

NEWS VS. FAKE IN THE INFORMATION SYSTEM

INTERIM REPORT
SECTOR INQUIRY "ONLINE
PLATFORMS AND THE NEWS SYSTEM"



AUTORITÀ PER LE
GARANZIE NELLE
COMUNICAZIONI

DEPARTMENT OF ECONOMICS AND STATISTICS



AUTORITÀ PER LE
GARANZIE NELLE
AGCOM COMUNICAZIONI

Department of Economics and Statistics

November 2018

TABLE OF CONTENTS

EXECUTIVE SUMMARY	i
INTRODUCTION	1
1. THE NEWS PRODUCTION	6
1.1. INTRODUCTION	7
1.2. THE QUANTITY AND QUALITY OF NEWS PRODUCED IN ITALY	9
1.3. TOPICS OF INFORMATION	19
2. THE PRODUCTION OF DISINFORMATION	23
2.1. INTRODUCTION	24
2.2. IMPORTANCE OF DISINFORMATION PRODUCED IN ITALY	28
2.3. TOPICS OF DISINFORMATION	33
3. THE DISSEMINATION OF REAL AND FAKE NEWS	47
3.1. INTRODUCTION	48
3.2. THE LIFECYCLE AND DISSEMINATION OF REAL NEWS	51
3.3. MODALITIES OF DISINFORMATION DISSEMINATION AND THE LIFECYCLE OF FAKE NEWS: SUBJECTIVE AND OBJECTIVE SCOPE	62
4. THE PROPAGATION OF REAL AND FAKE NEWS ON ONLINE PLATFORMS	73
4.1. INTRODUCTION	74
4.2. THE NEWS CONSUMPTION ON ONLINE PLATFORMS AND POLARIZATION	76
4.3. THE IMPLEMENTATION OF THE DISINFORMATION STRATEGIES ON ONLINE PLATFORMS	85
4.4. THE INTERACTION MECHANISMS WITH THE FAKE CONTENTS ON ONLINE PLATFORMS	91
METHODOLOGICAL APPENDIX	104



EXECUTIVE SUMMARY

The adoption of an innovative method based on the direct observation of millions of data

This Report is a first result of the sector inquiry on “Digital platforms and the information system” and has the purpose of designing industry scenarios in the Italian news system, based on the direct observation of millions of data.

Remembering that the AGCOM mission is the protection of news-media pluralism that inspires the study, the Report has the goal of carrying out an evaluation about the strength of the information system and the performances on online platforms and of other media in the news production and disclosure, as well as in better understanding the dimension, peculiarity, and method of the spread of disinformation.

This study has involved an experimental method, focused on the use and integration of many data sources on the main components of the information and disinformation systems: news publishing, sources of disinformation, journalists and citizens/users of news.

In particular, the overall analysis of the entire ecosystem was carried out on big amounts of data, coming from different and unique *datasets*, mainly attributable to:

- the entire text content taken, by means of the platform (developed by *Volocom Technology*), with which AGCOM was equipped, from almost 35 million of documents created in Italy in a period of more than two years, from 1.800 information sources (television and radio channels, newspapers, websites of traditional editors, exclusively online publishing and relevant sites and *accounts* of social network), and disinformation sources (websites and sites/social *accounts*), considered as such by external people, specialized in *debunking*;
- a sample of almost 700 main news (real and fake), representing all categories of genres, considering a period of a year;
- information gathered in the framework of the 2nd edition of the Agcom Observatory on Journalism, on the methods of news production and contents by a big and representative sample of professionals (almost 2.000) working in Italy;
- the results of a survey on the news consumption, carried out for AGCOM by GfK Italia on a sample of more than 14.000 individuals, representing the Italian population;
- tens of millions of public social accounts, where models of *big data analytics* were applied, in order to examine the methods of news consumption and the mechanisms of interaction of the users through the online platforms. The part of

the Report dedicated to these specific aspects was written on the basis of a research task, given from AGCOM to Walter Quattrociocchi, who has been dealing for years with scientific analysis of the phenomenon and publishing national and international studies.

The Report, founded on such a methodological approach, falls under the supervisory and monitoring activities of the media sector, starting from the evidences gathered by AGCOM in other occasions. So, the study, with consideration to the already carried out relations of interdependence between the information system and the system of disinformation, can only dwell the deep analysis of both components.

In fact, more than one time AGCOM has observed how the most serious phenomenology of disinformation occurs at the national and global level in places where the online news system fails:

- issues of content monetization and the persisting reduction of investments in the news production and distribution; the stressed and connected regression in the use of verification mechanisms in the journalistic profession; the extremely short timing of both the news production and consumption (which risks to affect accuracy from the journalists' side, and lower levels of attention on the users' side) are compromising the suitability of the news supply with regard to reliability, trustworthiness and media coverage. In broader terms, they are at the basis of the spread loss of reputation and trust by citizens to the traditional news system.

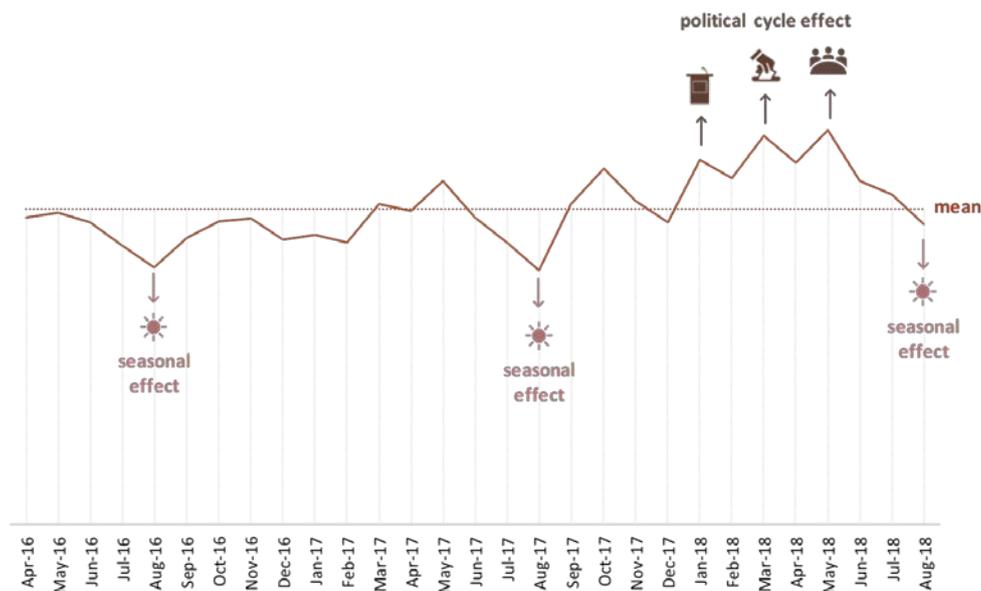
In this context, citizens risk to give trust always to alternative and not-qualified information sources, which are often at the basis of online disinformation strategies.

The evidences on the production and news dissemination

The dynamic analysis on the news quantity produced in Italy shows an increase of the average amount of the information volume placed in the system, reaching its maximum value in 2018. On the other side, the monthly trend shows how it is possible to track the determinants of the “news supply” from the combination of at least two factors, derived by the media:

- the first one comes from a recurring event (seasonality), manifesting in the reduction, the information volume suffers every year in summer;
- the second one is more contingent (as it comes from the political cycle) and was found in the increase of news production during the electoral campaigns, in the development of the last political elections in Italy (March 2018) and in the subsequent formation of a new government.

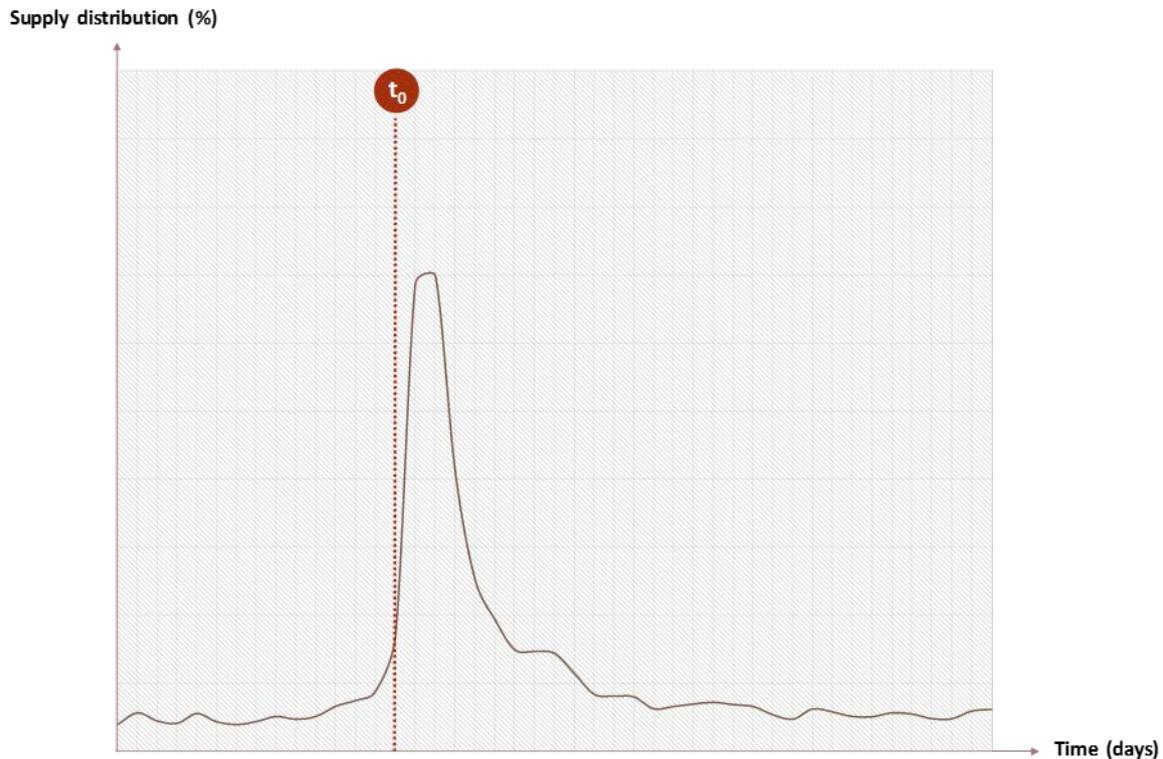
Monthly trend of the news production in Italy



The average daily trend of news contents produced around a single story (news or event), identifies the news lifecycle, which can start exactly on the day (t_0) when the event concerning the news itself occurs. However, the largest disclosure starts from when the greater distribution in the “news supply” occurs:

- the concentration of news contents around a single story, which is potentially subject to an increase by the t_0 , registers its maximum in the two days after the one when the event has occurred, when all media can disclose it.

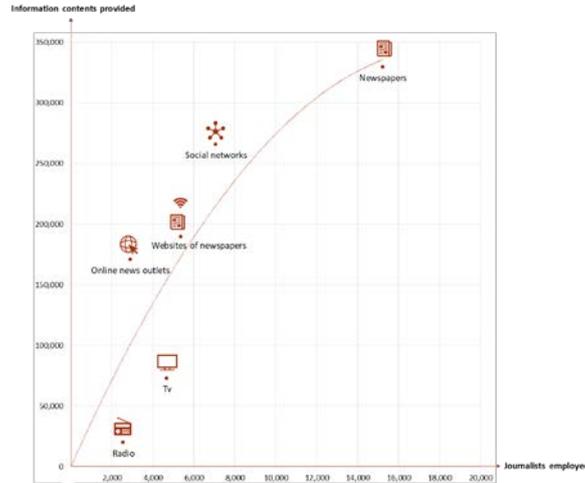
Lifecycle of a news



The compared examination of the quantity of produced information contents and of the used journalistic resources allows also to make a difference between the media:

- from one side, the online information sources (newspaper sites, online publishing, social network), which - in comparison with used professional sources and other conditions - produce a larger “news supply”, stressing an overuse of the used journalistic force, which has negative consequences on the accuracy and deepening of the content generated around a news. In addition to this, in the case of social networks, we can see the rapidity of passing from a news to another, which shows a bigger superficiality in the exposition of the facts, in comparison to the others.
- on the other side, newspapers, TV and radio, representing a lower production intensity of the journalists, used in relation to the quantity of the offered news contents and, so, a bigger expected level of quantity and deepening of the produced information which, in the case of daily publishing and of broadcasters, is accompanied by a disclosure and a daily coverage of the single news higher than the average.

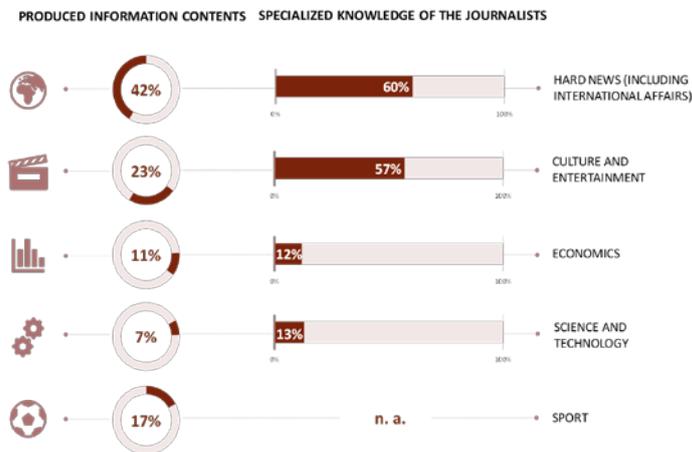
News production function



With regard to the topics covered in the online news, it must be noticed that:

- not only the specialized information is offered in a more reduced way than the more general one, but it is also produced for its biggest part by professional journalists who do not have specialized expertise in the topics of reference.

Variety and level of specialized knowledge of the treated topics



In addition to this, comparing the news supply to the demand, we identify new significant unbalances:

- if for “hard news” (politics, report, international affairs) and soft news (culture and entertainment) there is excess supply, for such categories that require specialized knowledge as “science and technology” there is excess demand.

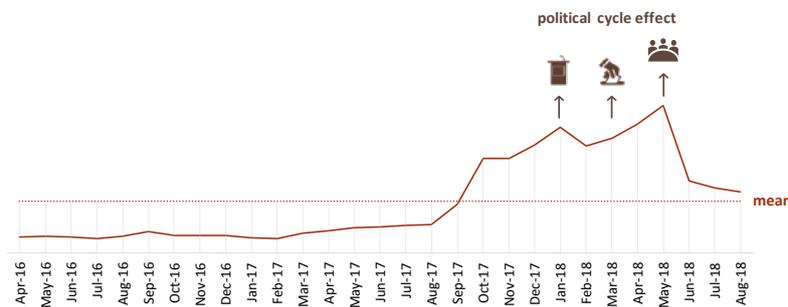
More specifically, the current “news supply” seems to fail to meet the demand of more specialized contents, whereas the request by the citizens is not efficiently satisfied nor quantitatively nor qualitatively with regard to the produced contents.

The evidences on the production, disclosure and propagation of disinformation

The dynamic analysis carried out on the disinformation production showed a national news system suffering the presence of a limited volume of fake contents, which:

- seems to have stabilized on a value averagely higher than in the past,
- reached the maximum level during the last political cycle, on the occasion of the political elections of 4th March 2018.

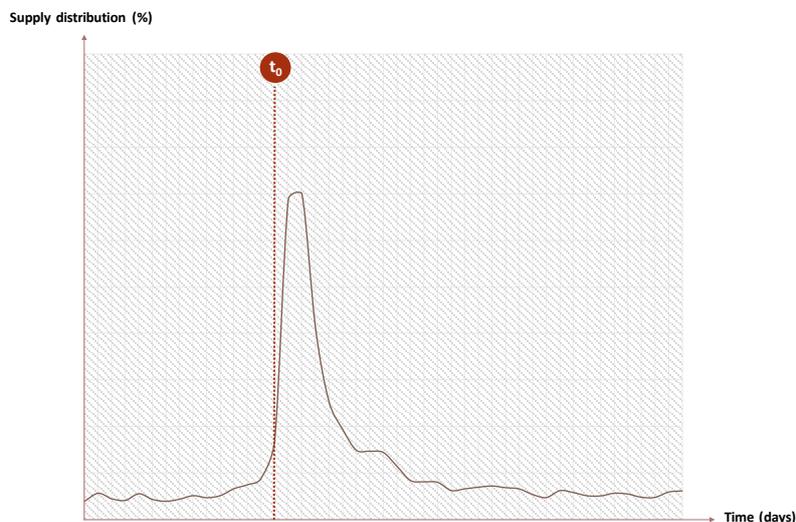
Monthly trend of disinformation produced in Italy



The lifecycle of single fake news is essentially characterized by:

- the almost absence of anticipations (i.e. fake contents disclosed before the t_0) on the facts object of the fake news;
- a sensibly shorter duration compared to the lifecycle of a real news, with a higher concentration around the t_0 , reaching the peak the day after the t_0 , and then lower quickly to values around the zero.

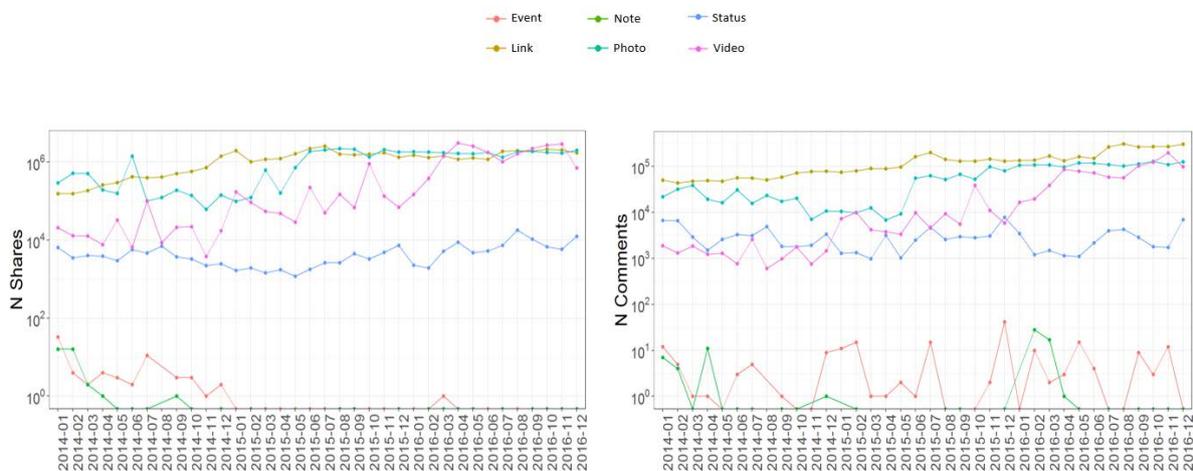
Lifecycle of a fake news



In this context, it is possible to notice that:

- the sources of disinformation (websites, pages and social accounts, detected as such by debunkers) assume a role consisting mostly in conferring an impulse to the fake contents inside the national system. The shortness of the lifecycle of a single fake news disclosed by means of these sources and the concentration in few days after the disclosure of the relevant contents are the indicators inside of carrying out a strategy of disinformation, preferring to treat many different news, avoiding to deep their contents;
- after the triggering, the fake news is put and launched in the system of online platforms, also by means of the unaware contribution of users sharing and commenting it on social networks.

Monthly trend of disinformation posts and relative users' interactions, according to the kind of content

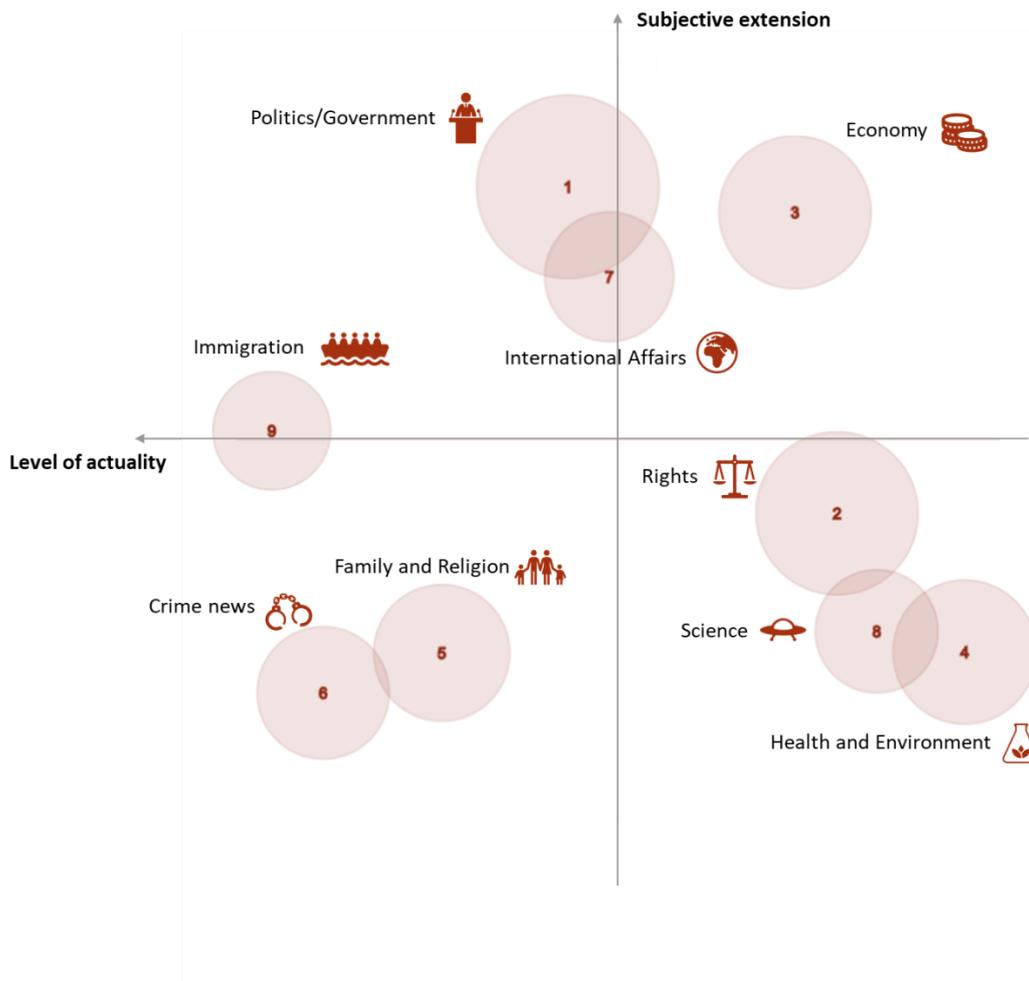


Also the choices of the treated topics by the sources of disinformation is collimating with the intention of activating mechanisms of viral propagation on online platforms:

- in Italy, 57% of the fake contents production focuses on of politics and reports, while almost 20% topics concern the scientific knowledge; all topics presenting a strong emotional impact and potentially divisive;
- the textual analysis of fake contents produced by the disinformation websites in 2018 let detect 9 main topics (politics, rights, economy, health and environment,

family and faith, reports, international affairs, science, immigration), the coverage of which in the news is peculiar: the topics are discussed superficially and in a sensationalist way, with the purpose of stimulating the people's emotional state of mind.

Main topics of the fake contents in 2018



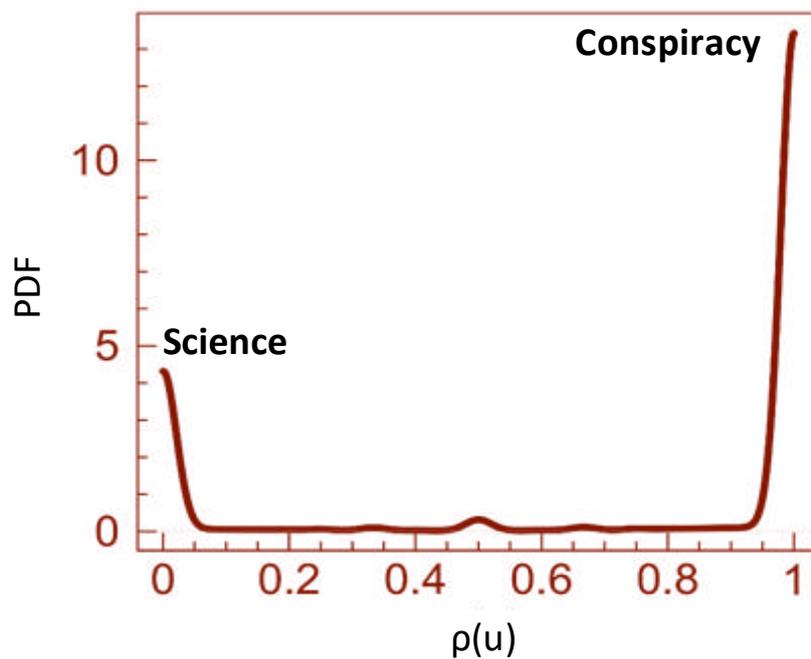
The viralization of fake contents is made possible by the modalities of news consumption on online platforms which, most of all for disinformation news, occurs in the framework of closed distinguished communities:

- the models of news consumption and the interaction of the users with the news on online platforms are characterized by the trend to polarization, selective exposition, homophily and the onset of echo chamber. Therefore, the users tend to select information which is coherent with the own system of beliefs, forming

polarized groups of people with similar ideas on shared narrations, where different news contents are ignored;

- these results show how the process leading mechanisms of acquisition and elaboration of news is due to cognitive phenomena but also by cultural and social practices.

Italian users polarization distribution on Facebook for “science” and “conspiracy”



INTRODUCTION

Media pluralism and the freedom of expression are the essential principles of a democratic society. As the matter of fact, they are an essential requirement for the exercise of the fundamental right of freedom of expression, recognized by the constitutional traditions of the different countries, as well as at European level.

The protection of such principles is one of the funding objectives of the regulatory actions, which AGCOM pursues within its scope and mission, observing a constant technological and market evolution. According to its institutional role, AGCOM has been also having an intense activity of supervision and monitoring of the news and media system. In parallel to this proceeding activity, since 2013 the AGCOM Department of Economics and Statistics has launched observatories¹, hosted events (meetings, seminars, workshops) and promoted other regulatory initiatives, such as the establishment of technical roundtables² on several aspects of the online information panorama, with the aim at promoting, *inter alia*, innovative methods and approaches to the regulatory issues at stake through the collaboration with the national and international scientific community.

The preferred policy approach is the so-called “know-it-first” principle that is based on a deep understanding of phenomena as a precondition to meet the challenges arising from the evolution dynamics of the information markets and by the rise of pathological

¹ These include the [Observatory on communication](#), with a suitable deepening section on media; the [Observatory on journalism](#), focused on the evolution and problems of the journalistic profession; and the [online headings](#), dedicated to the analysis of the sector of the editors operating exclusively on the Internet.

² AGCOM has recently established a [Technical Roundtable for safeguarding media pluralism and news fairness in the online platforms](#), which has the purpose of promoting the self-regulation of the platforms and the exchange of good practices for the identification of the online disinformation strategies. The Roundtable, representing a single example on the world, assumes the participation of the representatives of the online platforms (Google, Facebook, Wikipedia), the most important national editors (of newspapers, radio and TV), of the journalists, of the advertising component, of the category association, including the ones representing the consumers as well as academic institutions and research centers. The purpose of the Roundtable is the promotion of the sharing of information, the comparison, the surfacing of suitable methods of detection and the identification of transparency instruments, as well as of the most suitable rules and forms. In this sense and with reference to the fight against online disinformation in electoral campaign, in the framework of the Roundtable, the [Guidelines for the equal access to the online platforms during the electoral campaign 2018 were adopted](#).

forms of disinformation, AGCOM carries out specific analysis³, surveys⁴ and Reports⁵ having as an object the different components of the online news system: the users (that, in the end, are citizens), the editors (of all mass media), the journalists (in every aspect), the contents (with the relevant characteristics in terms of quality and typology of offer), the sources of financing and the relevant business model (both traditional and innovative ones), the new platforms (social networks and search engines) and the mechanisms of creation and disclosure of news (including algorithms).

In these activities, there is also the sector inquiry concerning “[Digital platforms and information system](#)”, and in particular this Report, which is a first result in this sense. The Report has the purpose of examining the online news system by an unusual perspective, starting from the static and dynamic analysis of the information produced in Italy, both under the profile of quantity of news contents both the relative disclosure by a mean, and under the one of the quality of news and of the treated topics. The analysis combines “input” data, about journalists and their activity, and information on news “output” (Chapter 1).

In other terms, the main object of the AGCOM study in this case are the news content and their production modalities. The focus is put not only on the news as products in the strict sense, concerning news and real facts, but also about fake, unfounded, manipulated or not truly reported news (for a definition of these, see the recent Report by AGCOM, “[Le strategie di disinformazione online e la filiera dei contenuti fake](#)”). As a consequence, a specific analysis is dedicated to the disinformation, to the volume and variety of contents generated by fake news active in Italy (Chapter 2).

The following part of this Report (Chapter 3) regards strictly the aspect following the news production contents, such as the lifecycle (input moment and persistence inside the information system), the modalities of treating and disclosure of (real and fake)

³ See for example the activities of [TV and radio monitoring on political and social pluralism](#), of the analysis of the data concerning the [edition of the daily headings](#), on monitoring of the media markets by means of the [Economic System Information \(ESI\)](#) and the [Register of Communications Operators \(ROC\)](#).

⁴ See the survey by AGCOM on the “[Settore dei servizi Internet e sulla pubblicità online](#)” (2014), on “[Informazioni e Internet in Italia. Modelli di business, consumi, professioni](#)” (2015) and the one which are now ruling in matter of “[Informazione locale](#)” and of “[Big data](#)” (of which AGCOM has published the first evidences in an [interim report](#)).

⁵ See, among the others, the Report “[L'informazione locale in Italia: il ruolo del servizio pubblico](#)” (2017) and the Reports published by AGCOM on the request of information: “[Il consumo di informazione e la comunicazione politica in campagna elettorale](#)” (2016) and more recently, “[Rapporto sul consumo di informazione](#)” (2018).

news. And right the comparison between real and fake news considering the modalities of disclosure of the contents show significant peculiarities of them.

For this purpose, the fourth and last Chapter focuses on the news contents propagation and, most of all, on the production of fake ones. In particular, disinformation feeds on elements like the polarization and viralization, characterizing the news consumption of citizens on online platforms (and especially social networks).

A study like this assumes the use of an experimental method (of which a detailed description is proposed in the Appendix), focused on the use and implementation of several data sources on different components of the information and disinformation sources: information headings, sources of disinformation, journalists and citizens/users of news.

The first data source is represented by an enormous database, composed by the entire textual content of about 35 millions of documents created in more than two years (29 months, from April 2016 till August 2018), by 1.800 information sources referable to TV and radio channels, newspapers, websites of traditional editors, exclusively online headings and relevant accounts of social network, as well as websites (and social accounts) included in the lists of fake news⁶, written by sources specialized in *debunking activities*. For the creation of the databases, where most of the following activities were done on the platform developed by *Volocom Technology*, which offers the possibility of extracting the repetition of every key word (or group of words) and of examining entire documents regarding news on every topic or fact.

In order to make a rigorous analysis on the disclosure of information and disinformation, a sample of about 700 main (real and fake)⁷ news was found, which represent all categories of treated topics, considering a period of a year.

