# Report on online disinformation

SPECIAL ISSUE ON CORONAVIRUS

#### Table of contents

#### 1. What information and disinformation offer about the coronavirus

1.1 Information: media coverage of the coronavirus in Italy	
1.2 Disinformation: space dedicated to the coronavirus	<b></b>
1.3 Information vs. Disinformation about the coronavirus in social networks	<b></b>
1.4 Weight of disinformation on the coronavirus online news	<b></b>
1.5 Information vs. Disinformation: narratives about the coronavirus - First period	<b>&gt;</b>
1.6 Information vs. Disinformation: narratives about the coronavirus - Second period	<b></b>
1.7 Main fake news on the coronavirus spread around the world	<b></b>

#### 2. What Italians watch online about the coronavirus

- 2.1 Online searches: how much information Italians look for about the coronavirus
- 2.2 Interactions of Italians on social media: how they react to coronavirus contents
- 2.3 Social engagement: how much and what Italians comment on the coronavirus
- 2.4 Online video: the most viewed contents about the coronavirus

#### 3. How internet consumption changes during the epidemic in Europe

- 3.1 Online news: the consumption in Europe during the epidemic
- 3.2 Social networks: the consumption in Europe during the epidemic
- 3.3 Instant messaging: the consumption in Europe during the epidemic

#### 4. Cybersecurity threats and coronavirus

- 4.1 Cyber attacks: most frequent types and techniques
- 4.2 Cyber attacks related to the coronavirus worldwide
- 4.3 Cyber attacks related to the coronavirus in Italy

#### Methodology

In this Report we refer to the term «coronavirus» in a broad sense to describe all topics related to the current epidemic emergency. More precisely, the scientific name of "new coronavirus" (namely, the virus belonging to the coronaviruses «family» that had never been found in humans before) is "SARS-CoV-2", while the term "Covid-19" identifies the infectious disease caused by the virus

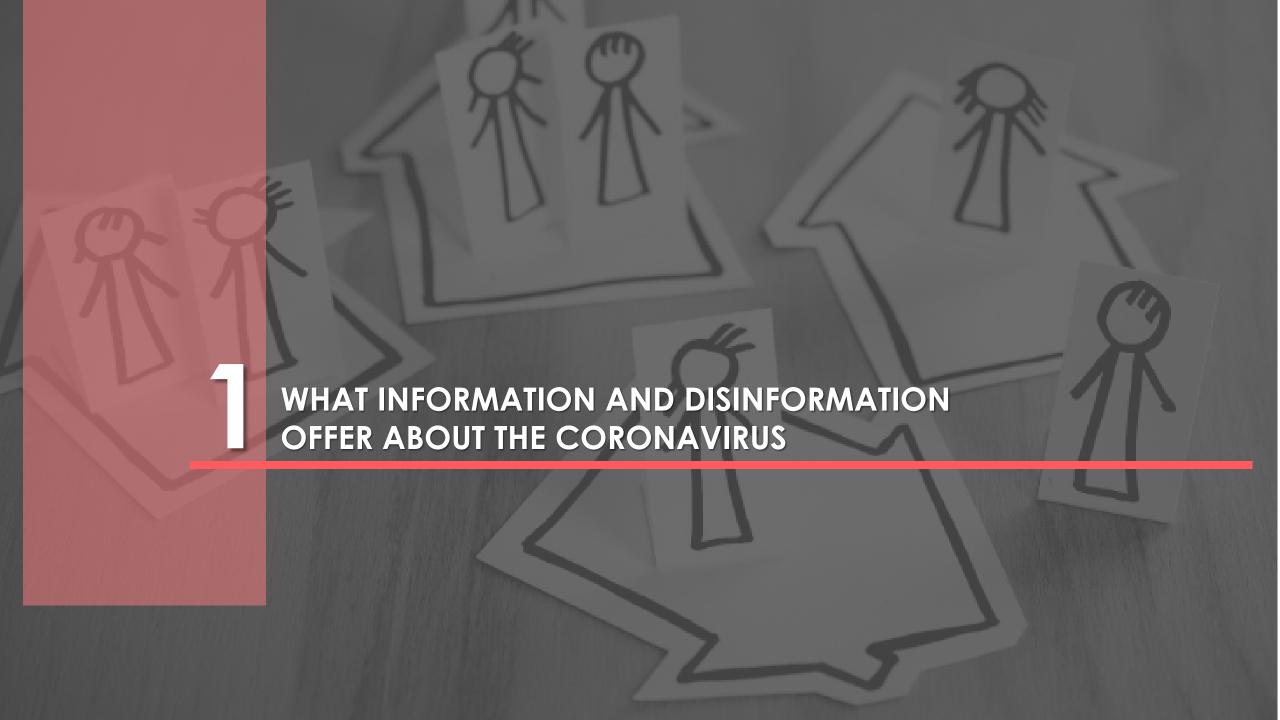


This issue of the Report on Online
Disinformation is dedicated to the
analysis of information and
disinformation production on the Covid19 theme, consumption of coronavirus
news, as well as cyber security threats.

The figures shown are updated to April 20, 2020 in order to outline the information scenarios that characterize the Italian system two months after the beginning of the medical-health emergency in the country.

