis and interactivity for topic 10

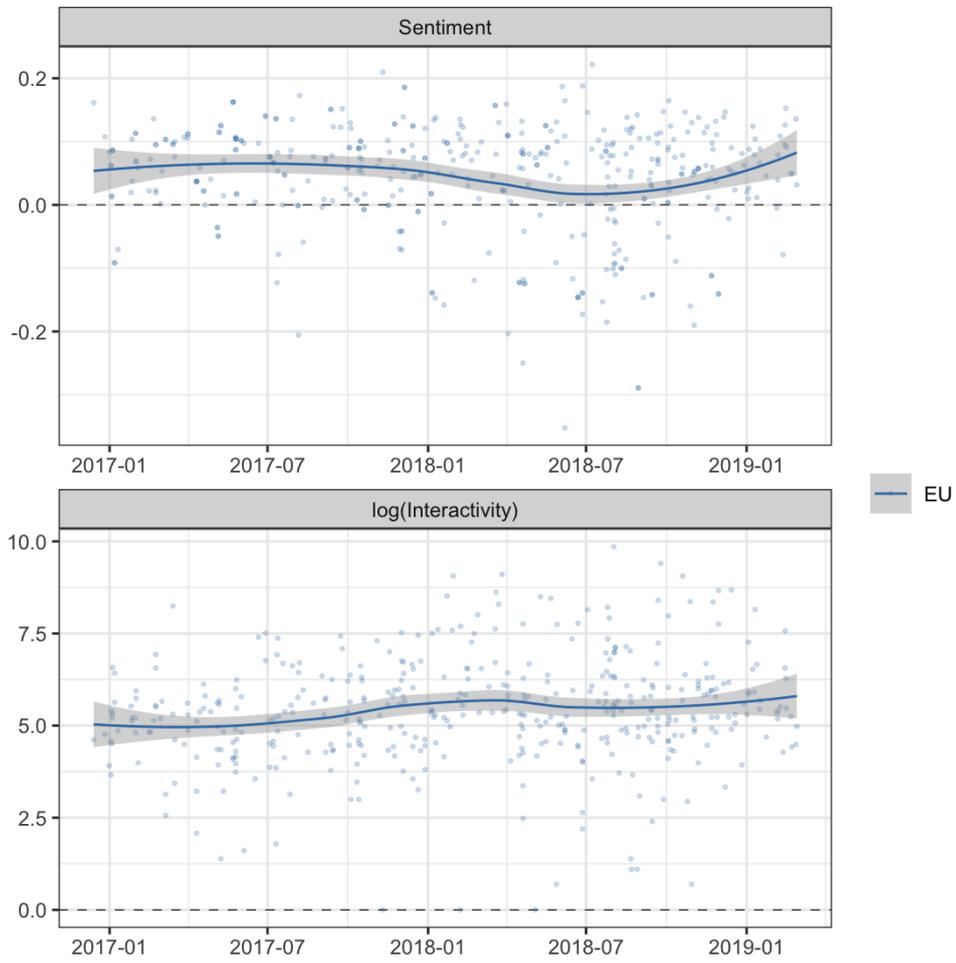
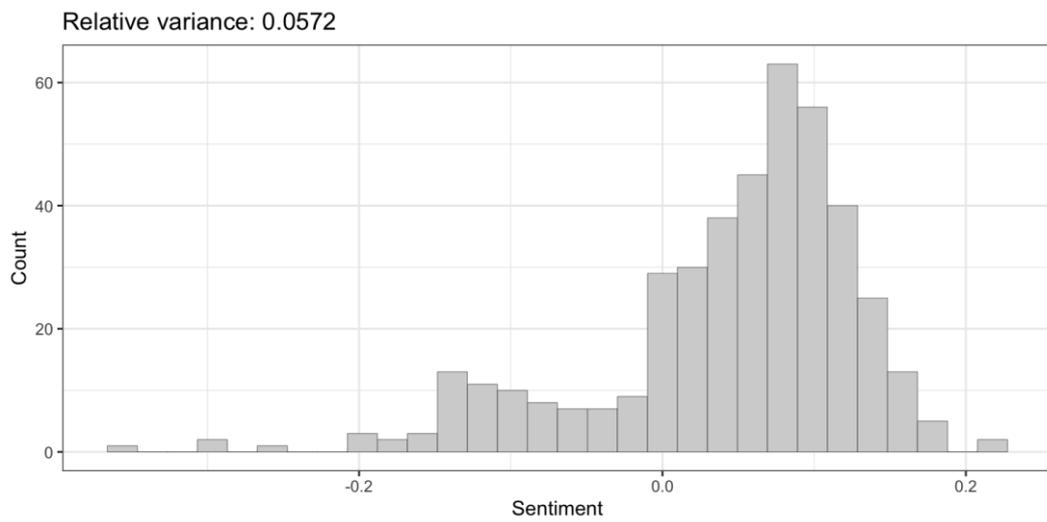


Table 24. Distribution of sentiment in topic 10



As we can see in the figure above, the sentiment in topic 10 is positive with relatively few articles characterized by negative sentiment.

Topic 14. Public health & technology

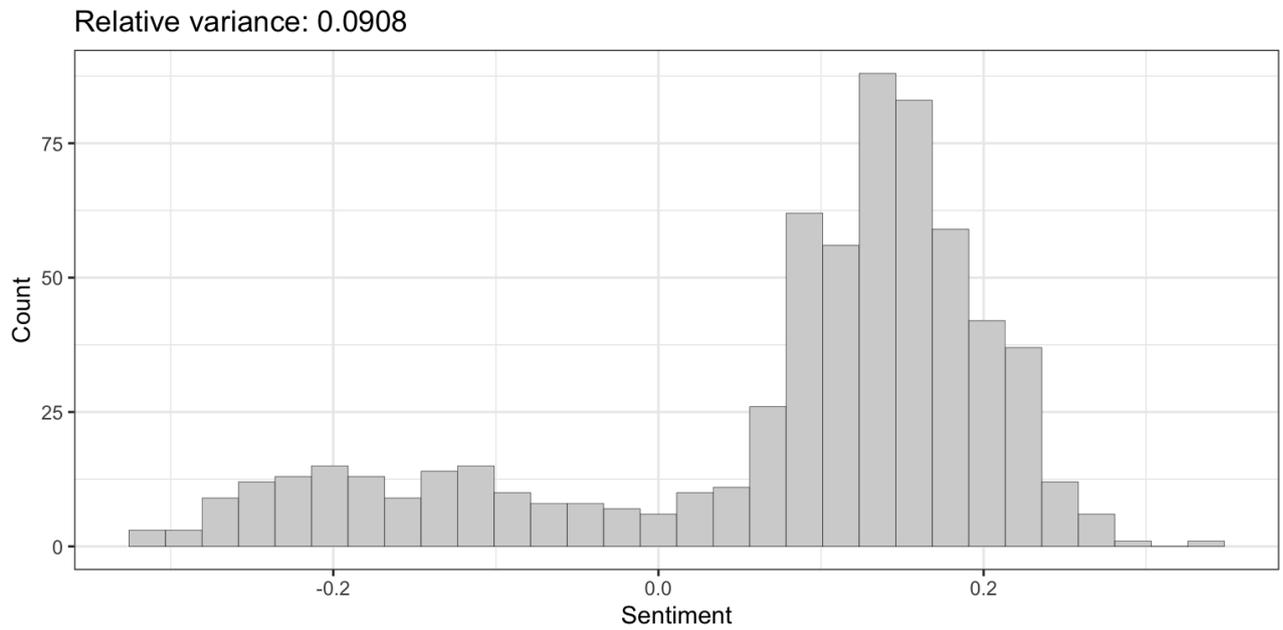
Table 23: 15 the best fitting articles in topic 14

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
The Parliament Magazine	eCigarettes are part of the solution to one of Europe's most serious health problems: tobacco smoking	0.2095	1	0	0	1	0.990
The Parliament Magazine	The EU cannot let big pharma call the shots on research	0.1621	0	2	0	32	0.954
The Parliament Magazine	Robotics will shape the future of work	0.2267	0	0	0	0	0.954
The Parliament Magazine	Cancer care: Recognising the nursing contribution	-0.3228	1	0	0	1	0.936
The Parliament Magazine	It's time to extend EU's digital transformation to health and care	0.1908	0	0	0	0	0.890
EURACTIV	Life-course vaccination – leave no one behind	0.2421	2	0	0	2	0.861
EURACTIV	Combination vaccines: more protection for all with less injections	0.2009	1	0	0	1	0.852
EURACTIV	Are you ready for the future of work?	0.0237	1	1	0	17	0.843
The Parliament Magazine	How the tech revolution is transforming the future of work	0.1402	2	1	0	18	0.824
EURACTIV	Data is the new blood	0.2199	0	0	1	4	0.823
The Parliament Magazine	Europe needs a vaccination action plan	-0.1850	0	0	0	0	0.807
The Parliament Magazine	Policymakers must tackle work-related stress	-0.2545	0	0	0	0	0.787
The Parliament Magazine	Rare disease research likely to lose out from Brexit	0.1692	0	0	0	0	0.786
EURACTIV	Successful data innovation starts with enthusiastic curiosity, rather than technology	0.2158	1	1	0	17	0.782
Euronews English	View: Is vaping safer than smoking cigarettes?	0.1498	61	16	13	369	0.781

Table 24 10 worst fitting articles in topic 14

Source	Title	Sentiment	Likes	Shares	Comments	INI	Fit
Euronews English	How can protecting intellectual property rights protect your brand?	0.1904	6	0	0	6	0.272
Euronews English	Estonia 'first country in world able to function without physical land'	-0.1128	325	57	30	1357	0.272

Figure 26 a,b. Sentiment and interactivity in topic 14

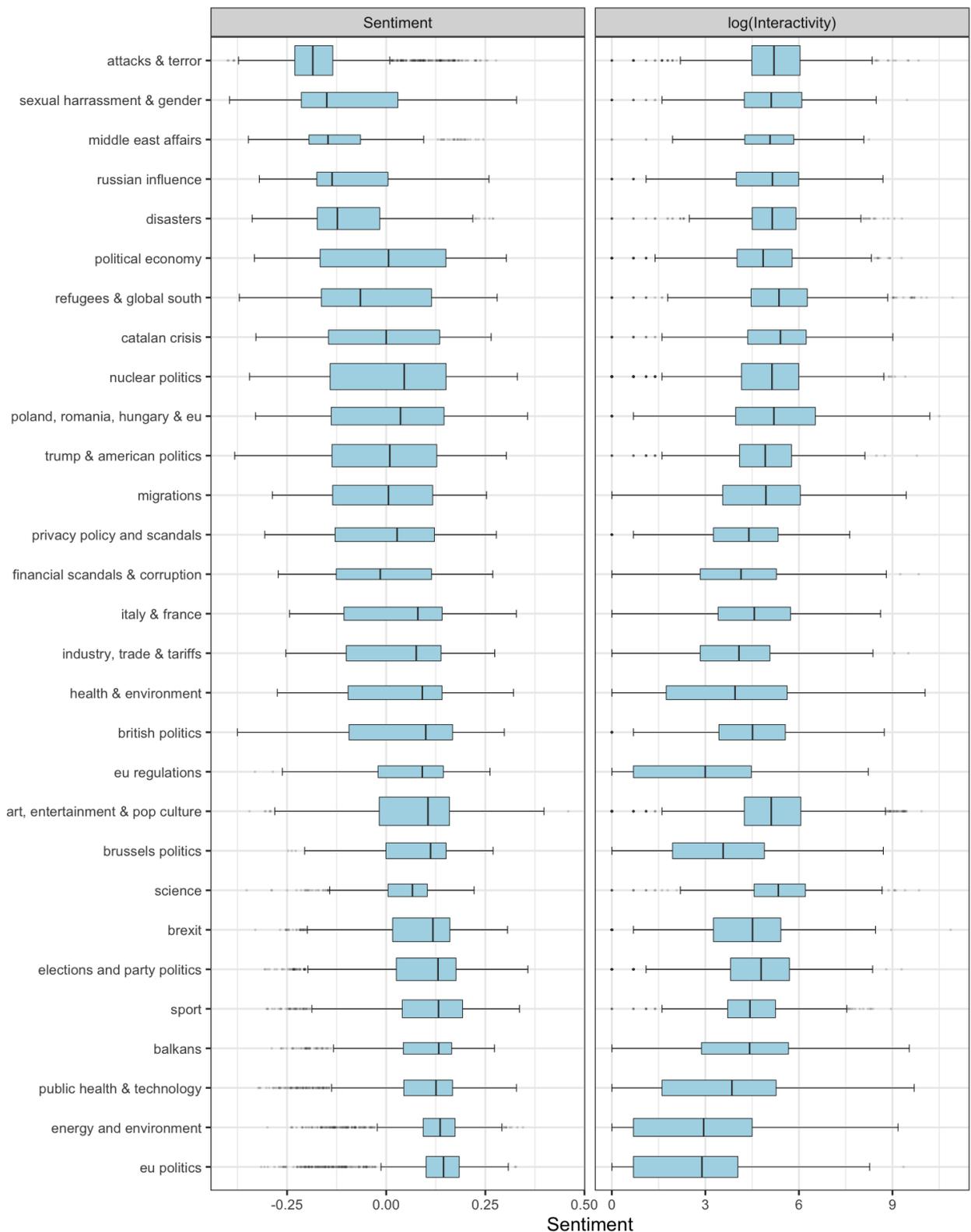


As we can see in the figure above the dominant sentiment in topic 14 is definitely positive.

Analyses across topics

In this section we present the overall results for all topics. The plot below shows distributions of sentiment and interactivity over all topics sorted by the 1st quartile (ascending order). It is worth noting a negative correlation between sentiment and interactivity (topics with low sentiment tend to have higher interactivity.)

Figure 27 Sentiment and interactivity for all topics

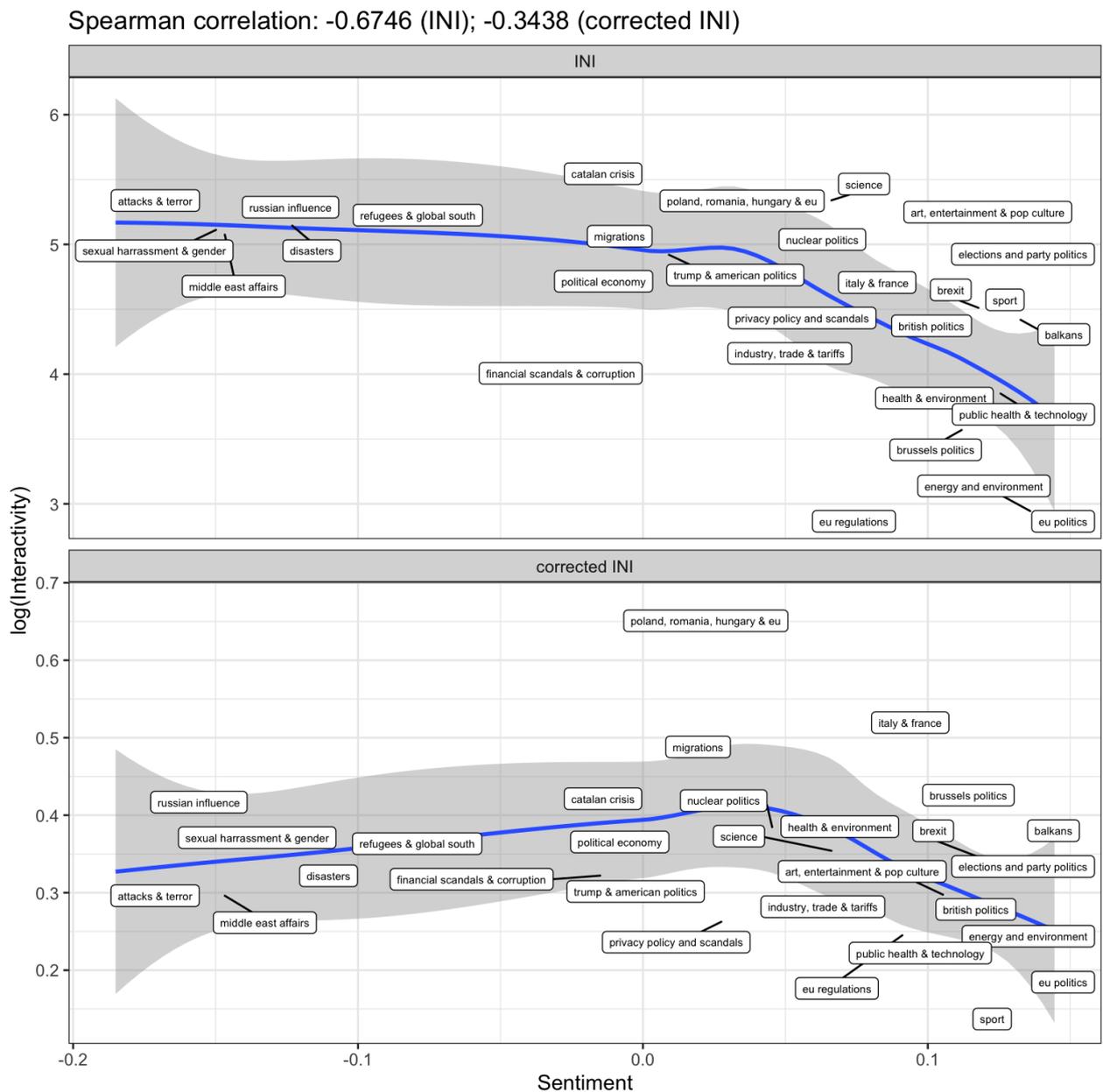


As we can see, the most positive sentiment is associated with EU politics, energy and environment, public health and technology, Balkans and sport. The most negative: attacks and terror, sexual harassment and gender, Middle East affairs, Russian Influence, disasters, refugees and global south.