In addition to this, the study considers information acquired by the AGCOM Department of Economics and Statistics on the modalities of the news production by a broad and

⁶For the detection of these sites, without prejudice that the role of AGCOM does not include the verification whether a news is real or fake, reference was made on the lists proposed by external sources (in particular, debunking sites, such as *butac.it* and *bufale.net*) used by several scientific studies on online disinformation (see among the others, M. Del Vicario, W. Quattrociocchi, A. Scala, F. Zollo (2018), "Polarization and Fake News: Early Warning of Potential Disinformation Targets", *arXiv preprint arXiv:1802.01400*).

⁷Also in this case, for the purpose of the identification of the sample of fake news to be included in the analysis, reference was made to external sources, in particular to articles on debunking published on specialized sites, such as *butac.it*.

representative sample of journalists working in Italy⁸ (about 2.000 professionals, equal to more than 5% of the relative reference population).

With regard to the news consumption by citizens, AGCOM used the results of a survey based on a sample of more than 14.000 individuals, representative for the Italian population⁹. This survey regards the news consumptions by the Italian citizens on all media (TV, radio, newspapers, websites, search, social). On the other side, with reference to the modalities of news consumption by online platforms, the results come from models of big data analytics, carried out on ten million of social accounts (see Chapter 4).

In this sense, the method uses big amounts of data coming from different and unique datasets, implementing them in an overall analysis of the entire information ecosystem, both in the production phase of the information - in the double meaning journalistic activity both in the one of offering the final product - and in the one of consumption.

For the implementation of suitable methods for the treatment of big amounts of data, the Agcom Department of Economics and Statistics, was supported by the scientific collaboration of the Università degli Studi di Napoli “Federico II”, according to the Framework Agreement binding the two institutions¹⁰. In addition to this, the part considering the news consumption and online disinformation in online platforms (according to Chapter 4) was made on the basis of a research task assigned by AGCOM to [Walter Quattrocchi](#)¹¹, who has been analyzing the phenomenon for several years, publishing national and international studies.

In the following pages, we will illustrate the results of the mentioned innovative analysis, with the purpose of delineating - thanks to the direct observation of millions of data - the information scenarios of the Italian system. Considering the protection of pluralism as an explorative motivation, the Report has the goal of carrying out an evaluation about the robustness of the news system, and the performances of the different media in the production and disclosure of information, as well as the one of making a further step towards the comprehension of the extent, the peculiarities and modalities of disclosure

⁸ Data of the 2nd edition of the [Observatory on Journalism](#), see

⁹ See the [“Rapporto sul consumo di informazione”](#), see

¹⁰ In particular, we thank Professor Maria Gabriella Grassia for the scientific collaboration in the definition of the analysis methods on the disclosure of real and fake news. Moreover, we thank Professor Marina Marino and Dt. Alessandro Grieco.

¹¹ See [Results of the comparison procedure for the assignment of an individual research task on “Informazione e piattaforme digitali”](#)

of disinformation. The relevant evaluations are essential in order to orientate correctly the regulatory and policy path of AGCOM as guarantee of fair, free and plural information in the news media system.

THE NEWS PRODUCTION



1

1.1. INTRODUCTION

In general, the news production identifies the process of creating - and offering to the public - news contents having as an object facts, happenings, events, in other words news of every kind. So, it is process from which the quantity, variety and quality of information depends, which reaches the citizens, on the basis of which they form their own opinions and points of view.

The news production in every media (TV, radio, newspapers, Internet) occurs by means of editors and professional figures, employed in the different editorial structures, first of all the journalists.

In details, the editors, owners of information sources, finance production and news distribution by means of business models, expecting only in some cases (for example for newspapers) the sale to the final users of their publishings. Most of all, the sale of advertisement blocks to an advertiser is the main part of the financing source, which can be added to some forms of public and private contributions¹².

Journalists are the main productive factor (“input”) used in the process of producing news. The journalists’ work includes the finding, analysis, deepening of news, which is followed by the actual composition of the information content (articles, television and radio services, but also posts/tweets on online platforms). The journalistic profession has also its own peculiarities, connected to the media use, which comply the features of the information content offered on the media itself¹³.

Under the technical profile, the journalistic output can be addressed to different tools (printed paper, television, radio, pc/tablets/smartphone) and have different formats, also mixed (text, video, audio, Figures). Moreover, every news content has to adapt to the place, made available on the media. As a consequence, the news products can have differences in the length, frequency of updating, style and graphic set out.

Under the content profile, the information products are characterized also by an editorial line, grade of deepening, particular language.

¹² In 2017, in the television sector, the advertisement was the first income source with 41% of the total revenue; percentage which increases to 43% for newspapers and to 78% for the radio. See “[Relazione Annuale Agcom 2018](#)”.

¹³ For a detailed description of the job, including the activity of journalists, of the used means and instruments, of the necessary competences, see the [Observatory of journalism](#), see

Finally, under the economic profile, news take the features of a “meritorious good” (i.e. worthy, also in the absence of market failures, of a special protection by the State), of a “creative product” (containing a substantial creative effort), as well as of “good of experience” (being able to be evaluated only after its actual consumption). Moreover, the news have assumed in time the connotation of “(almost) public good”, having the two characteristics of the absence of rivalry and - always more - of the excludability in the consumption. As the matter of fact, the digital revolution made the information content a good which can be reproduced infinite times, consumed in group, often freely, losing it from a physical support and from the binds of lacks¹⁴.

Considering the features of the news production process, and of the technical and economic nature of the news product, in the next part of the Chapter, we will present and discuss the result of the analysis carried out on the quantity, quality and variety of the information produced in Italy by all the mass media.

¹⁴ See Agcom, (2015), “[Informazione e Internet in Italia. Modelli di business, consumi, professioni](#)”, Chapter 1 - L’informazione.

1.2. THE QUANTITY AND QUALITY OF NEWS PRODUCED IN ITALY

In a media pluralism and diversity perspective, considered that it is necessary that the news produced can actually reach the public, including all the people using media to seek information¹⁵, it is widely assumed that the most quantity of information available in a news media system the better for citizens and democracy¹⁶.

In this sense, the news quantity (and, at least potentially, available for all citizens) represents a first indicator of the state of mind of the information system of a country. A bigger quantity of offered contents can be an indicator (or at least increases the possibility) of increased transparency, plurality, coverage of facts, topics and visions, offering to the citizens a better opportunity of learning and comparing different news and sources, also within the same media.

An analysis of the characteristics of the information produced in the Italian information system cannot ignore a preliminary reconnaissance on the quantitative aspect and on the possible factors, regarding the process of production of the information.

A first study is indicated in the Figure 1.1, showing the dynamic evolution of the volume of information produced in Italy in the last three years. In the carried-out analysis, we considered the overall amount of information documents produced every month by all media of communication: newspapers and television channels, radio channels and websites of information (editors active exclusively online and editors acting also on different media), as well as online platforms (sites and accounts of editors and influencers). In detail, the information documents give to every article (in the case of newspapers and websites of information), television and radio service, posts/tweets (in the case of online platforms), issued by about 1.800 sources of information (see METHODOLOGICAL APPENDIX).

Moreover, reference was made to the information in this more general meaning, including not only the information products with object the so-called hard news (like the ones about reports and politics, which - because of their importance- require the immediate publishing), but also the so-called soft news, i.e. that news with topics with a lower grade of information urgency and which do not require a prompt publication.

¹⁵ On the problems connected to the risks of exclusion and medial marginalization of certain parts of population, see Agcom, "[Rapporto sul consumo di informazione](#)", see

¹⁶ To this regard, see A. Prat, D. Stromberg, (2013), *The Political Economy of Mass Media*, in a cura di D. Acemoglu, M- Arellano, E. Dekel, *Advances in Economics and Econometrics: Volume 2, Applied Economics*, Cambridge University Press.

Therefore, the analysis included all categories of produced information: report, politics, economics, international affairs, science and technology, culture, shows and sport.

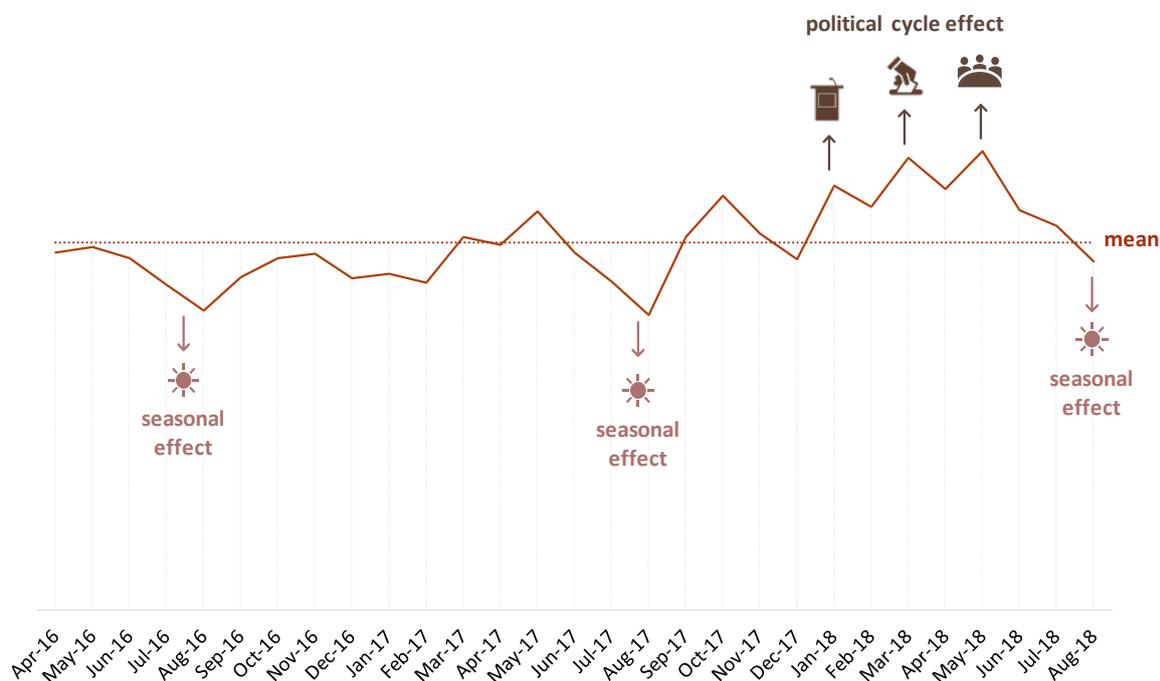
On the basis of the above mentioned method premises, Figure 1.1 shows a trend of the produced information, which, if it does not seem constant in every month, is not subjected to significant variations (upwards or downwards) in comparison to the average of the period. To this regard, it must be stressed that, especially for some media, there is bounds (of space, time, realization and publication of the product, ...), there is the existence of flexibility of adjustment (upwards) of the offered quantity in comparison to the request and to the exigence of require coverage of facts, which can occur every day. In any case, if we consider the annual trend, we observe an increase of the average amount of information put in the system, reaching the maximum value in 2018.

On the other side, the monthly dynamic analysis shows hot it is possible to track the determinant elements of the actions of at least two factors on the quantity of the “news supply”, produced by the media: the first one with recurrent nature (seasonality) and the second one with contingent character (political cycle).

In the first case, it consists in the negative effect suffered every year by the information volume in summer, in particular in the month of August, when there is a lower request by the public, by the closing of different public and corporate structures and by a less intense engagement of the working force in different sectors, the editorial one included.

A second effect, with an opposite mark, occurs during the start of a new political cycle. As the matter of fact, in the period of the electoral campaign, the last Italian political elections (in March 2018) and the formation of the new government (occurred in May 2018) the highest peaks of produced information are registered. This evidence shows the big importance given by the media to politics and to institutional happenings, also in order to meet the citizens' exigencies, which create their own opinions from the media, transforming them in their votes. It is enough to think that, as detected by AGCOM in the last [Report on the consumption of information](#), 85% of the Italian adult population declares to look for news on politics in the media.

Figure 1.1 - Monthly trend of the news production in Italy



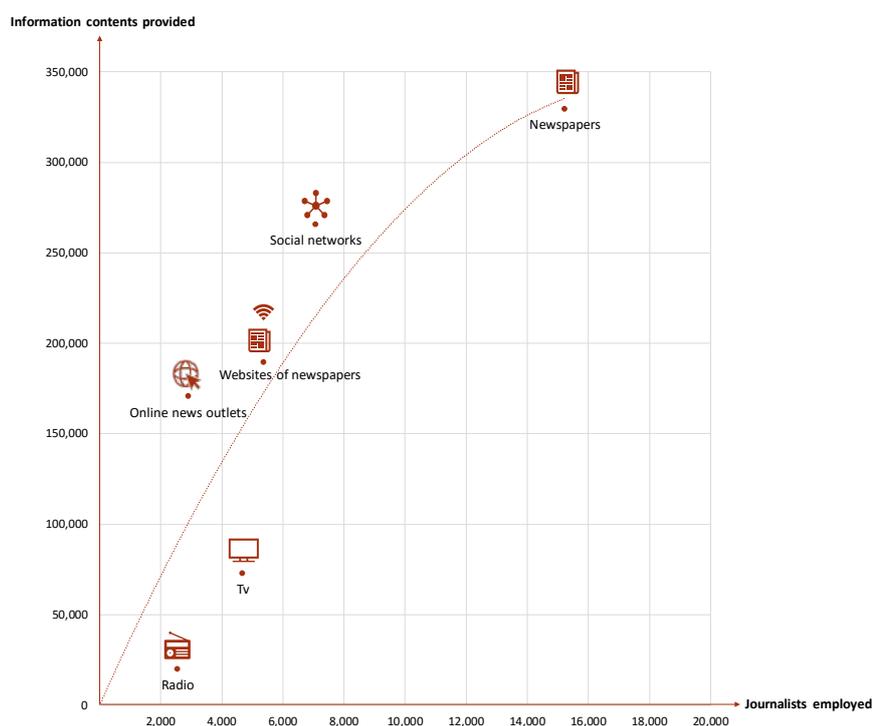
Source: Agcom elaboration on Volocom data

If we want to make a difference among media, considering only the quantitative point of view, Figure 1.2 shows the different contribution of the media to the national information system. More specifically, the biggest contribution in terms of produced output comes from (national and local) newspapers, which - without considering the advertisements, offers contents completely dedicated to the general or specialized information. It is followed by the Internet, which includes the editorial and social component. A smaller quantitative contribution comes from television channels and - even at a smaller level - from radio broadcasters, for which the information contents are only a part of the proposed offered, in favour of entertainment programs.

However, the quantity values, in particular when the product in question is the information, has to be considered from a broader point of view, which considers also other aspects of the production, such as the editorial efficiency and the quality of the offered contents. In order to meet this analytical exigence, in the Figure 1.2, an evaluation was made concerning the function of production of the information, understood as the curve which, in comparison with other conditions, expresses in every point the relation between the used unities of productive factor (the input, on the

horizontal axis) and the offered amount of information contents (the produced output, on the vertical axis). The evaluation of the production of the information, more in particular, was taken from the interpolation of the known values of the produced information contents and journalists employed averagely in a month by every media.

Figure1.2 - News production function (evaluation, monthly average values)



Note: for TV and radio, national broadcasters were considered. For these media, the amount of the offered information contents, considering the programming hours dedicated to newscasts and other programs of information is calculated on the basis of the average duration of a service. For newspapers and information websites, the information contents are identified with published articles, while for social networks to published posts/tweets.

Source: Agcom elaboration on Volocom data and corporate ones (for the offered information contents), and *Osservatorio Agcom sul giornalismo - II edizione* (for the used journalists)

First of all, it is possible to notice that the function of news production is increasing and has decreasing returns of scale, for which an increase of the number of employed journalists determines a less proportional increase than the quantity of the produced information contents. It follows that, in line with what AGCOM found in the framework of the sector inquiry on [“Informazione e internet in Italia. Modelli di business, consumi, professioni”](#), the perfect minimum size in an editor – where, in full use conditions of the productive capacity, the average unit cost of the input is minimum (it does not decrease

according to the increase of the size) – takes a low value (which decreased in the last years¹⁷ and changes according to media¹⁸), providing the use of reduced number of journalists.

Secondly, the positioning of the points identifying the combinations of input/output of the single means in relation to the curve of the function of production delivers indications about the use intensity of the journalists for the news production (see also Figure 1.3) and so, indirectly, also on the qualitative level of the created information content.

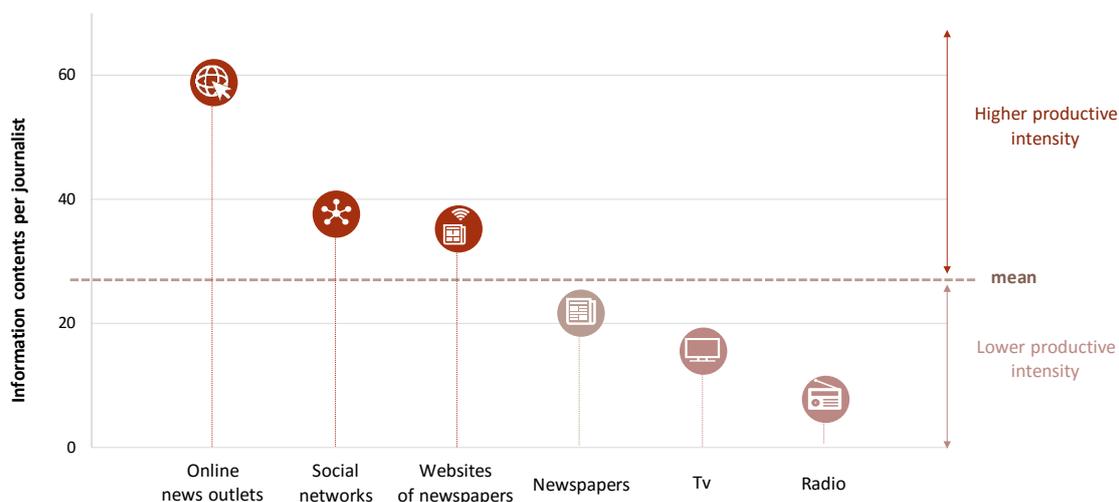
To this regard, it is possible to distinguish three kinds of media, according to the fact if they are in line, over or under the function of productions.

More specifically, the information sources referable to Internet (websites of newspapers, online headings, social networks) are over the production curve and so, in comparison with the employed journalists and with other conditions, they produce a higher offer of information. This circumstance shows an overuse of the employed productive factor (journalists), having a negative effect on the quality of the final product. In this case, the journalists, producing the information contents, should harmonize the development of the own tasks, on the one hand with the quickness of the updating time required by the online information and, on the other one, with the logics of monetization of online contents, which are often based on the attraction of a higher number of clicks. The short timescale becomes practically a lower attention in the collection of pieces of information and in the verification of the sources, as well as in a lower deepening level, to the detriment of the quality itself of the offered information (see also *infra* par. 3.3).

¹⁷ In this sense, “*The reorganization of the information system regarded also the organizational complex aspect and so - at editorial level - the editors. In the whole world, the reorganization brought to the birth of new professional figures (such as web designers and web editors), who work with and support the journalists, the overcoming of some traditional figures, and in particular the reorganization of all professional figures working in the production of the information product. This implied an overall rethink of the organization of the information chain (which redefined the editors also from the architectural and logistic points of view) which was accompanied by a process of downsizing of the productive structures. The editors are on the one hand downsized, while on the other one the incorporated in a single place the different modalities of a reference group.*” (Agcom, “[Indagine conoscitiva Informazione e internet in Italia](#)”, see par. 513, p. 215).

¹⁸ “...the average dimension [of editors] goes from 53 employees (of which 16 journalists, 35 external collaborators and 2 other professional figures) in newspapers, to 21 in Tv and even 9 in radio”. (Agcom, “[Indagine conoscitiva Informazione e internet in Italia](#)”, see par. 514, p. 215).

Figure - 1.3 Productivity of the Italian journalists in an average month



Note: the productive intensity of the journalists is calculated for each media, with the ratio between the offered contents and the employed journalists in the average month.

For TV and radio, national broadcasters were considered. For these media, the amount of the offered information contents, considering the programming hours dedicated to newscasts and other programs of information is calculated on the basis of the average duration of a service. For newspapers and information websites, the information contents are identified with published articles, while for social networks to published posts/tweets.

Source: Agcom elaboration on Volocom data and corporate ones (for the offered information contents), and *Osservatorio Agcom sul giornalismo - II edizione* (for the used journalists)

A second group of media includes the television and radio broadcasters, which are under the function of production, having an opposite situation than the online sources. The journalists have a lower production intensity, even if on the one hand (because of their not only informative nature of the media) can be attributed to their assignment but also to different tasks than the journalistic routine actions, and for another part it can be an indicator of higher accuracy and deepening of the transmitted information.

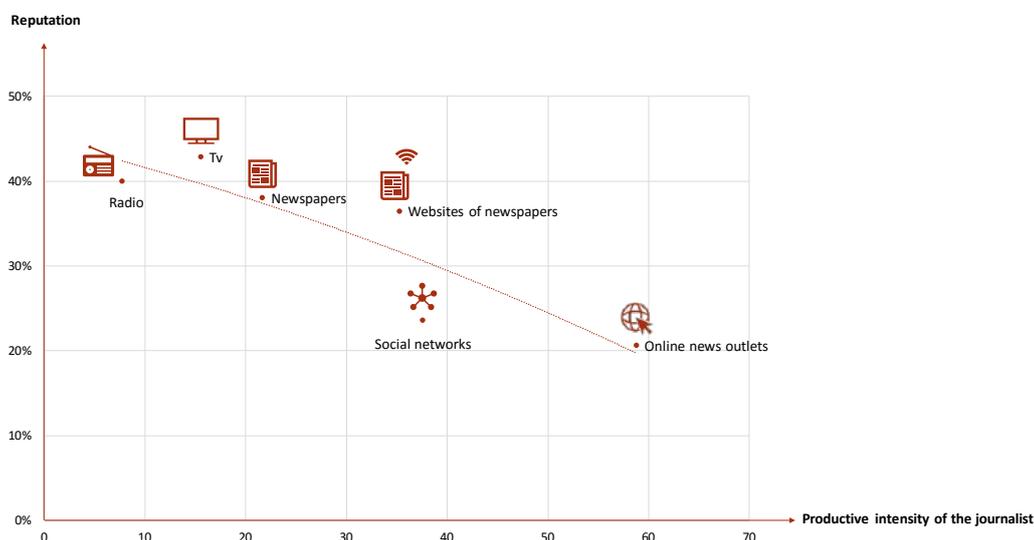
Differently from the previous media, the positioning of newspapers is almost in line (somehow below) with the production function. Newspapers are therefore the means with a productive intensity nearer to the average, as it is possible to see in the Figure 1.3. Inside the mean, journalists are employed essentially in the most typical tasks of the profession (as newspapers thought for information) and do not suffer the short timeline set by the necessity of constant publishing in the day, having the possibility of dedicating a higher attention to the qualitative aspect.

The mentioned considerations are further confirmed by the relationship between the productive intensity (as number of information contents by media of products in a

month) of the journalists employed by a media company and the reputation of the media itself among the public, i.e. the reliability felt by those who use it to find news. The grade of reliability recognized to every media, which can be considered an indicator of satisfaction by the users about the consumptions they actually do of the information disclosed by it. Since, as said before, the information is a “good of experience”, whose value can be determined only after the use of it, a subjective evaluation about the availability of an information source directly reflect the level of perceived quality.

Figure 1.4 shows clearly how the increase of the productive intensity of the journalist, the quality perceived by the disclosed information decrease with the traditional media, that, against a lower offered quantity of information, maintain the highest reputation, and on the contrary, the online sources, producing the highest number of information contents have the lowest reliability.

Figure 1.3 - Relationship between productive intensity of journalists and media reputation



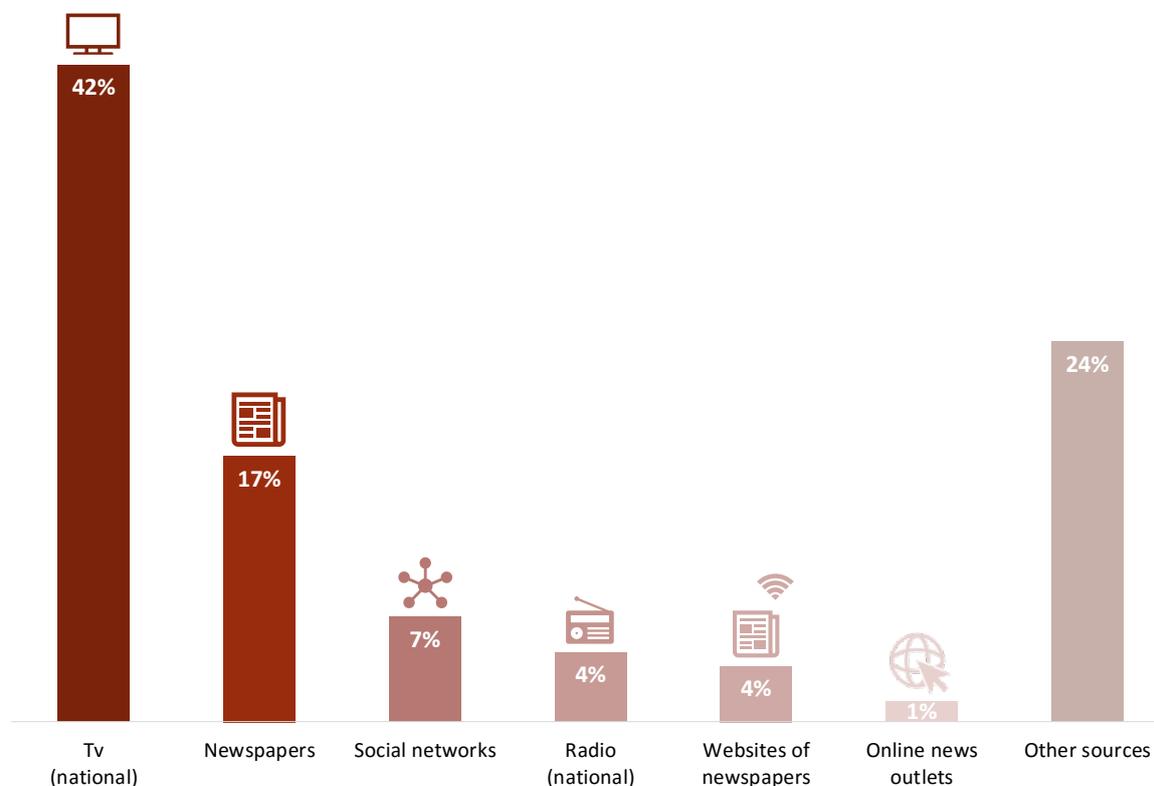
Note: the productive intensity of the journalists is calculated for each media, with the ratio between the offered contents and the employed journalists in the average month.

The reputation of the media is calculated as a percentage of the users of the media for informational purposes, who consider it as “reliable” or “very reliable”.

Source: Agcom elaboration on Volocom and corporate data; *Osservatorio Agcom sul giornalismo - 2nd editions* and AGCOM (2018), *Rapporto sul consumo di informazione*

When individuals decide to access media to get informed, it is possible that they do this because they consider the source reliable, on the basis of the reputation of which this last one and the own personal experience. It is not a case that in the research of information, the source which the individuals consider the most reliable is often the traditional one. In the last [Report of information consumption](#) published by AGCOM (see Figure 1.5), it is possible to notice that the (national) tv broadcasters are considered the most source to get informed by 42% of the population, followed by newspapers, indicated by 17% of the citizens, whereas the online sources have clearly lower satisfaction level, between 1% and 7%.

Figure 1.4 - Main information sources among the Italian citizens (% population)



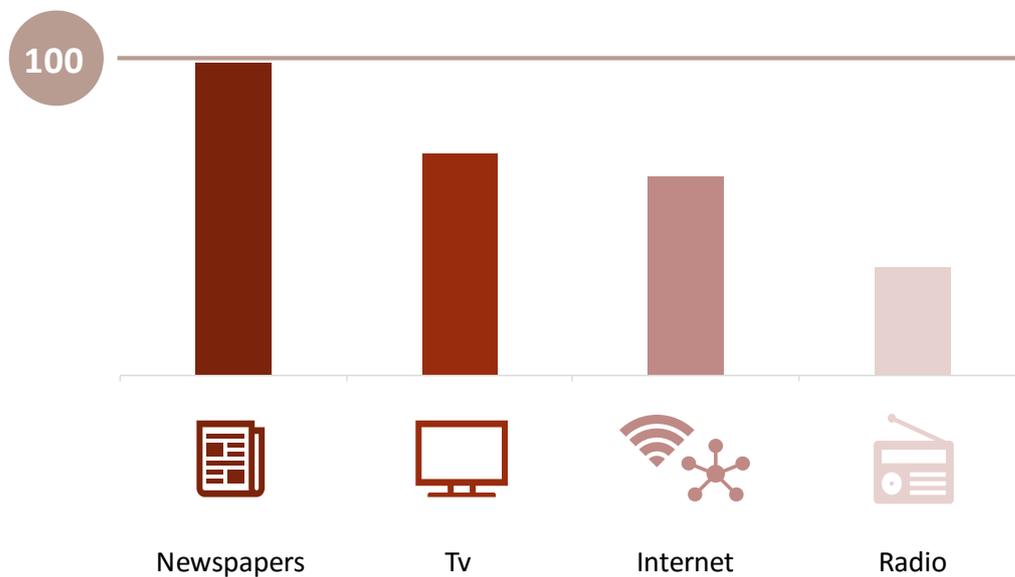
Note: among the other sources of information, you find local television and radio broadcasters and online sources, as research engines, news aggregator and portals, blogs.

Source: AGCOM (2018), *Rapporto sul consumo di informazione*

Even if we consider the actual consumption, which is attentive and not superficial, the informative value given to the newspapers in particular and to television emerges clearly. Figure 1.6, which indicates an evaluation of the actual consumption of information, calculated as ratio between the percentage of those considering a means the most important source to get informed and the ones having access to the same media for informational purposes, has for newspapers a value next to the maximum one (100), showing the big care and reliability, the readers give in reading the articles.

On the contrary, with the always more importance of Internet as information mass media, we always associate low levels of attention and reliability, so much so that an actual consumption of information on the media regard less than two thirds of the users getting informed online.

Figure 1.5 - Evaluation of the actual news consumption



Note: the actual news consumption is estimated how the relationship between those considering a means the most important source to get informed and the percentage of those having access to it for informational purposes.

Source: AGCOM (2018), *Rapporto sul consumo di informazione*

1.3. TOPICS OF INFORMATION

In order to design in a more exhaustive way the framework of the features characterizing the news production in the Italian system, a further aspect to be studied regards the variety of genres and topics treated by the media and offered to the public.

To this regard, the analysis of millions of information contents produced by communication media allowed to track the distribution of the information according to five telematic categories: “*hard news*”, “culture and sport”, “economics”, “science and technology”, and “sport”.

More specifically, (see Figure 1.7), it is possible that, in an average month, more than 40% of information produced in Italy regards *hard news*, i.e. report (as the one connected to current happenings, for example natural events, and to criminal or judicial events: crimes, homicides, robberies, violence, incidents, ...), politics (including news referred to elections referendums, to institutional or politic questions) and facts of international importance (wars, human rights, international politics, ...).