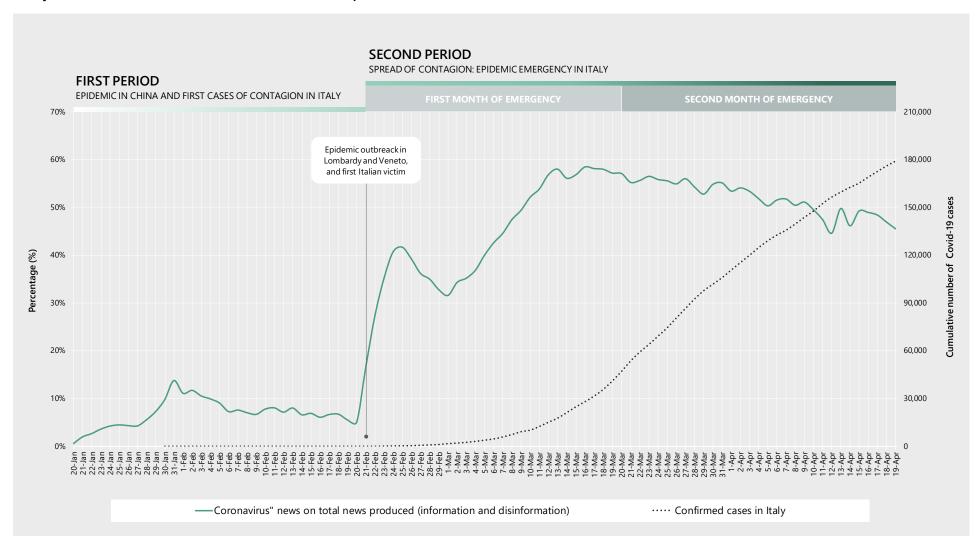
The special issue on the coronavirus of the Report on Online Disinformation is part of the activities of the Roundtable on Digital Platforms and Big Data - Covid-19 Emergency set up by AGCOM to contribute, among other things, to the fight against online disinformation on issues related to the epidemic.

www.agcom.it





#### Daily incidence of coronavirus news on total news produced



After reaching the highest level by the end of the first month of the epidemic outbreak, the media coverage of the coronavirus news in Italy has settled on decreasing average daily values

# CORONAVIRUS NEWS ON THE AVERAGE DAY

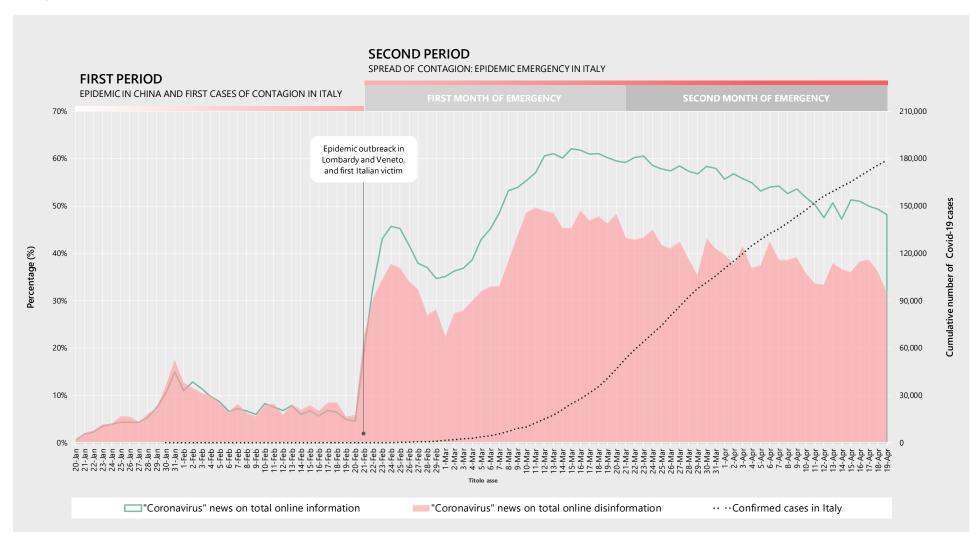
57% MARCH, of the total 16-22

**★48%** APRIL, of the total 13-19

The quantity of daily news on the coronavirus is still high. Overall, there are over 1 Million contents published by the regarding the epidemic during the second period, and more the 600 thousands in the last month



#### Daily incidence of coronavirus news on total disinformation: comparison with online information



The attention attributed by the sources of disinformation (websites, pages and social accounts) to the coronavirus remains high, even if on the average day it is lower than those observed between 10 and 20 March

# CORONAVIRUS NEWS ON THE AVERAGE DAY

46% MARCH, of the total 16-22

disinformation

**¥** 37%

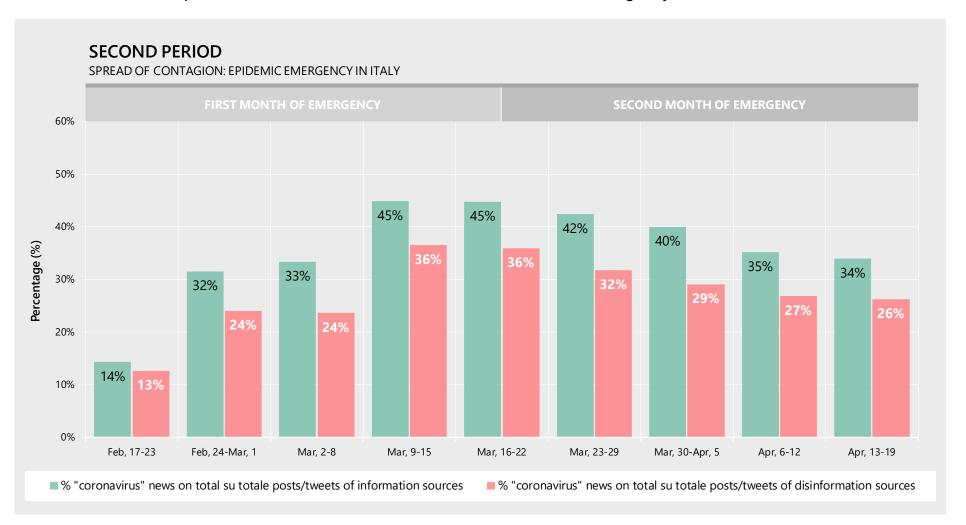
**37%** APRIL, e total 13-19

of the total disinformation

Comparing the daily trend of the incidence of coronavirus-related news on the total, there are confirmed higher values for online information than for disinformation. In the last week considered, in fact, the space dedicated to the epidemic by the information component is on average 50%