As can be seen in the table on the right, the negative topics elicit much higher interactivity than the positive topics.

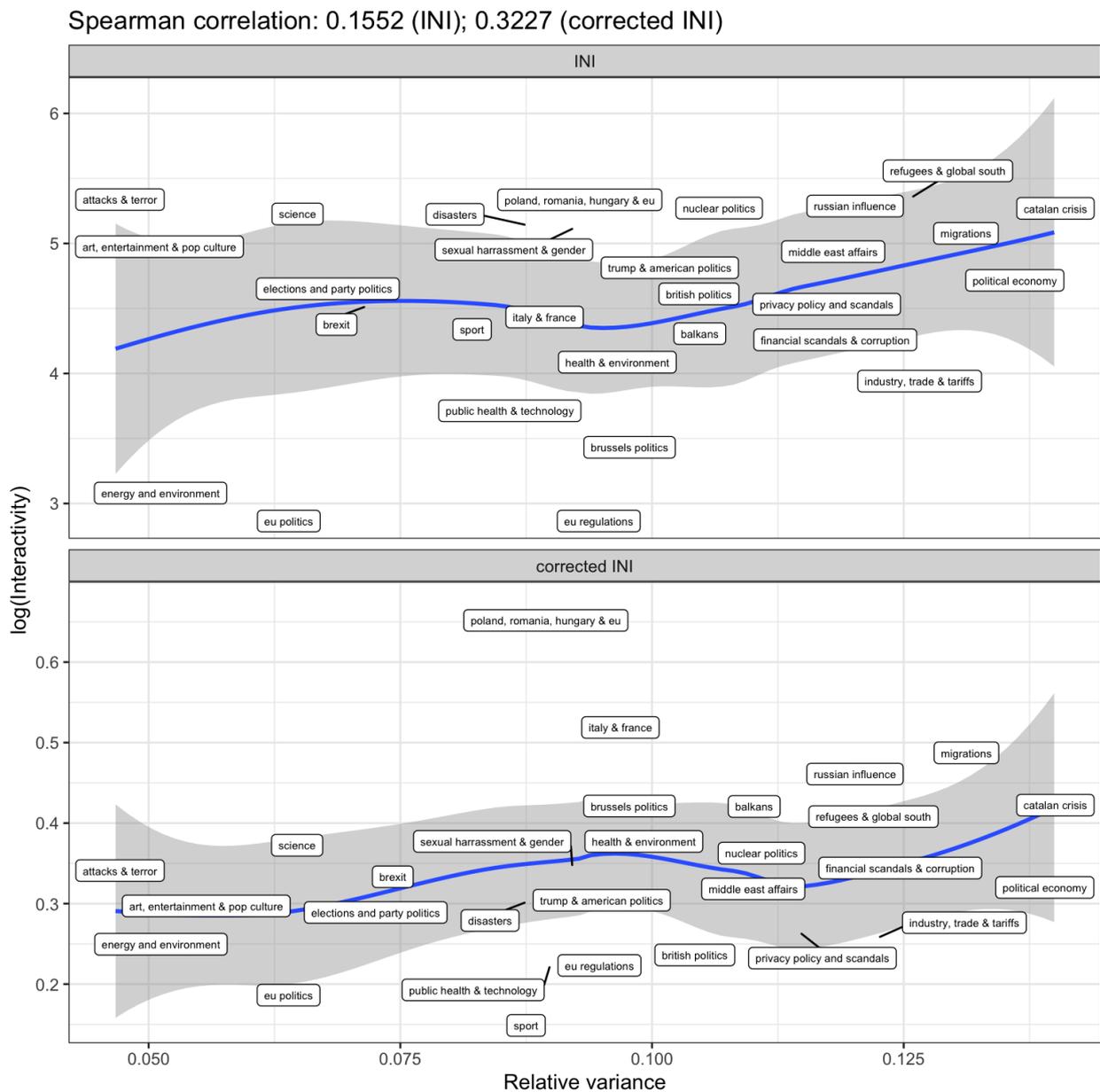
Correlograms of sentiment and INI / corrected INI provide better view of the above-mentioned correlation. It is clear that for both measures topics with lower average sentiment raise more user activity. The trend is particularly sharp in the tail of low sentiment topics. The correlation is particularly high for standard INI, but is retained also in the case of corrected INI. It means that topics generating negative sentiment attract more attention and stronger reaction.

Figure 28.



A similar trend can be observed also in the case of interactivity and polarization. More polarizing topics have more user activity. Both of these results are congruent with literature pointing to polarizing effects typical for the new media ecosystem. However, based on the data, it is not clear whether the effect can be attributed purely to social media or rather to the media ecosystem consisting of social media and traditional publishers as such, since our data consists of content produced by traditional publishers and social media reactions.

Figure 29 a,b. Relative variance and interactivity



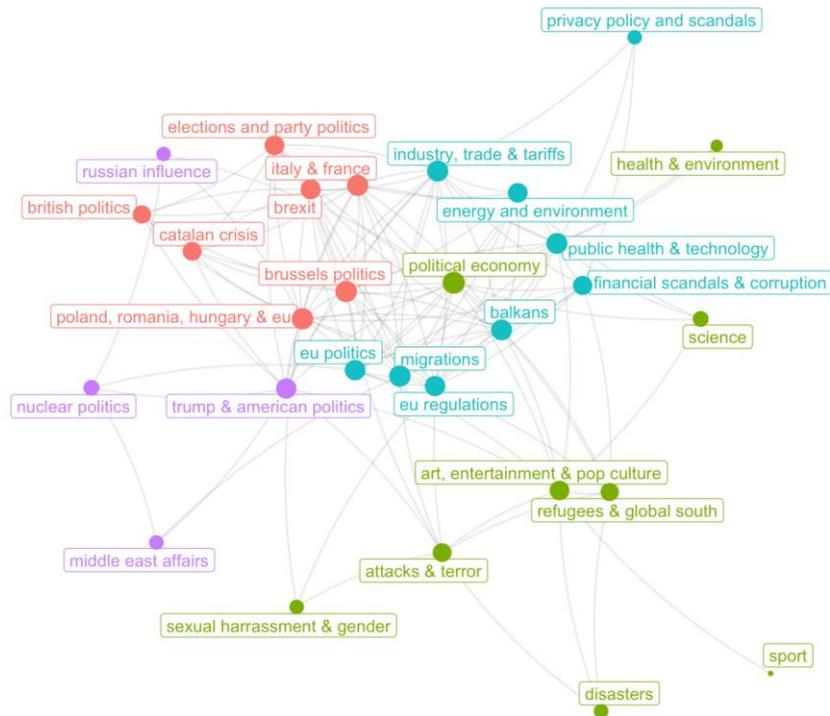
As we can see in the graph 29 a and especially b above, the more polarized is opinion on a topic the more interactivity, so the more attention it attracts and it generates more reactions.

It is also important to note that in both cases we found correlations between semantic sentiment (derived from linguistic construction of articles) and user activity. Hence, we see a clear connection between semantics of produced content (output of traditional publishers) and user reactions. This perhaps suggests that both sides play a role in supposed increase of polarity in the new media landscape. Articles with negative sentiment and articles on more polarized issue attract more attention, what increases negativity and polarization.

The plot below shows frequencies of topics in time. It should be noted that the estimates for the beginning of 2016 may be in some cases unstable due to data sparsity.

Semantic web of topics

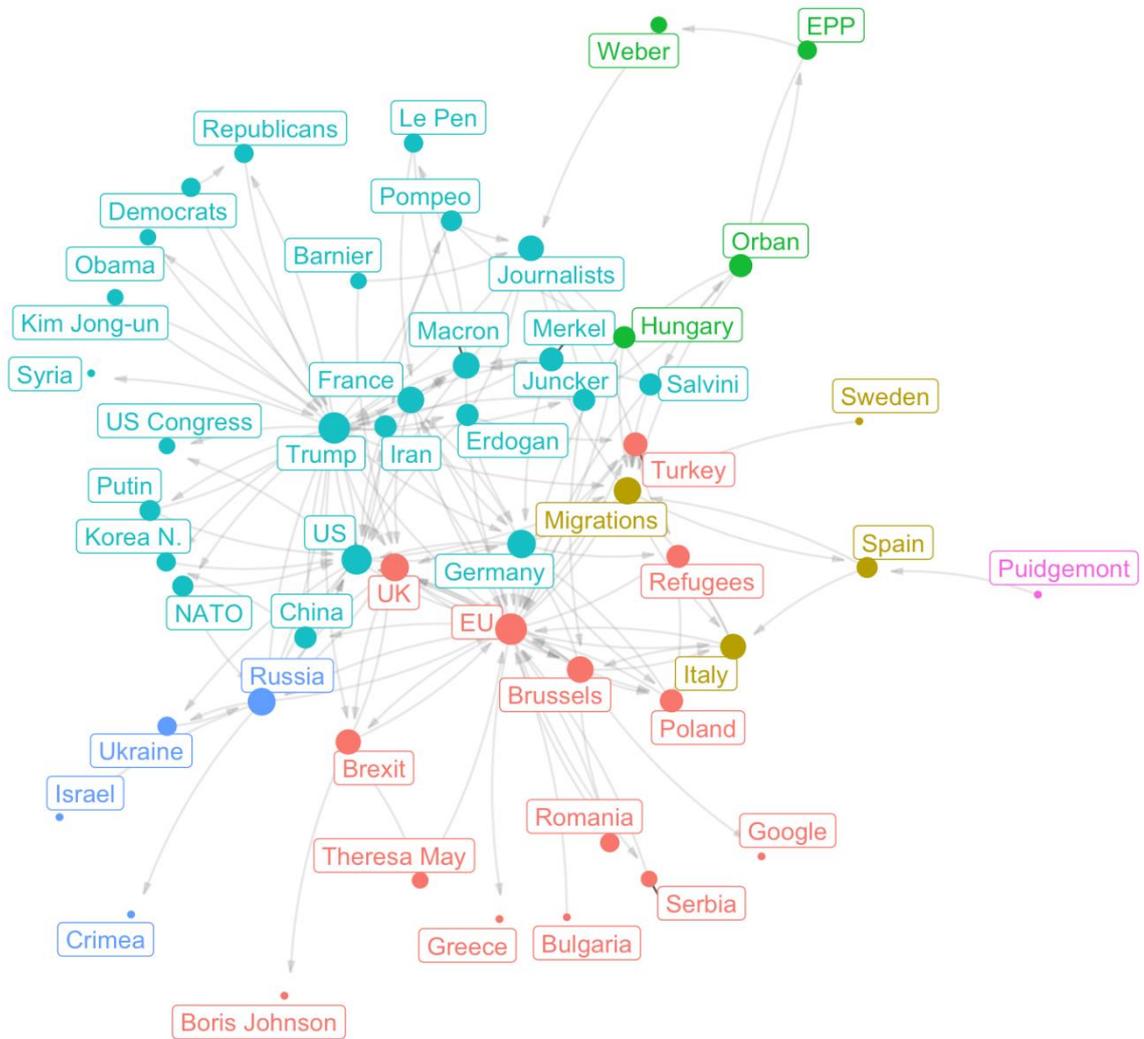
In the case of topic network we also see four main clusters that are relatively similar to the clusters in the web of words.



Narrative analysis

In this section we focus on the narrative analysis based on relations between particular actors detected using Named Entity Recognition (NER) and Part-of-Speech (POS) tagging. For the sake of clarity of visualization only 25% of the most frequent relations are shown and the graphs are limited to the largest connected components. Edges point from active agents (subjects) to passive agents (objects) connected by an action (verb). Such triples are called subject-verb-objects (SVO) triplets. The analysis in this part is limited to 57 most frequent unique entities of various types (public persons, political entities, nation states, market agents and events). This is due to the need of manual correction and grouping of entities automatically detected by the deep neural network provided by *Spacy* (the natural language processing library we used).

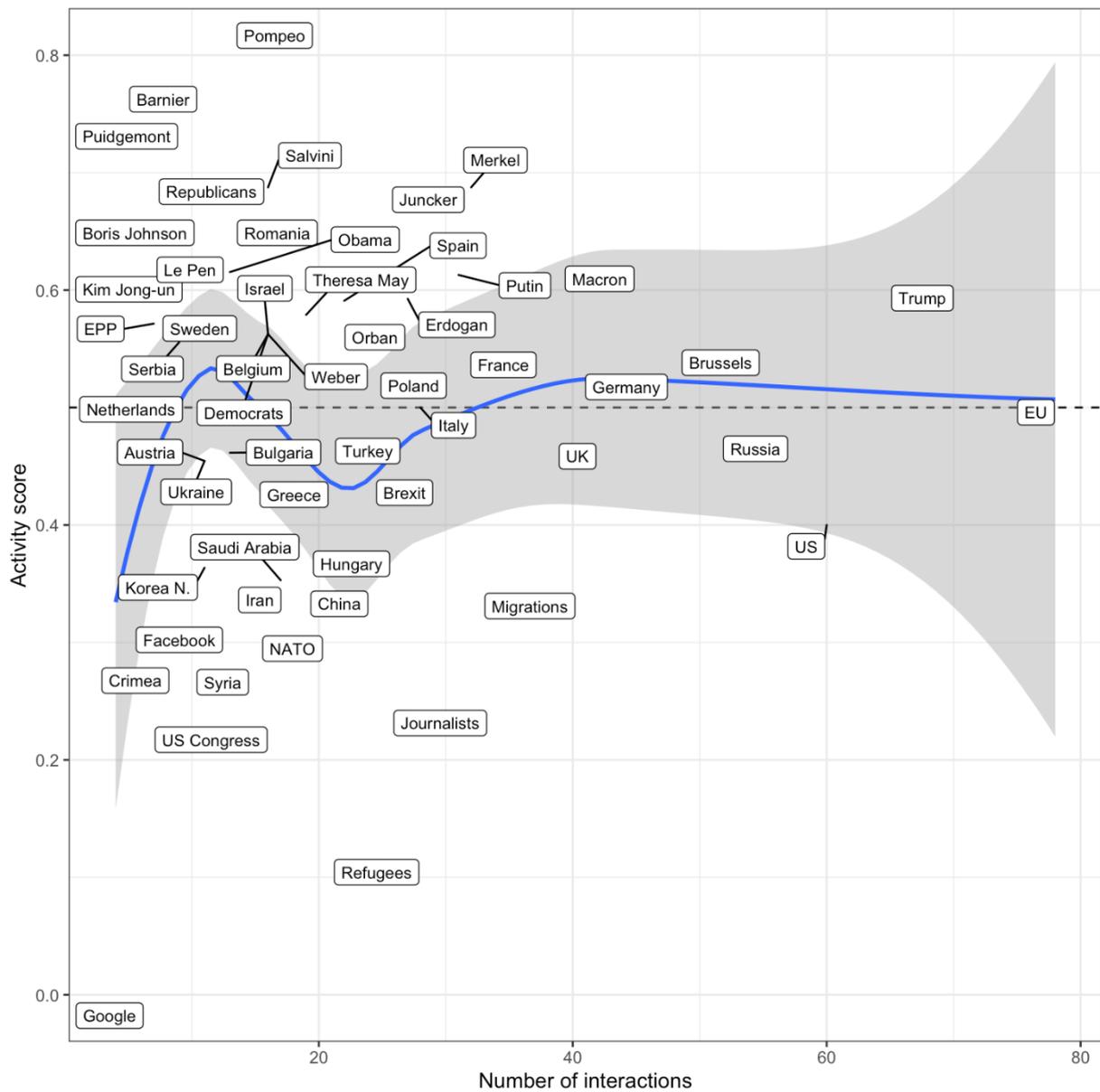
Figure 32. Network of entities and their relation to topics



Activity scores (out-degrees / in-degrees)

Based on the main SVO graph we can derive activity scores for agents and relate them to total volumes of interactions. Activity score is defined as the fraction of out-degrees of a node (the number of interactions with a node being an active part) to the total number of interactions.