After that, we find categories more connected to the entertainment, such as “culture and show”, representing almost a fourth of the overall informational volume, and “sport”, which is 17% of offered information.

The lowest contributions in terms of quantity of contents put in the information system is detectable in the typologies associated to a more specialized information on the one hand and “science and technology” on the other one. These are information categories which concern topics with a certain relevance, suitable to affect the ideological sphere and decisions on investing and consumption of citizens. The economical information affects in fact topics like corporate transactions, finance, macroeconomic performance; the information on science can include among the others also medical information (vaccines, medical discoveries, therapeutic treatments ...), news on environment and climatic change, news on technological innovations.

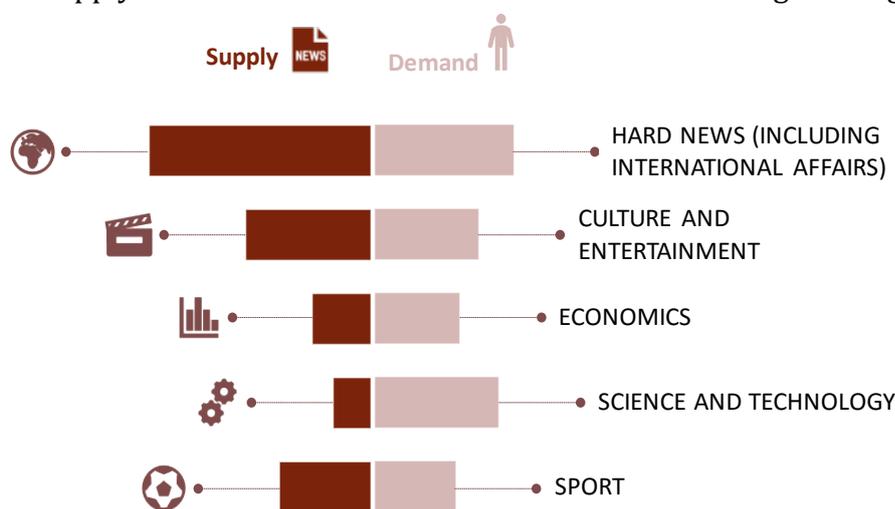
In this sense, from Figure 1.7 you can understand that the categories with a bigger relevance (next to 60%) between the qualification and the topics treated by the journalists every day of their profession are the ones concerning the “*hard news*” and “culture and show”, categories for which, as we have seen, there is also the highest volume of offered contents.

At the same time, it must be noticed that treating topics like “economics” and “science and technology”, which require a higher reservoir of specialized skills, is the one where the presence of specialized journalists is less significant (somewhat more than 10%).

Therefore, not only the specialized information is offered in a more reduced way than the more general one, but it is also produced for its biggest part by professionals, who do not have specialized competences in the topics of reference.

In addition to this, comparing the distribution of the “news supply” to the demand one, we identify new significant unbalances (see Figure 1.8): if for “*hard news*” and news of “culture and shows” there is an excess of offer, the categories linked to the specialized offer, “science and technology” in particular have an excess of request.

Figure 1.7 - Supply and demand distribution of the news according to category



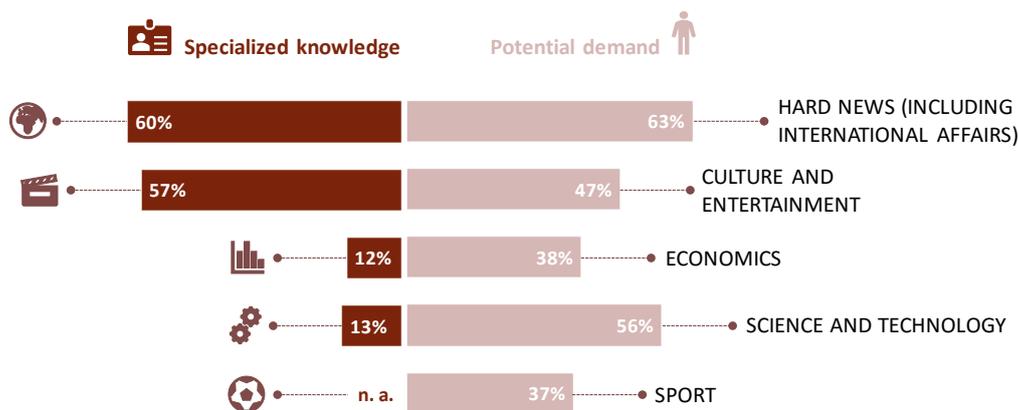
Note: the category “hard news” include news on report, politics and the one of international importance.

Source: Agcom elaboration on Volocom and corporate data (for the offer) and Reuters Institute for the Study of Journalism, *Digital News Report 2017* (for the request)

The current “news supply” seems to lack in particular the offer of more specialized contents or of new interests, whereas the request by the citizens is not efficiently satisfied nor quantitatively nor qualitatively with regard to the produced contents.

As Figure 1.9 clearly shows, the level of specialized knowledge for information contents concerning economics, finance, science and technology seems to be not enough. A potential request, equal to 56% of the population in case of news on science and technology and to somewhat more than 40% in the case of news on economics, corresponds to a low qualification of the offer, for which the specialist topics are often treated by journalists without a specialized general formation. And this problem is also more present in a moment where the journalists’ job is constantly judged by the readers, who have always more specific professionalisms and interests.

Figure 1.8 - Level of specialized knowledge of journalists and potential demand according to category



Note: the category “hard news” include news on report, politics and the one of international importance.

The level of specialized knowledge of journalists is given by the percentage of those with a specialist training (understood as university studies) concerning a certain number of journalists producing contents with as an object the same topic.

The potential request for every information category is calculated as a percentage of the population declaring to be very or extremely interested in news of that category.

Source: *Osservatorio Agcom sul giornalismo - II edizione* ed elaborazioni Agcom su dati Reuters Institute for the Study of Journalism, *Digital News Report 2017*

THE PRODUCTION OF DISINFORMATION

2



2.1. INTRODUCTION

When we speak about disinformation, we refer to the disclosure of fake, unfounded, manipulated or not truly reported information contents, created in order to look like plausible in the media context.

As much as the disinformation could regard all media and the disclosure of this kind of contents have always been considered a possible way of affect the public opinion, the revolution of Internet (of online platforms, in particular) was the cause why this phenomenon became so widespread.

So, the analysis reported in this Chapter are focused right on the examination of the online production of disinformation. In the recent Report "[Le strategie di disinformazione online e la filiera di contenuti fake](#)", written in the framework of the activity of the [Technical board for the guarantee of pluralism and of the correct information on digital platforms](#), AGCOM defined online disinformation and found some distinctive elements. These are elements concerning both the objective (i.e. regarding the object of the transmitted content) and subjective sphere (i.e. the individuals involved in the creation, production and disclosure of the content) and interest all phases of the chain, leading to the putting into practice of the disinformation strategy.

In brief, the existence of online disinformation is characterized by the presence of six main elements:

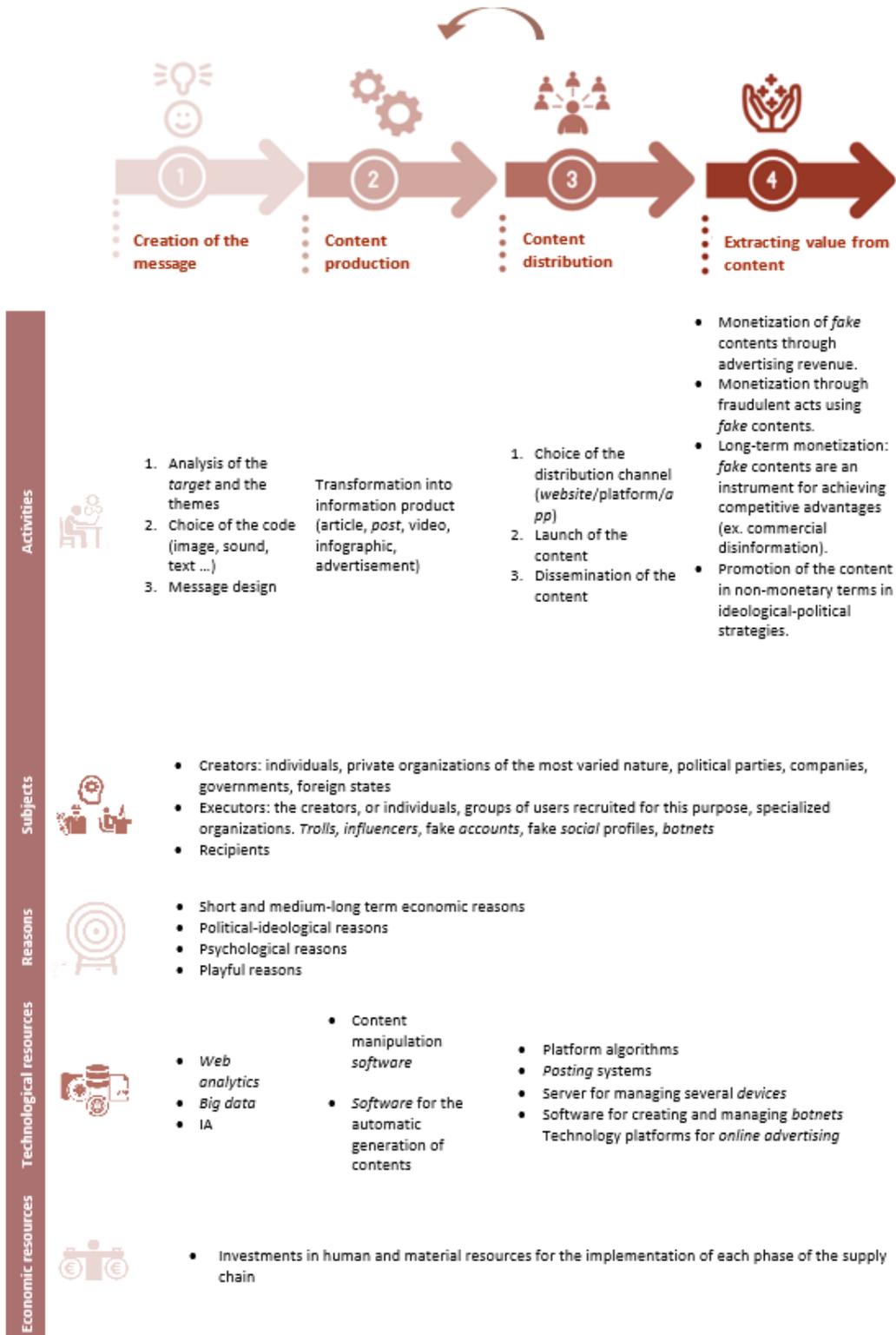
- fake contents;
- the aptitude of transferring users' emotional states or perception or affecting the receivers' behavior;
- the malicious intent of their creation;
- the political/ideological or economical motivation of those who create and disclose them;
- their massive disclosure;
- the aptitude of creating an impact for the information pluralism (so of generating effects on the building of the citizens' opinion)¹⁹.

The entry into the information system of fake contents occurs essentially in three steps (see Figure 2.1), which are the creation the message to be transmitted, the production

¹⁹ On the basis of the mentioned elements, we clearly exclude every satirical elements or the fake ones, which do not have any criminal intent and which often characterized the conversation on socials. This kind of fake contents is part of the concept of mis-information, see Chapter 2 of the Report "[Le strategie di disinformazione online e la filiera dei contenuti fake](#)", see

of the contents where the message is incorporated and transformed in an information product and at last its disclosure (after these phases, you find the valorization).

Figure 2.1 - The chain of fake contents



In the phase of creation, the message is elaborated which is to be transmitted by means of the fake content; this has different features according to the purpose of the creators and according to the target of the disinformation strategy. In general, the message is thought in order to activate the audience which is thought for, involving it also in the further disclosure of the content.

In the phase of production of the content, the message is transformed in an information product, which can have the form of a text (for example a post or an article), of a Figure, of a video or of a combination of these.

At last, in the phase of disclosure, the fake content is published online and so made available. The disclosure is carried out by means of one or more channels, typically a website or an online platform (most of all of social networks) and allows to put the fake content in the chosen media context (which is the network of contents circulating on different online and offline media around and/or together with the fake contents), which has an important role especially in making the message reliable.

Generally, several individuals participate in the realization of fake contents. In particular, we distinguish the creator of the content or of a whole disinformation campaign (single individuals; editorial company and not only, organizations with cultural, ideological, political, criminal purposes; services intelligence, governments, States), and the executors of the different activities of the chain. The executors are those contributing directly to the creation and production of the fake content and sometimes coincide with the creators of the initiative. They can be single individuals, groups of users recruited for this purpose, actual organizations specialized in the design and implementation of disinformation campaigns.

Moreover, the individuals pursuing disinformation strategies can be very often act using automatic mechanisms like bots, which allow to publish and disclose fake contents by means of many fake accounts or fake social profiles. Also the addressees of the contents themselves can contribute to the disclosure of the fake contents, whereas - also unaware of it - they share them and contribute to their disclosure.

Considering the characteristics of the chain leading to the introduction of disinformation in the Italian ecosystem, this Chapter presents an analysis on the quantities of produced fake contents and on the truth of topics subject to disinformation in Italy.

2.2. IMPORTANCE OF DISINFORMATION PRODUCED IN ITALY

Regardless of other conditions, the efficiency of an information system is big as much as less extended the volume of disinformation on is. In other terms, from the moment when the disinformation leads the citizen to have distorted views, formulate wrong speculations and so to take not informed economics, political and social decision, the socially perfect quantity of disinformation put in the information system is equal to zero. As the matter of fact, if the information is an economic and social good with the characteristics described in the previous Chapter (see par. 1.1), the disinformation is inherently a bad thing, or in other words a good thing, whose bigger availability let reduce the satisfaction (utility) of the consumer.

Consequently, not being evaluable the qualitative aspect (the fake content has already from the beginning a negative connotation), the quantitative data becomes a fundamental element of the analysis on the production of disinformation from the point of view of the protection of pluralism, an immediate indicator of the size of the phenomenon and of the own evolutionary dynamic.

The next analysis on the volume of online disinformation produced in Italy consider the overall number of information documents produced every month by websites and pages/social network accounts (Facebook and Twitter), included in the lists written by external sources, specialized in the activity of *debunking* and usually used in scientific studies on these topics. More in detail, the considered documents belong to every article (in the case of websites), post or tweet (in the case of social networks) published by 335 sources (see METHODOLOGICAL APPENDIX)²⁰.

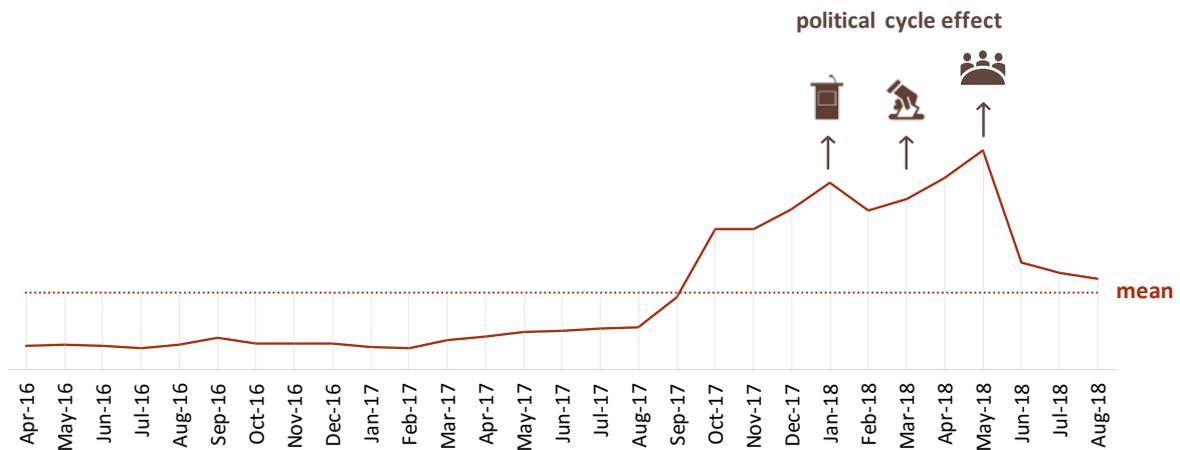
Figure 2.2 shows the monthly trend of the mentioned quantity of fake contents, produced in Italy from April 2016 till August 2018. The registered trend allows to ideally divide the graphics in two parts: the first one, arriving till August 2017, where the evaluation of the level of disinformation produced in a month, even if positive, has a rather constant value and lower than the average of the entire period; the second one,

²⁰ This method allows to obtain an evaluation of the production of fake contents, produced in Italy in an average month. The value calculated this way represents the productions of news by all those sources, which were detected by workers in the sectors, such as creators of fake contents. In this sense, it is a subjective evaluation (calculated starting from fake producers) and probably by defect (not being possible to include all producers of fake news in Italy). Similarly, this value does not include the cases of disinformation with incidental character, produced by headings producing actual information. On the other side, this evaluation includes all information contents coming from fake sources, but also the ones which could be true. In Chapter 3, the exercise will consider the objective character of disinformation, analyzing a sample of real and fake news, regardless from the individual who produced and disclosed the relative contents.

which includes the last year, where the evaluation of the quantity of the produced fake contents has a significant increase and has every month a value higher than the average of the entire period.

The value shows an evident criticality in the system because not only the produced volume of disinformation is not equal to zero, but it increases in time in the own average level.

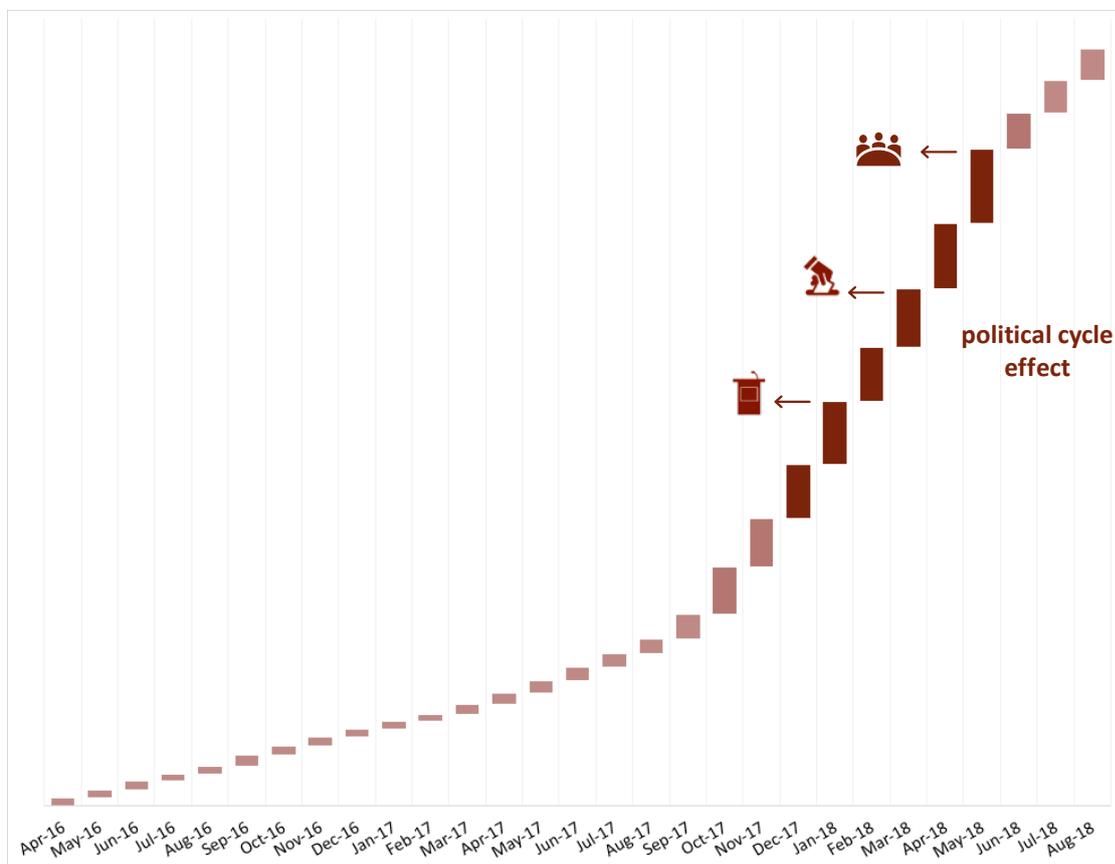
Figure 2.2- Monthly trend of disinformation produced in Italy



Source: Agcom elaboration on Volocom data

Substantially, if we consider the amount of fake contents, produced cumulatively in the detected period (see Figure 2.3), we can clearly observe an always increasing curve, expression of the sum of monthly amounts, never equal to zero.

Figure 2.3 - Incremental amount of fake contents produced in a month



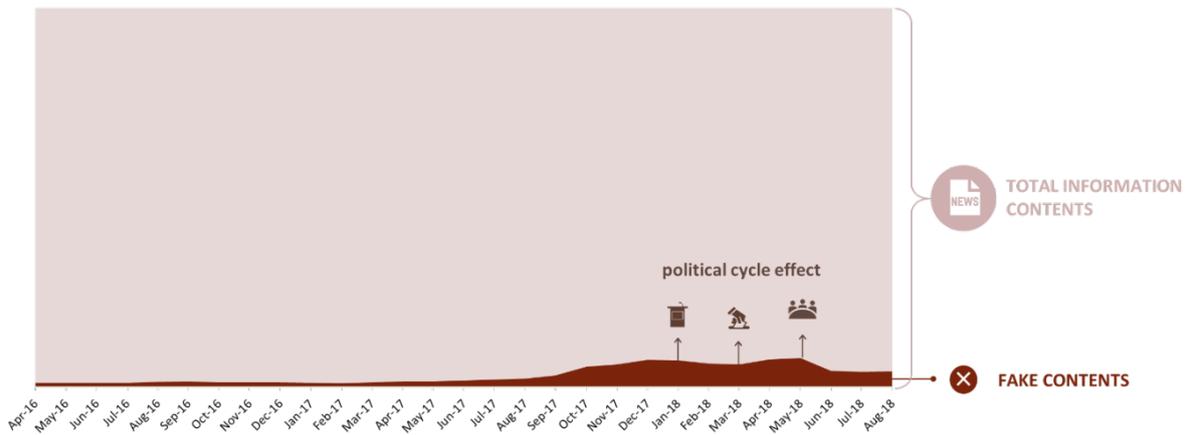
Source: Agcom elaboration on Volocom data

On top of that, in the last year, the amount of disinformation is generally increasing both in terms of absolute terms and relative terms. To this regard, as Figure 2.4 shows, the evaluation of the average impact of the fake contents on the information contents of the national system has an increase, going from 1% of the period till August 2017 (2% if we consider only the online contents) to about 6% of last twelve months (10% of online contents).

Moreover, these considerations has to consider the fact that the carried out analysis do not include the documents put into circulation in the information system by single user or by traditional headings (see note 20), that, in the sharing activities carried out by means of the online platforms, can contribute not only to the disclosure of *fake* contents, produced in Italy by other sources (even if modifying them and transforming them in

other contents), but also introducing new ones, increasing the volume of disinformation (on the role of online users see Chapter 4).

Figure 2.4 - Incidence of fake contents in the information system



Source: Agcom elaboration on Volocom data

The analysis of the results exposed in the previous Figures show an even bigger criticality from the point of view of information pluralism, if we consider the coincidence of the increase of disinformation with the political happenings in the country. The new political cycle (from the dissolution of parliament of the previous government till the establishment of the current government, going through the political elections and the relative electoral campaign) produced a propulsive effect on the quantity of issued disinformation, which registered peaks right at the beginning of the electoral campaign, during the holding of elections and of the establishment of the new government (see Figure 2.2 and Figure 2.3).

As observed in par. 1.2, also in the case of the information, the contingency of the political cycle created an increase of the offered contents. However, the increase of the effect of fake news on the overall information system (see Figure 2.4) right during the political cycle is an indicator of the increase of the disinformation more proportional than the increase of the information.

How we are going to deep in par. 2.3, the criticality of the scenario designed under the quantitative aspect is confirmed also from the point of view of the content, whereas the topics concerning politics and government represent the most recurrent object in the fake contents produced in 2018 (see Figure 2.6).

In brief, the detected evidences show an information system suffering the presence of a volume of disinformation which, on the one hand, seems to have stabilized on an averagely higher value and, on the other one, reached the maximum level during the last political cycle. The risks associated to these trends are clearly connected to the attitude of the fake contents, equal to the one with as object real news, to affect the knowledge, the opinions and the points of view of citizens about various questions. And when the treated topics concern the politics, mostly during the electoral period, the risk is that fake, manipulated or unfounded news create the expressions of vote of the citizens and, by means of them, to affect the electoral results, from which the political lines adopted at national level depend.

Particular importance is also given to the actual exam of the genres and to the topics of the fake contents, which, considering the (political/ideological/economical) motivations generally hidden behind the disinformation strategy, produce a negative impact on the pluralism as big as the involved topics are able to affect the public opinion. For this reason, we decided to dedicate the next paragraph to the specific analysis of the topics of disinformation produced in Italy.

2.3. TOPICS OF DISINFORMATION

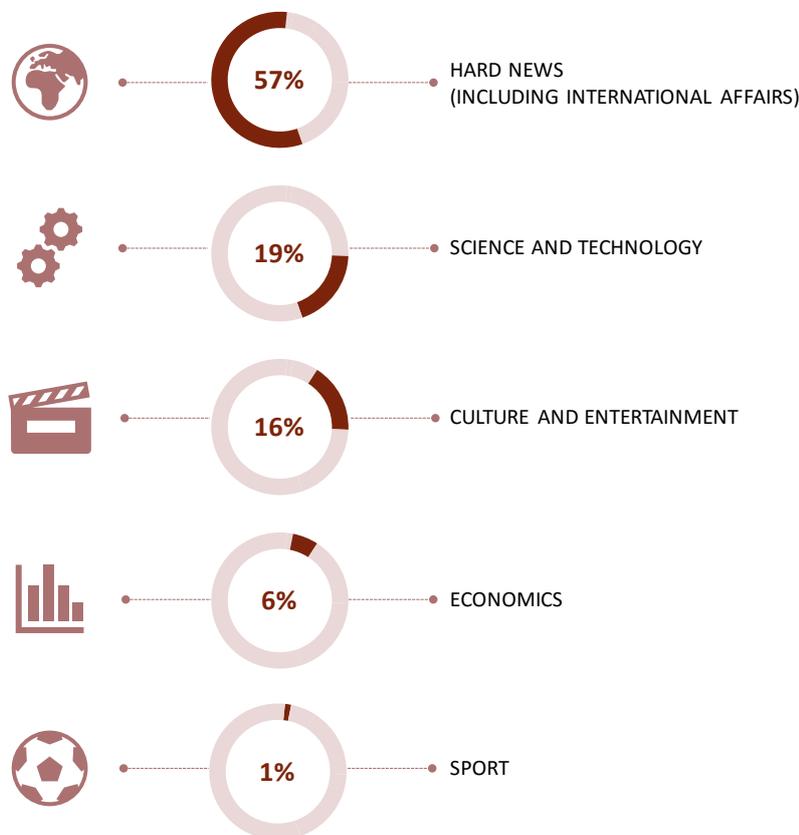
A first indication on the topics which in Italy are more commonly treated in fake news is given from the division according to category of the produced disinformation.

To this regard, Figure 2.5 shows how in average, in the entire analyzed period, more than the half of the contents disclosed by considered fake sources, regarding right the news with a higher value from the point of view of the pluralism (“*hard news*”), referable to facts of crimes, politics and happening with international importance.

The category of scientific and technological news follows, object of 19% of the fake contents. They are news suitable to produce effects on the ideological sphere of citizens and which are of great interest in the population, as shown by Figure 1.8 and Figure 1.9. As already said, this interest is not suitably satisfied by the offer of information nor in terms of quantity (concerning only 7% of the produced output, with a significant lack of request, see Figure 1.8) nor in terms of level of specialized formation of the journalists taking care of it (see Figure 1.7), and which risks more than the other categories to be overcome by the fruition of fake news.

Considering the other categories of content (“culture and show”, “economics” and “sport”, they cover together less than a fourth of the overall offer of disinformation, whereas the sportive news represent the least portion, also in terms of their easy verifiability by the citizens and so of their less suitability to be manipulated or artfully falsified.

Figure 2.5 - Supply distribution of fake contents according to category



Note: the category "hard news" include news on report, politics and the one of international importance.

Source: Agcom elaboration on Volocom data

In any case, besides the distribution according to category, the importance of the implications associated to the spreading of disinformation requires a more specific study of the topics of the fake contents, allowing to verify what the topics with a higher concentration of fake news are.

For this purpose, starting from the group of all documents produced in 2018 (from 1st January till 31st August) by the websites²¹ detected as source of fake news and included in these analyses, we adopt a methodologic approach, referable to the so-called *topic modelling*. A *topic model* is a statistic model for the automatic individuation of topics

²¹ In the carried out analysis, we did not consider documents referable to pages and social network accounts, because the reduced length of the contents (posts, tweets) make them few suitable to be treated with topic modelling technics.

appearing in a collection of documents. In particular, the classification of the text together to the group of fake contents in determined topics was obtained by means of the use of a LDA-model (*Latent Dirichlet Allocation*) - an algorithm of not-supervised automatic learning considering the frequency and the context of the terms used in the collection of documents. The same is at the basis of the interactive visualization system LDAvis²², with which the results exposed in the following Figures are exposed. More specifically, the LDAvis system supplies an overall vision on the identified topics (and of the way in which they differ), allowing at the same time a deep analysis of the most salient words, associated to every one of them. In fact, the system offers a compact visualization where, on the one hand, the prevalent topics are shown, considering their frequency and their positioning in the space, and on the other the most salient words are stressed (the first 30), which have the highest number of occurrences in the relevant topics, indicated in decreasing order according to frequency.

The application of these textual analyses technics to fake contents allowed of making a mapping of the main thematic areas object of disinformation in Italy in 2018. More specifically, the next Figures offer a representation of the 9 arguments for which the highest presence in fake contents was found. Every topic is represented by a circle, which amplitude shows the occurrence frequency in the entire group of the examined documents. Moreover, the topics have a position in the Cartesian coordinate system, being around the horizontal and vertical axes, representing the two main dimensions and explaining the variability of the data. In detail, in this case, the first quantity is put on the horizontal axis and expresses the level of actuality of the news, referable to a topic, which assumes an increasing value going from the right to the left hand side (from less to more actuality). The second quantity, indicated on the vertical axis, expresses the subjective extension of the news concerning a certain topic and has a growing value going from the bottom to the top (from a very small extension, circumscribed to single subjects, to a very broad extension, which involves institutions or countries).

Figure 2.6 proposes an overall vision of the predominant topics in fake contents, many of which belong to the category of *hard news*. Particular importance is given to the topic “Politics/Government”, which becomes even bigger, as much as that many words, which

²² See C. Sievert, K. E. Shirley (2014), “LDAvis: A method for visualizing and interpreting topics”, *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pp. 63–70. Cfr. anche D. M. Blei, John Lafferty (2009), “Visualizing Topics with Multi-Word Expressions”, *arXiv:0907.1013v1* [stat.ML]; J. M. Bischof, E. M. Airolidi (2012), “Summarizing topical content with word frequency and exclusivity”, ICML; J. Chuang, C. D. Manning, J. Heer (2012), “Termite: Visualization Techniques for Assessing Textual Topic Models”, *Advanced Visual Interfaces*.

belong to the most frequent 30 ones in the entire dataset, are linked to the political and institutional happenings in Italy. This evidence is at least an indicator of a certain concentration of disinformation around political questions in the first 8 months of 2018 and contributes to sharpen the criticalities already shown in the previous paragraph, whereas there is a higher production of fake contents during the electoral period. The result suggests that not only in the months immediately before and after the Italian vote, disinformation registered a significant increase, but also the disclosed fake contents have had as object news strictly connected to the political sphere.