#### Incidence of coronavirus posts/tweets on total: information and disinformation (% on average day of the week)



Note: the figures indicated refer to the posts and tweets published by the pages and social accounts (Facebook and Twitter) of the sources of information and sources of disinformation identified as such by external debunking specialists. Therefore, they do not include users' posts/tweets

# POSTS/TWEETS ABOUT THE CORONAVIRUS ON THE AVERAGE DAY

of the total social information sources

34%

APRIL,

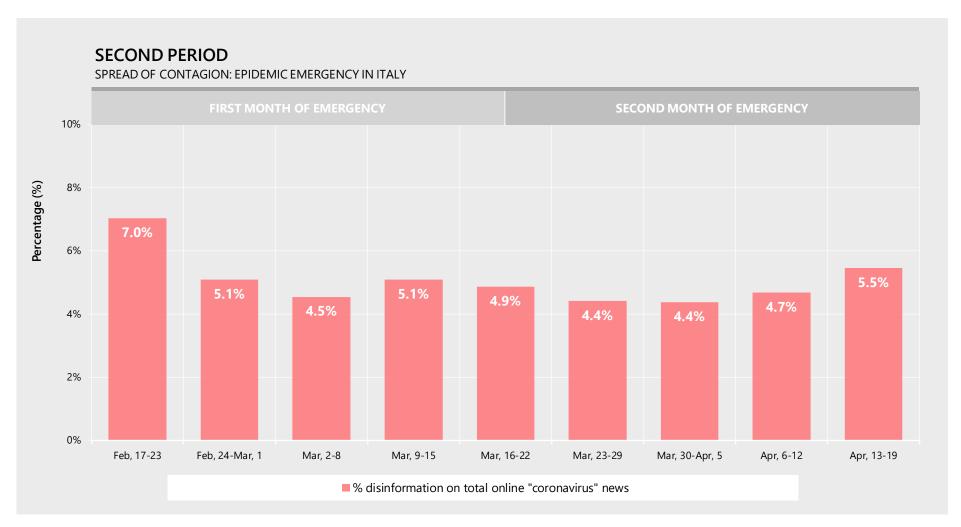
13-19

of the total social disinformation sources 13-19

In the second month of the epidemic in Italy, although still considerable, the percentage incidence of coronavirus posts and tweets on the total of those published shows a decrease. The decreasing trend is found both for social sources of information (-11 p.p. in the last week compared to the week from March 16 to 22) and for those of disinformation (-10 p.p. in the last 5 weeks)



#### Fake news on total online coronavirus news (% on average day of the week)



Note: the figures indicated for the first 5 weeks may differ slightly from those reported in the previous issue of the Report, in view of the inclusion in the analysis of further sources of disinformation identified by external debunking specialists, which made it possible to make adjustments also with reference to previous periods

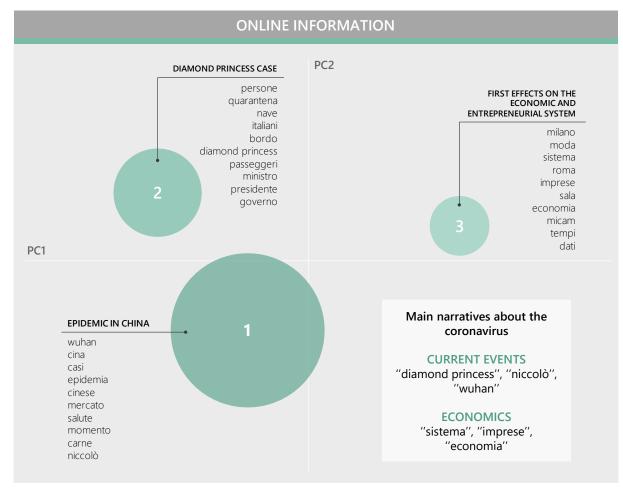
In the last two weeks under examination, the incidence of disinformation on the total news concerning the coronavirus has increased again, to over 5%

# INCIDENCE OF DISINFORMATION ON THE AVERAGE DAY

of the total "coronavirus" online news

Except for the week in which the epidemic began in Italy, where the burden of misinformation on coronavirus-related news reached 7%, that of the week between 13 and 19 April is the highest value detected in the second period