Figure 33.

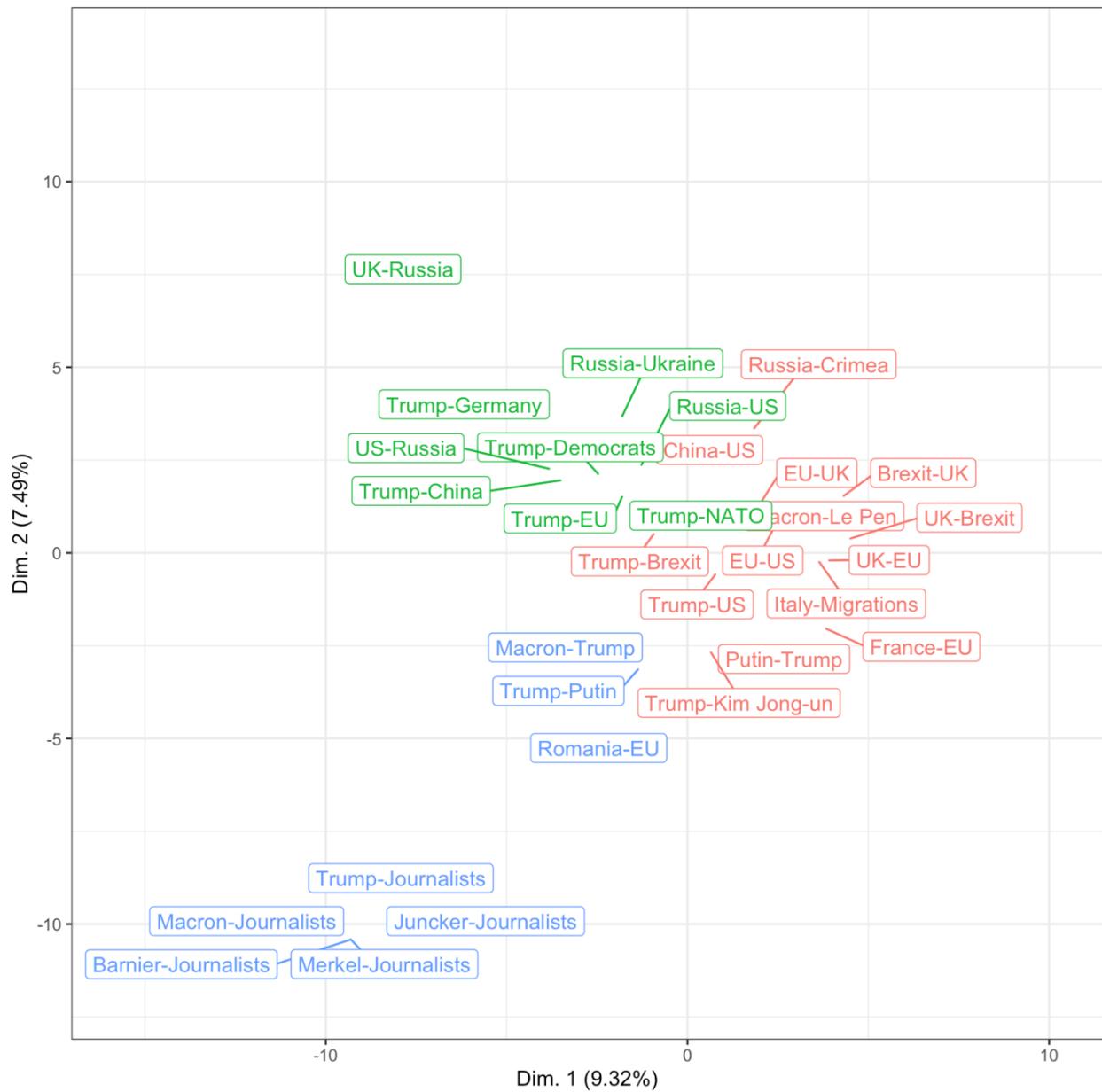


Semantic space of dyads

In this section we derive a semantic space of relations between directed dyads of actors. To do so we use the 300D vector space representation of verbs (actions) connecting actors. For each dyad we compute an average verb vector and based on this each dyad can be positioned in the 300-dimensional semantic vector space. Then, we reduce the space to a 2-dimensional plane based using PCA. This allows us to derive a crude spatial representation of semantics of actions connecting the actors. For the sake of clarity of visualization we limit the results to 5% of the most frequent dyads.

The plot below shows the resulting semantic space. Colors encode cluster assignments based on the K-means method.

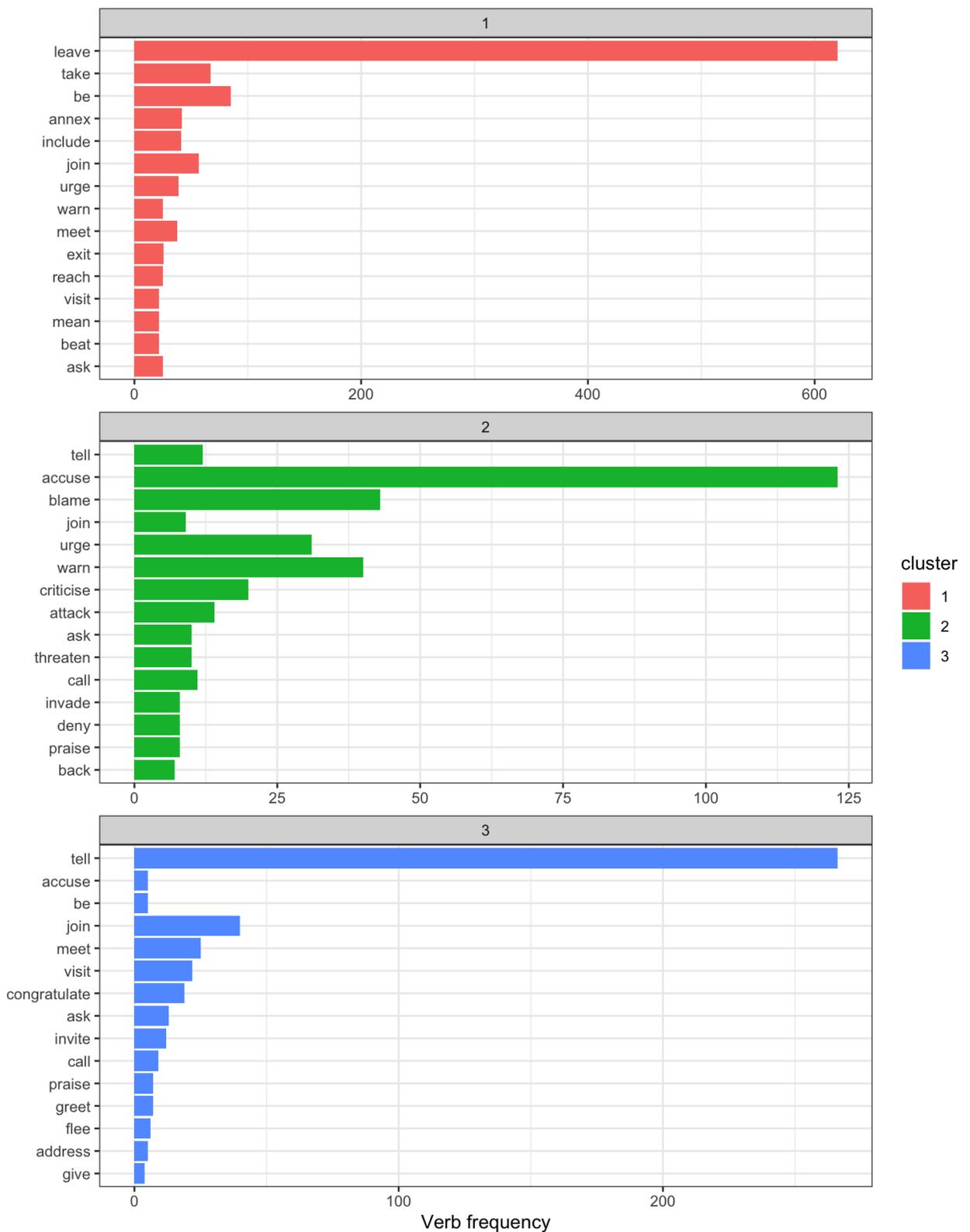
Figure 34. Semantic space of dyads



Most typical verbs in the semantic clusters

To facilitate interpretation below we show most frequent verbs for each cluster (the color coding is analogous to the one on the semantic space plot).

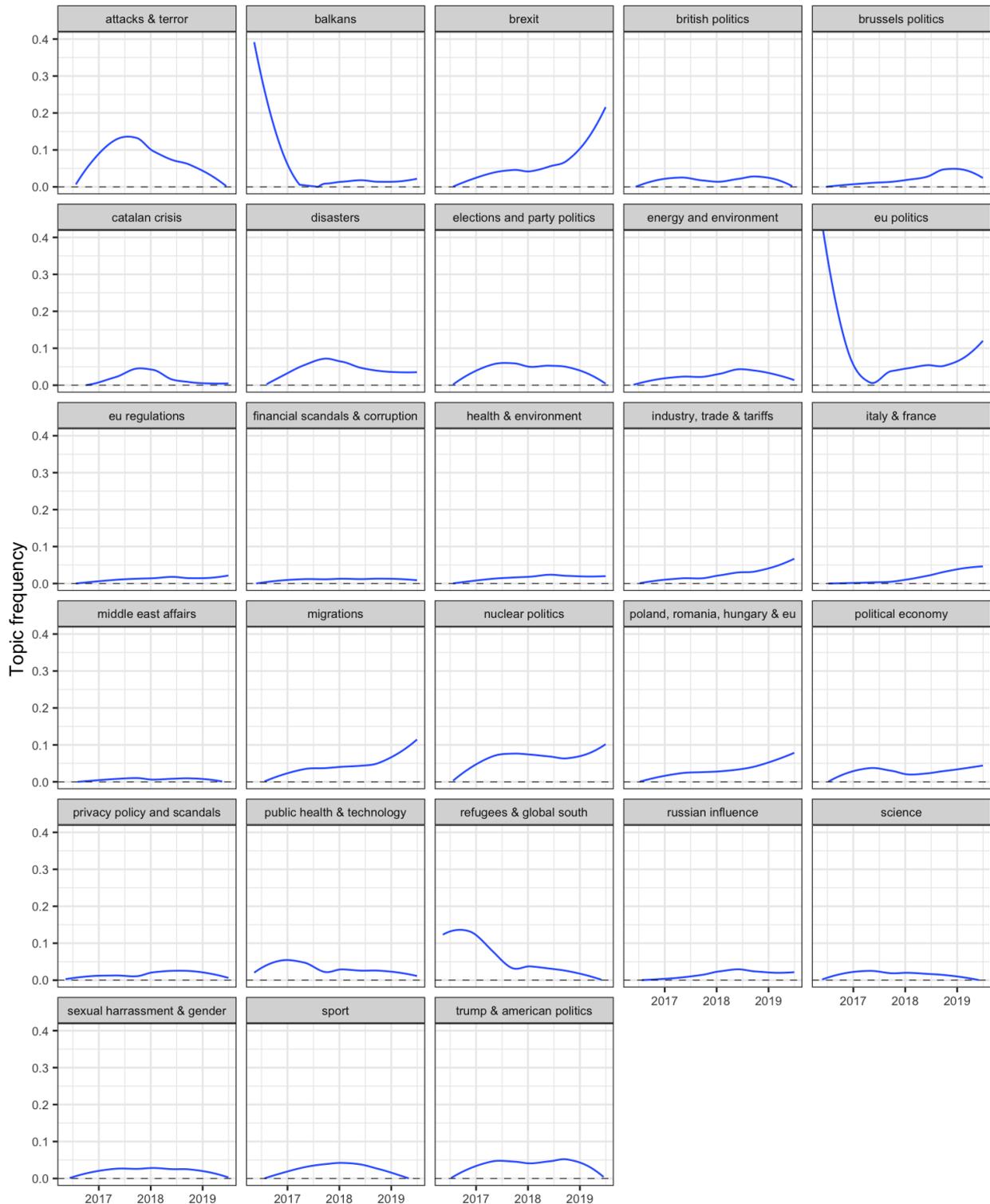
Figure 35.



To see the details, we visualized the relationship between sentiment and interactivity for all individual topics based on the estimated model. Topics are colored based on their median INI (groups are based on deciles of the distribution over all topics).

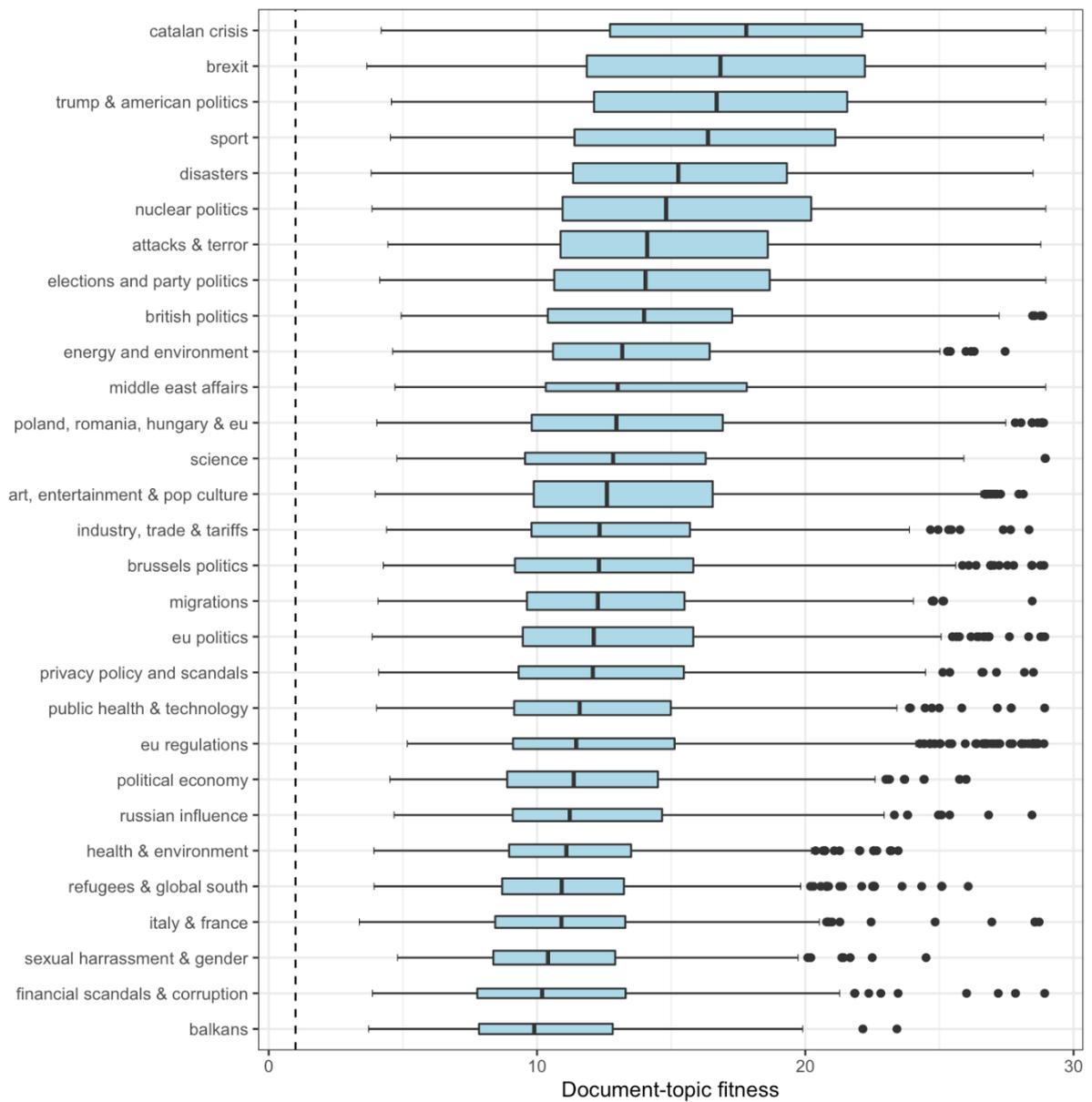
This subsection presents results on the quality of topic model in terms of topic coherence. Here, by topic coherence we mean simply a distribution over goodness-of-fit measures of documents assigned to a topic. Goodness-of-fit of a document is the probability (as estimated in the model) that it belongs to a particular topic. It should be noted in our case (30 topics) a naive prior over such probabilities is a uniform distributions. This means that even probability of (0.3) gives a 10-times improvement in comparison to pure guessing.