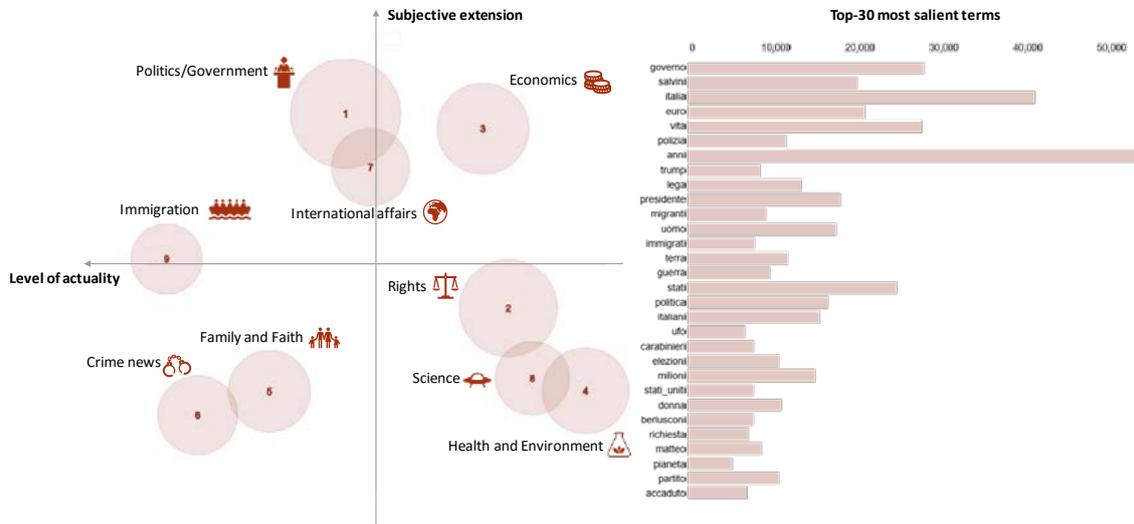
Beside politics Figure 2.6 indicates in the prevalent macro-topics of disinformation in Italy further topics belonging to the category of the *hard news*, such as “Report”, “Immigration” and “International affairs”.

In the most fake contents, big importance in terms of frequency distribution is given also to topics connect to social aspects, such as “Rights” and “Family and Faith” and to the one of “Economics”. Moreover, in compliance with what Figure 2.5 detected, a big part of disinformation concerns the scientific-technological topics, with particular regard to the one regarding the healthcare and environmental segment.

Generally, it is possible to observe how the 9 detected macro-themes have as a common element the fact of recalling polarizing topics²³, which can create or sharpen the separation of individuals in distinct groups (this aspect will be treated in Chapter 4, to which reference is made). So, these discussed and particularly interesting topics with a certain level of contagiousness transfer users’ emotional and perception states of mind. As a consequence, they are topics which often become object of viral disclosure, most of all by means of online platforms (also by users themselves boosting and sharing them).

²³ For ideological polarization we mean the individual result of that social process of separation and fragmentation of population in different groups, divided and not communicating each other on dividing topics. See Sunstein, C. R. (2002), “The Law of Group Polarization”, *Journal of Political Philosophy*, 10(2), pp. 175–195; Idem (2017), *#Republic. Divided Democracy in the Age of Social Media*, Princeton University Press. The level of polarization of a content can clearly be associated to its capacity of generating phenomena of *echo chamber* or *confirmation bias* in certain user groups (see Quattrociocchi, W., Scala, A. & Sunstein, C. (2016), “Echo Chambers on Facebook”, *working paper*).

Figure 2.6 - Main topics of the fake contents (2018)



Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

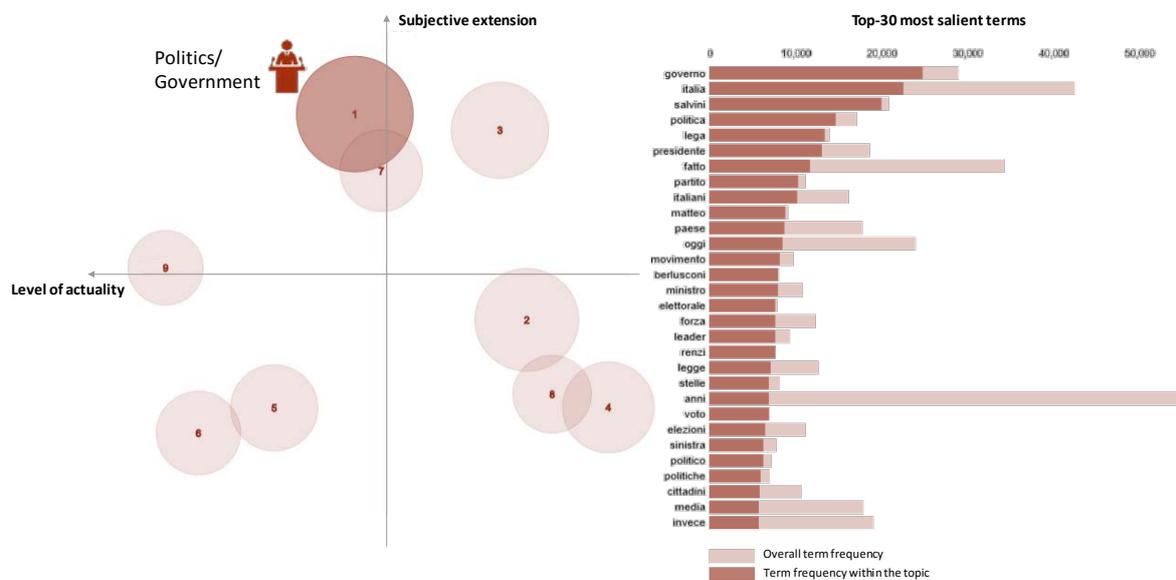
Increasing even more the level or granularity of the analysis, it is possible to examine the composition of every topics of disinformation, obtaining - according to the salient words associated to them - indications about the specific object of the disclosed news.

In this sense, Figure 2.7 explores the composition of the topic “Politics/Government”. The analysis of the most common words shows that the attention of the fake contents, referring to this topics is given to three main aspects: *i)* the political topics, in particular the parties (let consider words like “partito”, “lega”, “movimento”, “stelle”, “forza”, “italia”), exponents and candidates in the elections of the 4th March 2018 (“salvini”, “renzi”, ...); *ii)* the institutional bodies (“governo”, “ministri”, “presidente”); *iii)* the development of the political elections of 4th March and the political participation (“elettorale”, “voto”, “elezioni”, “politiche”, ...).

These aspects are clearly the ones assuming more importance under the profile of pluralism, even more considering that the most part of the period, which the analysis refers, was involved in the explication of the new political cycle, whereas the citizens

were called to vote and were exposed to the risk of taking false information, on which they found their expression of vote.

Figure 2.7 - Most salient terms in the topic “Politics/Government” (2018)

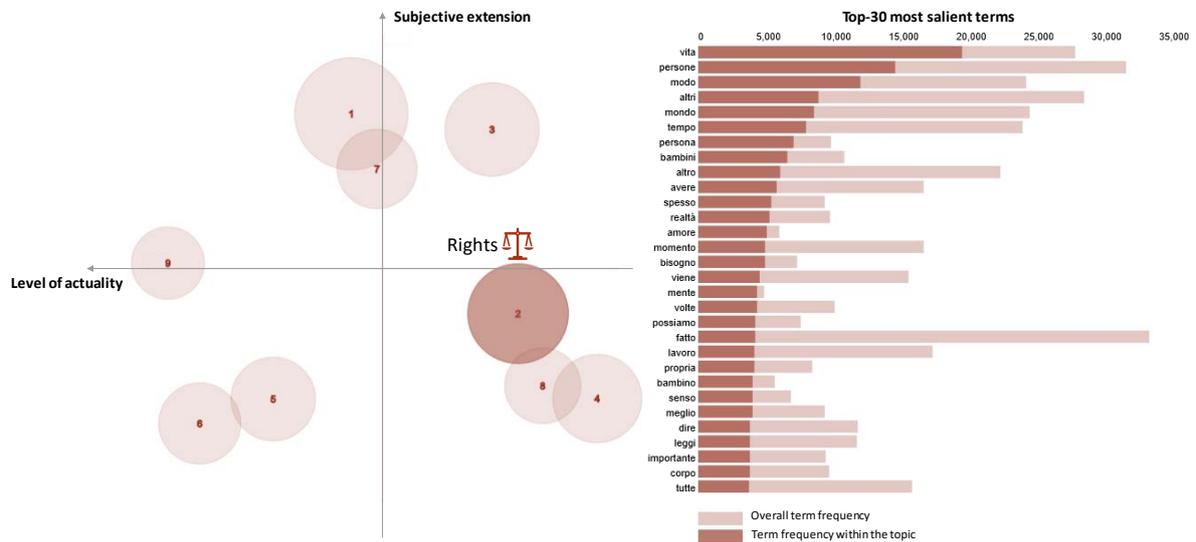


Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

The second more extended topics in the framework of fake contents concerns the “Rights”, understood in a wide sense.

The analysis of the salient words offered by Figure 2.8 stresses the statement that very often the news connected to the rights and person’s needs (“vita”, “persone”, “bambini”, “amore”, “lavoro”, ...) and the exigence of protect them (“legge”, “bisogno”, ...) are the lever of disinformation. Disinformation which, in cases like this, is the result of strategies with the purpose of affecting the emotional sphere and the direct involvement of the citizens.

Figure 8. - Most salient terms in the topic “Rights” (2018)

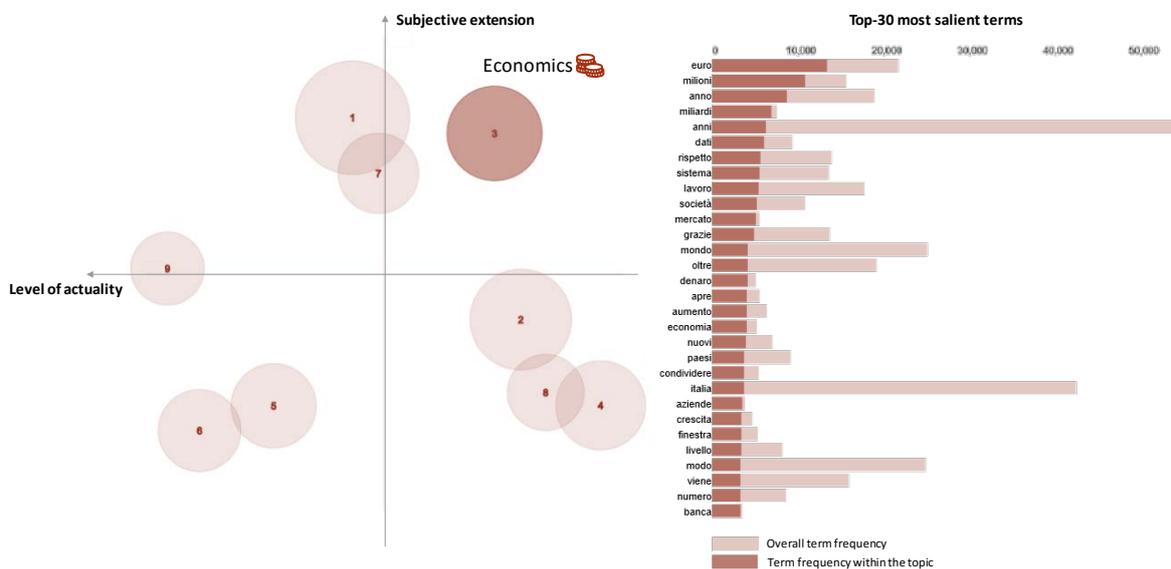


Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

The fake contents on the topic “Economics”, as Figure 2.9 shows, seem to be focused on the general economic trend (“sistema”, “crescita”, ...) and on the aspects of financial nature (“euro”, “milioni”, “miliardi”, “sistema”, “mercato”, “banche”, ...), on the other side on questions of corporate (“società”, “aziende”, ...) and work context (“lavoro”, “livello”, “aumento”, “finestra”, ...).

It is interesting to notice how often the citizens feel these with particular interest, addressing to their economic conditions and the destiny of their savings and how they are in the center of political intents and programs, also during the electoral campaign, becoming decisional variables, relevant for the preferences of the individuals. This circumstance worsens without doubt the risk of negative spillover effects, associated with the disinformation.

Figure 2.9 - Most salient terms in the topic “Economics” (2018)



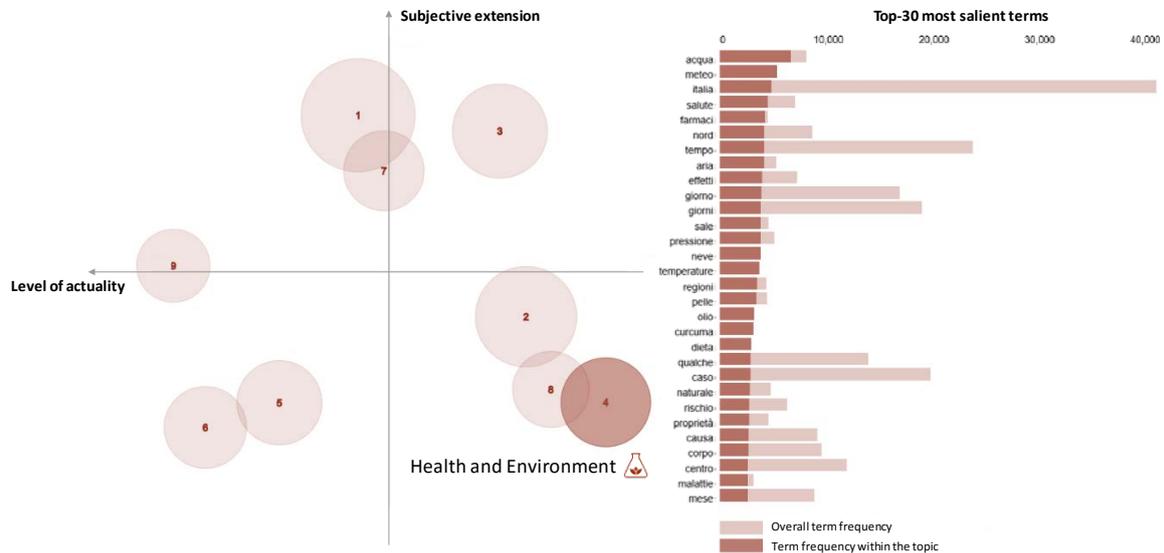
Source: Agcom elaborations on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

A further impulse for the production of fake contents comes from topics which, even regarding different topics, involve people’s ideology.

To this regard, the topics “Health and Environment” (see Figure 2.10) covers surely news affecting the citizens’ orientations and points of view about facts inspiring apprehension, like the health and the (negative) health and environmental conditions.

Evidence of it is the high frequency of words recalling question on which individuals feel a state of worry: “salute”, “farmaci”, “dieta”, “pelle”, “corpo”, “malattie”, or “acqua”, “tempo”, “aria”, “temperature”, which are also associated to words, such as “effetti”, “rischio”, “causa”, stressing furtherly their negative connotation.

Figure 2.10 - Most salient terms in the topic “Health and Environment” (2018)

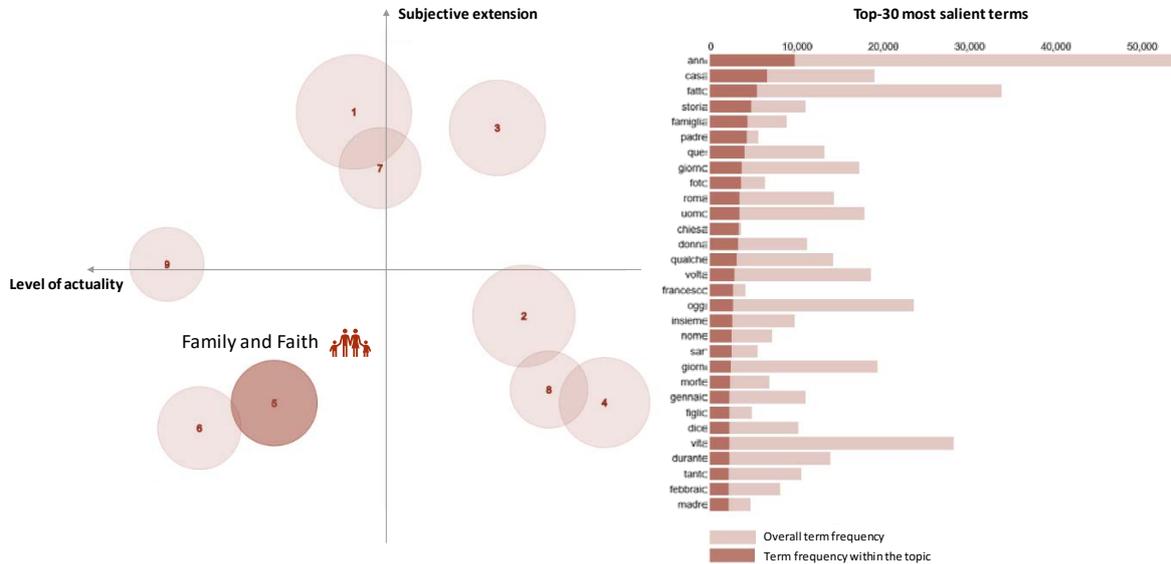


Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

Similarly, the ideological profile is at the basis of disinformation strategies concerning the religious faith. Many fake contents of 2018 concerned news on “Family and Faith”, as much as the topic is the fifth most treated result on websites, detected as disinformation sources.

In particular, the highest number of fake, manipulated or false news concerning the topic in question, regarded the church (“chiesa”, “francesco”, “san”, ...), but also esistenial questions (such as “vita” and “morte”), as well as family (“casa”, “famiglia”, “uomo”, “donna”, “insieme”, ...), which have always been considered a founding value of the religious communities (see Figure 2.11).

Figure 2.11 - Most salient terms in the topic “Family and Faith” (2018)

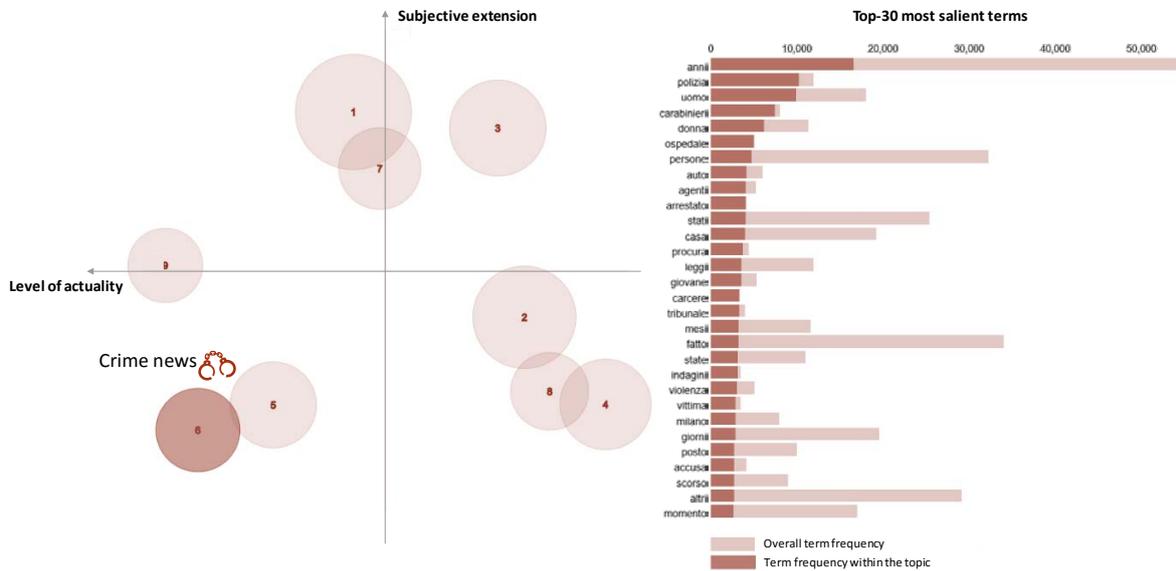


Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

With regard to disinformation produced on the topic “Report”, it must be observed that, even if it seems to be focused on one or few specific facts, which assumed relevance at national level in 2018, it is mainly focused on happening about crimes and judicial news.

From the analysis of the salient words indicated in the Figure 2.12, we can notice the abundance of news regarding the modalities of a crime (“persone”, “ospedale”, “violenza”, “vittima”, ...), as much as the intervention of police forces (“polizia”, “carabinieri”, “agenti”, “arrestato”, “indagini” ...) and the judicial facts connected to them (“procura”, “leggi”, “tribunale”, “accusa”, “carcere”, ...).

Figure 2.12 - Most salient terms in the topic “Crime news” (2018)

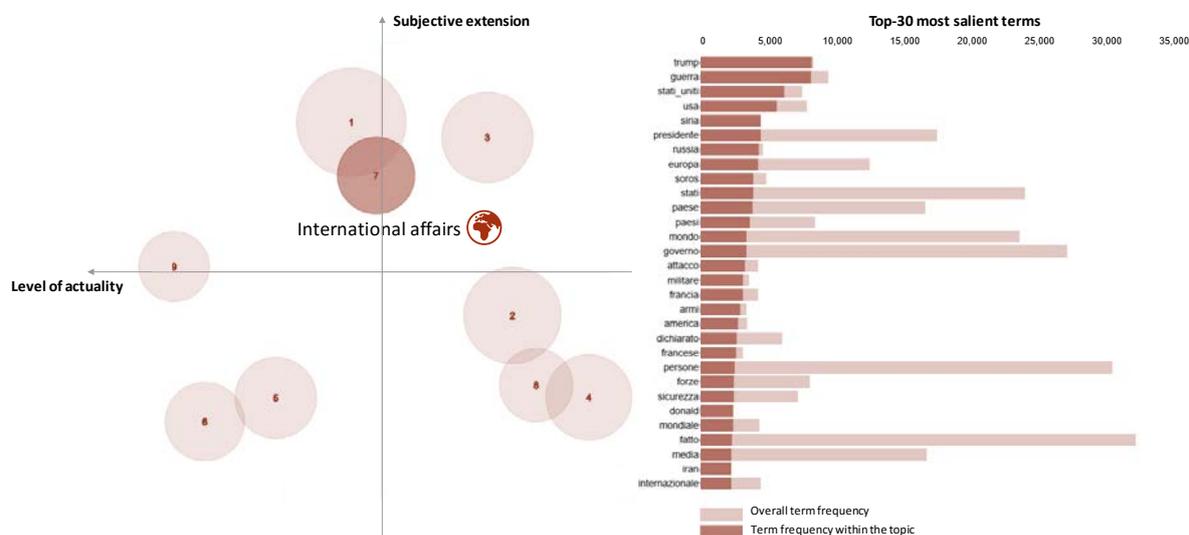


Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

Concerning the international happenings (“International affairs”), Figure 2.13 shows how the fake contents produced in Italy in the eight months of 2018 has as object mainly news concerning conflicts and military attacks (“guerra”, “siria”, “attacco”, “militare”, “armi”, “iran”, ...).

Particular importance is given to the happenings in the US, as shown by the frequency of occurrence of salient words like “trump”, “stati_uniti”, “usa”, “america”, “donald” ...).

Figure 2.13 - Most salient terms in the topic “International affairs” (2018)



Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

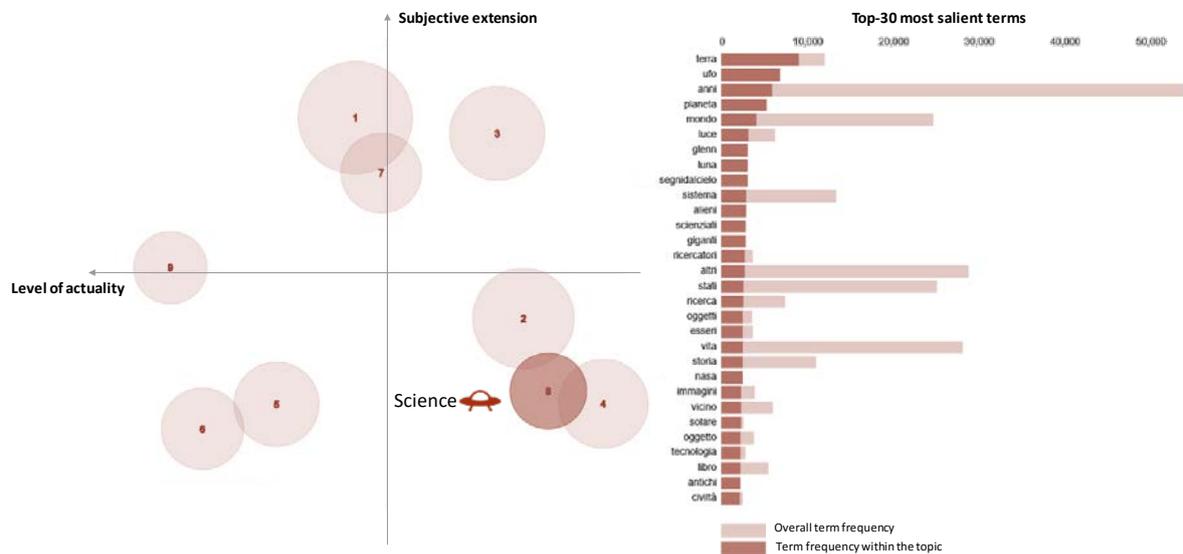
About disinformation in scientific field, it must be said that it concerns news on health and environment, but it seems to consider very significantly questions and phenomena about astronomy and space.

In this sense, Figure 2.14 stresses how the topic “Science” involves mostly news on planets and celestial bodies (“terra”, “pianeta”, “sistema”, “solare”, ...); UFO and alien presence in the space (“ufo”, “alieni”, “vita”, ...).

It proves also how the salient ones seem to be words recalling directly the scientific research work, probably with the intent of giving truthfulness to the theories exposed in the articles and often subject to strategies of disinformation, which follow conspiracy logics²⁴.

²⁴ On the study of the online disclosure of fake contents with as object scientific news, see among the others, A. Bessi et al. (2015), “Science vs Conspiracy: Collective Narratives in the Age of Disinformation”, *PLoS ONE* 10(2); A. Bessi et al. (2016), “Homophily and Polarization in the Age of Disinformation”, *The European Physical Journal Special Topics*, 225(10); M. Del Vicario et al. (2016), “The Spreading of Disinformation Online”, *Proceedings of the National Academy of Science* 113(3); F. Zollo et al. (2017), “Debunking in a World of Tribes”, *PLoS ONE* 12(7).

Figure 2.14 - Most salient terms in the topic “Science” (2018)



Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

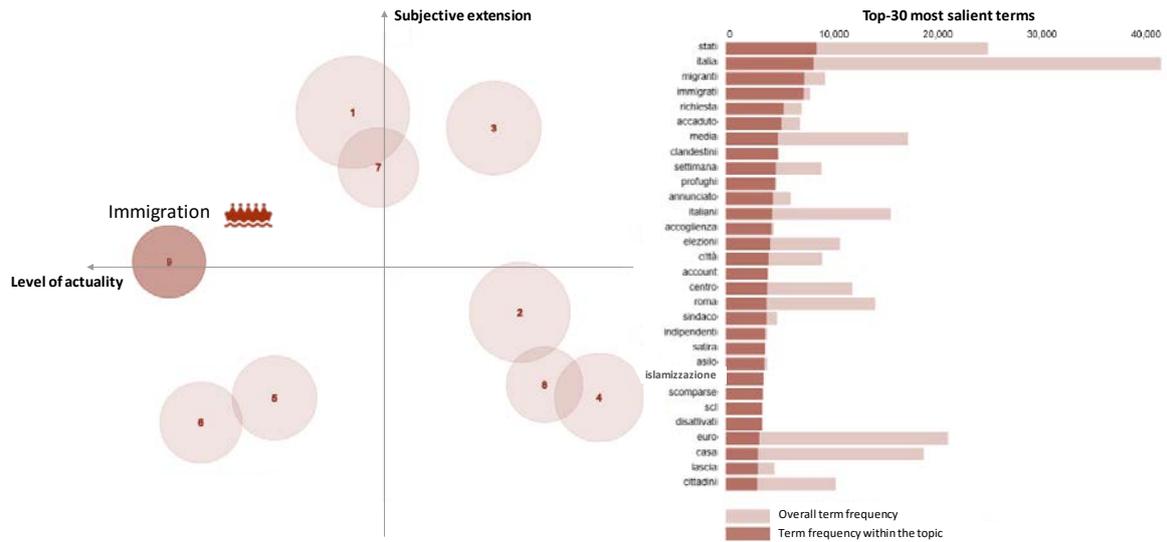
At last, the main topics of disinformation there is also, as we said before, the “immigration”. During 2018, many news on the topic characterized the policy debate and sharpen the tones of the electoral campaign. The report facts (among which the emblematic one about the shootings against the migrants in Macerata) and the measures adopted by the new government (such as the closing of ports to the humanitarian ships) divided the public opinion.

In this scenario, the proliferation of fake contents found fertile ground. Most time, the motivations for the production of disinformation on the topic of immigration are of ideological nature and the disclosed contents have the purpose of starting campaign of hatred (in particular with racist purposes) and instill intolerance.

As represented by Figure 2.15, disinformation on immigration in the examined period was built mostly on news about reception of migrants by Italy and other countries. Among the salient words on this topic, you find lots of them referable to evocative words

in this sens, such as “stati”, “italia”, “migranti”, “immigrati”, “richiesta”, “asilo”, “profughi”, “clandestini”.

Figure 2.15 - Most salient terms in the topic “Immigration” (2018)



Source: Agcom elaboration on Volocom data with visualization system LDAvis (Sievert and Shirley; 2014)

THE DISSEMINATION OF REAL AND FAKE NEWS

3



3.1. INTRODUCTION

In the previous chapters, the information system was examined under the aspect of the production, understood as quantity, quality and variety of contents created by the sources, active in Italy on different media, such as newspapers, tv, radio, websites and online platforms (in particular, social networks).

Starting from a subjective method (i.e. starting from the identification of the information and disinformation sources), we identified the characteristics and stressed the criticalities of the “news supply” proposed by every media (also in relation to the request and to the used productive resources, see Chapter 1), as well as of the disinformation produced and put in the national system (see Chapter 2).

Changing necessarily the prospective on analysis, which means starting from the news, so from the object of the information contents and false products, the objective of this chapter is the identification of the peculiarities characterizing the disclosure of the information and disinformation by the mentioned sources.

In the following paragraphs, we are going to study the propagation modalities of single, real and false news, in time and space defined in all media composing the information system. The attention will be given in particular to the definition of the lifecycle of the news (permanence in the information system) and to the comprehension of distinctive elements (duration, actual presence, coverage) of them, according to the kind (genre) of news and means of disclosure.