# MAIN TOPICS OF THE CORONAVIRUS NEWS

# FIRST PERIOD

UNTIL FEBRUARY, 20

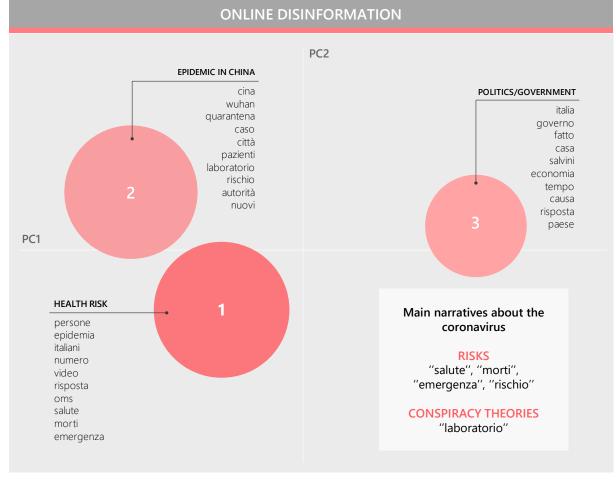
#### **INFORMATION**

Epidemic outbreak in China

Diamond Princess cruise ship case

First effects on the Italian economic
and entrepreneurial system

Main narratives: current events and economics



#### **DISINFORMATION**

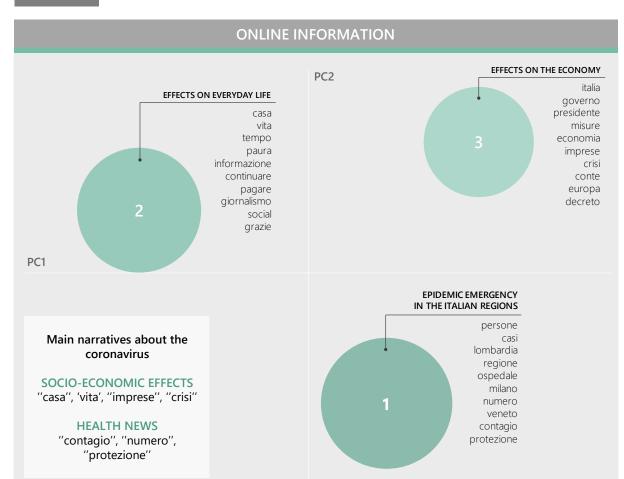
Health risk related to the epidemic
Epidemic outbreak in China
Political issues and activities of the Italian
Government related to the epidemic
Main narratives risks and conspiracy theories

The main macro-themes are identified through the **textual analysis** of coronavirus news of information and disinformation sites (*topic modeling*), which allows to obtain groups of **terms** frequently cooccurring within articles

5

# Information vs. Disinformation: narratives about the coronavirus - Second period







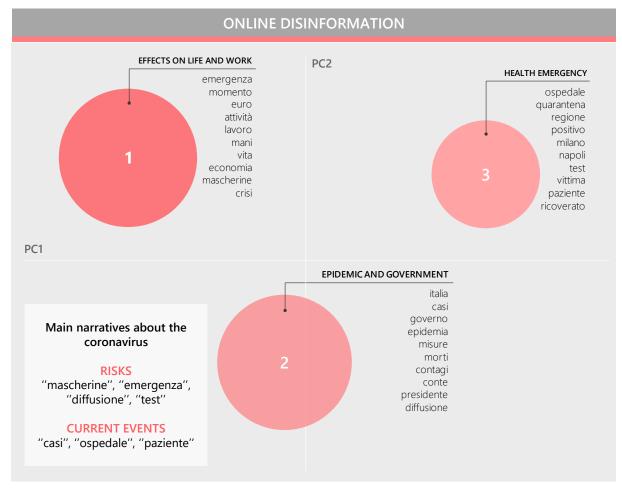
SECOND PERIOD

SINCE FEBRUARY, 21

#### **INFORMATION**

Epidemic emergency in the Italian regions due to the spread of contagion Effects of the epidemic on everyday life Effects of the epidemic on the Italian economy

Main narratives: socio-economic effects and health news



#### DISINFORMATION

Effects of the epidemic on everyday life and work

Epidemic in Italy and government actions Health emergency situation

Main narratives: risks and current events

The <u>textual analysis</u> of the coronavirus news highlights how, in the second period, information and disinformation focus on the Italian medicalhealth emergency situation and the impact on citizens' living conditions. In terms of **disinformation**, news communication appears to be based on the recurrent use of terms capable of stimulating negative emotions

Department of Economics and Statistics

7



# THE COVID-19 VIRUS WAS STOLEN OUT OF A CANADIAN LAB BY CHINESE SPIES



#### **FALSE**

There is no evidence that the COVID-19 virus was stolen by Chinese spies from a Canadian lab

#### THE COVID-19 VIRUS CONTAINS HIV-LIKE SEQUENCES



#### **MISLEADING**

The similarity with HIV was deduced from a study, later withdrawn, based on extremely short sequences, which are present in many other organisms

# THE COVID-19 PANDEMIC WAS PREDICTED IN A SIMULATION



#### FALSE

The Event 201 pandemic preparedness exercise involved a fictional coronavirus, with different characteristics than the COVID-19 virus

# A GROUP FUNDED BY BILL GATES PATENTED THE COVID-19 VIRUS



#### **FALSE**

The patent in question covers a different strain of coronavirus that affects only chickens, not humans

# THE COVID-19 VIRUS IS A MANMADE BIOWEAPON



#### **FALSE**

Scientific studies show that the new coronavirus is not a laboratory construct or a purposefully manipulated virus

# 5G CELL PHONE TECHNOLOGY IS LINKED TO THE CORONAVIRUS OUTBREAK



#### **FALSE**

There is no evidence that the health effects of the COVID-19 virus are related to 5G cell phone technology

#### COLLOIDAL SILVER CAN CURE COVID-19



#### **FALSE**

There is no evidence of the effectiveness of dietary supplements in preventing or curing Covid-19, and colloidal silver can have serious side effects