Figure 36.



The plot below shows the distributions of document goodness-of-fit. It can be interpreted as the coherence of topic.

Figure 37. Goodness of fit of articles to topics



As it can be seen in the graph above, the highest coherence is in topic *Catalán crisis* and Brexit, the lowest in financial scandals and corruption and Balkans. This indicates that articles on these topics are quite varied.

Chapter 7 – AN ARTISTIC APPROACH

Counteracting the Spread of Online Disinformation

7.1 Introduction

Through the SMART project, we have tried to map the state of the art of those disciplines, techniques, methods, tools and approaches which can be used to study the complex, networked phenomena that take place in the public infosphere, to gain better understanding about how these influence social, political, economic and psychological dimensions of society.

While all of these techniques, methods and tools are dedicated to researchers and experts, we felt the need to go beyond this objective, since we believe that the most radical innovation which was brought about by digital transformation is the possibility and opportunity for society to observe and understand itself together with scientists and researchers. This is an epistemological transformation which potentially has important implications for the possibility to implement open, democratic, inclusive societies.

This kind of issue (“who observes who”) is also a typically artistic one.

Therefore, the possibility to bring science into society through creative processes that can trigger empathy, collaboration, participation, shared reflections and inclusive construction bringing together experts, policy makers and the general public seemed something very important to us, with a view of addressing not only the technical implications of technologies and of networked communication, but also their psychological ones, which are crucial in understanding how people perceive and position themselves in communication and interactive ecosystems, and to collaboratively work in shaping these interactions and communications in constructive ways, so that a more democratic and self-aware scenario can be achieved.

This is why we experimented with the creation of the To Fake or not To Fake data-theatre performance.

For the SMART project we produced a social media observatory that has been used to monitor public communication about the European Elections in 2019 on Twitter and on a selection of major news outlets, in multiple languages. The data coming from the observatory is completely anonymised and/or aggregated, so that it is possible for other researchers to safely visualize, download and use it from the project's website.

The data includes social media posts and web articles dealing with EU Elections 2019, as well as the social interactions which they triggered (shares, appreciations, comments, mentions) and, thus define a dynamic network of relations which evolve through time.

7.2 The Data Theatre Performance

The cognitive complexity of the phenomena we face in today's globalised and interconnected world is more than a single person can process.

Moreover, the only way we can experience global issues like climate change, migration, energy, poverty, as well as the networked phenomena of the infosphere, is through enormous amounts of interconnected data.

Sensemaking is about the exploration of what's “real”, and of our relationship to “real”.

When “real” becomes interconnected with data of the types, quantities and interconnections that we're talking about, a peculiar phenomenon takes place: multiple computational agents constantly extract data from human behaviour and from the environment, and use it to recognize and classify recurring patterns that, when used to make business, political, economic and communicational decisions, change reality.

If Wittgenstein talks about the fact that “we are spoken by language”, today we might say that “we are thought by data and Artificial Intelligence”. Through this sentence, we mean that these computational, data-based agents who constantly and continuously think for us and about us transform reality: notifications, alerts, suggestions, recommendations, filters, and all those other elements of our digital lives which shape the way in which we perceive and understand reality, are also transforming our relationships, the information we access and knowledge we acquire.

In the SMART project we investigated how this type of phenomena influence our capacity to understand the public communication and information spheres, and how these influences undermine the civil, democratic processes, for example through disinformation.

We wanted to bring these understanding to the general public, in ways that would trigger participatory, constructive and inclusive action.

For this purpose, we designed publications, made conferences, created browser plugins that integrated visualizations and representations of this data and its meanings into peoples' everyday lives.

But, still, they were expert tools, most of the time adopted by those people that already have a specific sensibility for these types of phenomena.

We decided that we had to try to work on culture, to create bridges between our biological and psychological limits and the evolution produced by our adoption of technology.

What was needed, in our opinion, was a form of pragmatic knowledge that could go beyond the logos, that is language, thought and calculation.

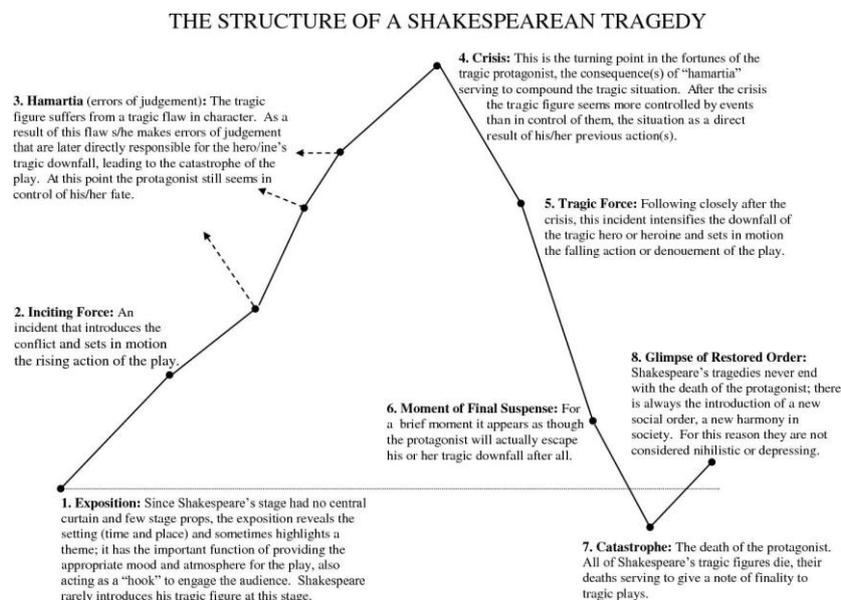
Tragedy, both in our daily lives and in artistic processes, includes everything which cannot be recomposed by logos. Tragedy is irreducible and, thus, it is our access door to complex phenomena.

For this reason, we decided to produce a data-driven tragedy, in the form of a data-theatre performance, called "To Fake or not To Fake" (TFonTF), through which we tried to answer a series of questions:

- Is it possible to use data-based creative practices (such as data visualization, data representation, data-theatre, information aesthetics, data-driven performance and more) in a narrative context in order to create usable knowledge about disinformation phenomena, which is not only usable and accessible to large audiences, but also desirable and engaging?
- How is it possible to create virtuous loops that include scientists, technologists, artists and designers in this process, so that the result has both the formal characteristics of science and the psychological and social effectiveness of art?
- How is it possible to use this kind of tools to trigger wider processes, including education, collaboration and action, aimed at achieving a more civil, just, inclusive society?

7.3 To Fake or not To Fake

TFonTF is a data-theatre performance with a Shakespearean structure.



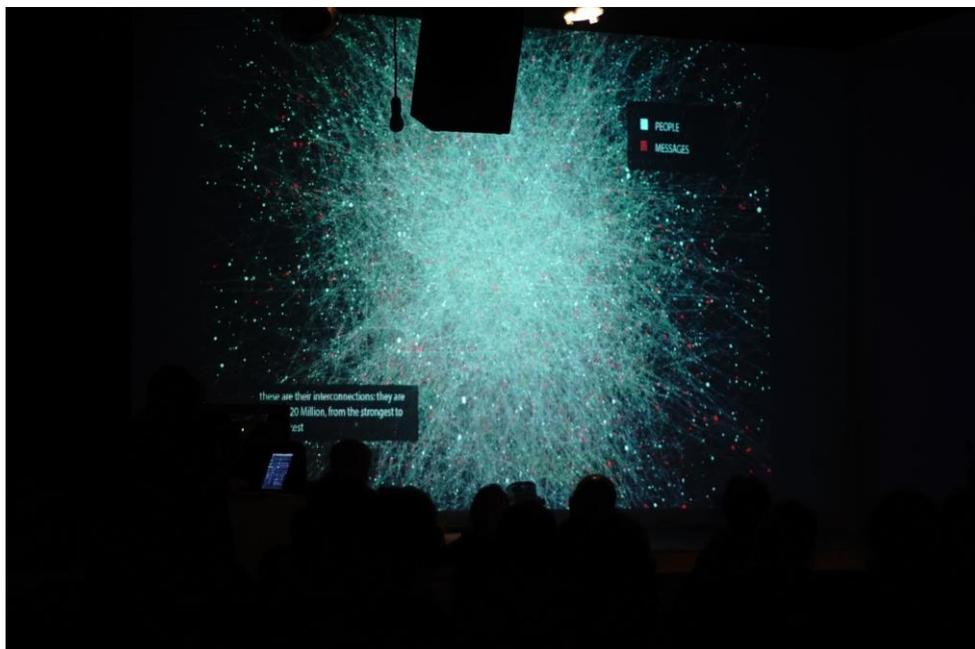
It narrates the life on Infosphere, a god-like entity which animates the global communication ecosystem. The turning point of the tragedy is a disinformation phenomenon, as it unfolds online. Using her god-like powers of communication, Infosphere freezes the communication ecosystem and pulls one of its participating subjects out of it, to let him experience disinformation from an ecosystemic point of view, to understand the complex network

phenomena that constitute disinformation. As actions of fact-checking and debunking start being deployed, the tragedy unfolds, through polarization and violence. In the end, it is uncovered that the subject which was pulled out of the communication ecosystem was a malicious BOT.

In the play, actors move in front of a digital scenography which is generated by visualizing the data harvested through the Observatory.

The juxtaposition and interplay of actors and visualizations, together with the fact that “one explains the other” has an effect which makes the narrative concrete, palpable, without being didascallic and banal: the performative element takes the visualizations and bring them to the audience with urgency and constructive critical approaches, while the scenographic visualizations and the soundtrack create an ambience of wonder, tension and mystery that is progressively resolved as the performance goes by, further enhancing the psychological and cognitive effects.

The narrative was constructed by following the Shakespearean structure, to be able to achieve the psychological engagement which we felt was needed to bring people into the condition of attention and concentration which is crucial for them to experience and understand complex phenomena.



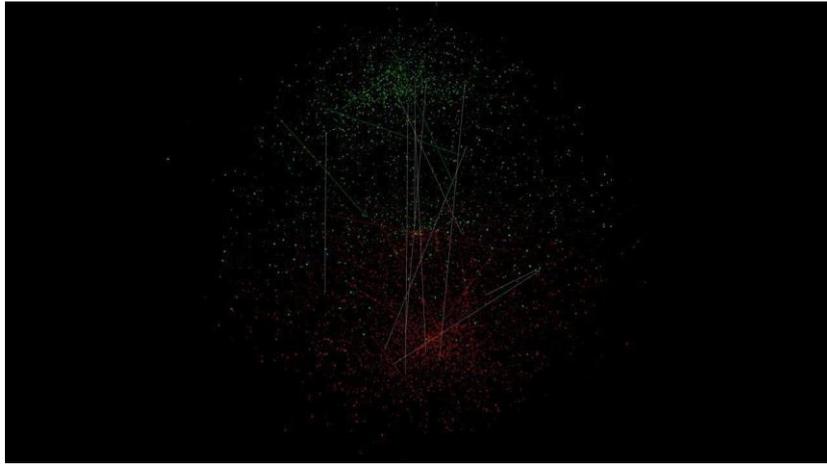
Cautious work was performed on language, to abandon jargon usually associated with these themes, and to adopt terms, figures and metaphors which are suitable for broader, generalistic audiences.

We have also dedicated attention to children and younger generations, by working in constructive ways (instead of the merely dystopian ones which are often featured in articles) and using multimedia to confront with reduced attention frames.

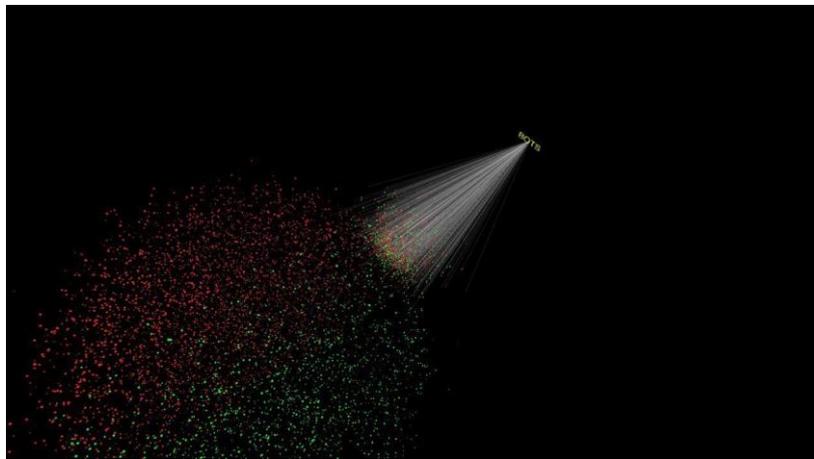
Additionally, particular attention was dedicated to staying on a very concrete, practical and pragmatic level. These issues are usually dealt with in very abstract, conceptual and intellectual ways: this is suitable for researchers and experts, who are able to bridge the concepts towards the pragmatic reality of communication phenomena, but not for general publics, who most of the time lack the scientific approaches, cultures, skills and competences to create these bridges.

For example, to remain on practical, concrete domains, the focus was placed on the emergence of a single disinformation phenomena (for example HERE), which dealt with the notion that, at the time, Google might have been algorithmically penalizing conservative sites and social media profiles in their search results.

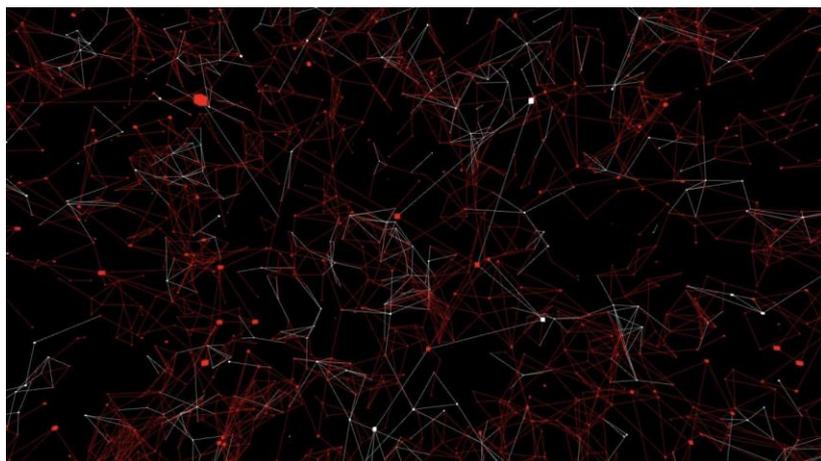
During the performance, the audience has the opportunity to experience disinformation in front of their eyes, in an accessible way.



For example, in the image above, they can see how two polarised opinions confront each other, while the actors use the narrative structure to describe the characteristics of the communication phenomenon.



In this other image, the audience sees how bots can be recognised and studied, to understand their influence in communication.



In this other image, people can see before their eyes how disinformation spreads: nodes and edges of the graph become red when people are exposed to it, and highlighted borders which embrace entire communities appear as reinforcement scenarios take place.