Under the methodological profile, as we said, the analysis exposed in the followings are carried out applying an objective approach. In this sense, *datasets* were created for two samples of news: one for real news and one for fake ones²⁵.

²⁵ At the passing from the first to the second method of analysis, a change is implicit which has to be very good understood by the reader. The subjective method (Chapters 1 and 2) contained the information and disinformation sources. Being the first ones well defined, the analysis of the information was carried out almost as if it would be a census. On the other hand, for the second one we considered sites and disinformation pages/accounts, identified in the scientific community. These represent necessarily a subgroup of all active disinformation sources. In this chapter, which has the purpose of analyzing the dynamic path of the single news, we necessarily identified a sample of real and fake news, using the approach described in the text. So, the analysis not only refers to the objective character of information and disinformation, but uses also a sample inevitably limited (even if, as we are going to show, highly representative) of all news of the ecosystem. As a consequence, the analysis is affected by the selected sample and by the considered period: in fact, the main news can belong every year to different categories according to the events happened in that period (political elections, reports of crime, international events, sport event ...).

In order to ensure robustness and significance of the results, particular care was given to the definition of samples of news, which were selected considering their relevance in the information context and their representativeness, according to the category of treated genre, regardless of the individuals who produced the information or fake contents.

In detail, considering the real news, we identified the main happenings occurred in a period of 12 months. For every month we included in the sample the most relevant 5 news of every category (report, politics, international affairs, economics, science, culture, show and sport), noting the exact date (t_0) where the fact occurred and defining for each of them the group of words, most suitable to create the research *query*²⁶, in order to extract the relevant daily occurrences from the used documental database (see METHODOLOGICAL APPENDIX).

Similarly to what is broadly indicated in the literature, the relevance of the news was evaluated by the study of the trends of the researches of the Internet users, interpreting the prompt growths of researches (and so of interest) for a certain topic by the users like the reflection of the trend of a news, concerning a fact or happening occurring in Italy or in the world. In general, for every month, they were inserted in the sample of news, which in the framework of a category registered the highest research rate²⁷.

Otherwise, for the selection of the sample of fake news, we used external debunking sources, specialized in the application of scientific approaches for the discover and

²⁶ The *query* is the elaboration of a sequence of words by means of logical connectors (Boolean operators), according to the rules described in the language *Apache Lucene Query Parser Syntax*.

²⁷ In particular, for the analysis of the research trends, we referred to the instrument *Google Trends*, which allows of recognizing the frequency of online researches of a word or a group of words, from which it is possible to go back to the news of reference. In detail, the results on the research trends were taken by means of the function "Explore" on *Google Trends*, allowing of circumscribing the analysis to a specific geographical area (in this case, Italy), to a specific period (in this case, a single month), and to select the them categories according to which the research trends were grouped, as well as the source by means of which the users carry out their research, which is "Google research" and "Google News".

In the scientific literature, it is common to use this approach and in particular *Google Trends*, in order to quantify and/or foresee phenomena regarding economics (see for example H. Choi and H. Varian, 2012, "Predicting the Present with Google Trends", *Economic Record*), finance (T. Preis, 2013, "Quantifying Trading Behavior in Financial Markets Using Google Trends", *Scientific Records*), medicine (S. V. Nuti, B. Wayda, I. Ranasinghe, S. Wang, R. P. Dreyer, S. I. Chen, K. Murugiah, 2014, "The Use of Google Trends in Health Care Research: A Systematic Review", *Plos one*), l'ecosistema digitale (L. Kristoufek 2013, "BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era", *Scientific Reports*). Recently, the method was extended to the measure of the users' interest for news (see for example, L. T. P. Nghiem, S. K. Papworth, F. K. S. Lim, L. R. Carrasco, 2016, "Analysis of the Capacity of Google Trends to Measure Interest in Conservation Topics and the Role of Online News", *Plos one*).

denial of fake news²⁸. More specifically, for the individuation of fake news we collected the articles published during a year about fake, unfounded, manipulated or manipulated or not truly reported facts, having an actual “newsworthiness”, which means they could be transformed in news and perceived by the media context and by the public as such. In the choice of the news to be included in the sample, we considered also the category of treated genre, according to which the exam of the offer distribution showed (see Figure 2.5). Like for real news, also for everyone of the fake ones, we noted the date of origin of the treated fact (t_0) and formulated the research *query* most suitable to extract the corresponding occurrences from the documental database used for the analysis of this Report.

After having defined the samples of real and fake news according to the approach mentioned above, we created the two datasets containing the daily occurrences, registered by every news on every information source, in the period of 45 days corresponding to the 15 days before the fact (t_0) and to the 30 days after it ²⁹.

The definition of t_0 allowed to achieve all news with reference to a coordinate, in order to make possible the comparison between the different news and to have direct indications about the characteristic lifecycle of the news, their persistence in the information system and their mediatic coverage in time.

The evidences emerged from the analysis carried out on the datasets of real and fake news are presented in the following paragraphs, dedicated respectively to the exam of disclosure of information and disinformation.

²⁸ For the identification of fake news we referred in particular to the site butac.it, recognizing, in any case, like the most part of fake news object of the articles produced by this source were generally object of different debunking websites.

²⁹ Do not forget that very often before the fact, it is possible to find a certain information “rumor” on the topic (see infra, par. 3.2, Figure 3.1).

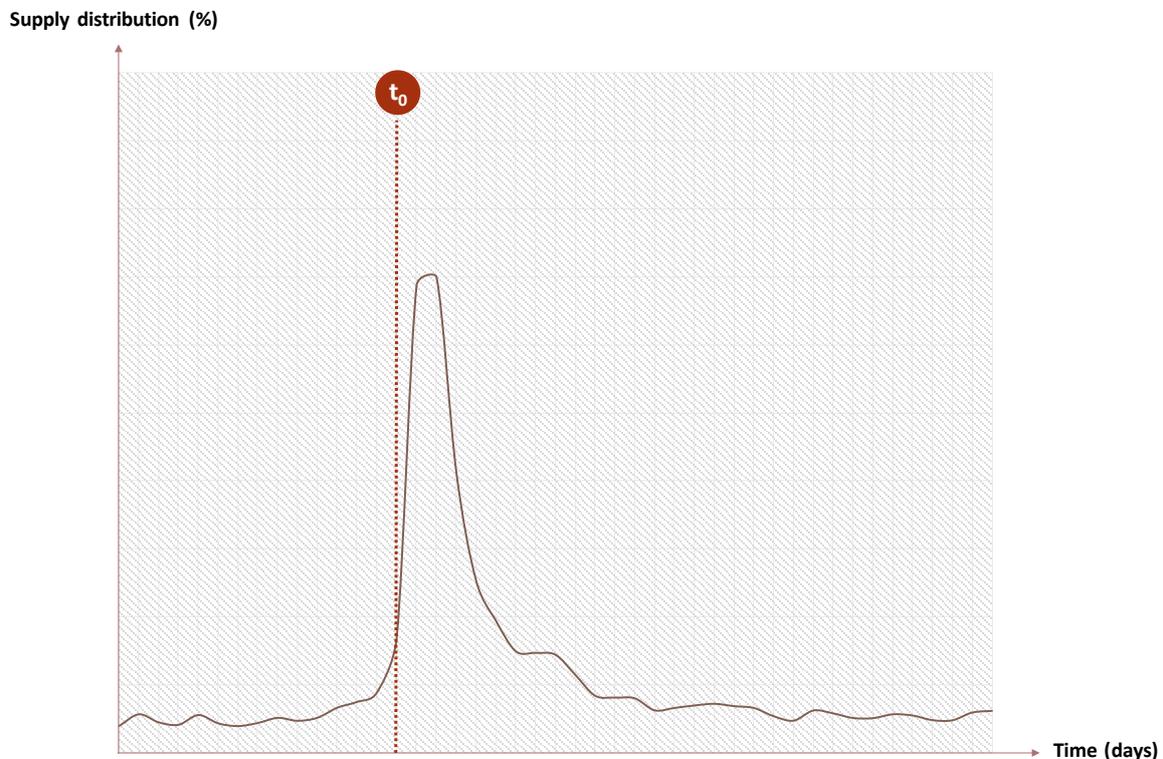
3.2. THE LIFECYCLE AND DISSEMINATION OF REAL NEWS

The extent of the Italian information system, in terms of produced information contents, was object of the evaluation of Chapter 1, in the framework of which the characteristics were analyzed which refer specifically to the news production (not only quantity, but also quality and variety). According to what was examined and discussed in the first part of the Report, following the trend of a numerous sample of main news (selected on the basis of the above mentioned criteria, see par. 3.1), this paragraph examines in detail how a single news is averagely treated and disclosed inside the information system.

A first indication on the approach of disclosure of the news is offered by the representation of their lifecycle, understood as average daily trend of the quantity of information contents regarding the news, disclosed by different media.

How Figure 3.1 shows, the disclosure of a news can start also in the exact day (t_0) when the fact itself occurs. In this sense, the kind of news is very relevant, which can be linked to a prompt or unpredictable fact, or to a programmed and foreseeable event. In this case, it is obviously easy to detect the coverage of the news also in the days before the t_0 . However, the biggest disclosure starts from t_0 when the biggest distribution of “news supply” occurs: In particular, the concentration of contents on the news, which tendentially is subject to an increase by the t_0 , registers its maximum in the two days after the one when the fact occurred (assuming in each of the two days a value next to 14% of the total), when all media can disclose it. In fact, as we are going to discuss later, every media has its own peculiarities and covers the news with timing and approaches, which can differ each other according to the frequency of updating of the own offer (for example, the newspapers, being published once a day, can cover the news starting from the day after the t_0 , except in the cases where the fact occurs within the first hours of the day), as well as to the chosen editorial line and the space for the information.

Figure 3.1- Lifecycle of a news



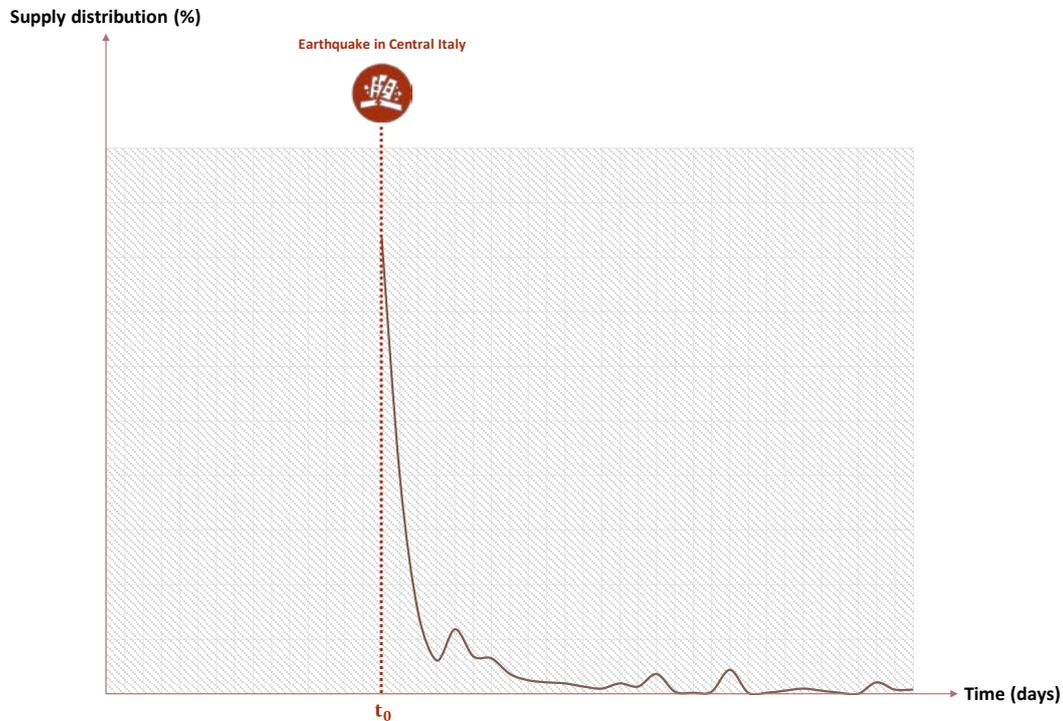
Source: Agcom elaboration on Volocom data

So, t_0 indicates surely the essential moment, around which the lifecycle of a news develops. More in general, we distinguish the news, for which the t_0 is “absolute” from those for which it is “relative”. In the first case, the t_0 indicates the beginning of the lifecycle of the news. This occurs for news having a trend such as the one we indicated as an example in the next Figure.

In detail, Figure 3.2 exposes the lifecycle of the news on the earthquake occurred in the night of 24th August 2016 in the center of Italy. In this case, it is a not-foreseeable fact. Therefore, the news does not have any anticipation, but it is put in the information system in the instants immediately after the fact, with a cover reaching its peak on t_0 and let presume that the treating by all media during t_0 .

A lifecycle with analogue features repeats always for news of crimes, natural disasters, terroristic attacks, ...

Figure 3.2 - Lifecycle of the news “Earthquake in Central Italy”
(August - September 2016)

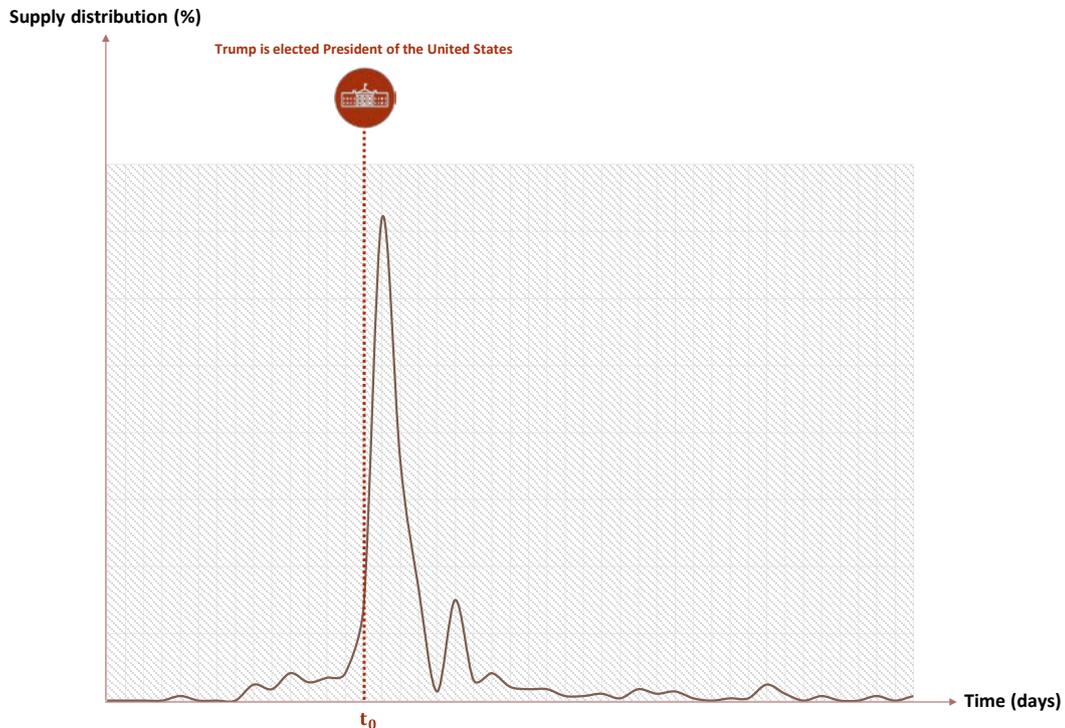


Source: Agcom elaboration on Volocom data

On the contrary, when the t_0 is relative, lifecycle of a news starts before the event when it occurs. An example of news with occurrences in the information system before the t_0 is indicated in the Figure 3.3, referring to the election of Mr. Donald Trump in the United States. Of course, the news, being connected to a foreseeable event, has anticipations linked in particular to the treating of the scenario expected in case of victory of the candidate at the elections.

It is common to have anticipations in the news production contents on a news in the case of political competitions, but also of sportive, cultural events or manifestations. The same applies also for company transactions or astronomic events and, in general, for all those organized or expected events, where the media cover before the t_0 can have the purpose to announce the event.

Figure 3.3 - Lifecycle of the news “Trump is elected President of the United States” (November - December 2016)



Source: Agcom elaboration on Volocom data

Increasing the level of accuracy of the analysis, it is possible to identify specific indicators concerning the disclosure of the news, in order to design punctually the modalities of disclosure of the information by different media (see Figure 3.4 and Figure 3.5).

In detail, we considered three main indicators, referable to:

- the “average duration of a news”, corresponding to the distance occurring averagely between the first and the last days, in which a news registers at least an occurrence or, in other words, is object of at least an information content. In short, this indicator quantifies the length of the lifecycle of the news, understood as days of permanence of the news in the national system;
- the “actual presence of a news” inside the information system, corresponding to the average number of days, even not consecutive, where a news was treated at least by an information document. This measure, compared with the average

duration, supplies an indicator about the extent of the portion of lifecycle of the news concretely interested by the presence of information contents on a fact. Therefore, the indicator is an important element to quantify the period where the citizen can potentially be exposed to the direct information on a fact and so entering in contact by means of the fruition of the media disclosing it;

- the “coverage of a news” in the average day, corresponding to the index expressing the ratio (multiplied for 100) between the number of daily occurrences registered averagely by a news on a specific means and the number of daily occurrences registered averagely by a news on a whatever media. As a consequence, values higher than 100 have a coverage higher than the average, values equal to 100 indicates a level of coverage in line with the overall average of the information media and value lower than 100 indicate a coverage lower than the average. The coverage defined this way is also a relative quantity of the way a news is followed by a mean, according to the editorial line, the interest and the relevance of the facts for the public.

The tree indicators allow to qualify and quantify the diffusion of the information following the treating of the news from a static (how much it is treated in an average day) and dynamic point of view (how many days the news remains in the information system and how many days it is actually treated).

To this regard, Figure 3.4 show how in average, among all media, the online editorial sources have the biggest amount of days of permanence of a news (25), during which the happening is actually object of treating for about 60% of the period. Moreover, for the same sources, also in consideration of the high frequency of updating, there is the highest rate of daily coverage of the news. In any case, on the one hand, it must be stressed that the result of the aggregation of different sources (such as websites of newspapers, of radio, Tv and exclusively online headings) and that, as said in par. 3.1, the selection of the used sample of news is based on the *trends* of the researches made on the Internet, circumstances affecting positively the values referred to the online sources (which could lead to carry out a partial overrepresentation of the phenomenon).

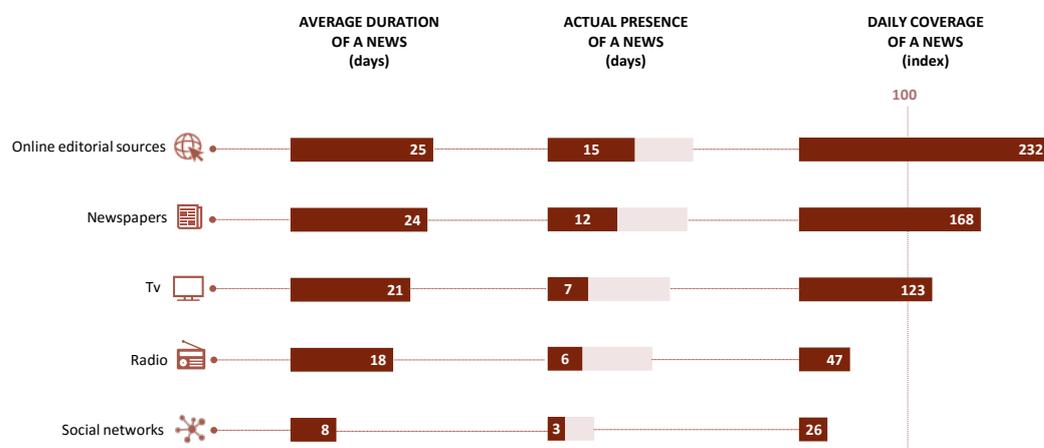
On the other hand, the diffusion indices have to be analyzed together with the provisions of Chapter 1 (see Figures 1.2, 1.3, 1.4), whereas, for the online editorial sources, against a big news production contents, we found an overuse of the used journalistic power, with the subsequent risks of negative results on the accuracy and deepening on the contents generated around the news.

For newspapers and tv, which have a lower production intensity than the journalists in relation to the quantity of offered information contents, a bigger level of expected quality is accompanied by a daily coverage of the news higher than the average and a length of the lifecycle of the news, averagely higher than 20 days. Between these two media, also in relation to their quite completely informative vocation, the newspapers have higher values, also because of the days of actual presence (treating) of the news. 12 (over the half of the average duration) against 7 of television (a third of the respective average duration).

In the case of the radio, the overall low “news supply” is associated to a low coverage of the single news, both daily and in time. However, as shown before (see Figure 1.2), this means, similarly to the tv, has a low productive intensity of the journalistic resources.

For social networks (pages and accounts of information sources and influencers), the disclosure of the news occurs in a very shorter period. The average duration of a news is of only 8 days, in 3 of which there is the actual treating of the news, together with an average daily coverage very lower than in the other media. These results, if compared with the high values of the quantity of overall offered information contents give an indication about the rapidity which posts and tweets go from the treating of a news to the other ones. So, it is also a question of the nature itself of the information product, typical of online platforms and of the space where it is circumscribed, of a treating of the contents more superficial than in other media, as shown also by the low level of quality, as analyzed in the first Chapter.

Figure 3.4 - Diffusion indices of a news



Note: for TV and radio, national broadcasters were considered. The online editorial sources include websites of newspapers, of radios, tv and exclusively online headings; while social networks include pages and accounts of information sources and influencers.

The “duration of the news” is calculated as average distance between the first and the last days where a news is registered at least once.

the “actual presence of a news” expresses the average number of days, also not consecutive, where a news registers at least an occurrence.

The “coverage index” expresses the ratio (multiplied for 100) between the number of daily occurrences registered averagely by a news on a specific means and the number of daily occurrences registered averagely by a news on a whatever media. Therefore, values higher than 100 have a coverage higher than the average, values equal to 100 indicates a level of coverage in line with the overall average of the information media and value lower than 100 indicate a coverage lower than the average.

Source: Agcom elaboration on Volocom data

The analysis of the diffusion indicators of the information by singles media can be furtherly deepen examining how the distribution of a news is concentrated in time. For this purpose, we used a suitable graphic representation, the so-called *heat map* (see Figure 3.5), which allows at the same time the aggregation of statistically homogeneous distributions, obtained by means of *clustering* procedures (left part of the graphics) and the different grade of concentration of daily occurrences, registered for a news on every media. This last aspect is registered by the coloration assumed in every single point, for which the level of concentration increases going from the vivid red (minimum) to the brilliant green (maximum).

In general, Figure 3.5 contributes to confirm and further qualify the features resulted from the study of the previous indicators.

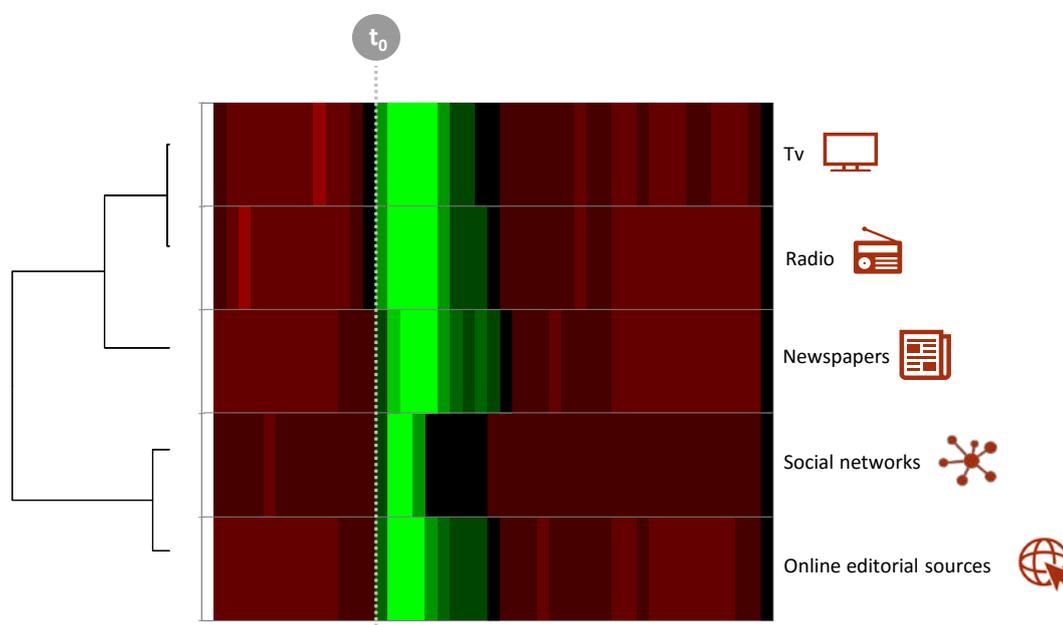
In detail, we stress a similar trend for television and radio broadcasters, with a distribution of occurrences according to news, which begins to intensify shortly before t_0 and then concentrates more in the following 4 days.

The newspaper, having as known an updating frequency of a day, have a distribution of the news in the peculiar time, which begins to intensify a day after in comparison with the other media, and then to assume the maximum value (with regard to the total amount of the media) in the 3 following days. Moreover, the newspapers - compared with the other media - have the highest number of days characterized by a green nuance, indicating that the distribution of the news is more spread out in time, with a less acute variance.

On the contrary, the distributions with which news propagate on the Internet have in common a stark variance in the levels of concentration, reached shortly before the t_0 and the ones on the following days (and are aggregated by the static process of clustering). In fact, in the day after t_0 , the distributions reach a concentration peak assuming a value higher than the registered one by other media in correspondence of their maximum, and then remaining on high percentage of occurrences also the day after. The trend seems to be even more acute for social networks, to the extent that the passage from the maximum to the minimum concentration of information contents regarding a news occurs in a more quickly way than in the other media of information.

The same indicators (average duration, actual presence and average daily coverage of the news) used for the analysis for the diffusion of a news on different media can be used to study the characterizing elements of the disclosure of the different categories of news, not forgetting during the analysis of the data that the indicated values are expression of average trends, obtained for a sample of main news (even if big and suitably representative).

Figure 3.5 - Concentration of a news distribution over time



Note: for TV and radio, national broadcasters were considered. The online editorial sources include websites of newspapers, of radios, tv and exclusively online headings; while social networks include pages and accounts of information sources and influencers

The Figure (so-called *heat map*) represents the concentration of the distribution of daily occurrences, averagely registered for a news on a mean. The level of concentration increases as going from the vivid red (minimum) to the brilliant green (maximum).

Source: Agcom elaboration on Volocom data

In this sense, Figure 3.6 shows how in average there is no very significant registration between *hard news* (considered in their complexity) and specialized news, with reference to the first two indicators: average duration of the news and extent of the period of actual treating.

The news on politics and economics are different essentially because of a longer lifecycle. Compared with the others, they begin to be treated before the t_0 and remain in the information system till several days after the t_0 , whereas the news on politics have (in absolute and relative value) the biggest actual presence inside the disclosed information contents.

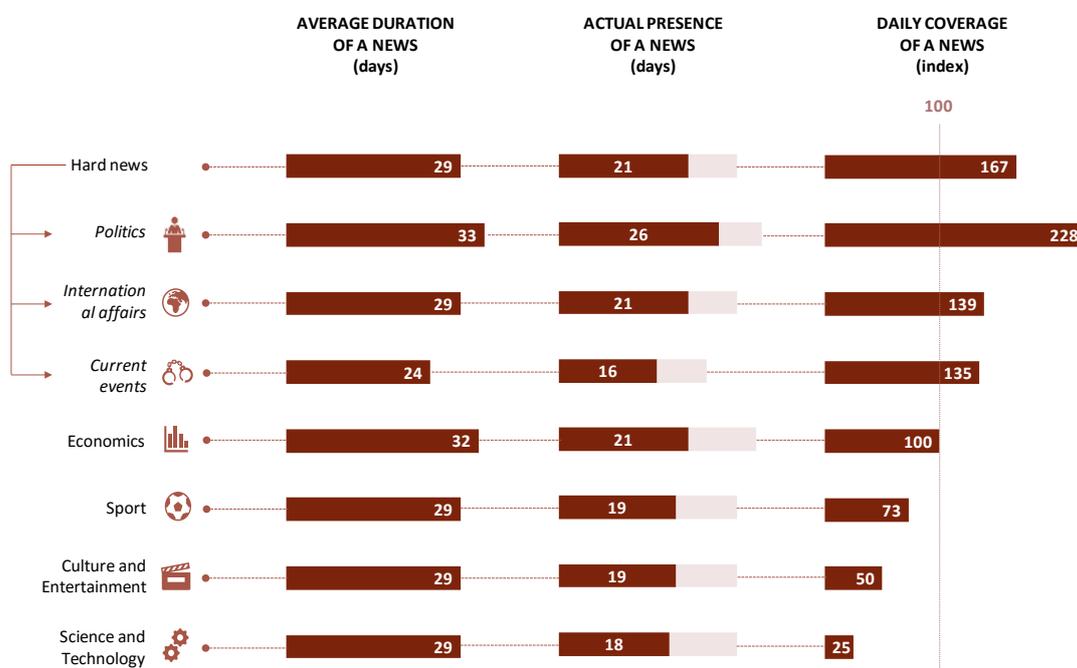
On the opposite, the news of report is characterized because of short lifecycle, which - for the most facts they refer to - begins shortly before the t_0 and ends in a moment not very far from the fact, in comparison with the other genres of news.

The comparison among the categories of news (shown in Figure 3.6) supplies a more acute diversification with relation to the third indicator, which is the level of coverage referred to the different topics. In particular, for the three genres of *hard news* (politics, report and international affairs), which under the profile of protection of pluralism have a more significant relevance, there is a coverage higher than the average, in particular for the news on politics. This data confirms, also at level of single news and single days, the abundancy of information produced for these categories, which was already observed at level of overall information category in Chapter 1 (see Figure 1.7), among other detecting the highest level of specialization of journalists dealing with it.

The news of economics, if generally find not much space in the informative offer and are treated by journalists with a low level of specialized education, have however a daily coverage in line with the average.

Low daily coverage in comparison with the average is however noted for the remaining categories of news, both sport and culture and show - which, as already said in the first chapter, affect the overall offer of information contents each for a rate next to 20% - as well as science and technology. This last category, for which the total offer of information contents is lacking at quantitative (lack of offer compared with the request) and qualitative level (low specialization by the journalists), is few treated also in terms of disclosure of information contents with single news as an object.

Figure 3.6 - Diffusion indices of a news, according to category



Note: the “duration of the news” is calculated as average distance between the first and the last days where a news is registered at least once.

the “actual presence of a news” expresses the average number of days, also not consecutive, where a news registers at least an occurrence.

The “coverage index” expresses the ratio (multiplied for 100) between the number of daily occurrences registered averagely by a news on a specific means and the number of daily occurrences registered averagely by a news on a whatever category. Therefore, values higher than 100 have a coverage higher than the average, values equal to 100 indicates a level of coverage in line with the average and value lower than 100 indicate a coverage lower than the average.