#### MIRACLE MINERAL SOLUTION CAN CURE COVID-19



#### FALSE

There is no reliable evidence supporting its use for COVID-19 or any other disease, and ingesting it can cause serious side effects

#### GARLIC CAN CURE COVID-19



#### **FALSE**

According to the WHO, there is no evidence that garlic consumption has protected people from the new coronavirus strain

# THE EFFECTIVENESS OF DOSES OF VITAMIN C FOR COVID-19 HAS BEEN PROVEN



#### **FALSE**

It is true that a clinical study is underway in China, but the claim that it is a proven treatment is not supported by scientific evidence

Source: Newsguard, analysis of J. Gregory e K. McDonald (Coronavirus Misinformation Tracking Center)



For more information, go to the Fact-checking section of the Roundtable on Digital Platforms and Big Data - Covid-19 Emergency

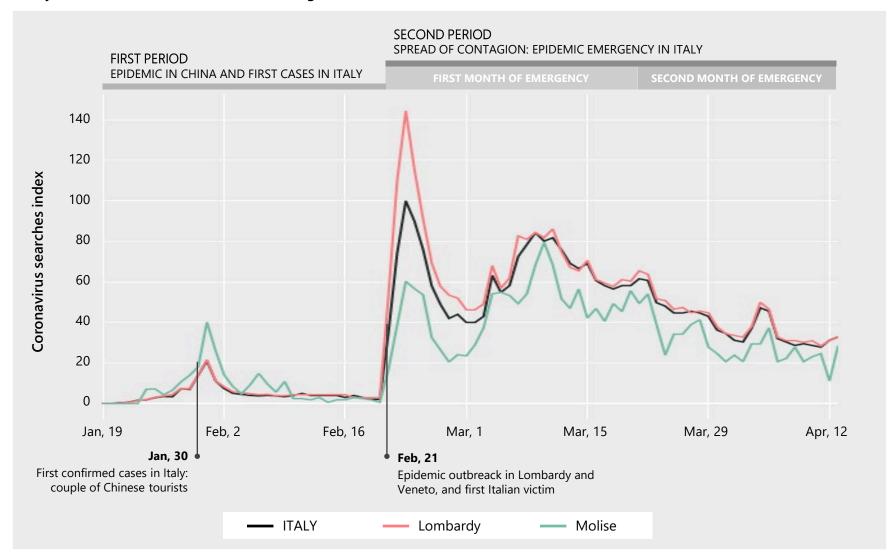
Department of Economics and Statistics



### Online searches: how much information Italians look for about the coronavirus

#### AUTORITÀ PER I GARANZIE NELI AGCOM COMUNICAZIO

#### Daily index of coronavirus searches on Google



Note: for Italy, the index is the value of searches normalized on a scale from 0 to 100; for regional indices, the scale of values is based on the relative importance with respect to the national figure

Source: Agcom Data Science Task Force elaborations on Google Trends data

The searches carried out by users are a clear **indication of the information needs** of the Italian population on the Covid-19 epidemic. In this sense, the **searches index** expresses how much on average the user of a given area is looking for information on the topic

#### **SEARCHES TREND IN ITALY**

The interest of Italian users in the topic of coronavirus begins to manifest itself at the end of January. The peak is recorded in the days (21-23 February) that mark the beginning of the epidemiological emergency in the country. After 9 March (national lockdown), searches for the topic gradually decreased, with a trend, in the second month of the emergency phase, similar to that of the coronavirus information supply

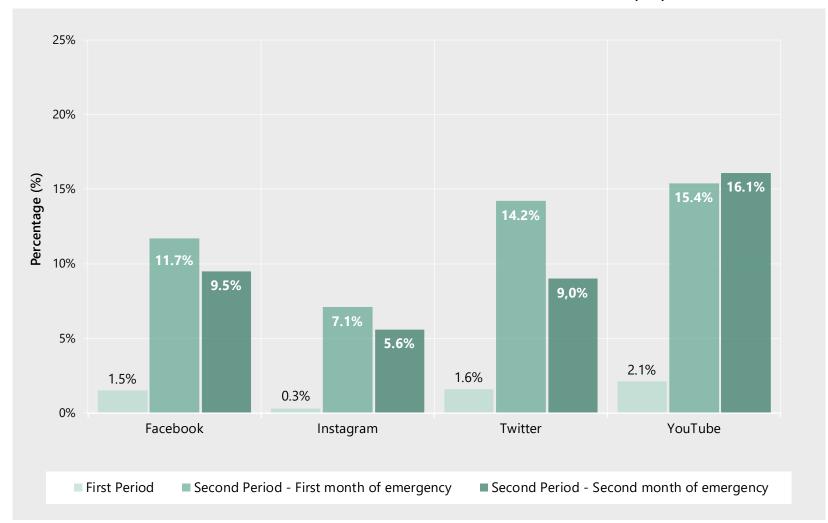
#### SEARCHES TREND IN THE ITALIAN REGIONS

For the regions of Lombardy and Molise (respectively the regions with the highest and lowest average searches index), the trend follows the national one. Comparing the two periods, it can be observed that, before February 21, interest in coronavirus information in Molise is on average higher than in Lombardy (and the Italian average). On the contrary, in the second period, with the spread of the contagion especially in Lombardy, the concern of the citizens of the region affects the need to find information on the epidemic and research updates on the emergency situation