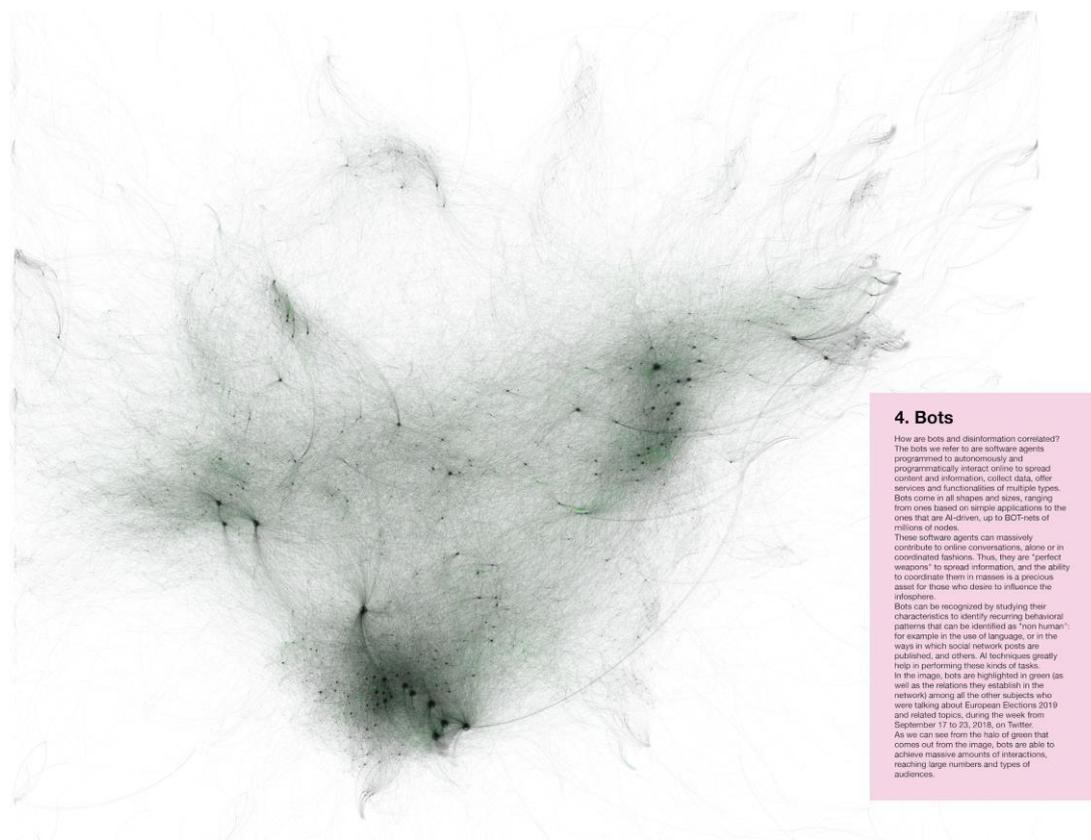
All of these visualizations are based on the actions performed by the SMART consortium regarding the analysis of the state of the art of social media studies in relation to their impact on society. The consortium selected the most important scientific publication in the field, and tested the tools, methods and approaches they proposed, obtaining a reusable, updatable toolkit which can be used to further study disinformation phenomena.

This toolkit was also used to generate the data about how disinformation spreads, the role of BOTs, memetics, the different types of contagion, and more. This data was visualised, for example by creating animations of how these phenomena evolved in time, thus producing the scenographic elements of the performance.

The exhibit

The cultural program also included a data-driven exhibit.

Some of the most interesting phenomena were isolated and represented in beautiful large prints which are suitable for gallery exhibition, or for events, festivals and more.



For example, the image above was the one dedicated to the role of BOTs.

The exhibit was always accompanied with reference materials (for example the consortium's reports) and with the indications about when/how to participate to the education initiatives.

The education programme

The performance and the exhibit always go together with an education programme, whose objective is to use the emotional and cognitive relationship formed through the artistic processes:

- to suggest the urgency and necessity for people to understand how to confront with disinformation and
- to provide them with the skills, tools and capacities for social collaboration and relation which are needed to confront with the social and psychological impacts of disinformation

As described in the previous sections, even if technical actions like counter-narratives, debunking and fact checking produce tangible results, they are basically unable to deal with the psychological effects which derive from living in

such a complex and problematic communication and information ecosystem. Cultural and aesthetics actions must be taken for this, focusing on people's sensibilities and with their capacity for solidarity and collaboration.

The education programmes directly originate from the artistic ones, and in their capacity to convey scientific concepts.

For example, the visualizations from the exhibit or from the performance, would be used to explain to people how they were generated from public communication, and then used to interactively explore communication phenomena, allowing them to materially experience filter bubbles, community reinforcement processes, the action of BOTs and son on.

Then, a constructivist approach would be used, to enable people to construct their analysis, in their community. A practical example would be chosen, preferably among the more relevant ones for the actual community, and some tools would be used to show people how they could observe the phenomenon, visualize it, and use the findings from the analysis and visualization to socially agree on a communitarian strategy.

7.4 Feedback and Conclusions

TFonTF was performed 2 times, together with its supporting elements (the exhibit and education programme):

- During the winter 2018 edition of the HER She Loves San Lorenzo festival for art, data and artificial intelligence, in Rome
- During the "Countering online disinformation" initiative of the European Commission in Brussels, in January 2019



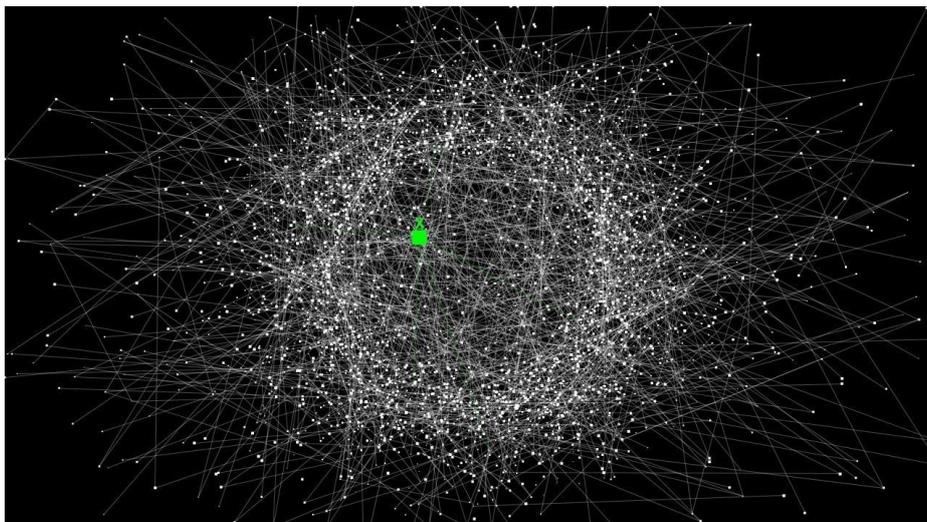
The HER She Loves San Lorenzo festival is a very popular initiative in Rome, in which the whole neighbourhood is transformed into an open-air exhibition space, with local people actively involved. Each year it brings together thousands of visitors around the ways in which art can help us to better understand our relationships with data and AI, during a program that lasts about 1 month.

Around 150 people were present at the TFonTF performance, and a lively discussion followed afterwards, with citizens, students, researchers, activists, journalists, debating for more than an hour, and then contacting us multiple times in the following days to understand how they could participate.

Another moment for public discussion was held at the CNR, with the participation of multiple consortium partners (for example Fiorenza Lipparini, Guido Caldarelli, Gianni Riotta), for an audience of around 90 researchers and journalists.



Then there was the education programme, called "Dati di Borgata", in which 40 people came together to explore how a neighborhood could confront with disinformation, starting from a set of news that had spread locally, and that then resulted fake or ambiguous: they understood how to use social media analysis tools, how to analyse and visualise data, and how to use it in their community to design countermeasures and solutions for solidarity, culture and common life.



In Brussels, the 300+ people that experienced the performance also saw the exhibit and provided us with precious feedbacks.

Considering both events, we can attempt to list the things that worked and the things which did not, if compared to our initial questions.

First, a general feedback that was common between the two types of audiences: the initiative was seen as valuable and effective, especially if connected with initiatives in the education and culture areas.

As for the first one of our research questions, dealing with the possibility to use artistic practices in order to create usable knowledge about complex communication phenomena, the response was different in the two types of audiences (more generalist and communicational in the Rome event, more expert oriented in the Brussels one). While in Brussels some of the people manifested the desire to see a more formal information visualization approach even in the artistic performance (for example introducing graphic elements which are difficult to include in a performance, like references to the dataset, literature etc., which we thought was best to include in the information about the performance and in the education processes), the audience in Rome expressed multiple times how the possibility to focus on a sensorial experience within a narration that helped to make sense, augmented the desire to act and to participate to the other initiatives, like the education ones.

About the second question, concerning the creation of virtuous loops between scientists, artists, designers and the general public, the audience in Rome expressed was very positive on such collaborations and synergies, also due to the fact that one of the principal researchers of the SMART project was there, live, participating to the performance, giving a powerful valence to the action. The idea of using art as a way to bring science and technology into society was widely approved by the vast majority of the audience, and it also incited people to participate to the following initiatives. In Brussels, response was positive, too, but maybe in more future-oriented ways: we were approached multiple times by different researchers and technologists who were really interested in designing new forms of immersive, creative, accessible data visualizations.

The third question was possibly the more problematic of the three: how can this type of action trigger synergies between science, technology, institutions and society?

The answers were many and varied. They ranged from those who were positive about the fact that institutions and governments should provide protection against malicious communication phenomena, to those who expressed the need to work with schools, to those who powerfully questioned current business models connected to communication (along the lines of "in current models, polarization means more money"), up to the more advanced ones, which expressed the desire to have tools and technologies that can be freely used at different levels of society, benefitting both from the infrastructures provided by digital ecosystems, as well as from the possibility for radically distributed models, in which knowledge is produced and used worldwide in interconnective ways.

For all of these reasons, this has been an incredibly rich experience, which can lead to interesting and usable developments about the ways in which society can collaborate with researchers, technologists, companies and institutions, to lead to a more inclusive, just, democratic, constructive communication and information ecosystem.

PART 3 – CONCLUSIONS

Future Areas of Research and Policy Recommendations

Chapter 8 – CONCLUSIONS

8.1 Introduction

The recent revolution in communication technologies has finally made its effects evident not only on the media industry but on the entire fabric of society. Social media are definitely changing not only the way we obtain, consume and disseminate information, but also how we relate to others, exchange and compare different opinions. After a period of enthusiasm for the remarkable possibilities that these new technologies allow, the general advice has moved towards a negative reception of their impact on social dynamics, particularly in relation to the exercise of democracy and individual freedoms. We are still far from having an in-depth understanding of the implications that the paradigm shift caused by the digital revolution has brought to the dynamics of information dissemination in our society.

This shows how, for any systemic change, the negative effects are faster to manifest than the positive ones. A greater understanding of the phenomenon will be needed before the inevitable positive effects can be systematically exploited. Precisely for this reason, deepening the multidisciplinary study of the mechanisms that give rise to the dissemination of information on social media is so important. Gathering a set of shared knowledge, from which a constructive debate can develop, is of fundamental importance, in particular in this moment of global emergency due to the pandemic. There is a dire need for shared arguments on which to pivot, both to allow the debate to evolve and to take steps to counter the general crisis of truth and expertise the world is experiencing in recent years. In Part 1 of this report, we analyse in detail the state-of-the-art in research from different perspectives and approaches. As also remarked by [Lazer et al, 2018], there are surprisingly few scientific answers to basic questions such as the extent of the phenomenon of fake news and its impact on individuals. As already mentioned, [Alcott et al., 2018] estimated that the average American encountered between one and three stories from known publishers of fake news during the month before the 2016 election, and that "post-election surveys suggest that many people who read these stories believed them to be true (Allcott and Gentzkow 2017; Guess et al. 2018)". Even after the post-election decline – equal to over 50% -, Facebook engagements with fake news sites still averaged about 70 million per month in July 2018 and this is likely to be a very conservative estimate, as interactions with "only" 570 sites were examined. Most recently, cybersecurity firm SafeGuard Cyber found that half the European population might have been reached by Kremlin-backed disinformation campaigns about Europe between the 1st and the 10th of March 2019, propagated by 6,700 so-called "bad actors" including both troll and bot accounts.

We also know from [Vosughi et al, 2018] that fake-news tends to spread more rapidly and widely than mainstream information, and that this trend can be greatly magnified by the intervention of bots [Ferrara et al. 2017]; however, our knowledge of patterns followed by false information, particularly of visual and audio contents, is still very limited, or nearly non-existent for closed networks such as Snapchat or WhatsApp. Indeed, as remarked by [Lazer and al, 2018]: "in the absence of methods to derive representative samples of bots and humans on a given platform, any point estimates of bot prevalence must be interpreted cautiously".

As for the impact of micro-targeting and computational propaganda, information is once again limited: while scientific research on micro-targeting applied to consumers' behaviours [Kosinski and al, 2013; 2016; Matz et al, 2016; 2017] proves beyond any reasonable doubt its effectiveness; results from research on computational propaganda are mixed: so, for instance, [Guess et al; 2018] found that during the 2016 US presidential campaign "fake news consumption was heavily concentrated among a small group — almost 6 in 10 visits to fake news websites came from the 10% of people with the most conservative online information diets". Most recently, [Grinbeg et al. 2019] examined exposure to and sharing of fake news by registered voters on Twitter and found that engagement with fake news sources was extremely concentrated. Only 1% of individuals accounted for 80% of fake news source exposures, and 0.1% accounted for nearly 80% of fake news sources shared. Individuals most likely to engage with fake news sources were conservative-leaning, older, and highly engaged with political news. A cluster of fake news sources shared overlapping audiences on the extreme right, but for people across the political spectrum, most political news exposure still came from mainstream media outlets.

While studies on micro-targeting and computational propaganda largely draw from psychology and sociology in determining the efficacy of personalised content, we are not aware of scientific studies trying to link diffusion patterns and impact of fake-news in light of their narrative plausibility.

Finally, when it comes to measuring the effectiveness of countermeasures such as fact-checking, most scholars seem to agree on their ineffectiveness: both [Chavalarias, 2018 and Guess et al; 2018] show that debunked news very rarely reach fake-news consumers. [Lazer and al, 2018] highlighted the lack of evidence on the efficacy of specific education and training programmes on new media, and the debate around if and how to regulate tech-platforms seems also scarcely supported by scientific evidence [Lazer and al, 2018].

A fundamental problem which is largely preventing the scientific community from fully investigating phenomena related to the spread of disinformation online consists in the difficulty to access relevant data owned by technological platforms. For the time being, it is nearly impossible to determine the representativeness of data released through APIs on major platforms (Facebook, Twitter, Instagram), including in terms of quantities, date ranges, age and gender clustering, geographic information, not to mention website links being shared, words and sentences being used by users, geographical locations being affected and timelines - all key pieces of information when trying to understand what is going on social media. Search engines are also sharing very little information on how information is filtered online. In [Lazer and al, 2018] words: "There is little research focused on fake news and no comprehensive data-collection system to provide a dynamic understanding of how pervasive systems of fake news provision are evolving. It is impossible to recreate the Google of 2010. Google itself could not do so even if it had the underlying code because the patterns emerge from a complex interaction among code, content, and users. However, it is possible to record what the Google of 2018 is doing. More generally, researchers need to conduct a rigorous, ongoing audit of how the major platforms filter information".

Furthermore, it has now been made evident that automatic form of (dis)information as those represented by bot on Twitter reached such a level of sophistication, that even supervised algorithm cannot spot the difference from legitimate users [Cresci et al. 2017, Artoni et al 2019].

8.2 Areas for further research and consideration

It is clear from the first part of this report that the study of how social media networks have changed the way information spreads and how this change has impacted society is only in its infancy. Lots of research is still needed to understand better how information spreads on social media, the impact and role of narratives and cognitive sciences as well as understanding emerging threats and ways to tackle the spread of disinformation.

Based on desk research and consultation with experts (scientists, policymakers, media organisations and relevant civil-society organisations) and on the original research performed by the project team, the following topics have been identified as key areas that would demand more research and where the European Union could and should play a key role:

8.2.1 Better understand the impact of social media on society and information flows and how they affect our cognitions, emotions and collective behaviour.