Source: Agcom elaboration on Volocom data

3.3. MODALITIES OF DISINFORMATION DISSEMINATION AND THE LIFECYCLE OF FAKE NEWS: SUBJECTIVE AND OBJECTIVE SCOPE

In the previous paragraph we examined the modalities with which a real news discloses in time inside the national information system. The obtained results, analyzed together with the scenario designed under the profile of production (see Chapter 1), allowed a more precise identification of the peculiarities, the satisfying aspects and the criticalities under the profile of pluralism of the “news supply”, both at the general level of the overall generated contents and at particular level of coverage given to single happenings.

In a similar manner, the objective of this paragraph is completing the framework defined in Chapter 2 concerning the production of disinformation, deepening how fake news disclose in the system by means of different sources, which - it should be stressed, not considering the action of sharing of contents by users by means of pages and personal profiles (the treating will be discussed in the next chapter).

So, the disclosure of disinformation in the national system should be studied in both components, concerning respectively:

- the *subjective framework*, namely, the behavior adopted in the treating of news by individuals, identified as sources of disinformation (see Chapter 2);
- the *objective scope*, namely, the disclosure in time of single fake news.

The *subjective framework* can be researched by the exam of the same diffusion indicators introduced in par. 3.2 (average duration, actual presence and daily cover of the news). In particular, circumscribing the analysis only to the sources of fake contents (specific websites, pages and accounts of social networks identified as such by external sources), it is possible to evaluate how and how much they are followed real and fake news. To this regard, it must be stressed that the sites producing disinformation do not include only fake news, but they combine real news and fake ones, right with the purpose of cheat and manipulate the reader of the news.

The exercise is indicated in Figure 3.7. Concerning the diffusion of real news on disinformation sources, the first comparison can be done with regard to the values of the same indicators assumed by other media (see Figure 3.4). If compared to the other information sources, you observe how the fake contents treat news more superficially (for less time and less coverage intensity) compared to all other media (with the exception of social networks, which have information purposes and distinctive diffusion structure). For the sources of fake contents, unlike the other ones, the production of the

offer is not the result of a journalistic work, for which, also when the disclosed contents regard real news, the information is characterized nor by accuracy nor by a deepening. In addition to this, it is not seldom that these sources are created around real facts, eventually inserted in fake contextualization, being phenomena of bad information.

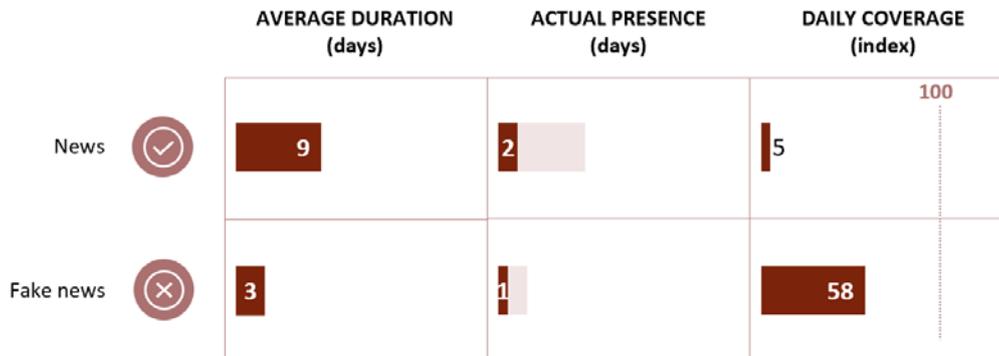
Moving the comparison on the different typologies of news disclosed on sources of disinformation, Figure 3.7 shows how the average duration for fake news is longer than the one of real news (3 days for the first ones and 9 days for the second ones), against of an amount of actual days of presence which is very low for both kind of news. This suggests that the distribution of contents created around news is very concentrated when the object of the disclosed contents is a fake news, averagely the totality of occurrences occurs in a single day.

The most significant difference between real and fake news is registered with regard to a daily average cover given to disinformation sources, which is definitely higher for fake news, noticing that the overall offer proposed and disclosed by these sources is composed of few real news (treated superficially or in order to disclose intentionally messages of bad information) and of many fake news.

Moreover, the rapidity with which single news is treated (real or fake ones) and a daily intensity of coverage, which remains in any case low in relation to the overall information system, if compared with the high quantity offered generally by disinformation sources (see Chapter 2), supply a first evidence of the way these one give start to many different news, letting that they have a massive disclosure by other channels, until they become viral. To this regard, let us think about the combined action of automatic mechanisms (algorithms) underlying the functioning of online platforms and of sharing of contents, carried out by users but, always more often, by bots.

In this sense, the production and treating of disinformation are substantially different from the information process, based on the journalistic work (see Chapter 1). In this case, a lot of news is produced (see Chapter 2) that is dealt with briefly and superficially. The purpose is not the deepening of a topic but the disclosure as broadly as possible of a fake information. In online disinformation, the moment of distribution and sharing by means of social networks of a fake news is so very important. Therefore, Chapter 4 will treat and deepen these aspects.

Figure 3.7 - Diffusion indices of real and fake news on disinformation sources



Note: the “duration of the news” is calculated as average distance between the first and the last days where a news is registered at least once.

the “actual presence of a news” expresses the average number of days, also not consecutive, where a news registers at least an occurrence.

The “coverage index” expresses the ratio (multiplied for 100) between the number of daily occurrences registered averagely by a news on sources of fake news and the number of daily occurrences registered averagely by a news on a whatever media. Therefore, values higher than 100 have a coverage higher than the average, values equal to 100 indicates a level of coverage in line with the overall average of the information media and value lower than 100 indicate a coverage lower than the average.

Source: Agcom elaboration on Volocom data

Shifting attention from the subjective to the objective component, through the following Figures, we are going to analyze the other aspect of the diffusion of disinformation (in this case including also phenomena of bad and disinformation³⁰, as some contents considered fake by debunkers are born or disclosed by traditional sources), namely the lifecycle of single fake news and the modalities with which they are disclosed inside the overall information system.

The *objective framework* can be examined starting as first from the representation of the average lifecycle of a fake news. To this regard, Figure 3.8 shows the curve (in red) expressing the daily average trend of the distribution of the contents produced regarding a fake news. The evidences which are clearer also in terms of comparison than the lifecycle of a real news are essentially referable to two observations.

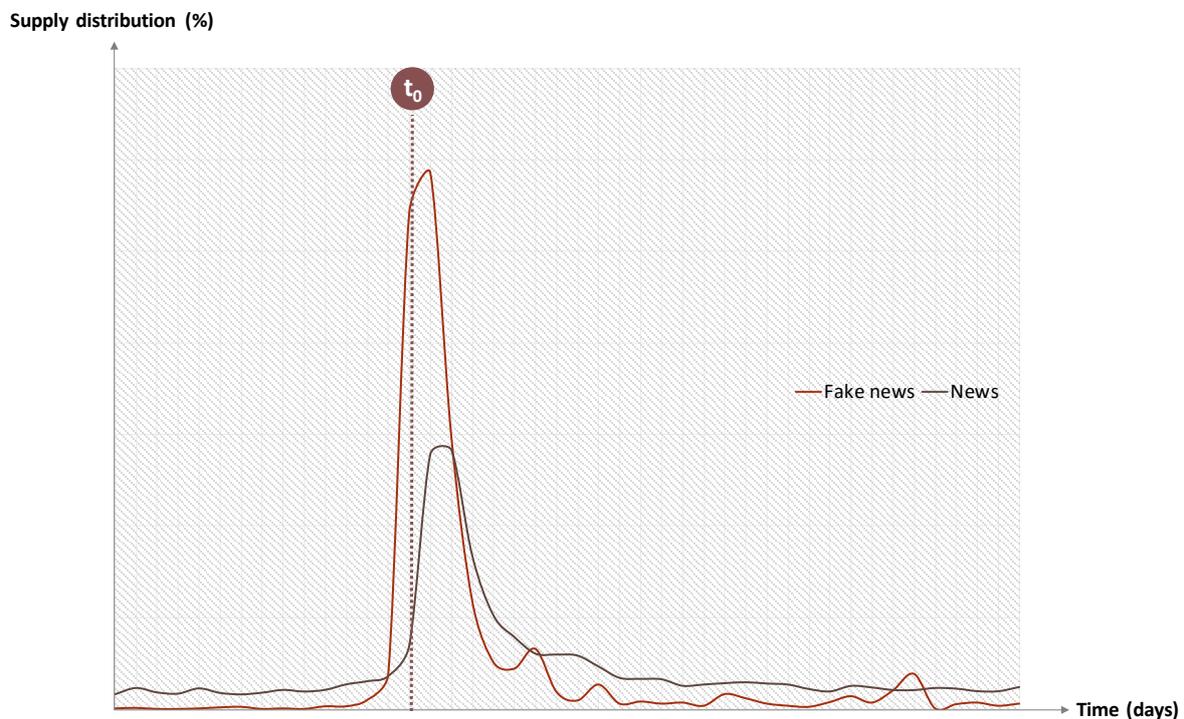
³⁰ See Chapter 1, Report “[Le strategie di disinformazione online e la filiera di contenuti fake](#)”, see

At first, the quite complete absence of anticipations, compared with the facts of the fake news. In fact, in the case of fake news, t_0 tends to be absolute and the happening, being fake, are seldom foretold.

Secondly, the lifecycle of fake news is significantly more focused on t_0 , reaching the peak the day after the t_0 and then decreases quickly towards value next to zero. In other terms, the rapidity of treating of fake news observed with reference to sources of fake contents (see Figure 3.7) can be generalized at system level. In fact, also if the contents about a fake news should be disclosed by other information media, the treating of the news tends to be quickly abandoned, as soon as verified in the framework of the authors' editorial activity.

Briefly, we can state that the two disclosures of the lifecycle of the real or fake news is statistically different the one from the other, and the fake contents are therefore easily recognizable because characterized by a "skewed" distribution and focused on t_0 .

Figure 3.8 - Lifecycle of a fake news

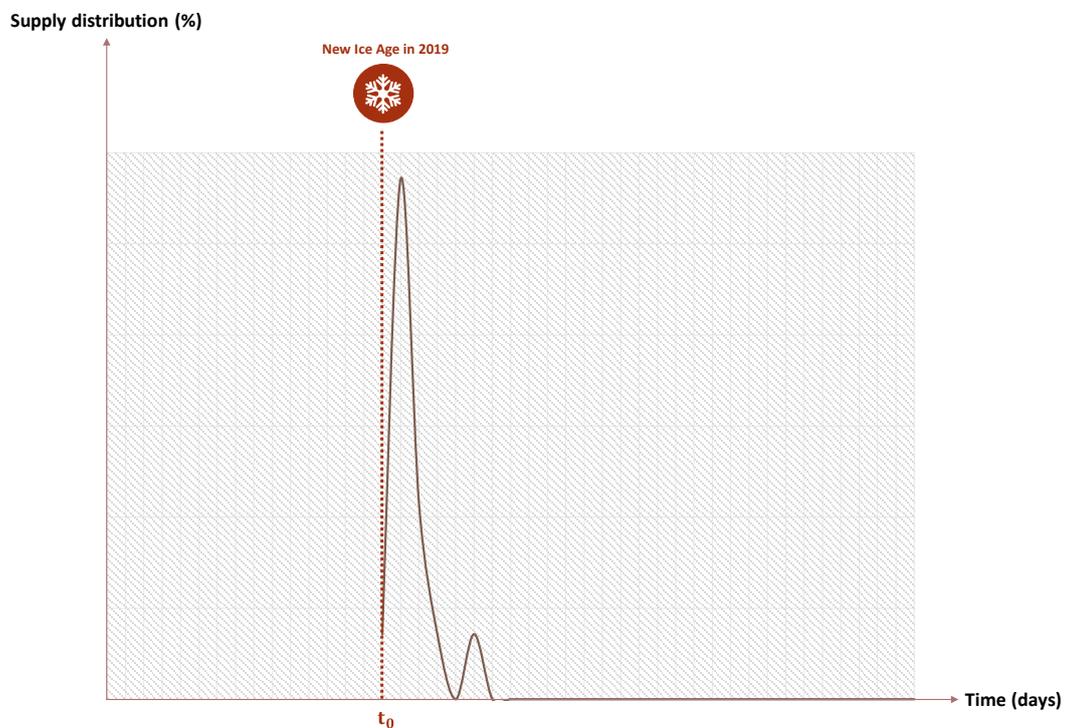


Source: Agcom elaboration on Volocom data

In Figure 3.9 there is an example of a specific fake news, for which the lifecycle begins in correspondence of t_0 , which coincides with the moment itself in which the first occurrence of the news is registered.

In detail, the news concerns the publication of assumed scientific study which could foresee the earth glaciation in 2019. As shown in the Figure, the lifecycle of the news, clearly unfounded, runs out in some days, even if it occurs in the period after t_0 , both in the sources of fake contents and in the other sources of online or printed information.

Figure 3.9 - Lifecycle of the fake news “New ice age in 2019”



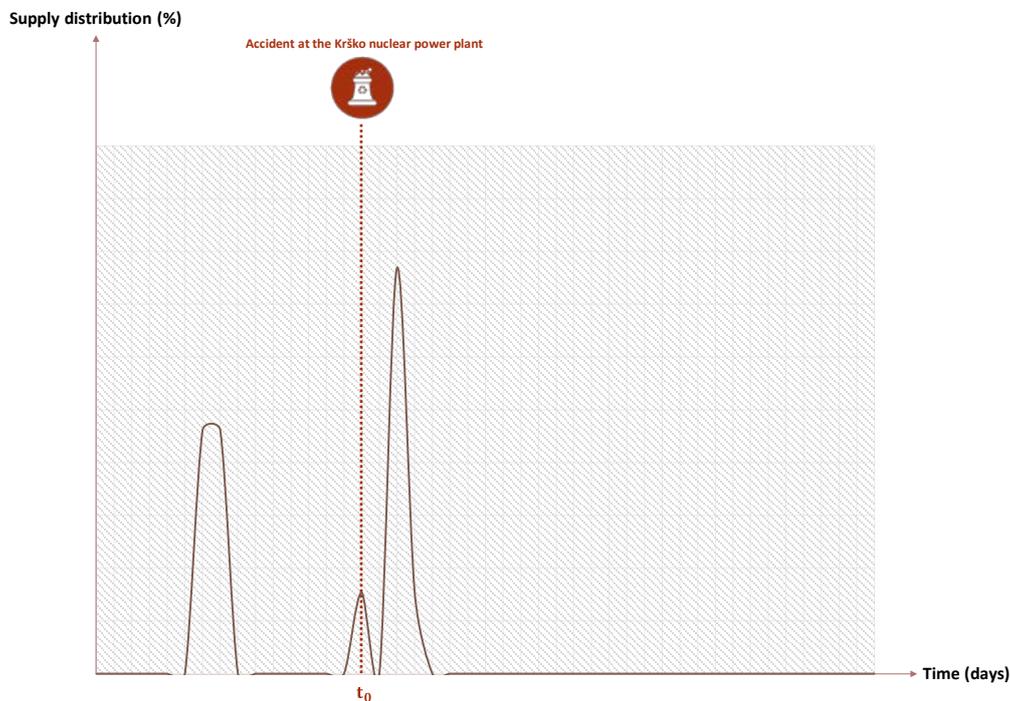
Source: Agcom elaboration on Volocom data

A further example of diffusion of a fake news in the information system is exposed in the Figure 3.10. The news refers to the fact, which did not happen, of an incident at the nuclear power station in Krško, Slovenia, not very far from the Italian border.

This fake news was launched online by a source of fake contents, with the obvious intent of leveraging the alarmism of citizens with certain ideological positions on the use of nuclear energy and at the same time of starting a clickbaiting process.

The news, which was echoed also in other information sources, has a peculiar lifecycle. In fact, occurrences are registered even before t_0 . In this case, the reason of the presence of anticipations is due to the fact that, some days before the fake news spread, a real event happened: in line with common practice, a plant of the nuclear station of Krško was stopped because of verifications on a problem detected in the security systems. The first contents appeared in the information system mentioned the incident hypothetically, creating the opportunity to invent fake contents, disclosed afterwards.

Figure 3.10 - Lifecycle of the fake news “Accident at the Krško nuclear power plant”



Source: Agcom elaboration on Volocom data

What the graphic representation of the curve identifying the lifecycle of fake news shows, is confirmed by the numeric data found for the indicators of news diffusion in the information system.

In this sense, Figure 3.11 shows in a comparative way the values of average duration and actual presence of a fake news in comparison with a real one. The difference is significant. If for a real news the average distance between the first and the last occurrence is of 30 days, for a fake one the period is 5 times lower (6 days). Consequently, the whole distribution of the offer of a fake news is concentrated

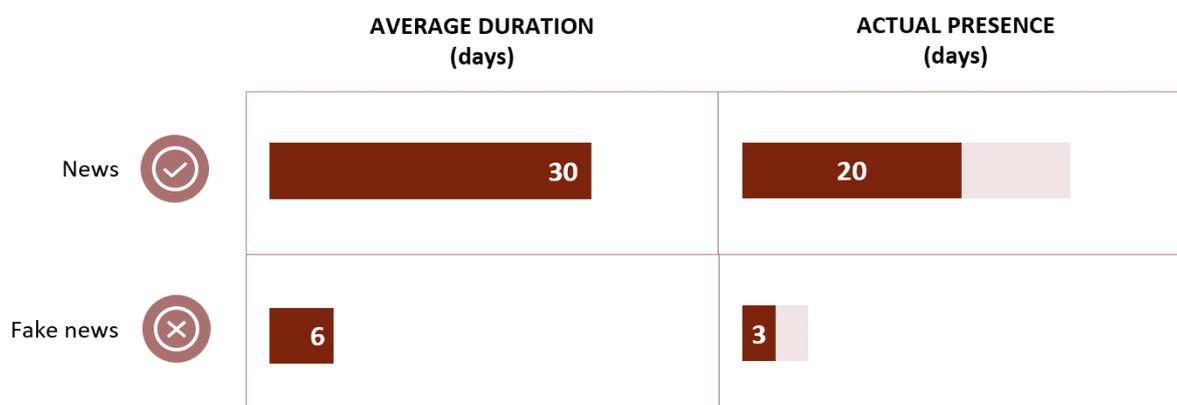
averagely in 3 days, whereas the effective disclosure of contents concerning a real news is distributed in a period of 20 days.

It is evident that the obtained results reflect also the nature itself of the information and disinformation strategies. The purpose of the diffusion of real news is to inform the public on certain events, following an editorial line, the interest of the request and the relevance for the public opinion on events. Differently, the diffusion of fake news - excluding the incidental inclusions in information sources - complies with different logics and motivations (of ideological, political or economic nature). Even if using the same cognitive mechanisms initiated by the information, it is moved in the most cases by a criminal intent of confounding, discrediting and instilling false beliefs, which is also difficultly to harmonize with exigencies of deepening and necessity of dedicating more time and space to a single news, which would risk of confute more than confirm the fake news.

Moreover, also when the fake news goes through traditional sources (in Figure 3.11, unlike Figure 3.7 we included all sources, not only the ones of fake contents, since the focus of the analysis is the object - the fake news - and not the subject) the duration is brief because of the activation of verification mechanisms regulating the information system. At last, the fake contents assume in some cases an informative value as such and are treated and discussed by the information system in the own meaning of fake news.

This does not exclude the existence of a discrete number of fake information contents in the traditional information system (mis-information), which are often the indicator of a regression in the use of such verification mechanisms; in this sense, in the framework of the Observatory of AGCOM of journalism, it was observed that “*journalists feel particularly encumbered by the necessity to operate in a unstable context, with decreasing income, without time for deepening and verification of the sources*” (see Agcom, 2017, [Osservatorio sul giornalismo: 2nd edition](#), par. 194, p. 79).

Figure 3.11 - Diffusion indices of real and fake news in the news system



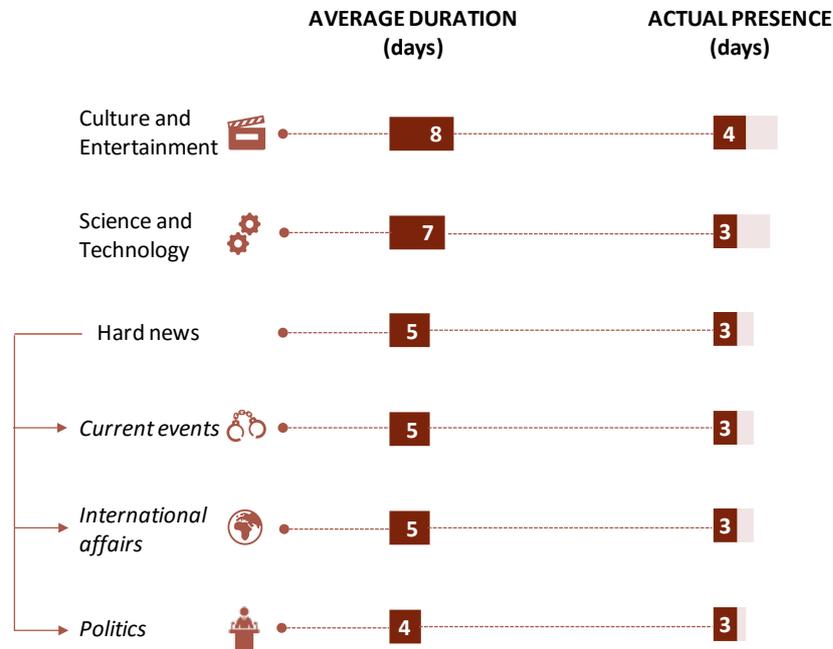
Note: the “duration of the news” is calculated as average distance between the first and the last days where a news is registered at least once.

the “actual presence of a news” expresses the average number of days, also not consecutive, where a news registers at least an occurrence.

Source: Agcom elaboration on Volocom data

Moreover, there are significant differences than the detected average value, if the analysis is included in the thematic category of fake news. Figure 3.12 notes how the average duration of a fake content is not much higher than the average in the case of culture and show and science and technology, reaching to a permanence of respectively 8 and 7 days in the information system. On the contrary, the duration of the lifecycle of fake hard news is slightly lower than the average, which is not longer than 5 days. However, they show an actual presence in the information system equal to the other ones (and so in percentage higher, if compared with the respective average duration).

Figure 3.12 - Diffusion indices of a fake news, according to category



Note: the “duration of the news” is calculated as average distance between the first and the last days where a news is registered at least once.

the “actual presence of a news” expresses the average number of days, also not consecutive, where a news registers at least an occurrence.

Source: Agcom elaboration on Volocom data

In short, the analysis on disinformation carried out in this paragraph and in the previous chapter showed that, concerning the subjective aspect, the disinformation sources assume a role which mostly consists in conferring impulse to fake contents inside the national system.

From the objective point of view, the shortness of the lifecycle of single fake news and the diffusion of the relative occurrences in few days are the indicator inside of carrying out a strategy of disinformation, preferring to treat many different news, avoiding to deep their contents. Very often, the fake news put in the informative system are characterized by the falseness of the narrated fact, as well as for their contagiousness in transferring emotional states and perceptions, which contributes, as said in par 2.1, to the easy creations of disinformation strategies. In relation to these strategies, the massive diffusion which actually leads to the economic valorization of fake contents occurs by means of the activation of further distribution channels, which directly recall

to the consumption and information actions, carried out by the users on online social platforms. The study of this aspect, introduced in several recent Reports of AGCOM³¹ will be treated in detail in the next Chapter, considering only the results of empirical study, made on big masses of data.

A last notation regards the relation of interdependence between the information system and the disinformation system.

The diffusion of fake contents seems to be connected to several criticalities, which concern always more the first one and that were treated the analysis and evaluation by the Authorities in other occasions. In this framework, it is necessary to recall synthetically the main questions linked to the examined pathological phenomena:

- the *continuous* reduction of the investments in information, which “*risks ... to begin a negative spiral, where the reduction of the quality of the information leads to a contraction of the income*” (Agcom, 2015, “[Indagine conoscitiva su Informazione e Internet in Italia](#)”, see 520, p. 217);
- the stressed and connected regression in the use of verification mechanisms in the framework of the journalistic profession (cfr. Agcom, 2017, [Osservatorio sul giornalismo: 2nd edition](#));
- the narrowness of times of online information, both in the production (see Chapter 1) and in the consumption (see Agcom 2018, [Rapporto sul consumo di informazione](#)), which risks to create less reliability in the collection of information by redactions, as well as a less level of attention by the user, despite of the level of quality of the information itself;
- the lack of (quantitative and qualitative) deepening of these topics, such as in particular the scientific and technological ones, which have a continuous lack of “news supply” in the framework of the traditional information;
- the subsequent decline of the reputation of the traditional information system in its complexity (phenomenon which, like above, characterizes not only the Italian ecosystem, but more in general the global one)³².

³¹ See Agcom, “[Rapporto sul consumo di informazione](#)”, see; “[Big data Interim report nell'ambito dell'indagine conoscitiva di cui alla delibera n. 217/17/CONS](#)”, see; “[Le strategie di disinformazione online e la filiera di contenuti fake](#)”, see

³² The annual Edelman’s report on the state of reliability of the citizens (“[Trust Barometer](#)”) shows that the media are the institution with the lowest reputation. In 2018, at global level (among the 28 biggest

In this framework, the citizens risk to give trust always to alternative and not-qualified information sources, which are often at the basis of disinformation strategies. So, the next Chapter will the modalities of online news consumption and the role of citizens and of online platforms (and of the relevant algorithms).

countries), only 43% of the population trusts the media. In Italy, this value is equal to 45%, reducing of 3 percentage points than the previous year.

THE PROPAGATION OF REAL AND FAKE NEWS ON ONLINE PLATFORMS



4.1. INTRODUCTION

The evolution of Internet and of the web technologies radically changed the paradigm of consumption of news, giving to formation a new scenario where people participate actively not only to the disclosure of contents, but also to the production of them. We went from a model where information was supplied by a defined group of official sources and by experts and journalists to the actual environment disintermediate and re-intermediate by algorithmic platforms, constituted by a heterogeneous mass of news, combining and mixing with the traditional flux.

In this context, online platforms carry out a crucial role for the information of users and, more in general, for the social life and the political and civil world. Only to make some examples, every 60 seconds on Facebook people create 3,3 million posts, publish 510.000 comments and update 293.000 states; on Twitter people send 350.000 *tweets*; on Whatsapp change 29 million of messages; on Google make 3,8 million researches. In other words, the users deal with a continuous and nonstop flux of information, real and fake news which coexist and make be confused. On online platforms, next to official information sources, unfounded and often not real rumors multiply, which can actually affect the creation of the public opinion. It is not a case that since 2013 the World Economic Forum has been stressing the global danger of a big digital disinformation, considering it as a part of technological and geopolitical risks, which include terrorism, IT-attacks, as well as the failure of global *governance*³³.

Disinformation, as we observed above, tends to exist where the information system fails: difficulties of monetization of the contents and contractions of the investments, reduced specialized training of the professional resources in certain subjects, exigence of quickness of updating of the information contents (mostly online) compromise the suitability of the online “news supply” at the level of the offer, of the deepening and of the coverage of the news. In broader terms, they are at the basis of the spread loss of reputation and trust by the citizens to the information system.

When people are misinformed, they have the inclination of defending the own beliefs, ignoring the actual proofs. When they do not have trust in the information system, their attitude of mistrust leads them to resist the facts³⁴, and the corrections by official

³³ See W.L. Howell (2013), *Digital Wildfires in a Hyperconnected World. Tech. Rep. Global Risks*, World Economic Forum.

³⁴See J.H. Kuklinski et al. (2000), “Disinformation and the Currency of Democratic Citizenship.” *The Journal of Politics*, 62 (3), pp. 790-816.

sources (even if founded scientifically) can be not able of reducing the wrong idea, sometimes having also a backfire effect³⁵.

In a situation like this you register the inclination of individuals to get informed by means of the own online net of contacts and to give reliability tot the contents and sources confirming the own speculations, to share and to state personally the own point of view or ideological orientation. And right these topics the subjects carrying out the strategies of disinformation use to create the effect of viralization, which they can contribute to share, together with the technological characteristics and the algorithms of personalization of the online platforms.

After having considered the aspects linked to the production (Chapter 1 and 2) and to the diffusion (Chapter 3) of real news and of fake contents by information sources and by the ones considered by the debunkers as disinformation sources, the next paragraphs will be dedicated to the analysis: *i)* of the consumption of news on online platforms and of the factors leading to phenomena of polarization; *ii)* of the modalities through which the strategies of disinformation are carried out and how they diffuse massively on online platforms, also using the information actions of the users; *iii)* of the relation between the polarization of the users about certain topics and themes, object of disinformation.

In order to face correctly the study of these topics and to understand the mentioned phenomena, we considered necessary to follow a methodological approach based on the exam of big masses of data, which, by using experimental instruments, can consider socio-cognitive factors, intervening in the studied processes. In this sense, most part of the analysis exposed in this Chapter are the result of the research collaboration of Walter Quattrociochi, who - in the carrying out of the tasks AGCOM assigned to him and in the framework of the sector inquiry of this Report³⁶ - produced a first summarizing contribution on the most recent scientific evidences, detected on the online disinformation.

³⁵ See B. Nyhan, J. Reifler (2010), "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior*, 32 (2), pp. 303-330.

³⁶ See the [results](#) of the comparison procedure for the assignment of an individual research task on "Informazione e piattaforme digitali", ex [Law No. 78/17/SG](#).

4.2. THE NEWS CONSUMPTION ON ONLINE PLATFORMS AND POLARIZATION

Globally, online platforms have definitively become integral part of the daily information diet of citizens. To this regard, in the last [Report on information consumption](#), AGCOM found out that Italians using Internet to look for information and mostly by means of online platforms. The consultation of at least an online platform to get informed regards in fact 55% of the population, whereas the fruition of the editorial online sources stops at 39%. In particular, social networks and search engines reach bigger portions of population, each equal to 37%. Also when the informative purpose is political/electoral, the detection of news and points of view on the Internet goes mostly by means of online platforms, rather than through websites or apps of newspapers or other online information sites: 28% against 8% of the major population.

More in general, platforms are always oftener actual *gatekeepers* for the access to information. In this context, characterized by the “unpackaging” of the informative product and by a fragmented fruition of the contents (articles, comments, videos, *posts*, etc.), platforms are new intermediaries for the access to online information by the individuals, whereas the access is often the result of incidents or cases of the discover of news by the citizen himself, who also risks of not being aware of the nature and origin of the information³⁷.

In order to analyze how users actually use news on online platforms and the relative features, we propose in the following the results of scientific researches, carried out on datasets of millions of data. The followed methodological approach combines sciences of nets and statistic mechanics, in order to find from a macroscopic point of view physical features of the interaction of users with the information.

A first analysis is focused on the modalities of consumptions adopted by the users in the fruition on social networks of news produces by official information sources. In detail, we examined the models of consumption of 376 million of users on Facebook, in a period of 6 years³⁸. For every Facebook page of an information source, we extracted the data concerning information actions carried out by individuals (likes, comments) in relation to the news (*posts*) published by the sources, as exposed in Table 4.1.

³⁷ For example, “*the user accessing to Facebook can find news or, more probably, find them accidentally, by means of several sources: pages of newspapers, native Facebook pages, links to news shared by friends with the relative comments and discussions*”, pp. 39-40; [Rapporto sul consumo di informazione](#), see.

³⁸ See A.L. Schmidt et al. (2017), “Anatomy of News Consumption on Facebook”, *PNAS*, 114 (12).