# Interactions of Italians on social media: how they react to coronavirus contents



#### Social contents dedicated to the coronavirus: the share of actions/views on total interactions per platform



Note: the percentage of actions (for Facebook, Instagram and Twitter) and views (for YouTube) indicates the share of actions (comments + reactions) or views reached by the main contents dedicated to the coronavirus (max. 5,000 for each platform) compared to the total contents published on each platform analyzed in the 3 months under review

Source: elaborations on Sensemakers-Shareablee data

#### **ACTIONS/VIEWS ON THE CORONAVIRUS**

First Period vs. Second Period (first month of emergency)

Although YouTube reaches the highest percentage share of views achieved by the coronavirus contents in both months, the highest percentage (+1,832%) and absolute (+46 million) increase in coronavirus actions is observed for Instagram

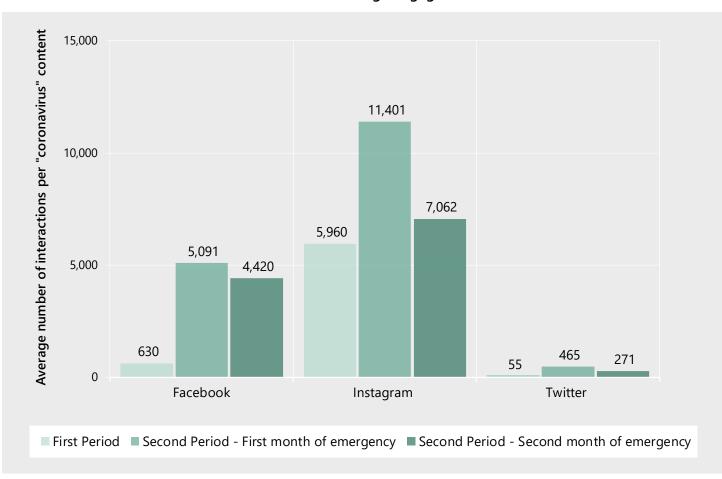
Second Period: first vs. second month of emergency

Although the second month of the second period coincides with the so-called total lockdown, the percentage share of actions decreases on all social networks (only the percentage of views on YouTube increases). The biggest decrease of actions on the coronavirus (-42%) is observed for Twitter, the absolute decrease (-13.7 million) for Instagram

In the second period, in any case, at least 1 action out of 10 relates to coronavirus contents. Given its greater use in fields far from information and news, Instagram is the only social network in which interactions related to coronavirus contents are around 6%

#### AUTORITÀ PER L GARANZIE NELL AGCOM COMUNICAZION

#### Social contents dedicated to the coronavirus: the average engagement



Note: the average engagement of contents on the coronavirus is calculated by dividing the total number of actions (comments + reactions) related to content contents (posts, images, videos) on the coronavirus by the total number of contents dedicated to the theme (max. 5,000 for each social network). It was not possible to calculate the average engagement of the coronavirus content published on YouTube

Source: elaborations on Sensemakers-Shareablee data

#### **TOP 20 CORONAVIRUS CONTENTS**





In the first period the contents with the highest engagement are those published by editorial organizations and satirical pages. By contrast, in the second period the contents published by influencers and entertainment pages (even if not strictly related to health or news topics) stand out. Institutions and politicians find the interest of users in the second month

#### INSTAGRAM



Also on Instagram, there is a clear difference between the first and second period, with a shift from politicians and journalists to influencers, advocates of charity projects, and entertainment pages, the latter increasingly at the center of users' attention in the second month of the emergency

#### **TWITTER**



Twitter is the only social network where the statuses (texts) find space among the most engaged content. As noted by numerous scientific and market analyses, Twitter is the place of choice for **journalists** and **politicians** interested in reaching their followers

#### YOUTUBE



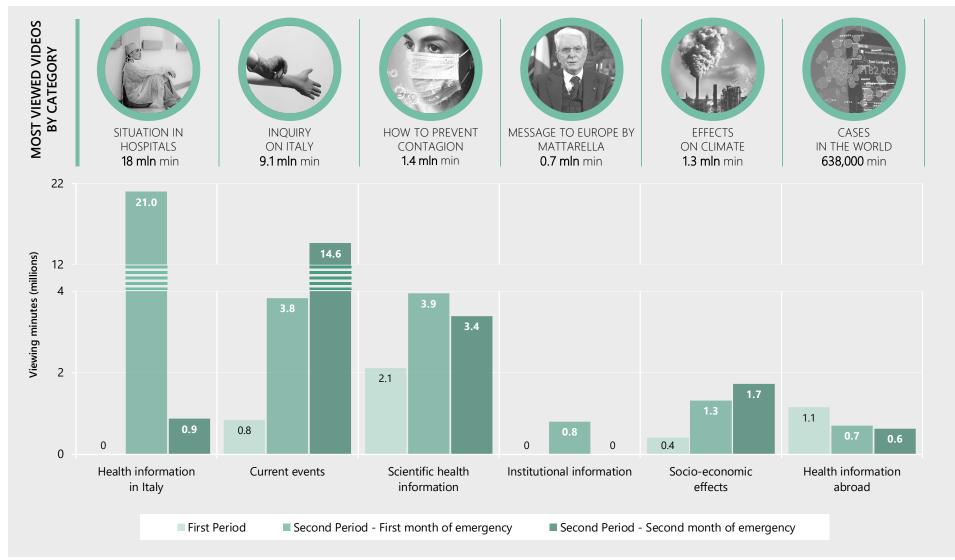
In the **first period**, views related to coronavirus content focus mainly on **institutional subjects** and **publishers**. In the **second period**, in addition to **publishers**, the interest of users (particularly in the second month of emergency) is also felt by **influencers** and **companies producing consumer goods** 

Department of Economics and Statistics

#### Online video: the most viewed contents about the coronavirus

#### AUTORITÀ PER LE GARANZIE NELLE AGCOM COMUNICAZION

#### Most popular online videos by total viewing time (top 15 videos for each period)



Note: the values refer to the total number of minutes of viewing (Total Time Spent) of the single video content on demand or of the video category through the online platforms of the publishers RAI, Mediaset, Sky Italia and La7

Source: elaborations on Auditel data

# TIME SPENT FOR CORONAVIRUS-RELATED VIDEOS

In the first month of the epidemic emergency, Italians have paid increasing attention to online video on the coronavirus: the related viewing is 8 times higher than in the first period.