Social media and online platforms have a significant impact on the creation of opinion and identities in societies as well as in created (or destroying) social cohesion and trust. Better understanding these mechanisms and how the current incentive models affect society will be very important in developing a more systemic understanding of how our media ecosystem affects society.

8.2.2 Promoting a system approach combining the science of network and narratology

Disinformation functions in information ecosystems. These ecosystems consist of one or more senders (information sources), media channels and recipients, but they are also constituted by shared values, symbolic codes, language, media-related habits and attitudes etc.

The analysis of narratives described above demonstrated that similar facts or events may be framed and interpreted differently if presented through different narratives. Ideological and political clashes endangering the common European project are often caused by such a narrative incompatibility. Let us consider two discussants engaged in a debate on the migration crisis. One of them perceives it through the lens of "escape and rescue" narrative, while the other one uses the image of the "invasion of barbarians" linked to the "decline and fall" story. They may even agree on basic facts, but first of them will always see the migrants primarily as victims, while the other considers them to be invaders.

Even a very limited pilot study performed on a relatively consistent corpus of pro-European, English language media demonstrated the existence of different corpuses of articles regarding same or similar issues, clearly separated by vocabulary, sentiment and dominant narrative patterns. A larger study comparing various media outlets would certainly result in discovering more of such interpretative patterns.

This calls for certain action from the communicators and policymakers.

- **We need to create pro-European campaigns, integrating narratives consciously.** While communicating verified, reliable facts still remains the main duty, it is important to consider carefully also the narrative through which these facts are communicated. A first important step towards that is to realise that we do need narratives. Pro-European attitudes may suffer serious blows from the Eurosceptics nowadays because of the narrative deficiency. This term has been coined by Randy Olson, an experienced science communicator, to describe "not enough comprehension of narrative and how it works". Narrative deficiency seriously weakens the message, its attractiveness, clarity and "stickiness".
- **Efficient narratives may be easily reduced to a core picture.** Our analysis demonstrates that there are clear, repetitive patterns recurring in most analysed articles from a given topic (set). They may be distilled to very simple spatial, temporal or performative models. To create a powerful narrative, it is necessary to "find the core of the idea", as the authors of the popular bestseller "Made to stick" explain. Many experts advise thinking of these patterns in strictly geometrical terms. Is Europe a circle in space – borders that divide us from whatever is outside? Or maybe it is a line or an arrow in time – symbolising progress?
- **The new media increase the importance of the narrative structure.** The very core of the story (the basic shape) is often the only part that gets transmitted through social media. After a series of paraphrasing, quoting, and adding layers of irony the simple spatial/temporal model may be the only thing that survives from the original story. Stories without a clear narrative structure are also much more difficult to translate to the language of infographics, tweets, memes or YouTube videos.
- **We have to understand and research the dominant anti-European narratives.** Research shows that the anti-European political narratives share a lot of common elements with conspiracy theories and pseudoscience. They present a very clear and attractive image of the world in which a given nation state and/or Europe is besieged by dominant hostile forces. These recurring Eurosceptical narrative patterns must be researched further in order to create powerful and meaningful alternatives.

A combination of complex network and narrative methods – also implying a multidisciplinary approach bringing together hard sciences and social sciences - would allow to better understand these complex phenomena. A number of research questions could be explored, for instance: is adherence to specific narrative schema related to disinformation? To what extent and in which narrative communities? How is credibility of information established in social media, and to what extent narrative plausibility makes information trusted? What is the role of narratives in shaping the opinions of society, and how new media differ from traditional media in this respect? To what degree are narratives openly revealed vs. hidden and implied and how the use of social media and the interactions across different media are affecting this balance? What is the role of public and shadow narratives in influencing the public opinion? Can narrative approaches help to tackle echo-chambers phenomena and stir genuine debate on social media? To what degree individuals read material representing opposite political view and how to integrate completely divergent views, integrate opinions of opposing echo-chambers?

8.2.3 Support the creation of interdisciplinary communities of researchers and practitioner using a shared toolkit

Through the SMART Observatory, we started building a toolkit that can be used by any researcher to start a data capture on a communication phenomenon (the EU and Italian Elections in our case), allowing to export the data in one command and execute a standard set of tests (selected from the State of the Art) and browse/study the results. Building a critical mass of researchers who could use and enrich the platform performing further tests based on literature and make them available would bring considerable benefits, for instance, in terms of:

- ease of use and quickness: the data capture could be configured minutes after the phenomenon is identified, and after the data capture the tests could be run immediately;
- visualisation of trends; network theory through community detection can provide the most efficient way to represent the structure of society;
- standardised tests: a community of researchers could maintain the selection of tests (we now have chosen the most relevant ones, as emerged from the State of the Art), to have a standardised set of tests that would make the results credible and comparable;
- transparency and accountability: the tests are selected from the most important and relevant publications in the research domain, and use software which is fully inspectable (open source), which means that they are supported by the entire research community and that they are completely transparent and accountable

This describes a model in which the international community of researchers in this area agree on the sets of tests to be performed, and these tests are made available on easy to use, standardised testing facilities that are completely shared, transparent and accountable, so that the results of these analyses can be shared, compared, and used as evidence.

The use of the SMART plug-in would allow adding a bottom-up approach, allowing both researchers and generic users to help setting levels of trustworthiness of different news and sources, which in turn could allow starting data-capture processes base on the outcome of this crowd-fact-checking exercise.

8.2.3 Monitoring threats, measuring impact and understanding long-term trends

More research is needed to understand the patterns followed by false information, particularly of visual and audio contents, and on closed networks such as Messenger, Snapchat or WhatsApp. Emerging phenomena such as deepfakes and, more in general, the evolution of bots in their interactions with cyborgs and trolls, is another area which would deserve attention, including in terms of finding reliable methods to derive representative samples of bots and humans on a given platform, also considering that most work so far has been limited to Twitter. Longitudinal analysis of complete datasets would allow a dynamic understanding of how pervasive systems of fake news provision are evolving over time, including in terms of impacts on target audiences (concerning for instance the impact of micro-targeting and computational propaganda on individuals' as well as groups' behaviours). This would also allow to timely understand the efficacy of countermeasures such as fact-checking and automatic debunking.

8.2.4 Long-term cultural approaches

"How can we create a news ecosystem and culture that values and promotes truth?" [Lazer et al, 2017]. Long-term approaches aimed at changing the culture of data in our societies and economies, contributing to raising awareness about the wide range of phenomena linked to the circulation of disinformation on social and new media, are deeply needed. This broad area of research includes both the creation and evaluation of specific programmes and tools, designed and implemented by a broad range of stakeholders, including scientists, tech-platforms, civil society and media organisations as well as pedagogues, anthropologists, psychologists and artists. Common languages and strategies should be built to help collaboration among these stakeholders, and more research is needed to evaluate regulatory and non-regulatory scenarios and ethical and moral implications at stake. Sound impact evaluation systems should be designed in collaboration with the scientific community to understand the efficacy of different approaches to tackling disinformation at the micro, meso and macro level, so to allow for rapid adoption, replication, adaptation and scaling of effective measures.

8.2.5 A European Approach to Artificial Intelligence

Artificial intelligence (AI) is already a reality and it is expected to largely impact on economy, social behavior and, eventually, political equilibrium. In this context, many actors are asked to address regulatory and ethical issues, assure robust quality assurance process and deliver real business value. No country can act in isolation or lead this process, as well as no single industry should control the technologies or dominate the market.

European Union is the best candidate to provide the requested critical mass, guarantee the transparency of the process and the respect of the ethical values. Indeed, the European scientific and industrial excellence can drive fair competitiveness at global level. Science, industry and policy have the joint responsibility to make this happen, throughout an integrated ecosystem where funds, instruments, rules and measures should be coherent and effective.

In this regard, the European Union has already proved to be able to launch and manage ambitious initiatives in which scientific results, industrial capabilities and public authorities have been involved in efficient governance of this complex system.

To do so, and to provide socio-economic benefits for European citizens, it is mandatory to increase the competitiveness of Europe and contribute to establishing an ethical framework where Europe acts as a key player in research, policy and deployment of advanced products.

In the process towards the launch of Horizon Europe and the activities of the new European Parliament, a reflection among Institutions and Research Organizations on the future research developments in support to policy, in particular addressing Artificial Intelligence, is urgently needed.

8.3 Policy recommendations

The digital era has drastically changed the nature of the public space, the actors who control this space and who has access to the most effective tools. As shown in this report, although the study of the impact of the spread of (dis)information on social media is in its infancy, this phenomenon will have a lasting impact on society.

The question no longer concerns the possibility that this impact is real, but its modalities and how to minimise its effects on our shared model of ethical values and individual freedoms. The decisions that will be made in the coming years will be pivotal in defining how the digital media ecosystem develops and what the future of democracy will look like. Remembering the words of Winston Churchill: "We shape our buildings; thereafter they shape us".

Europe is in a unique position to be able to develop a European model for the digital world in line with the European values of human dignity, freedom, democracy, equality, rule of law, human rights and peace. It is clear from the above analysis that technological and computer science approaches will not be sufficient in tackling the problem of disinformation. A more systemic approach will be required, one that utilises both network science and narratology to ensure that the information exchanged is accurate and truthful, without limiting the freedom of speech that democratic institutions must provide.

Throughout the development of this report, several policy suggestions were brought forward by experts across disciplines and stakeholders. The most relevant are:

8.3.1 Organising and regulating the digital online media ecosystem.

Seeing the importance of the information ecosystem for society, a different ecosystem would be essential to preserve and strengthen our collective intelligence. As demonstrated by a long history of research, a multitude benefiting from a healthy diversity of opinions and experience can intelligently respond to difficult questions; this capacity is however undermined the more polarized a society becomes.

As underlined by this report, the communication ecosystem plays a key role for every society and today power is played on the digital social media. This creates the need to regulate this new space in order to ensure that the very positive aspects of social media are delivered as opposed to this becoming a new medium that can be harnessed to manipulate societies to push specific interests. Examples of how this is done today are clear not only with the Facebook-Cambridge Analytica scandal but also during the Irish referendum.

It is important to remember that the traditional media landscape is one of the most highly regulated precisely to safeguard the information ecosystem and in the interest of society at large. Today there are different rules for traditional and new media and the choices we make today in how the new media environment should be regulated will shape how our political landscape will develop for the decades to come.

Several scholars have posited that in order to change the thinking on how to organise and regulate the information ecosystem, we should change the way we approach the question itself – change the narrative. It is undeniable that the communication infrastructure of society is pivotal in shaping society and that a pluralistic and transparent space, allowing for a multitude of opinions and perspectives to reach citizens is a cornerstone of our way of organising society. Whilst today the space where public is confronted with information is privately owned by foreign global conglomerates with their own interests and incentives, Europe should think about how the public space could again become a “public good”, independent from media, company and government interests with the role of preserving the social diversity essential for democracy.

8.3.2 Access to Data

As seen above, limited access to data from private companies that today control Europe's data is very detrimental in being able to accurately assess the impact of disinformation and the role of various actors in the spread of fake news. As shown in Chapter 6, the current actions of platforms like Facebook do not seem to have any impact on effectively dealing with the problem of disinformation.

Policy measures could also include **new ways of looking at data ownership**, something that GDPR has already introduced in the EU regulation arena: "more than with specific technological solutions, societies and governments should come up with principles that inspire policies, and then evaluate the application of those policies against the guiding principles to make sure those are not betrayed by the implementation and that there are no clearly exploitable loopholes to circumvent the principles. It will certainly be an iterative process of trial, error and adjustment, but the starting point should be a humanistic declaration of rights of individuals with regards to the data they produce.

Data access to researchers is undoubtedly an issue. Another issue is that companies that today are the self-proclaimed keepers of such data are not held accountable to security standards. One the one hand, research in sensitive fields like cybersecurity has so far shown how security by obfuscation rarely works, on the other hand there are so many possible ways in which total freedom of access to people's data can be used to damage them.

For this reason, more than specific technological solutions, societies and governments should come up with principles that inspire policies and then evaluate the application of those policies against the guiding principles.

Concerns linked to privacy and compliance with regulations are also a potential barrier. IN this regard a potential solution could be the "privacy-by-design" approach suggested by GDPR.

Today, both private platforms and governments tend to see data as an "extractive" industry: we pull data out of people for good (to provide them with more personalised, integrated and cost-effective services) or for worse (to sell them more products, to manipulate their opinion), but there are still very few examples of projects which look at data as something co-created with people and machines, including bots. This is also because, as technology becomes increasingly ubiquitous and frictionless, understanding the full spectrum of the data we generate at any time – as we walk across sensors and cameras, search the internet for products and services, make financial transactions using our cards or smartphones, get our heartbeats recorded, interact with smart grids and buildings - becomes increasingly difficult. Indeed, even for the most gifted technologists, understanding both what data we generate, who uses it, how – with what other data is recombined and analysed – and for what purposes, is nearly impossible. In this context, **it is fundamental for public authorities at EU and national level, to help safeguard spaces where data and computation can be organised and performed as ever-evolving commons, reminding that a common is given not only by the data, but also by the relational ecosystem around the data and**

the fact that this ecosystem negotiates and shares practices and rules. This can only happen if we heavily invest in supporting people, small businesses and public administrations in becoming aware of the data they produce – and of the fact that sometimes they cannot be aware neither of the data they produce nor of the ways this is used – and if we are able to find easy, cost-effective and engaging ways to turn data into actionable information, which should then be followed by concrete actions. In fact, it seems difficult-to-impossible to imagine any solution to disinformation and to its effects on society without reflecting on the social architectures and processes that bring to these kinds of phenomena.

8.3.3 Who controls the public sphere?

As underlined by this report, the communication ecosystem plays a key role for every society and today, power is played out on digital social media. This creates the need to regulate this new space in order to ensure that the very positive aspects of social media are delivered, as opposed to this becoming a new medium that can be harnessed to manipulate societies to push specific interests. Examples of how this is done today are clear not only with the Facebook-Cambridge Analytica scandal but also during the Irish referendum.

It is important to remember that the traditional media landscape is one of the most highly regulated precisely to safeguard the information ecosystem and in the interest of society at large. Today there are different rules for traditional and new media and the choices we make today in how the new media environment should be regulated will shape how our political landscape will develop for the decades to come

8.3.4 A European Approach to Artificial Intelligence

The report underlines the importance of Europe's efforts in developing a European approach to AI. By 2020 AI ability to counterfeit media will largely surpass those of AI to identify such media. As highlighted above (8.2.5), AI is already here and is already having an enormous impact on the economy, the media ecosystem and society. As previously highlighted, human-centric AI may help us design novel platforms and mechanisms for public access to news and information, focused on counterbalancing the human tendency to confirmation bias. It is possible to imagine AI mechanisms helping individuals and communities become informed on controversial issues by offering multiple perspectives, connecting opposing views and conflicting arguments, therefore fostering critical thought. Advances in person-machine interaction models based on explainable AI have the potential to reach novel cognitive trade-offs between our confirmation bias and our curiosity of novelty and diversity, making it possible for more sustainable and humanised information ecosystems to emerge.

8.3.5 Alternatives to the current business model

Governments – and the European Commission in particular – should take a proactive approach building policies and funding programmes to create and grow alternatives to the current centralised platforms, rather than trying to make arrangements with them.

According to Piero Molino (Chapter 2), we have at least three aspects of the current circumstances that, combined, led to the current situation:

- a) On the one hand you have the absolute low cost of content creation and diffusion, as creating it can either be done by really cheap labour together with semi-automated processes, and the diffusion is fully automatable;
- b) The second factor is the fact that it's almost impossible to make the people responsible for the generation of disinformation accountable for their actions, because of a lack in jurisdiction together with a lack in practical instruments of law enforcement on the internet. Already pinpointing who is responsible for the initial spreading is difficult (but not impossible, there are studies and techniques of network science virology that have been applied with great success to the task, but never left the academic lab so far and have not been adopted by law enforcement to the best of my knowledge), but even once the responsible people are discovered, there could be other problems in making them accountable (international laws, the exploitation of mass labour in countries where labour is cheap, lack of local laws in those countries, and so on);
- c) Finally, the third point is the general crisis of meaning and truth that the flooding of disinformation has given rise to (anti-vax, flat-earthers and similar groups) that would defend the freedom of spreading disinformation like they would defend freedom of speech, which highlights how thin the distinction

really is and how, as a global society, we should probably start rethinking the meaning of freedom, truth, reality and opinion, if we want the current blur to be overcome.