Table 4.1 - Composition of the dataset for the analysis of the news consumption models on social networks

Variables	No. observations	Incidence on total
Facebook pages of information sources	920	
Post	12.825.291	
Like	3.621.383.495	
Comments	366.406.014	
Users	376.320.713	100%
Users who put a “like”	360.303.021	96%
Users who commented	60.115.975	16%

The collected data in Table 4.1 give a first indication on the volume of interactions, produced averagely by the users about a news on an online platform. The near totality of the users of social networks (96%) expressed at least a positive reaction for a post published on an informative source, creating in the considered period an overall amount of over 3,6 billion of likes. 16% of the users on the platform directly intervened in the discussion on a news, commenting it and producing in the whole period 366 billion of comments.

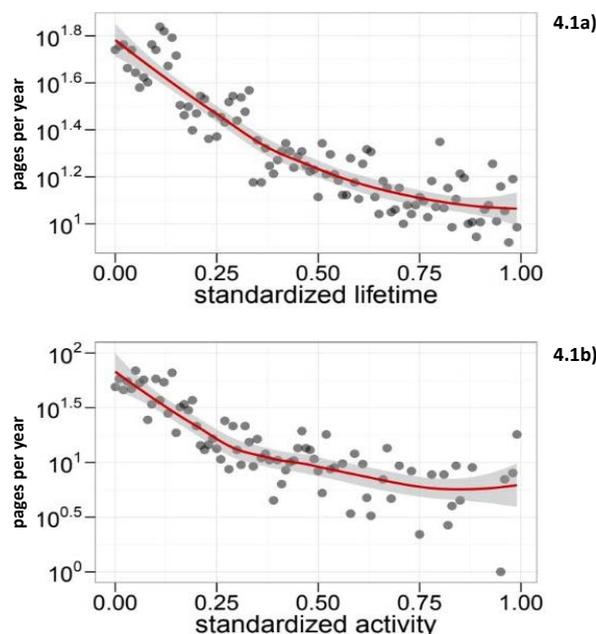
Observing the interactions of the users with Facebook pages of the informative sources, it is possible to evaluate the grade of heterogeneity of the occurred information consumption, namely verifying with how many different information sources the users interacts in general on the platform.

In this sense, Figure 4.1 shows how the number of pages of information sources, the individual interacts with (in a year) varies according to the variation of two factors: the duration of the *lifetime* between the first and the last interaction of the user (with regard to a post of the information sources); and the level of involvement of the user (in terms of quantity of likes expressed for a post, so-called *activity*). The values of both factors are standardized (*standardized lifetime* and *standardized activity*), for which are between a minimum of 0 and a maximum of 1.

In detail, Figure 4.1 shows that on online platforms a user tends to interact with a restricted group of information sources. Moreover, according to the increase of the duration of the interaction period as well as of the level of involvement, the number of information sources with which the user interacts reduces. In fact, while the users with a very low interaction duration and a level of involvement interact with about 100 pages in a year (30 in a month); the most active users and with the longest duration of interaction interact only with 10 pages in a year (less than 4 pages in a month).

These evidences, taken from the study on the attitude of hundreds of millions of users in several years concerning a big number of information sources, distinctly show the diffused trend of users of limiting the own information activity to a reduced number of social pages, referable to information sources, showing substantially the emergency of a selective exposition phenomenon in the consumption of news, which grows according to the activity of the users.

Figure 4.1 - Heterogeneity of the news consumption, according to lifetime (4.1a) and activity (4.1b)



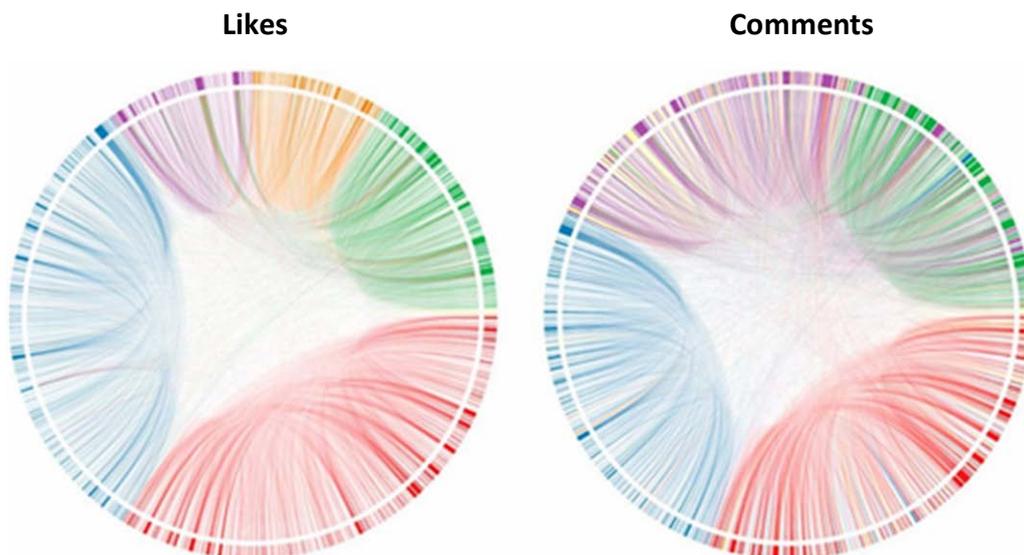
Note: the “duration of interaction” (*lifetime*) represents the period between the first and last interaction of a user with regard to a post of the information sources.

The “level of involvement” (*activity*) represents the quantity of likes expressed by the user for a post of the information sources.

The trend to interact with few pages of information sources can determine their aggregation in groups (so-called communities), on which the users focus their attention and from which they take information.

In order to verify this hypothesis, we at first represented the users' interaction with information sources, in order to define the structure of the community of pages, detected according to the activities of users. In this sense, in the graphics composing Figure 4.2, the nodes (along the circle) represent the pages and two pages are linked if a user expresses appreciation (or comments) at least a post of both of them. The importance of a link, which is the dimension of the bow, is determined by the number of users, the two pages have in common. The colors identify the belonging a node (page) to a specific community³⁹.

Figure 4.2 - Community structure, according to the kind of informative action



Note: the communities were detected by the algorithm Fast Greedy (FG). The colors identify the belonging a node (page) to a specific community.

The communities were detected by the algorithm Fast Greedy (FG). The algorithm of detection of the communities is used to identify groups of nodes in a net. The strategy is based on the modularity quantifying the division of a net in separate clusters and a high modularity corresponds to a dense connectivity among nodes in a community and scattered connections among the modules. See A. Clauset, M.E.J. Newman, C. Moore (2004), "Finding Community Structure in Very Large Networks", *Physical review E (APS)*, 70 (6).

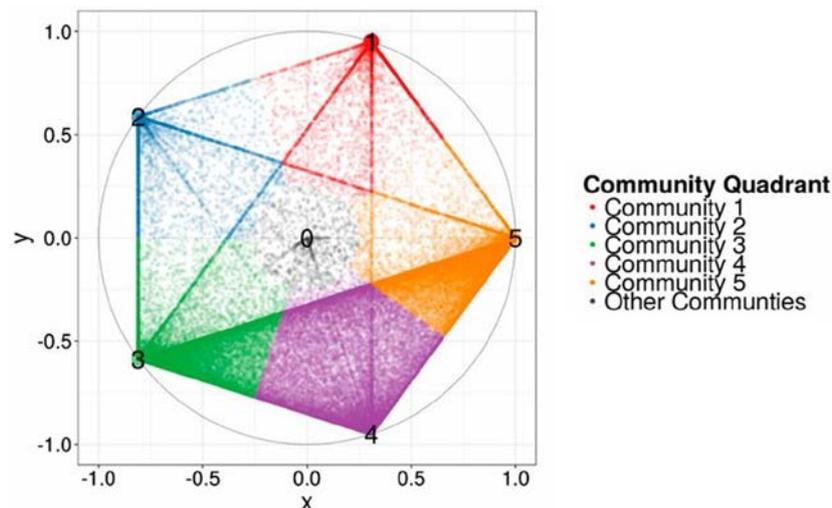
Examining in detail the activities of the users among the various communities and measuring how they extend between the pages, we find in fact that the most part of users remains confined in their own communities.

For a better understanding of the relation between the groups of pages and the attitude of the users, it is possible to quantify the fraction of activity which a user has in the biggest community and the one s/he has in any other community.

Figure 4.3 shows the activity of the users in the biggest communities. More in detail, the summits of the Pentagon represent the 5 biggest communities while the central point represents everything else. The position of every point is determined by the number of community, with which the users interact; the dimension and the transparency show the number of users in that position.

The Figure shows that the users are highly polarized and how their attention is limited to a single community of pages. In short, the interaction of the Facebook users with the sources of news indicated a structure of communities of pages, to which well-defined groups of users correspond. Since the users tend to focus the interactions towards a limited number of pages, the group of information sources on social networks is grouped in a specific communitarian structure.

Figure 4.3 - Users polarization



Note: the summits of the Pentagon represent the 5 biggest communities and the central point represents everything else. The position of every point is determined by the number of communities, the user interacts with. The dimension and the transparency indicate the number of users in that position.

Focusing the analysis on the consumption models indicated by the users towards a specific topic, a specific study was carried out with reference to the information on the Brexit.

The research considered all posts published on Facebook by official information sources. For every page, we collected all posts from 1st January to 15th July 2016 (the referendum on Brexit took place on 23rd June 2016), and all relative *likes*, well as all comments about it.

All pages of the information sources were divided in two groups: “pagine Brexit” including all pages engaged in the discussion about the Brexit and all others. In 81 pages (generating 303 thousand posts), 38 published at least a news on the Brexit (for a total of 5 thousand posts). The composition of the *dataset* is shown in Table 4.2.

Table 4.2 - Composition of the dataset for the analysis of the news consumption models on the topic “Brexit”

Variables	Total	(of which) Brexit
Facebook pages of information sources	81	38
Post	303.428	5.039
Like	186.947.027	2.504.956
Comments	38.182.541	469.397
Users who put a “like”	30.932.388	1.365.821
Users who commented	7.222.273	259.078

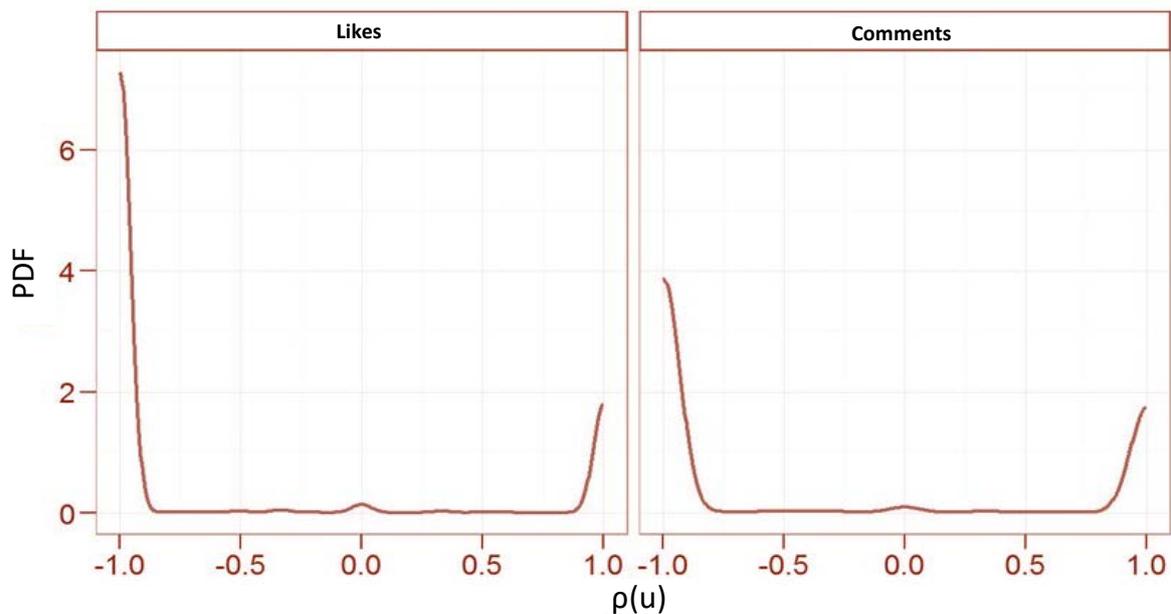
As first step, we identified all structural properties of interactions about the “pagine Brexit”, in order to verify if also this specific case is included in the phenomenon of polarization of users.

Figure 4.4 shows the graphics of interactions between pages and users, where the nodes are pages Brexit and two pages have a bow linking them if a user is active (makes an informative actions) on both. The importance of a link is determined by the number of users in common of the two pages.

Without making any categorization a priori of “pagine Brexit”, we clearly observe the spontaneous creation of distinct and separated communities (C1 and C2), where the

seem to be divided and put their attention to a limited group of specific pages, tending to focus on an only narration and ignoring the other one.

Figure 4.5 - Users polarization distribution towards “Brexit pages” community



In conclusion, the studies carried out on enormous quantities of data about the informative consumption by the citizens through algorithmic platforms, such as social networks, demonstrating that

- despite the presence of a very high number of informative sources (which are also free), the users tend to access to a reduced number of them;
- the most active users on social are the one, who access to less informative sources;
- this phenomenon, which seems to be paradoxical, is explained by the increase on social medias of very polarized communities, distinct and separated the one from another.

These elements produce a very significant effect for the information system in its complexity: despite the plurality of informative sources on the net, the models of online

consumptions do not lead to an enlargement of the cognitive space of the individuals, who risks also to be closed in the so-called echo chambers, which are very polarized. It is as if we would say that the information pluralism on the net and on the social media in particular risks of not producing the expected positive effects, connected with the enlargement of the points of view, the citizen has at his own disposal.

Moreover, in the framework of the [Report on information consumption](#), AGCOM observed that the polarization works already at the level of selection of the media for the access to the information: the internet users are in fact more polarized (ideologically) than those using the TV as information media.

Moreover, in the examination of the relation between the ideological polarization of the users and the informative activities they carry out in the net, we found that the polarization may have a significant effect on the biggest *engagement* towards that news disclosed on social networks, with evident reflections on the realization of diffusion phenomena on radicalized positions and on the creations of ideological bubbles.

In the following paragraph, where we focused on disinformation, we will deep empirically the relevance of the phenomenon of the polarization in the actuation of the disinformation strategies by means of online platforms.

4.3. THE IMPLEMENTATION OF THE DISINFORMATION STRATEGIES ON ONLINE PLATFORMS

The concept of disinformation and main features of the phenomenon have been already defined in Chapter 2, where we described how the entry in the system of fake contents and the diffusion of disinformation occurs, with the involvement of different individuals (creators and executors), through four following phases (for a detailed description of the chain, see AGCOM's report "[Le strategie di disinformazione online e la filiera dei contenuti fake](#)"; see also Figure 2.1). As said before, the main phases composing the chain of online fake contents can be referred to the creation of the message; to the production of the content where it is incorporated; to its distribution of the content; to the - monetary or not - valorization.

In the creation phase, the message - in order to be efficient - should be created so that the audience is reached and activated, involving it also in a further diffusion of the content. In the production phase, the communication code is furtherly refined, so that it is possible to find different "genres" of fake contents, for example according to the different level of "manipulation" conferred to the message (ideally, we go from the completely fake content, which is created *ex novo*, to the one which is based on a originally true information which was manipulated afterwards). In the distribution phase, you choose one or more channels, through which the content should be disclosed, and you find the mediatic context where the content is. In the last phase, the fake contents are valorized, namely can produce immediate monetary or not monetary incomes, or can reach the expected goals without necessarily creating an income flux, because they do not respond to ideological- political motivations. In particular, with reference to the income, mostly in short-medium term commercial strategies, there are two main sources of remuneration for the producers: advertising activities and, in some cases, the direct contribution of the users, obtained with criminal actions. In the framework of a longer term, incomes can come for example from disinformation campaigns which, damaging the image and the reputation of a concurrent company, goal to steal market shares. At last, we find the presence of hybrid strategies, where we have at the same time political-ideological strategies and purposes of economical nature, which can produce an alteration of the market assets, such to determine an enforcement of the economical position of some company to the detriments of other ones, creating economic advantages for the creators.

Observing the mentioned disinformation chain, the analysis on the phases of creation and diffusion of fake contents, carried out in the previous chapters showed, against an

elevated overall offer of disinformation sources, identified as such (websites and pages/*social* accounts), a low coverage in the diffusion of single news, both in terms of contents produced on the single fake news and in terms of duration of the lifecycle of the news. So, it followed that the role of the disinformation sources (not envisaging a journalistic job) is mostly the one of giving impulse to many different news, letting that they are massively diffused through other channels, till they become viral. To this regard, in several occasions AGCOM could prove that the conformation and the mechanisms of functioning of online platforms – providing the conjoint use of automatic personalization systems (operating on the basis of algorithms and acquired *big data*) and the possibility of user's interaction – make the proliferation of fake news and the viral propagation of polarizing fake news easy.

As a consequence, the models of informative consumptions of users on online platforms, the intensity and the modalities of their interaction with relation to news and narrations assume centrality in the ideation and the actual actuation of disinformation strategies. In detail, the selective exposition and the polarization the users broadly have (see par. 4.2), natural addressees of such strategies, can be skillfully exploited by individuals intervening in all phases of the disinformation chain, in order to activate the cognitive and emotional mechanisms of the users, both in the predisposition to receive the message and in the inclination in sharing the content.

Already in the phase of creation of the fake content, attention is given by the creators to the user, in the moment when s/he is seen as part of the potential audience of the disinformation strategy. More specifically, those who carry out the disinformation strategy look for the message, identifying three precise aspects:

- the “profile of the users and the reference target”: right the profiling of online users allows the preparation of more efficient *fake* message and contents and, during the distribution, the realization of a goaled diffusion of them;
- the “topics to be treated”, which are often included in the discussed topics, of particular interest and polarizing (which means they can create and increase the separation of individuals in distinct groups)⁴⁰;
- the “way in which individuals elaborate information”, in order to be able of acting on the cognitive *bias* of the individuals and satisfy not only the need of

⁴⁰ On the link between disinformation topics and polarization, see M. Del Vicario, W. Quattrociocchi, A. Scala, F. Zollo (2018), “Polarization and Fake News: Early Warning of Potential Disinformation Targets”, see

information, but most of all their expectations in terms of complying with the own beliefs (so-called *confirmation bias*), of emotional involvement and sharing of the view of the world.

Afterwards, during the distribution, the individuals pursuing disinformation strategies - individuals or groups of individuals or more or less structured organizations - can be very often act using automatic mechanisms like bots, which allow to publish and disclose disinformation contents by means of many fake accounts or fake social profiles. At this point of the chain, the mentioned mechanisms can be activated by the users, addressees of the contents, who - also unconsciously - share fake contents (eventually modifying them and creating new ones) and encourage their diffusion, which has as a consequence the amplification and acceleration of the so-called *snowball effect*.

In fact, the mechanisms of interactions on social platforms allow in general to everyone to be an active part of the promotion of a disinformation content, transferring the own emotional states, contributing to the processes of viralization.

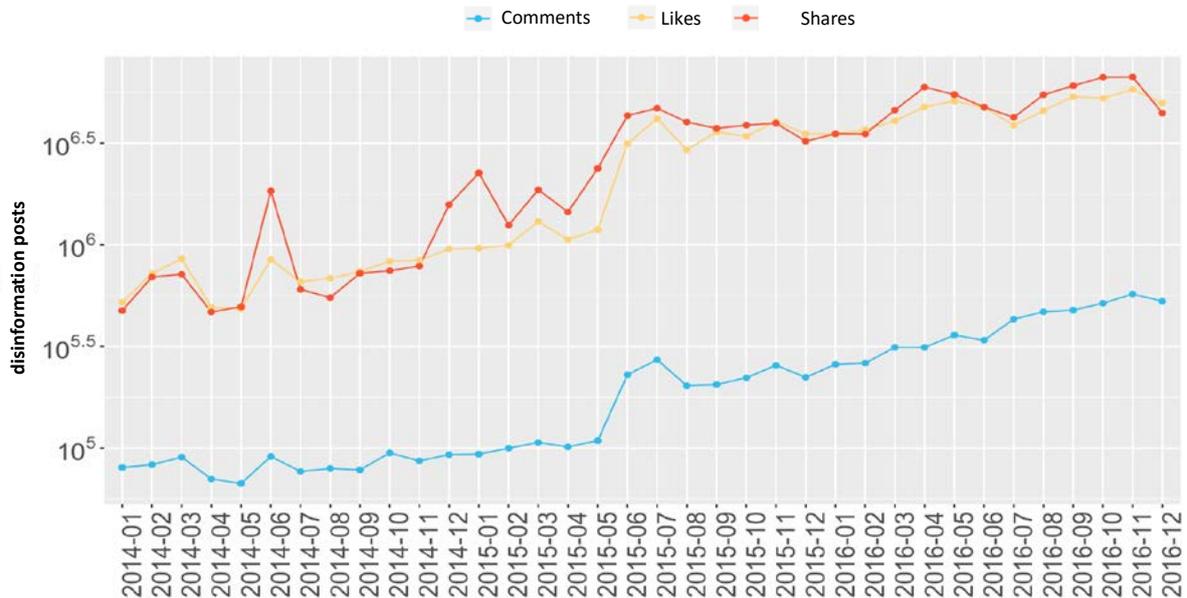
The following Figures supply a clear identification of the extent assumed in time of the users' interactions around disinformation contents on online platforms. The graphics refer to the interaction of the Facebook users in Italy to posts created by disinformation sources, identified as such by individuals specialized in debunking (namely by means of a subjective approach, see par. 3.3).

At first, it should be noticed that every informative action carried out by users, compared with a news is associated to a specific meaning⁴¹: while a like represents a positive reactions, of appreciation of the post, a sharing expresses the desire of increasing the visibility of a certain information; at last, a comment is the way to start or contribute to the increase of a discussion, taking form around the topic of the post.

That having been said, Figure 4.6 show a significant increase, if compared with 2014, of the amount of interactions concerning fake news, trend which occurs for every three kinds of actions. The values detected for likes and sharing are very elevated and similar. In detail, the data of sharing is quite significant for the strategies of disinformation, because they have the goal of increasing the visibility of the fake content and to encourage their effective diffusion.

⁴¹ See N.B. Ellison, C. Steinfield, C. Lampe (2007), "The Benefits of Facebook "Friends": Social Capital and College Students' Use of Online Social Network Sites", *Journal of Computer-Mediated Communication*, 12 (4).

Figure 4.6 - Monthly trend of users' interactions with disinformation posts



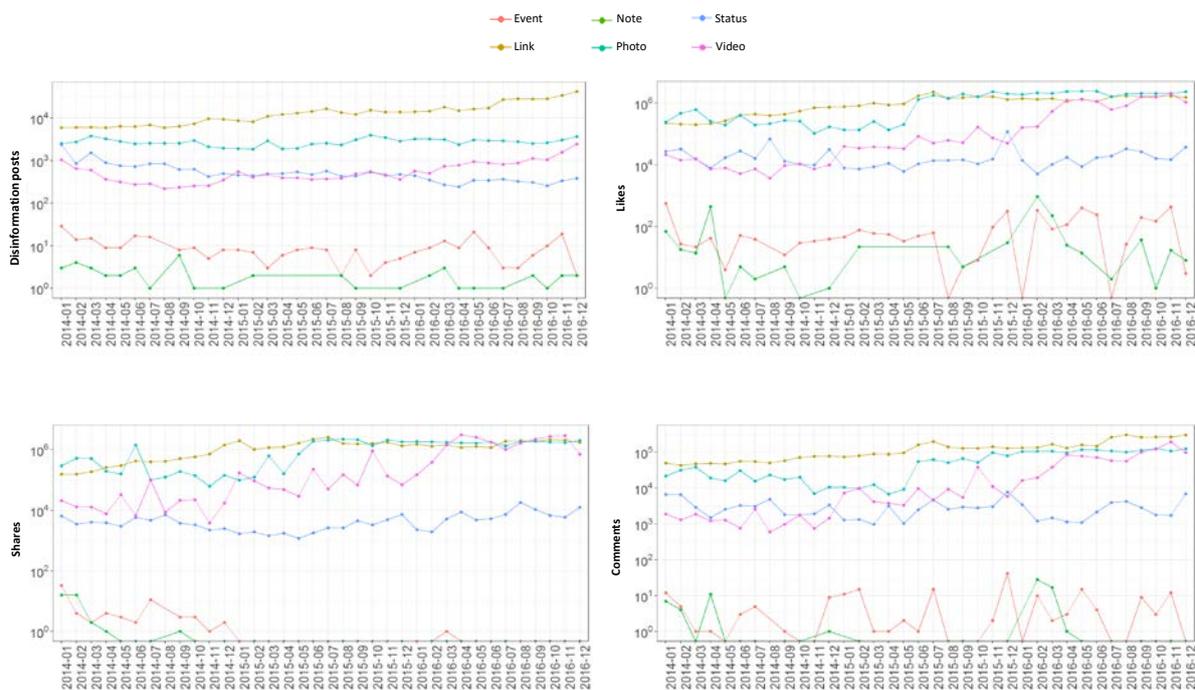
Note: the posts of disinformation are the one created in Italy on Facebook by pages, identified as disinformation sources by experts of debunking activity.

Clarifying the analysis according to the kind of disclosed content (link, event, personal note, state, Figure, video) by means of a post from the pages of the disinformation sources, Figure 4.7 indicates both the trend of the produced offer (first graphic on the top left hand side) and the trend of the interactions generated by such posts by the users (in the other 3 graphics respectively about likes, sharing and comments).

The links to an external page represent the most common form of disinformation content. In addition to this, the links, together with Figures and videos, belongs to the formats creating the highest number of interactions and the first ones in terms of received comments. These results contribute to confirm what we observed in Chapter 3 and is mentioned above, what concerns the function of impulse of the disinformation sources, compared to single fake news. The lacking coverage and diffusion given from this single news seems to be compensated and highly amplified by the interactions of users with contents. Therefore, on online platforms the impulse launched by disinformation sources propagate and reaches always more users, who appreciate,

comment and, most of all, share the *links* to the news (and the other formats of *fake contents*).

Figure 4.7 - Monthly trend of disinformation posts and relative users' interactions, according to the kind of content



Note: the posts of disinformation are the one created in Italy on Facebook by pages, identified as disinformation sources by experts of debunking activity.

In other words, the conjoint analysis of the evidences detected in this and in the previous chapters showed that the main phases of the disinformation system chain can be schematized in the following way:

- the profiling of the users and the selection of the topics to be treated: in Italy, 57% of the production of fake contents regards topics on politics and report, while about 20% scientific topics; all topics with a significant emotional impact can be decisive and are often not treated in a suitable way by the tradition information system;
- in particular, in 2018, there was a peak in the production of disinformation contents linked to the political elections of 4th March, where the topics

concerning parties, candidates, the proposed policies, the elections themselves, the economics and the report central. In addition to these ones, we detected lots of topics linked to science, health and immigration;

- the treating of these topics by disinformation sites is peculiar and distinctive if compared with the one occurring by means of the information activity; a fake content is promptly created, whose treating is concentrated in few days (not more than 3) and the topics are discussed superficially and in an impressionistic way, with the purpose of stimulating the state of mind of people.
- after the triggering, the fake news is put and launched in the system of online platforms, also by means of the unaware contribution of users sharing and commenting the news itself on social networks.
- The viralization of the fake content is made possible by the modalities of news consumption of social media which, most of all for disinformation news, occurs in the framework of closed distinguished and polarized communities:

The next and last paragraph will be therefore dedicated to deep the mechanism through which fake news is consumed and shared on the online platforms.

4.4. THE INTERACTION MECHANISMS WITH THE FAKE CONTENTS ON ONLINE PLATFORMS

In consideration to the relevance the users' interactions assume with the fake news (because of the reached volume and because of the level of involvement of the individuals) in the definition at first and in the actual disclosure of the disinformation strategy afterwards, is very important to understand the mechanisms. Therefore, the following analysis are orientated to examine, by means of millions of data, the modalities, through which the massive disclosure of disinformation on online platforms occurs, and how the polarization of users is involved in this process.

The objective is to characterize the role assumed by cognitive determinants (exhibition selection, polarization, *confirmation bias*), regulating the individual informative actions on online platforms and in the propagation of disinformation.

In this sense, an empirical study⁴² was made on the users' attitude with regard to different and specific narration, referable to: conspiracy theories (*conspiracy*) and scientific theories (*science*), so themes of disinformation on the one hand and of information on the other one. As we saw before, these topics are not suitably treated (in terms of quantity and quality) by the traditional information system and are largely discussed in the context of disinformation context.

In detail, the study focuses on the analysis of three aspects:

- i. the interaction of users with Facebook pages belonging to the two narrations, in a time interval of five years, in the Italian and in the American context;
- ii. the comparison of consumption and interactions models of users with regard to contents belonging to both categories (conspiracy and science) on different social media (Facebook and YouTube);
- iii. the answer of the users to three kinds of news, namely the ones complying with the own beliefs, the satirical/caricatural/derisory ones towards conspiracy

⁴² The description of the study and the presentation of the results base on the technical report of Prof. Walter Quattrociocchi. See also A. Bessi et al. (2015), "Science vs Conspiracy: Collective Narratives in the Age of Disinformation", *PLoS ONE*, 10 (2); F. Zollo et al. (2015), "Emotional Dynamics in the Age of Disinformation", *PLoS ONE*, 10 (9); A. Bessi et al. (2016), "Homophily and Polarization in the Age of Disinformation", *The European Physical Journal Special Topics*, 225 (10); M. Del Vicario et al. (2016), "The Spreading of Disinformation Online", *Proceedings of the National Academy of Sciences*, 113 (3); F. Zollo et al. (2017), "Debunking in a World of Tribes", *PLoS ONE*, 12 (7).

theories, and the one not complying with the own beliefs (for example, the attempts of debunking).

From a methodological point of view, for every category of narration we detected the pages disclosing the relevant news. The category “conspiracy” includes all pages disclosing news on conspiracy theories (alien attacks, chemtrails, geocentrism, as well as paradoxical effects coming from vaccinations). The category “science” includes all pages of scientific spread, among which the ones of institutions, organizations, scientific publishing. In addition to this, we detected two other categories of pages: the ones about “satire on conspiracy theories” (for the composition of the *dataset* on the Italian case) and the one of “debunking on conspiracy theories” with the purpose of correcting fake or false theories and news (for the *dataset* on the American case).

The individuation of the pages to be included in the analysis was carried out involving several experts active in the debunking activity, most of all on matter of fake news on scientific topics and conspiracy theories. All pages were manually verified observing their auto-description and the kind of proposed content.

The creations of the *datasets* for Italy and the US are respectively reported in the following tables. The whole procedure of data collection was made by means of API of Facebook Graph (Facebook 2017), using exclusively all publicly available information.

It should be at first observed that the examination of the collected data for the creation of the *datasets* allows the identification of some general trends. At first, with regard to the Italian context, Table 4.3 shows that, in proportion to the considered pages, the sources of scientific disinformation disclose a quantity of contents 3 times higher than the sources of scientific information, whereas in the US we observe the opposite situation (see Table 4.4).