Afterwards, there is a drop in consumption that are however higher (7 times more) than the values recorded before the epidemic emergency in Italy

#### **TOP 15 VIDEOS ON THE CORONAVIRUS**

The most viewed videos before the epidemic emergency in Italy were scientific and medical-information contents that highlight the risks of contagion in Italy and all around the World, as well as updates on the Chinese outbreak and the related containment measures.

By the **beginning of the epidemic spread** in Italia, there is an increasing attention around the **health-care system in Italy** and the situation in the most hit intensive care units.

In the second month of the emergency, much attention has been paid to current events and, more specifically on the journalistic investigations regarding the possible causes of the contagion

13

# HOW INTERNET CONSUMPTION CHANGES IN EUROPE DURING THE EPIDEMIC

# Online news: the consumption in Europe during the epidemic

# AGCOM COMUNICAZION

#### Total visits to general news sites and apps (index)



Source: elabor

Source: elaboration on Comscore Custom Reporting, Jan - Apr 2020

The attention to news about national and international affairs **gradually increases in all countries** with the beginning of the emergency and continues, reaching peaks in correspondence with the spread of the epidemic in different countries. **Starting from the week 30 March-5 April**, there is a **reduction** in user interest in these contents

#### **GROWTH IN TOTAL VISITS**



+90% ■ Italy
+71% ■ France
+68% ■ Germany

+59% ■ Spain
+55% ■ LIK

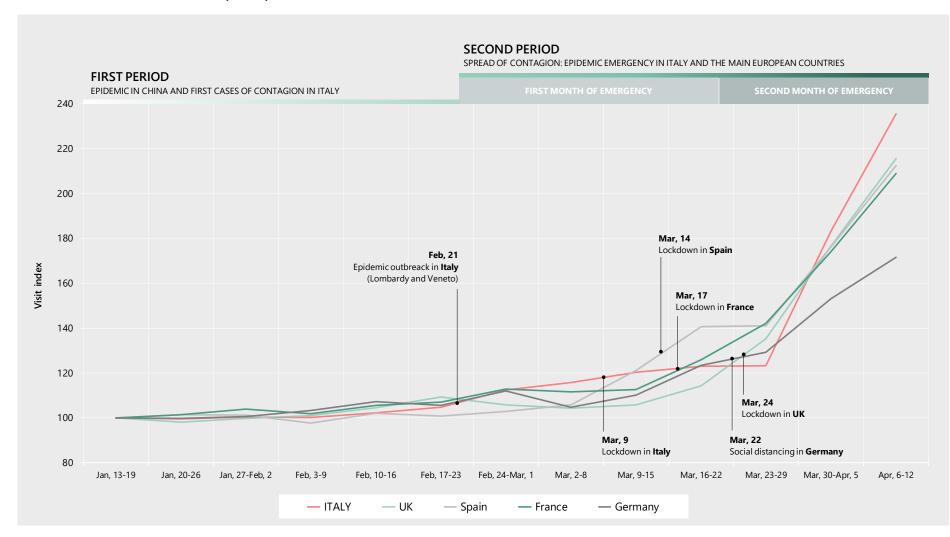
Note: for the second month of the emergency, data are available until the week 6-12 April

15

# Social networks: the consumption in Europe during the epidemic

#### AUTORITÀ PER LI GARANZIE NELLI AGCOM COMUNICAZION

#### Total visits to social networks (index)



seen above, in Italy devote great attention to the coronavirus, grow in all countries in the second period. Unlike general news sites and apps, user interest continues to grow even in the second month of the emergency, and even more so in the last 2 weeks

Also the visits to social networks, which, as

#### **GROWTH IN TOTAL VISITS**





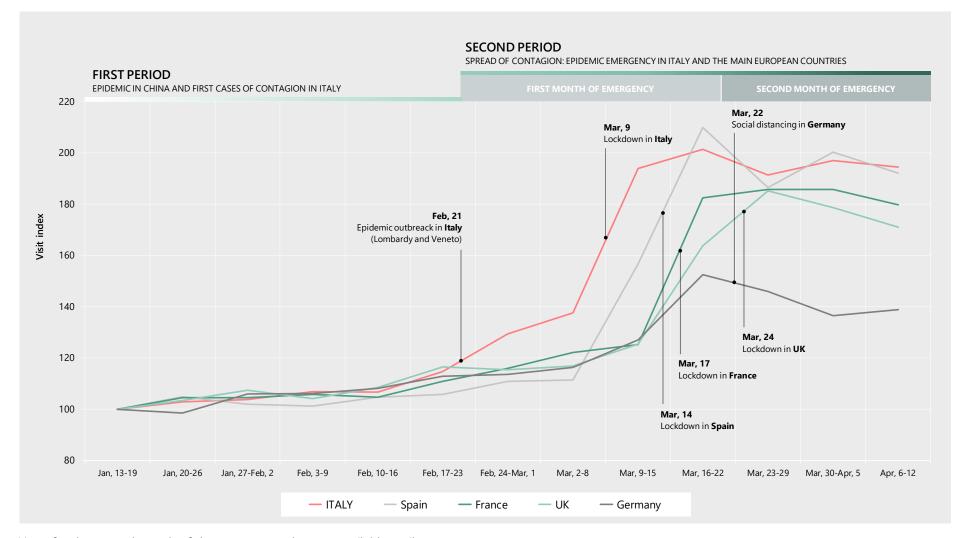
Note: for the second month of the emergency, data are available until the week 6-12 April