The military, cold-war like escalation of algorithms, BOTs, AIs and other computational cyberwarfare tools and techniques falls into this scenario and draws stamina from this scenario: in a world that is progressively more angry, aroused and polarised, escalation brings results that are communicable to the public, but which are not able to deal with the complexity of society. At the same time, the advancements bring on further advancements from the other factions. The result is a technical and methodological progression of those same mechanisms which are at the core of the problem, and a society which is progressively more divided and extreme, and in which people's rights and liberties are progressively eroded.

Therefore, solutions should be designed to tackle this scenario not only with the technical approaches which see the development of tools, techniques and methods able to progressively confront with the rising standards and achievements of disinformation operators. These solutions should go beyond the idea of implementing tools and services, and would assume the form of actions which take place in the public sphere, and that combine technologies, tools and engagement strategies in order to 1) promote culture and awareness of data and computation, 2) enact processes which stimulate social imagination and 3) adopt artistic and design/creative approaches to trigger solidarity, empathy, social cohesion.

As highlighted by prof. Guido Vetere, funding sdecentralised alternatives to the current media platforms might be far more effective than regulating them or trying to make arrangement with them: "the sdecentralisation of social platforms should be encouraged by active policies, which would also lead to overcoming monopolies.

The idea is that we should support a novel wave of better platforms, rather than try (or pretend) to control existing ones. A possible path starts from supporting the specification of functionalities, models and schemas (i.e. an Application Programming Interface) by calling academia, public research, developers, communities of practice on collaborative, open platforms such as GitHub. There's no need for huge investments for that, just some seed money to support the organisation and have a handful of right people involved. Then we give the right amount of resources to a lot of small nodes willing to implement the network based on that specification. It is reasonable to assume that each node could be financed with at most one thousand euros to start up. If we consider that an EU Flagship project sizes up to 1 Billion, a 'sDecentralised Flagship' could support thousands of independent nodes in each Member State. Attempts to go in this direction are already out there, such as Diaspora and Mastodon, but these are small and isolated initiatives. The W3C is at work on the problem of sdecentralised identifiers, and Tim Berners-Lee is striving to restore the original Web architecture: all these efforts should be encouraged and sustained. If the Commission launched a sdecentralised Flagship for social networking, this could really make a difference. Of course, a decentralised system is certainly more complex and less efficient than a centralised one: the way Facebook works on user inputs ('like' or whatever reaction) relies on the big architectural advantage of having all the data accessible by means of a single logical endpoint. Centralised systems are easier to create and manage. And critical mass is important too: there are over 2 bn people already on Facebook, which is really a heavy legacy. Moreover, this mass is a huge revenue generating asset, since if a company (say a European one) wants to know to whom they could advertise their products, the only option they have now is to buy profiling data from Facebook or Google. People have not a clue of what their data is worth. So, for sure, when we talk of decentralisation, we are not talking about an easy endeavour. At the same time, this is not impossible. Tech platforms are subject to fast-changing and disruptive trajectories.

It is crucial to come up with a business model which can open a market that currently doesn't exist.

This will probably happen, sooner or later, but if the Commission could accelerate this process by supporting and stimulating something that still seems to be far from being market-driven, this would be a key strategic move. To do that, we need all the wisdom we are capable of, this is really the best resource we can mobilise in Europe: our sense of liberty and sociality, our ability to integrate differences, our open research and critical thinking. The point is precisely to make sdecentralised social platforms a key ingredient of the European culture, to make them the object of a "social desire". There is so much room for EU policy here: electric cars are incentivised, so why not social decentralised platforms?" [SMART Expert interviews].

8.3.6 Intersectoral collaborations and collective intelligence

Other types of solutions should be designed to be able to address the rising separation between citizens, the need to produce the conditions for solidarity, empathy and the formation of an inclusive, high quality relational ecosystems in society.

These kinds of actions can be achieved by combining multiple disciplines and technologies. They should start with the conception and design of new education processes, professions and roles for citizens and professionals which are dedicated to "connecting the dots" in society, to obtain higher quality and more inclusiveness in relational ecosystems in cities and regions: using social and cultural actions, and also using data and computation in open ways, creating alternatives for the current "extractive" data and computation industries which bring data and computation out in the public sphere, where society can agree on rules, regulations and protocols in open, inclusive, readable, knowable ways.

Collaborations and interactions between sciences, technology, arts and design should take place to bring opportunities for social imagination in the public sphere, in which the results of scientific and technological innovation are not only showcased, but also act as the trigger for constructive, inclusive, collaborative imagination of future scenarios that are capable of bringing better relations, information, communication, knowledge and opportunities for exchange.

8.4 Final Statements

In conclusion, to put it in the words of Franklin Delano Roosevelt: "Democracy can keep alive only if the settlement of old difficulties clears the ground and transfers energies to face new responsibilities". The changes that created the current crisis of trust in the institutions are already present, and it is not possible to pretend that they have not happened. Addressing such systemic changes involves redesigning the whole complex ecosystem that shapes both the risks and most effective preventive actions.

A new paradigm for the dissemination of information and opinions has already arrived: just as it happens for the pandemic, our effort must be to study the dynamics it has created and try to build rules that allow us to maintain our values in a new system. We cannot simply go back to the dynamics that characterised human society before the digital revolution. What Europe can and must do is use its moral leadership and the strength of its founding values so that they can shape the change, and not be overwhelmed by it.

Shifting the focus from the content of information to the very architecture of the informational networks will make it possible to tame the change and adapt it to our model. During the necessary period of adaptation, it will be essential to use all reasonable measures to buffer the flaws that our democratic system will suffer. In the long run, however, such an approach is the only one that can guarantee that the substance of the values that guided the European dream will remain unchanged.

References

- Agarwal, N., Lim, M. and Wigand R.T. [2014], *Statistical Properties of Community Structure in Large Social and Information Networks*, Springer.
- Allcott, H. and Gentzkow, M. [2017], 'Social Media and Fake News in the 2016 Election', *Journal of Economic Perspectives* **31** (2), 211–236.
- URL: <http://pubs.aeaweb.org/doi/10.1257/jep.31.2.211>
- Anderson, B. (2006). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso Books.
- Arnaudo, D. (2017). Computational propaganda in Brazil: Social bots during elections. Computational Propaganda Research Project, Working Paper, (2017.8).
- Armbruster Heidi, Rollo Craig, Meinhof Ulrike H. (2003), *Imagining Europe: Everyday Narratives in European Border Communities*, "Journal of Ethnic and Migration Studies", Volume: 29. Issue: 5.
- Artoni, A., Ceri, S., & Pierri, F. (2019). [Investigating Italian Disinformation Spreading On Twitter In The Context Of 2019 European Elections.](#)
- Asch DA, Baron J, Hershey JC, Kunreuther H, Meszaros J, Ritov I, Spranca M. Omission bias and pertussis vaccination. *Med Decis Making* 1994;14:118–23
- Assmann, J. (2011). *Cultural memory and early civilization: writing, remembrance, and political imagination*. Cambridge University Press.
- Bal, Mieke (2009). *Narratology: Introduction to the theory of narrative*. University of Toronto Press.
- Bail C.A. et al, (2018) "Exposure to Opposing Views on Social Media Can Increase Political Polarization," *Proceedings of the National Academy of Sciences*, 115 9216-9221 <http://www.pnas.org/content/115/37/9216> (2018).
- Baker Jeffrey T., *The pertussis vaccine controversy in Great Britain, 1974–1986*, "Vaccine" 21 (2003) 4003 4010.
- Bakhtin, Mikhail. "Discourse in the Novel." *Literary theory: An anthology* 2 (1935): 674-685.
- Barabási A-L "The network takeover" *Nature Physics* 8, 14–16 (2011)
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber?. *Psychological science*, 26(10), 1531-1542.
- Barthes, R. (1975). An introduction to the structural analysis of narrative. *New literary history*, 6(2), 237-272.
- Barthes, Roland. (1975) "An Introduction to the Analysis of Narrative." *New Literary History* 6, 237–72.
- Bartlett, F.C., 1932, *Remembering*, Cambridge University Press.
- Bassett, Caroline. "The arc and the machine: Narrative and new media." (2013).

- Becatti, C., Caldarelli, G., Lambiotte, R., Saracco, F. "Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections" *Palgrave Communications* 5, 91 (2019).
- Benveniste Emile (1971) *Problems in General Linguistics*, Miami: University of Miami Press.
- Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*. *New York*.
- Bessi, A., Petroni, F., Del Vicario, M., Zollo, F., Anagnostopoulos, A., Scala, A., Caldarelli, G. and Quattrociocchi, W. [2016], 'Homophily and polarization in the age of misinformation', *EUROPEAN PHYSICAL JOURNAL-SPECIAL TOPICS* **225**(10), 2047–2059.
- Bishwas, J. [2018], 'Top 7 Platforms That Are Busting Fake News On Social Media'.
URL: <https://www.analyticsindiamag.com/platforms-busting-fake-news-social-media/>
- Blitz, M. J. (2018). Lies, Line Drawing, and (Deep) Fake News. *Oklahoma Law Review*, 71(1), 59.
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E "Fast unfolding of communities in large networks" *Journal of Statistical Mechanics (Theory and Experiments)* 10, P10008 (2008).
- Bonheim, H. (2000). "Shakespeare's narremes." In: Holland, P. [ed.] *Shakespeare Survey 53: Sakespeare and narrative*, pp. 1-11. Cambridge: Cambridge University Press,
- Bovet, A. and Makse, H. A. [2018], **Influence of fake news in Twitter during the 2016 US presidential election**, Technical report.
URL: <http://arxiv.org/abs/1803.08491>
- Bradshaw S. and Howard P.N. (2018), "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation" Oxford Internet Institute's Computational Propaganda Research Project, July 2018, <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf>.
- Brady, William J., et al. "Emotion shapes the diffusion of moralized content in social networks." *Proceedings of the National Academy of Sciences* 114.28 (2017): 7313-7318.
- Brendan Nyhan et al, (2014), "Effective Messages in Vaccine Promotion: A Randomized Trial," *Pediatrics*, 133 e835-e842.
- Bruner, J. (1991). *The Narrative Construction of Reality*. *Critical Inquiry*, 18(1), 1-21.
- Brunvand, J. H. (2001). *Encyclopedia of urban legends*. Abc-clio.
- Cadwalladr, C. [2017], 'The great British Brexit robbery: how our democracy was hijacked'.
URL: <https://www.theguardian.com/technology/2017/may/07/the-great-british-brexite-robbery-hijacked-democracy>
- Caldarelli, G. [2007], *Scale-Free Networks: complex webs in nature and technology*, Oxford University Press, Oxford, UK.
- Caldarelli, G., Chessa, A., Pammolli, F., Pompa, G., Puliga, M., Riccaboni, M. and Riotta, G. [2014], 'A multi-level geographical study of Italian political elections from Twitter data.', *PLoS one* **9**(5), e95809.
URL: <http://dx.plos.org/10.1371/journal.pone.0095809> <http://journals.plos.org/plosone/article?id=10.1371>
- Caldarelli, G., De Nicola, R., Del Vigna, F., Petrocchi, M., Saracco, F. "The role of bot squads in the political propaganda on Twitter" arXiv:1905:12687 (2019).

- Calvert, (2017), The Psychology Behind Fake News, <https://insight.kellogg.northwestern.edu/article/the-psychology-behind-fake-news>.
- Cameron, Allan (2008). Modular narratives in contemporary cinema. In *Modular Narratives in Contemporary Cinema* (pp. 1-19). Palgrave Macmillan, London.
- Canefe, Nergis, and Tanil Bora. "The intellectual roots of Anti-European sentiments in Turkish politics: The case of radical Turkish Nationalism." *Turkish Studies* 4.1 (2003): 127-148.
- Castells, Manuel. "The Rise of the Network Society." Wiley-Blackwell, 2010
- Chamberlain, M., & Thompson, P. R. (1998). Introduction in "Narrative and genre", Chamberlain, M., & Thompson, P. R. (Eds.). Psychology Press.
- Chan, M. Pui S., Jones, C. R., Hall Jamieson, K. & Albarracín, D. (2017), "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation" *Psychol. Sci.* 28, 1531–1546.
- Cinelli, M., Cresci, S., Galeazzi, A., Quattrociocchi, W., & Tesconi, M. (2019). The Limited Reach of Fake News on Twitter during 2019 European Elections. arXiv preprint arXiv:1911.12039.
- Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A. & Tesconi, M. "Fame for sale: efficient detection of fake twitter followers" *Decision Support Systems* 80, 56–71 (2015).
- Daniel Funke [2018], '16 ways to debunk hoaxes on WhatsApp | Poynter'.
URL: <https://www.poynter.org/news/16-ways-debunk-hoaxes-whatsapp>
- Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In: *ACM WWW Companion* (2017)
- Davis, C. A., Varol, O., Ferrara, E., Flammini, A. and Menczer, F. [2016], BotOrNot: A System to Evaluate Social Bots.
URL: <http://arxiv.org/abs/1602.00975>
- De Saussure, F. (2011). *Course in general linguistics*. Columbia University Press.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554-559.
- Diez Thomas (2001), Europe as a Discursive Battleground: Discourse Analysis and European Integration Studies, "Cooperation and Conflict" 36: 5.
- Durbach Nadja, *Bodily Matters*, Durham 2004;
- Durbach Nadja. *They Might As Well Brand Us': Working-Class Resistance to Compulsory Vaccination in Victorian England*, „Social History of Medicine" (2000) 13 (1): 45-63.
- Digital Wildfires in a Hyperconnected World* [2009], Technical report.
URL: http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/?doing_wp_cron=1540300175.5511839389801025390625
- Doshi, A. R., Raghavan, S., Weiss, R. and Pettitt, E. [2018], The Impact of the Supply of Fake News on Consumer Behavior During the 2016 US Election.
URL: <https://www.ssm.com/abstract=3093397>
- Eco, U. (1984). *The role of the reader: Explorations in the semiotics of texts* (Vol. 318). Indiana University Press.

- Edson C. Tandoc Jr., Zheng Wei Lim & Richard Ling (2018) Defining "Fake News", *Digital Journalism*, 6:2, 137-153, DOI: 10.1080/21670811.2017.1360143
- Eder, Klaus (2006), *Europe's Borders : The Narrative Construction of the Boundaries of Europe*, „European Journal of Social Theory" 2006 9: 255.
- Edell, A. [2018], 'I trained fake news detection AI with >95% accuracy, and almost went crazy'. URL: <https://towardsdatascience.com/i-trained-fake-news-detection-ai-with-95-accuracy-and-almost-went-crazy-d10589aa57c>
- Eikhenbaum, Boris (1974). How Gogol's' Overcoat is Made. in: *Gogol from the Twentieth Century: Eleven Essays*, 269-91.
- Elmas, T., Overdorf, R., Özkalay, A. F., & Aberer, K. (2019). Lateral Astroturfing Attacks on Twitter Trending Topics. arXiv preprint arXiv:1910.07783.
- Eom, Y.-H., Puliga, M., Smailović, J., Mozetic, I. and Caldarelli, G. [2015], 'Twitter-Based Analysis of the Dynamics of Collective Attention to Political Parties', *PLOS ONE* 10(7), e0131184. URL: <https://dx.plos.org/10.1371/journal.pone.0131184>
- European Commission [2018], A multi-dimensional approach to disinformation, Technical report, European Commission.
- Fenster, Mark. *Conspiracy theories: Secrecy and power in American culture*. U of Minnesota Press, 1999.
- Ferguson, Niall. "The square and the tower: Networks and power, from the Freemasons to Facebook." Penguin Books, 2019.
- Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A. [2016], 'The Rise of Social Bots', *Commun. ACM* 59(7), 96–104.
- Ferrara, E. (2019). Bots, elections, and social media: a brief overview. arXiv preprint arXiv:1910.01720. URL: <http://doi.acm.org/10.1145/2818717>
- Field M. and Wright M., "Russian trolls sent thousands of pro-Leave messages on day of Brexit referendum, Twitter data reveals", *The Telegraph*, 17 October (2018). <https://www.telegraph.co.uk/technology/2018/10/17/russian-iranian-twitter-trolls-sent-10-million-tweetsfakenews/>.
- Fireeye "Suspected Iranian Influence Operation: Leveraging Inauthentic News Sites and Social Media Aimed at US, UK, Other Audiences," Fireeye, 21 August (2018). <https://www.fireeye.com/content/dam/fireeye-www/current-threats/pdfs/rpt-FireEye-Iranian-IO.pdf>.
- Flanagin, Andrew J., and Miriam J. Metzger. 2007. "The role of Site Features, User Attributes, and Information Verification Behaviors on the Perceived Credibility of Web-based Information." *New Media & Society* 9 (2): 319–342. doi:10.1177/1461444807075015
- Fludernik, M. (2012). *An introduction to narratology*. Routledge.
- Fortunato S "Community detection in graphs" *Physics Reports* 486, 75–174 (2010).