Table 4.3 - Creation of the dataset for the Italian case

	Science (Information)	Conspiracy (Disinformation)	Satire on conspiracy
Pages	34	39	2
Posts	62.705	208.591	4.709
Likes	2.505.399	6.659.382	40.341
Comments	180.918	836.591	58.686
Users who put a "like"	332.357	864.047	15.209
Users who commented	53.438	226.534	43.102

Secondly, in Italy, for the scientific information and disinformation we can register an analogous level of interaction of the users. Averagely, single active users express 8 likes and 3 comments, in the case of information contents, and 8 likes and 4 contents in the case of disinformation contents.

In the US, the production as well as the reaction in terms of like created in the users seems to be higher concerning the contents of "science" (averagely, 11 likes per active user) than of the ones of "conspiracy" (7 likes per active user), against an average number of comments created by an active user which is quite the same (3 comments per active user).

Table 4.4 - Creation of the dataset for the US case

	Science (Information)	Conspiracy (Disinformation)	Debunking on conspiracy
Pages	83	330	66
Posts	262.815	369.420	47.780
Likes	453.966.494	145.388.117	3.986.922
Comments	22.093.692	8.304.644	429.204
Users who put a "like"	39.854.663	19.386.131	702.122
Users who commented	7.223.473	3.166.726	118.996

In short, we detect a completely different situation from Italy to US (see Table 4.5), for all the indicators of active involvement of the community of conspirators against the scientific one, whereas in the US the scientific community is more active, while in Italy the situation is even more complex.

What the indicator concerning the posts clearly show (which is equal to 2.9, so largely higher than 1; in view of 0,4 of the US), in Italy there is a production of scientific contents, in comparison to the analyzed sources, which is much higher with reference to disinformation than information. This confirms the evidences of the previous chapters.

However, if we consider other indicators, while, as we said, in the US the scientific community is much more active (in terms of likes and comments to produced posts), in Italy the level of involvement is more homogeneous for the two communities, whereas the conspirators are averagely more ready to comment (values higher than 1, for comments and commenting users) and the community following the scientific sites is more prepared to interact by means of likes (values lower than 1, for likes and users putting likes).

Table 4.5 - Involvement of the conspirator's community in comparison with the scientific one: Italy vs. US

Involvement by:	Italy	US
Posts	2,9	0,4
Likes	0,8	0,2
Comments	1,4	0,3
Users who put a "like"	0,8	0,3
Users who commented	1,3	0,3

Note: the involvement is calculated as ratio between the level of activity (post, like ...) per page (of the posts) and per posts (of likes, comments and users) of the community of conspirators, compared with the scientific one. Values higher than 1 (lower than 1) show a higher (lower) involvement of conspirators.

Two further *datasets* were created (starting from the group of pages according to Table 4.4, namely for the American samples) with reference to two different online platforms, Facebook and YouTube. In detail, we collected all *posts* published on Facebook by scientific information and disinformation sources containing a *link* to a YouTube video, as well as *likes*, comments and sharing of such posts. By means of the API of YouTube Data, for every YouTube video, linked to those posts on Facebook we extracted the date concerning the interaction of the users (like, comments and views). The composition of the two *datasets* is indicated in the following tables.

Table 4.6 shows that, in the case of Facebook, the reactions to the posts with links to YouTube videos is divided similarly for information and disinformation (also if in the US there will be a bigger involvement of the community following scientific sites; with the relevant value, if calculated, which is always under the unit value), with likes being the most frequent actions, followed by shares (which is the action which contributes at most to the viralization of contents).

Table 4.6 - Composition of the dataset concerning the posts on Facebook with links to YouTube videos (US)

	Science (Information)	Conspiracy (Disinformation)
Posts	4.388	16.689
Likes	925.000	1.000.000
Comments	86.000	127.000
Sharing	312.000	493.000

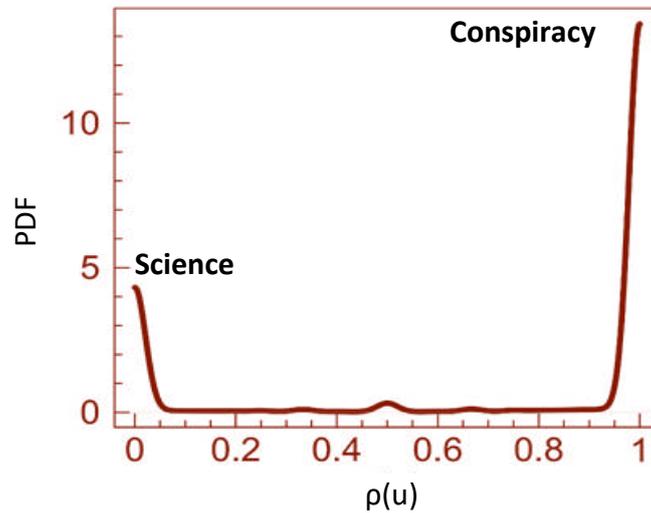
Carrying out a deeper qualification of the interactions made by users in relation to the detected narrations, it is possible to control whether the level of involvement (informative activity) for users for a specific kind of content indicates the level of polarization of the users sharing the same system of beliefs (namely the existence of *echo chambers*).

We suppose that a user is polarized against narration of “science” (information) or “conspiracy” (disinformation) if s/he put, respectively, more of 95% of his/her likes (or comments) on scientific or conspiracy posts.

Representing the function of density of possibilities (PDF) of polarization of users (according to their interactions) on Facebook in Italy, we observe a highly bimodal distribution, which has two peaks at the high of two values (see Figure 4.1). This result clearly indicates that the most part of users is polarized against the one or the other narration, which leads to the creation of two groups of well separated users who do not interact with each other.

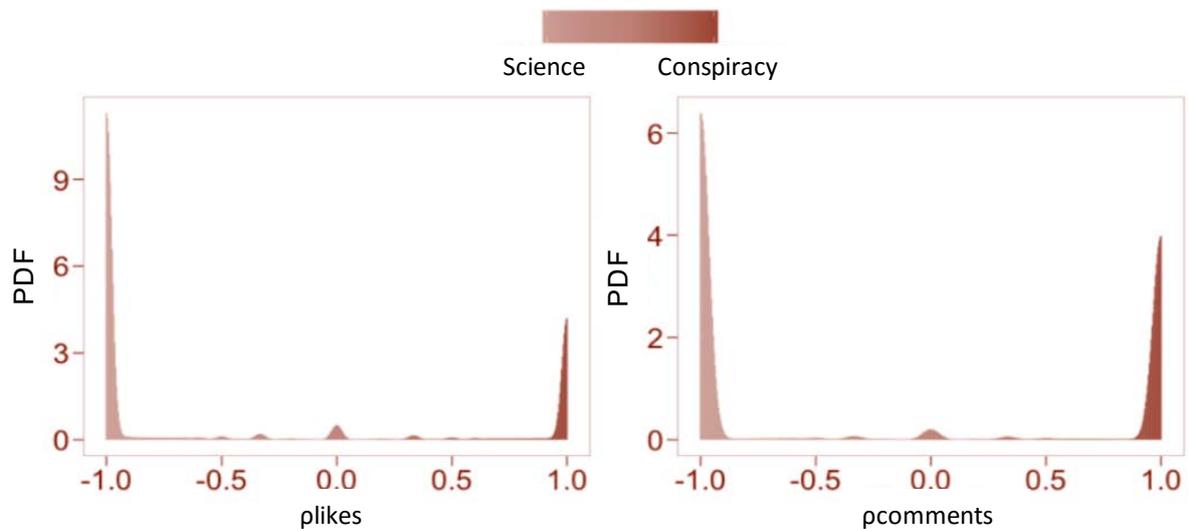
Moreover, the study shows that the level of activity of a user in his/her narration corresponds to a bigger number of friends on social networks sharing the same attitude. It follows that there is a high level of homophily characterizing social interactions of users on Facebook, for which individuals tend to associate and to enjoy people with the same preferences and similar beliefs. In fact, we notice that for polarized users the group of friends with the same polarization is very high and it grows according the increase of the information activity in the own community.

Figure 4.8 - Users polarization distribution on Facebook for “science” and “conspiracy” (Italy)



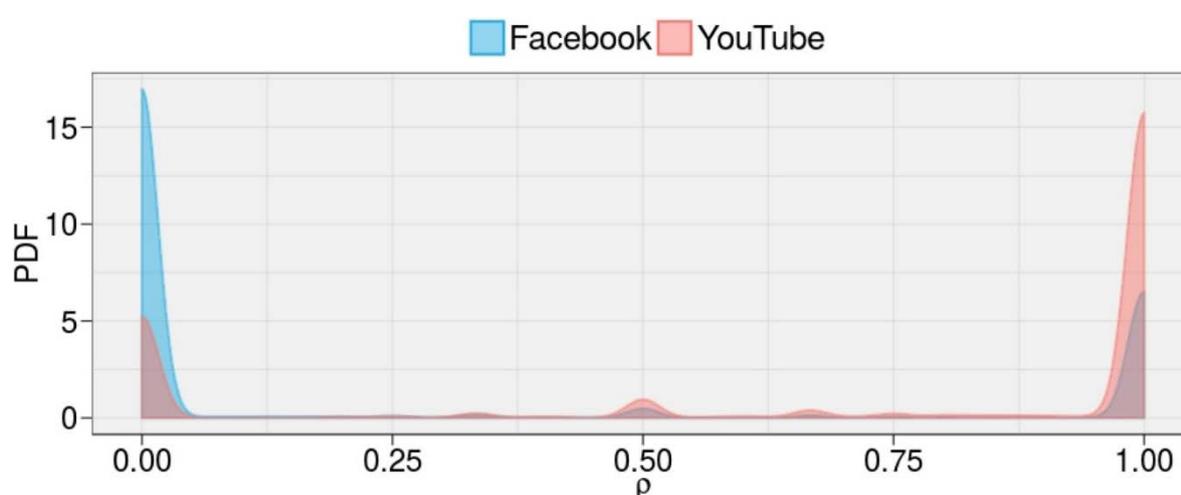
Similar models of consumption and interaction of users concerning information and disinformation contents are found also on the social platform in the US, with a distribution of the polarization which is highly bimodal at the level of the two different narrations, considering links as well as comments (see Figure 4.9).

Figure 4.9 - Users polarization distribution on Facebook for “science” and “conspiracy” (US)



Similarly, Figure 4.10 shows that both on Facebook and on YouTube the most part of users is polarized towards one of the two narrations. This shows that the phenomenon of polarization occurs quite often in the different kind of online platforms, whether they are social networks (Facebook) or more in general social media (YouTube).

Figure 4.10 - Users polarization distribution on Facebook and YouTube for “science” and “conspiracy” (US)



Briefly, the previous analysis stressed the emergency of models of news consumption and interaction of the users with real and fake news, having the same features (trend to polarization, homophily, onset of *echo chambers*), regardless of kind of narration (information and disinformation) and on the online platform (Facebook and YouTube). The contents concerning different narration aggregate users in different and polarized communities, where they interact with people sharing their own system of beliefs.

In Italy, in particular, the production of unscientific and conspiracy nature is very flourishing, while deepened tradition of informative nature on the same topics is lacking. The lack of an extended discussion (due to phenomena of polarization) and of an informative deepening create the conditions for a less receptive public opinion (for example, compared to the US) and more inclined to share news, which are without any doubt fake.

At this point, it is crucial to examine how cognitive *bias*, in particular the so-called *confirmation bias*, act in viral processes of diffusion of news. To this purpose, it is

necessary to detect from a statistic point of view the characteristics of cascades of users' interaction according to the typology of narration ("science" and "conspiracy", namely information and disinformation).

The first feature of a cascade is the extent of the time interval (in hours) passed between the first and last sharing of a post. The results show that, for both typologies, there are two peaks: one at a duration of about 1-2 hours and one at a duration of about 20 hours, which indicated that the time-sharing models are similar for information and disinformation.

However, if we consider the duration of the interaction according to the dimension of the cascade (which is the number of users sharing the post), we find a difference between the kinds of news. In fact, for scientific news, the variability of the duration of interaction increases according to the dimension of the cascade and bigger dimensions of the cascade correspond to a bigger variability of duration, namely to different durations according to the scientific news. For contents concerning conspiracy theories, the duration of interactions on a post shows a monotone increase compared to the dimension of the cascade. These results suggest that the scientific information is often assimilated (which means it reaches a higher level of diffusion) rapidly and evolves in variable way. On the contrary, the conspiracy news is assimilated more slowly and show a positive correlation between duration and dimension in all contents; so, the posts which remain longer are discussed by bigger communities. In other words, while for scientific news there is a bigger level of assimilation and passing to newer news, for conspiracy theories, some contents tend to remain in time and to be shared by always bigger communities.

Moreover, it is possible to observe that the most part of links between users sharing consecutively a post is homogeneous, which means both users share the same polarization and so belong to the same *echo chamber*. In particular, the detected evidences suggest that the diffusion of information occurs mostly inside homogeneous groups where all users share the same polarization. So, the contents tend to circulate and to be confined inside the respective *echo chamber*.

Summarizing, even if the consumption models compared to the narrations of information and disinformation are similar, the cascade dynamics of interactions differ. The selective exposition appears to be as main engine of diffusion of contents and generates the formation of polarized groups, each of them with its own cascade of interaction.

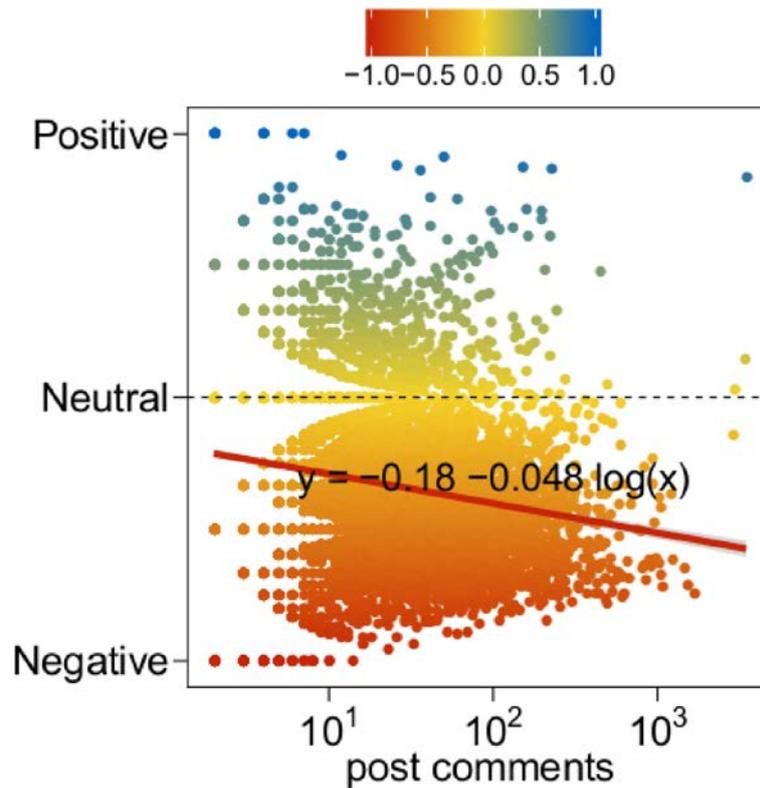
Furtherly extending the study, we can focus on the emotional dynamics which develop in different *echo chambers*. To this regard, we selected all posts representing the field where the discussion among users about “science” and “conspiracy” occurs. In particular, we collected all posts commented at least once both by a user polarized against “science” and by a user polarized against “conspiracy”. Precisely, we found 7.751 posts of this kind (out of 315.567), confirming the fact that the two communities are strictly separated and often do not interact each other.

Examining how the emotional state change when the number of comments of the posts increases, namely when the discussion becomes longer, Figure 4.11 shows the emotional state (at aggregated level) of these posts according to the increase of the number of comments. Of course, according to the increase of the number of comments and the continuing of the discussion, the emotional state becomes always more negative. So, the duration of the discussion influences the negativity of the feeling of the users, involved in the discussion.

Moreover, the exposition of the users to not-motivated statements (as the conspiring ones) affects their criteria of selections of the contents and increases their attitude of interacting with fake news. So, the *confirmation bias* has a very crucial role in the selection of the contents. Moreover, the results seem to be coherent with the literature, which indicates a relation between beliefs in conspiracy theories and the necessity of a cognitive closing, in order to find definite conclusions (also if irrational)⁴³. More in general, people preferring a heuristic approach to evaluate the evidences and create own opinions have higher possibilities of finding a coherent explanation with the existing system of beliefs.

⁴³ See among the others, P.J. Leman, M. Cinnirella (2013), “Beliefs in Conspiracy Theories and the Need for Cognitive Closure”, *Frontiers in psychology*, 4.

Figure 4.11 - Emotional state according to the number of comments to posts discussed by both narrations



At last, the models of information interactions of users can be studied with reference to dissenting news compared with the own beliefs.

To this regard, the pages of debunking on Facebook try to fight the diffusion of disinformation, supplying verified information on specific facts and topics. However, if the *confirmation bias* has a key role in the selection criteria, it is possible that the debunking contents appear to be only for users exposed to disinformation (“conspiracy”), like a new dissenting from their favorite narration.

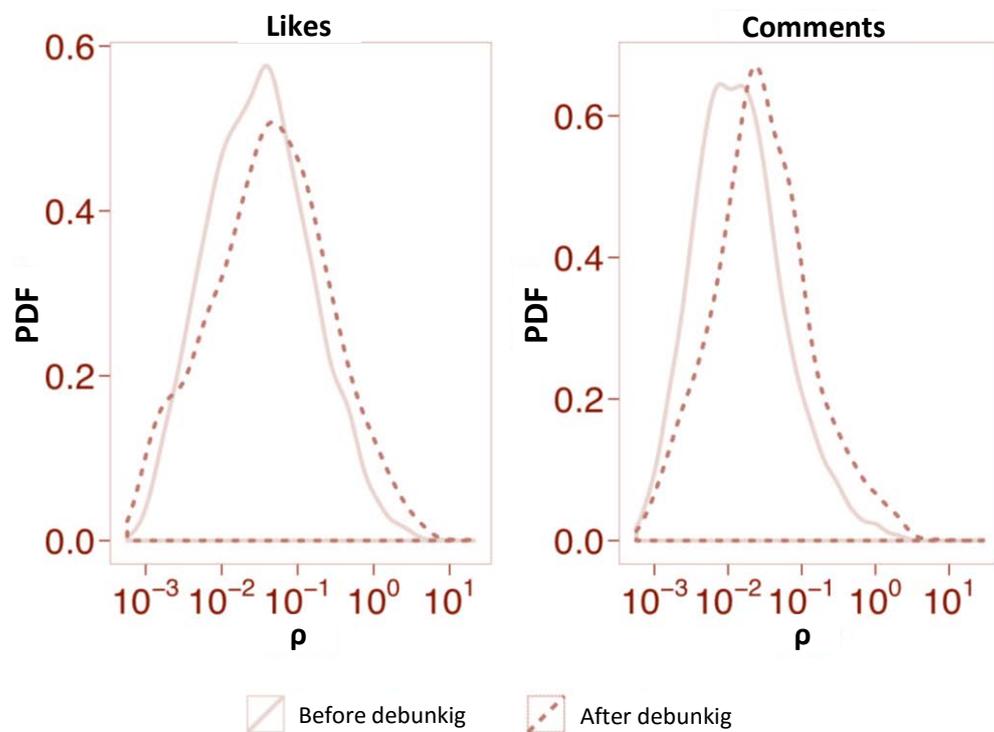
At first, Figure 4.12 shows how the *debunking* news is appreciated and commented according to the polarization of users. We observe that the most part of both actions (likes and comment) is carried out by polarized users in the category “science” (registering 67% of overall likes in the debunking news and 52% of all comments), while only a minority of interactions is composed by polarized users in the narration “conspiracy” (7% of likes and 4% of comments).

Figure 4.12 - Users' interactions with debunking news



Secondly, it is possible to make a last comparative analysis between the behavior of users belonging to the “conspiracy” community before and after their first interaction with a debunking news. Figure 4.13 shows the popularity and the rate of comment, which is the average amount of likes and comments on a conspiracy news on the day, before and after the first interactions with debunking posts. We could expect that the users recognize the correction and reduce their engagement in the conspiracy group. On the contrary, we stress that their likes and comments on the “conspiracy” posts do not change substantially (likes) or even increase (comments) after having interacted with a corrective news.

Figure 4.13 - Interactions of conspirators before and after debunking



In conclusion, the users tend to select information which is coherent with the own system of beliefs, forming polarized groups of people with similar ideas on shared narrations (so-called community), where different information are ignored, often making useless attempts of debunking.

This last result shows that the process leading mechanisms of acquisition and elaboration of information is due to cognitive phenomena but also by cultural and social practices. In other words, “*facts remain robust only when they are supported by a common culture, by institutions that can be trusted, by a more or less decent public life, by more or less reliable media*”⁴⁴.

⁴⁴ This is what the French philosopher thinks about the post-truth Bruno Latour, see A. Kofman, “Bruno Latour, the Post-Truth Philosopher, Mounts a Defense of Science”, *New York Times Magazine*, 25th October 2018.



METHODOLOGICAL APPENDIX

As stressed in the Introduction, the Report has adopted an approach for the purpose of the integrated analysis of data coming from different sources and datasets. In particular, we used qualitative and quantitative information, both on the request and on the offer of information.

For the first aspect, we made mostly reference to data coming from a specific survey on over 14.000 Italian consumers (carried out for AGCOM by GfK Italia), which were also used in the recent [Report on information consumption](#) in Italy. Reference is made to the relevant [Methodological appendix](#) for the description of the following approach (amount of samples, methods of stratification and reweighting, etc.).

Always with reference to the request of information, the analysis on the consumption of (real and fake) news by means of online platforms were done by means of big data analytics models, carried out by Walter Quattrociocchi (entrusted by AGCOM), another co-authors, on tens of millions of social accounts (generally Facebook and Twitter). These *datasets* are briefly described in Chapter 4.

With regard to the “news supply”, in this Report we used data on journalists and news production documents. On the journalists, the information was taken from the [2nd edition of the Observatory on journalism](#) and regard a sample of about 2.000 professionals in Italy. Also in this case, we refer to the Annex of the document, containing a detailed illustration of the methodological aspects of the research and of the mentioned sample.

At last, with reference to the analysis of the information products offered in Italy, the documental database used for the elaboration of this Report was created starting from the data taken by means of the platform (developed by *Volocom Technology*), with which AGCOM was equipped to carry out the studies on the diffusion and the content of the news published by every media.

Because of the absolute innovation of this dataset and the relative approach, in the following of this Appendix we illustrate the main characteristics of the platform and of the connected documental database.

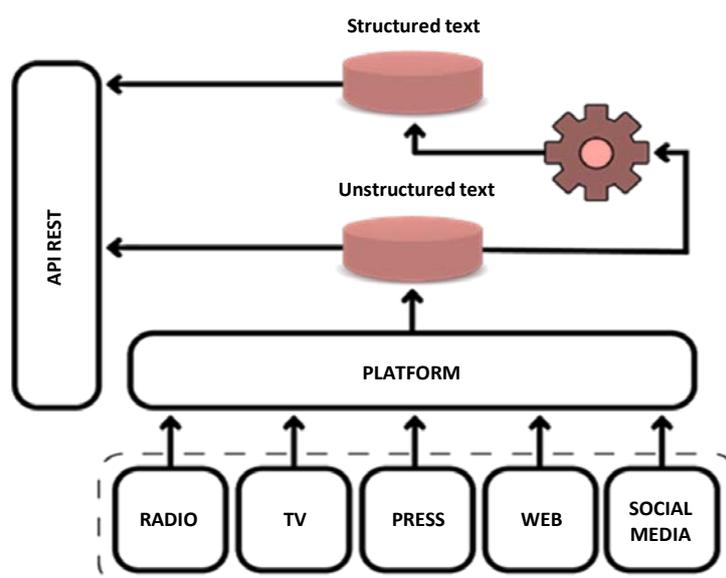
THE PLATFORM

The used Platform gives the possibility of extracting the recurrence of every key word (or group of words) and of examining the text of whole documents referable to news on every topic or happening.

Moreover, the platform, considering constantly the updating of the information fluxes of the sources, is able of issuing historical series of data and allows the carrying out of dynamic analysis.

Considering the functioning of the platform (see Figure A.0.1), the output supplied by itself consists in the not-structured text coming from the source (represented by single media) enriched by the metadata the text, such as source, data, category, etc. By means of the platform it is possible to have access to the historical record of the sources, in order to analyze them in time or along different axis corresponding to metadata. The actual system of metadata processing foresees the application of suitable algorithms to the not-structured text, in order to extract metadata describing the text.

Figure A.1 - Functioning of the platform

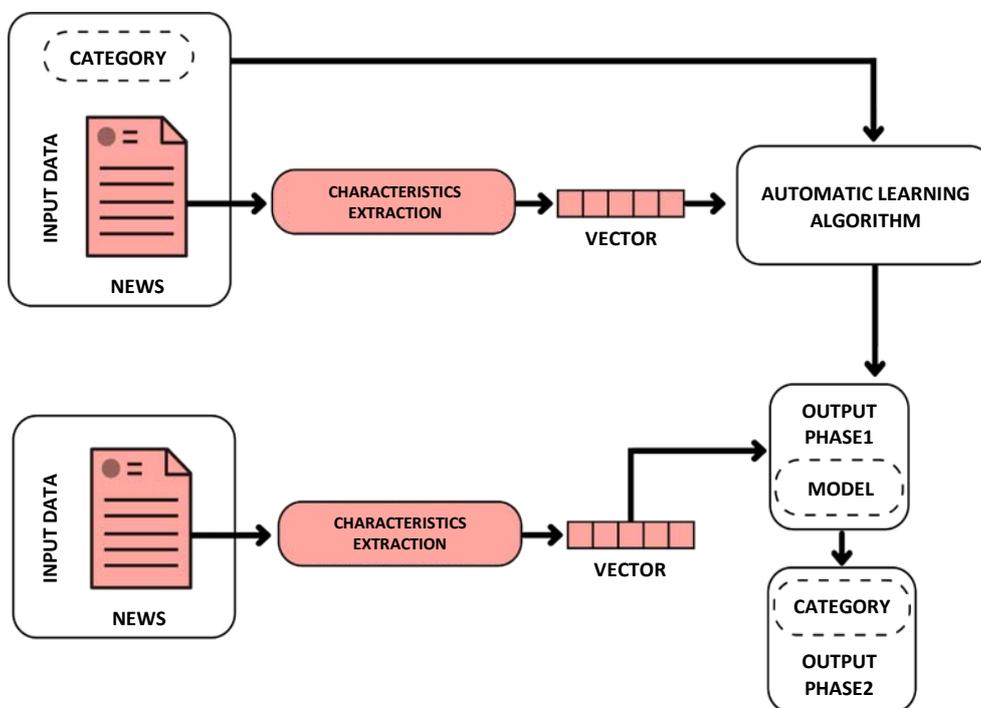


Under the technical-informatic profile, the services of the platform are made available by a programmatic interface based on API REST (*REpresentational State Transfer*). This allows the extraction of data on the platform with the maximum granularity possible, overcoming the limits imposed by a predefined user interface and allowing to make analysis with the data with the maximum flexibility. For the interchange of data, the platform uses the format JSON (*JavaScript Object Notation*).

The functioning of the platform uses also two relevant algorithms. The first one is a stemming algorithm, allowing the extension of the research to different versions of a chosen word. The second one is an artificial intelligence algorithm, by means of which the automatic classification of every news occurs with regard to a category of predefined algorithms (such as report, politics, international affairs, economics, international affairs, science and technology, cultures, show and sport).

It is an automatic learning algorithm of supervised typology. Its operational scheme, as shown in Figure A.0.2, consists in a preliminary phase where the algorithm is trained starting from input data (*training dataset*), and in a following operation phase with which, after the training, the algorithm is able to carry out a prediction of the belonging category for every new received document. The specific automatic learning algorithm, the platform uses, is a support vector machine (or SVM). Since the SVM algorithm is a binary linear classifier, the platforms carry out the algorithm with an iterative procedure in more steps, in order to obtain a classification of the information content in one of the possible categories.

Figure A.2 - Automatic classifier of the news content in categories



THE COMPOSITION OF THE DOCUMENTAL DATABASE

In creating the documental database subject to the elaborations of this Report, particular attention was given to the preliminary definition of the information sources, representative of every media. Particular attention was given to this aspect, because the robustness and the significance of the results obtained in the carried-out analysis depend on the completeness and representativeness of the defined group.

With the purpose of proceeding with the correct definition of the group of sources, we considered the entire information system, including the traditional media (TV, radio and newspapers) and online ones (websites of traditional media, exclusively online headings, online platforms).

For every of the mentioned media, we selected the broadest number of information sources, according to the availability of data in formats, suitable to be acquired and treated with the above-mentioned IT-techniques.

From the practical point of view, the selection of the single information sources occurred considering some specific criteria, referable to the relevance assumed by each of them in terms of:

- reached *audience*;
- diffusion on the territory;
- actual presence of information content inside the proposed offer; this way, for example, in the case of online platforms, we selected Facebook pages and Twitter account, referable to Tv, radio, newspapers, information sites and influencers.

Moreover, in the light of the actual conformation of the information system and the development of pathological forms of online disinformation, it was considered necessary, in order to study empirically such phenomena, to include in the media feeding the database also the websites and pages/*social* network accounts, included in the lists of the sources of “fake news”, arranged by external sources specialized in debunking activities (such as *butac.it* and *bufale.net*), used by several scientific studies on online disinformation.

Therefore, considering the structure and the criticalities of the Italian information system, of the values (in terms of access by citizens with the purpose of detecting information) assumed by these media, as well as the peculiarities of the used detection

platform, the group of information sources of the analysis carried out in this Report is composed as indicated in the next table.

Generally, the information sources feeding the documental database are about 1.800.

In detail, the database is composed by the entire textual content of all documents produced during a day by every information source. With document, we mean the entire article, in the newspapers and information websites; the transcription of a transmission segment (service), in the case of tv and radio: all tweets/posts in the case of online platforms.

Table A.1 - List of information sources feeding the database

MEDIA	INFORMATION SOURCES
Television	Ex. Rai 1, Rai 2, Rai 3, Rai News 24, Rete 4, Canale 5, Italia 1, TGCOM 24, La 7, SKYTG24, ...
Radio	Ex. Radio 1, Radio 2, Radio 3, Radio 24, ...
Newspapers	- National press (ex. La Repubblica, Corriere della Sera, Il Sole 24 Ore, La Stampa, ...) - Local press (ex. L'Eco di Bergamo, Gazzetta di Mantova, ...)
Sites of television channels/programs	Ex. rainews.it, tgcom24.mediaset.it,
Sites of radio channels/programs	Ex. radio24.ilsole24ore.com, rtl.it/notizie, ...
Sites of newspapers	Ex. repubblica.it, corriere.it, ecodibergamo.it, ...
Exclusively online headings	Ex. Il Post, Huffington Post, ...
Site of fake news	Sites included in the lists created by external sources specialized in debunking
Online platforms (social networks)	Facebook Pages and Twitter accounts of Tv, radio, newspapers, online headings, influencers and sources of fake news

The data composing the database refer to a very long period covering 29 months from April 2016 until August 2018. Therefore, the database is quite big, which gives big robustness to the obtained results. The analysis carried out in this Report base in fact on about 35 million of documents.