- Galasinski, Dariusz. Meinhof, Ulrike H. (2000) "Photography, memory, and the construction of identities on the former East West German border", *Discourse Studies*, 2(3): 323-53.
- Gangware, C. & Nemr, W. (2019), "Weapons of Mass Distraction: Foreign State-Sponsored Disinformation in the Digital Age" Park Advisors.
- Gelfert, A (2018), Fake news: A definition, *Informal Logic*, Vol. 38, No.1 pp. 84–117
- Genette, G. (1983). *Narrative discourse: An essay in method*. Cornell University Press
- Gilbert, Mark (2008), *Narrating the Process: Questioning the Progressive Story of European Integration*, *JCMS Volume 46. Number 3*. pp. 641–662
- Girvan M. & Newman M.E.J. "Community structure in social and biological networks" *PNAS* 99, 7821 - 7826 (2002).
- Good, B.H., de Montjoye, Y.-A. & Clauset. A. "The performance of modularity maximization in practical contexts" *Physical Review E* 81, 046106 (2010).
- Grassegger H, Krogerus M (2017) The data that turned the world upside down. Available at https://motherboard.vice.com/en_us/article/how-our-likes-helped-trump-win.
- Greimas, A.J. (1987) *On meaning: Selected writings in semiotic theory*. Minneapolis: University of Minnesota Press.
- Grice, Andrew. "Fake news handed Brexiteers the referendum." *e Independent*, 18 Jan. 2017.
- Grinberg N. et al. (2019), "Fake news on Twitter during the 2016 U.S. presidential election" *Science* 363, 374-378.
- Gualdi S, Cimini G, Primicerio K, Di Clemente R, Challet D "Statistically validated network of portfolio overlaps and systemic risk" *Scientific Reports* 6, 39467 (2016).
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. "Political science: Fake news on Twitter during the 2016 U.S. presidential election" *Science* 363, 374–378 (2019).
- Harari, Yuval Noah. "Homo Deus: A brief history of tomorrow." Random House, 2016
- Howard, Philip N., and Bence Kollanyi. "Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum." Oxford Internet Institute, Project on Computational Propaganda, 20 June 2016, <http://dx.doi.org/10.2139/ssrn.2798311>.
- Jakobson, R. (2010). *Two aspects of language* in: *Fundamentals of Language*, Walter de Gruyter.
- Jakobson, R. (1960). *Linguistics and poetics*. In *Style in language* (pp. 350-377). MA: MIT Press.
- Kim, J. W. (2018), "Evidence Can Change Partisan Minds: Rethinking the Bounds of Motivated Reasoning" Job Market Paper.
- Kollanyi, B., Howard, P. N., & Woolley, S. C. (2016). Bots and automation over Twitter during the first US presidential debate. *Comprop data memo*, 1, 1-4.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 201218772.
- Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological methods*, 21(4), 493.
- Kruglanski, A. W. & Webster, D. M. "Motivated closing of the mind: 'seizing' and 'freezing'" *Psychol. Rev.* 103, 263–83 (1996).

- Kudugunta, S. and Ferrara, E. [2018], Deep Neural Networks for Bot Detection.
URL: <http://arxiv.org/abs/1802.04289>
- Kuus, Merij (2002), *European Integration in Identity Narratives in Estonia: A Quest for Security*, Journal of Peace Research", vol. 39, no. 1, pp. 91–108.
- Iaconesi, Salvatore. (2017). "Interface and Data Biopolitics in the Age of Hyperconnectivity. Implications for Design.", in Design for Next 12th EAD Conference.
- Lacy, S. and Rosenstiel, T. [2015], Defining and Measuring Quality Journalism, in 'Media + the public interest initiative', Rutgers Schools of communication, pp. 1–66.
URL: <http://mpii.rutgers.edu/quality-journalism-in-the-digital-age/>
- Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.
- Lankina T and Watanabe K, (2018), "Russian Spring' or 'Spring Betrayal'? The Media as a Mirror of Putin's Evolving Strategy in Ukraine," Europe-Asia Studies, March 2018, <https://doi.org/10.1080/09668136.2017.1397603>.
- Lane, D., & Maxwell, R. R. (2005). Ontological uncertainty and innovation. *Journal of Evolutionary*
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J. and Zittrain, J. L. [2018], 'The science of fake news: Addressing fake news requires a multidisciplinary effort', *Science* **359**(6380), 1094–1096.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/29590025>
- Leach, E. (1989). *Claude Lévi-Strauss*. University of Chicago Press.
- Lévi-Strauss, C. (1990). *Mythologiques* (Vol. 4). University of Chicago Press.
- Mandler, J. M. (2014). *Stories, scripts, and scenes: Aspects of schema theory*. Psychology Press
- Mansky, J. [2018], 'The Age-Old Problem of "Fake News"'.
URL: <https://www.smithsonianmag.com/history/age-old-problem-fake-news-180968945/>
- Marchi, Regina. "With Facebook, blogs, and fake news, teens reject journalistic "objectivity"." Journal of Communication Inquiry 36.3 (2012): 246-262.
- Massey, Douglas S., and Magaly Sánchez. *Brokered boundaries: Immigrant identity in anti-immigrant times*. Russell Sage Foundation, 2010.
- Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the national academy of sciences*, *114*(48), 12714-12719.
- McAdams, D. P. (2006). *The Redemptive self: Stories Americans live by*. New York: Oxford University Press.
- McAdams, D. P. (2008). American identity: The redemptive self. *The General Psychologist*. 43, 20-27.
- Meister, Jan Christoph (2003), *Computing Action A Narratological Approach*,
- Metaxas, P., Finn, S. and Mustafaraj, E. [2015], Using TwitterTrails.com to Investigate Rumor Propagation, in 'Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing. CSCW'15 Companion, ACM NY USA'.
URL: <http://doi.acm.org/10.1145/2685553.2702691>

- Metaxas, P. T. (2018), "Technology , Propaganda , and the Limits of Human Intellect" arXiv Comput. Soc. doi:arXiv:1806.09541v1.
- Mielecinski, E. M., (1981). *Poetyka mitu*, trans. Dancygier, J., & Mayenowa, M. R., Państwowy Instytut Wydawniczy.
- Minsky, M. (1975) A Framework for Representing Knowledge ,in P. Winston (Ed.) The Psychology of Computer Vision, (New York: McGraw-Hill), 211-277. Moore, J. and Newell, A.
- Mitchell, A. and Page, D. [2015], The evolving role of news of Twitter and Facebook, Technical report.
- Mitra, T. and Gilbert, E. [2015], CREDBANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations, in 'AAAI Publications, Ninth International AAAI Conference on Web and Social Media'.
- Morozov, Evgeny. "The net delusion: The dark side of Internet freedom." Public Affairs, 2012
- Napiórkowski Marcin (2012), Porywacze psów i kebabowe podziemie. Współczesne mity kulinarne oobcych, „Kultura popularna” 4(34): 188-197.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton Century Crafts.
- Neumayer Laure, *Integrating the Central European Past into a Common Narrative: the mobilizations around the 'crimes of Communism' in the European Parliament*, Journal of Contemporary European Studies, 23 (3), pp.344-363.
- Newman, N., Fletcher, R., Kalogeropoulos, A., Nielsen, D., Levy, A. L. and Kleis, R. [2017], Digital News Report, Technical report, Reuters Institute.
- Nichols, T. M. (2017). The death of expertise. Tantor Media, Incorporated.
- Nicodemo, A. [2018], 'To stop fake news, researchers call for internet platforms to choose quality over quantity'.
- URL:** <https://phys.org/news/2018-03-fake-news-internet-platforms-quality.html#jCp>
- Nünning Ansgar (2003), Narratology or narratologies?
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Parkinson, H. J. "Click and elect: how fake news helped Donald Trump win a real election." *e Guardian*, 14 Nov. 2016, <https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies>.
- Pasupathi, M., Mansour, E., & Brubaker, J. R. (2007). Developing a life story: Constructing relations between self and experience in autobiographical narratives. *Human Development*, 50(2-3), 85-110.
- Pattabiraman, B., Patwary, M. M. A., Gebremedhin, A. H., Liao, W.-k. and Choudhary, A. [2015], 'Fast Algorithms for the Maximum Clique Problem on Massive Graphs with Applications to Overlapping Community Detection', *Internet Mathematics* 11(4-5), 421-448.
- URL:** <http://www.internetmathematicsjournal.com/article/1586>
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of personality and social psychology*, 62(2), 189.

- Poland, Gregory A., and Robert M. Jacobson. "Understanding those who do not understand: a brief review of the anti-vaccine movement." *Vaccine* 19.17-19 (2001): 2440-2445.
- Polkinghorne, D. E. (1991). Narrative and self-concept. *Journal of Narrative and Life History*, 1(2), 135-153.
- Prince Gerald (1982), *Narratology: The Form and Functioning of Narrative*, Mouton Publishers.
- Prince Gerald (2003), *Surveying Narratology*, in: Kindt and Müller (eds.), *What Is Narratology?*
- Propp, Vladimir (2010). *Morphology of the Folktale* (Vol. 9). University of Texas Press.
- Reich, R. B. (2005). The lost art of Democratic narrative. *Story time*. *New Republic*, 232(11-12), 16-19.
- Ritzer, George, and Nathan Jurgenson. "Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer'." *Journal of consumer culture* 10.1 (2010): 13-36.
- Robbins, Martin. "The myth that British data scientists won the election for Trump." *Little Atoms*, 30 Jan. 2017, <http://littleatoms.com/news-science/donald>
- Rosvall, M. & Bergstrom, C. T. "Maps of random walks on complex networks reveal community structure" *PNAS* 105, 1118–1123 (2008).
- Rumelhart, D. E. (2017). Schemata: The building blocks of cognition. In *Theoretical issues in reading comprehension* (pp. 33-58). Routledge.
- Saracco F, Straka MJ, Di Clemente R, Gabrielli A, Caldarelli G, Squartini T "Inferring monopartite projections of bipartite networks: An entropy-based approach" *New Journal of Physics* 19, 16 (2017).
- Sarbin, T. R. (1986). The narrative as a root metaphor for psychology. In T. R. Sarbin (Ed.), *Narrative psychology: The storied nature of human conduct* (pp. 3-21).
- Sarbin (Ed.), *Narrative psychology: The storied nature of human conduct* (pp. 3-21). Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.
- Schank, R. & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schmid Wolf (2003), *Narrativity and Eventfulness* in: Kindt and Müller (eds.), *What Is Narratology?*
- Schmid Wolf (2010) *Narratology An Introduction*, De Gruyter
- Shao, C., Ciampaglia, G.L., Varol, O., Yang, K-C, Flammini A., & Menczer F., "The spread of low-credibility content by social bots" *Nature Communications* 9, 4787 (2018).
- Shao, C., Hui, P. M., Wang, L., Jiang, X., Flammini, A., Menczer, F. and Ciampaglia, G. L. [2018], 'Anatomy of an online misinformation network', *PLoS ONE* 13(4), 1–28.
- Shao, C., Menczer, F. and Ciampaglia, G. [2017], 'Hoaxy FAQ: 2017'. URL: <http://hoaxy.iuni.iu.edu/faq.html>.
- Srnicek, Nick, and Alex Williams. *Inventing the future: Postcapitalism and a world without work*. Verso Books, 2015.
- Smith, D. (2017) et al. "Cooperation and the evolution of hunter-gatherer storytelling." *Nature communications* 8.1: 1853.
- Soll, J. [2016], 'The Long and Brutal History of Fake News'.

URL: <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535>

Stella, M., Ferrara, E. and De Domenico, M. [2018], Bots sustain and inflate striking opposition in online social systems.

URL: <http://arxiv.org/abs/1802.07292>

Stover, D. (2018). Garlin Gilchrist: Fighting fake news and the information apocalypse. *Bulletin of the Atomic Scientists*, 74(4), 283-288.

Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A. and Menczer, F. [2016], 'The DARPA Twitter Bot Challenge', *Computer* 49(6), 38–46.

URL: <http://arxiv.org/abs/1601.05140> <http://dx.doi.org/10.1109/MC.2016.183>

Sunstein, C. R. [2009], *Republic.com 2.0*, Princeton University Press.

Tafari, S., Gallone, M. S., Gallone, M. F., Zorico, I., Aiello, V., & Germinario, C. (2014). Communication about vaccinations in Italian websites: a quantitative analysis. *Human vaccines & immunotherapeutics*, 10(5), 1416-1420.

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153.

Thibodeau, Paul H., and Lera Boroditsky. "Metaphors we think with: The role of metaphor in reasoning." *PloS one* 6.2 (2011).

Tuan, Y. F. (1977). *Space and place: The perspective of experience*. U of Minnesota Press.

Tuckett, D., & Nikolic, M. (2017). The role of conviction and narrative in decision-making under radical uncertainty. *Theory & psychology*, 27(4), 501-523.

UK Parliament, (2018), "Russian influence in political campaigns," Disinformation and 'fake news': Interim Report,

<https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/363/36308.htm>.

US Department of State (1987), "Soviet Influence Activities: A Report on Active Measures and Propaganda, 1986-87, August.

<https://www.globalsecurity.org/intell/library/reports/1987/soviet-influence-activities-1987.pdf>.

US Office of the Director of National Intelligence (2017), "Background to "Assessing Russian Activities and Intentions in Recent US Elections,"The Analytic Process and Cyber Incident Attribution," 06 January 2017, https://www.dni.gov/files/documents/ICA_2017_01.pdf

Userfeed Protocol Whitepaper [n.d.], Technical report.

URL: [https://userfeeds.io/Userfeeds_Protocol_Whitepaper_\[Draft\].pdf](https://userfeeds.io/Userfeeds_Protocol_Whitepaper_[Draft].pdf)

Tufekci, Zeynep. (2014). "Engineering the public: Big data, surveillance and computational politics." *First Monday*, 19(7). doi:10.5210/fm.v19i7.4901

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036-1040.

Van der Linden (2017), S., Leiserowitz, A., Rosenthal, S. & Maibach, E. "Inoculating the Public against Misinformation about Climate Change" *Glob. Challenges* 1, 1600008.

- Varol, O., Ferrara, E., Davis, C. A., Menczer, F. and Flammini, A. [2017], Online Human-Bot Interactions: Detection, Estimation, and Characterization.
URL: <http://arxiv.org/abs/1703.03107>
- Waldrop, M. M. [2017], 'News Feature: The genuine problem of fake news.', *Proceedings of the National Academy of Sciences of the United States of America* **114**(48), 12631–12634.
URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5715799>
- Wallace, William. "The sharing of sovereignty: the European paradox." *Political Studies* 47.3 (1999): 503-521.
- Weedon, J., Nuland, W. and Stamos, A. [2017], 'Information operations and Facebook', *Facebook Publication* pp. 1–13.
URL: <https://fbnewsroom.us.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>
- Weng L., Menczer, F. & Ahn, Y-Y. "Virality Prediction and Community Structure in Social Networks" *Scientific Reports* 3, 2522 (2013).
- White, H. (1987). *The Content of the Form: Narrative Discourse and Historical Representation*. Baltimore: Johns Hopkins UP.
- Woolley, S. C., & Howard, P. N. (2017). Computational propaganda worldwide: Executive summary. *Computational Propaganda Project*.
- Wygotski, L. S. (1978). Mind in society. *The development of higher psychological processes*. Cambridge: Cambridge.
- Wyer Jr, R. S., Adaval, R., & Colcombe, S. J. (2002). Narrative-based representations of social knowledge: Their construction and use in comprehension, memory, and judgment. In *Advances in experimental social psychology* (Vol. 34, pp. 131-197). Academic Press.