Source: elaboration on Comscore Custom Reporting, Jan - Apr 2020

# Instant messaging: the consumption in Europe during the epidemic

#### AUTORITÀ PER LE GARANZIE NELLE AGCOM COMUNICAZION

#### Total visits to instant messaging sites and apps (index)



Note: for the second month of the emergency, data are available until the week 6-12 April

Source: elaboration on Comscore Custom Reporting, Jan - Apr 2020

Traffic to instant messaging sites and apps remains almost unchanged in the first period and increases with the spread of coronavirus contagion in Italy and other countries, as the respective containment measures are adopted. The use of these services, through which users also receive information about the coronavirus, begins to return as of March 30

#### **GROWTH IN TOTAL VISITS**



APRIL, 6-12 compared to week 13-19 Jan +94% ltaly

+92% **Spain** 

**+80%** ■ France

**+71%** ₩ UK

+39% Germany

CYBERSECURITY THREATS AND CORONAVIRUS

4

In collaboration with SOGEI



#### APT (HACKERS)



Targeted and persistent cyber attacks by subjects with considerable technical skills and resources

#### COMMAND AND CONTROL (C2)



It refers to the ability to influence, through dedicated servers, a compromised computer system by being able to control it

#### MALSPAM



Massive sending of emails containing malicious attachments in order to infect computer systems and steal confidential information

#### MALWARE



Abbreviation for "malicious software". Any form of code used to bring a computer system to perform unplanned operations

#### **PHISHING**



Email artifacts and/or malicious sites to induce the user to provide personal information or download files to infect the system

#### RANSOMWARE



Malicious program that restricts the use of the user device by encrypting the data and asking the owner for a ransom to unlock it

#### RAT



Acronym for "Remote Access Trojan". Malicious program aimed at capturing credentials and controlling the attacked user's machine

#### **SPOOFING**



In the context of email, masking for malicious purposes of the source address of an email

#### **SPEAR PHISHING**



Phishing aimed at specific categories of users. The emails are built ad hoc, with content targeted to a specific field

#### TROJAN



Type of malware that hides within an apparently useful and harmless application

# CYBER ATTACKS IN THE WORLD

+16% 1st BIMESTER 2020

compared to the 1st bimester 2019

Worldwide, in the first months of 2020, there has been a significant **increase** in cyber threats and attacks, many of which are based on the exploitation of the socio-psychological vehicle of the current pandemic.

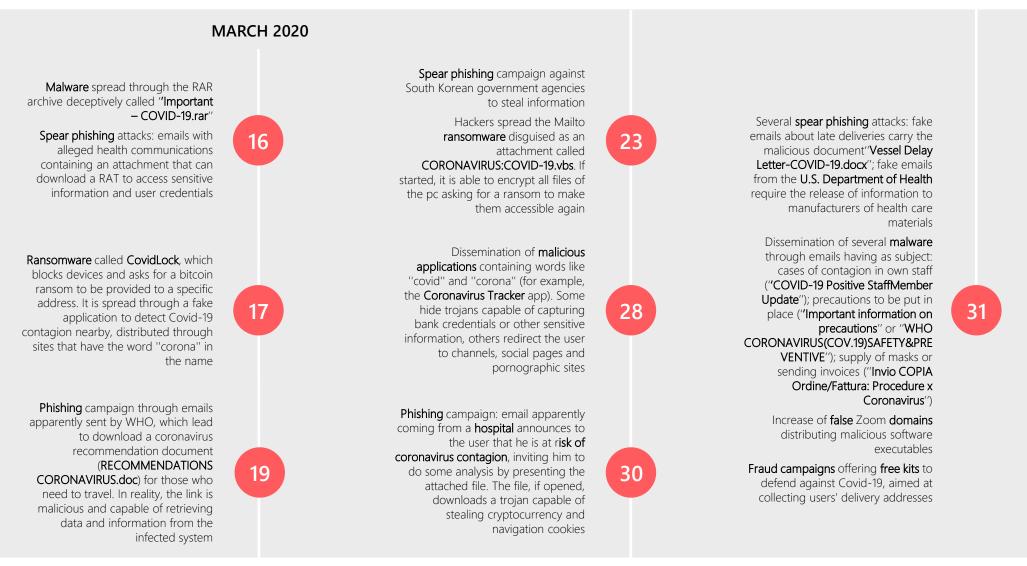
The apprehension and the need for up-to-date information make both users and companies involved in the economic and health crisis particularly vulnerable to attacks

Source: Sogei

# Cyber attacks related to the coronavirus worldwide



#### Timeline of the main cyber attacks detected in the world from March 15 to 31



# NEW INTERNET DOMAINS RELATED TO COVID-19

**16,000** 

FROM BEGINNING 2020

of which around 20% are malicious or with a low reputation

In March, several types of cyber attacks related to the coronavirus issue were detected. Among the sneakiest is the spread of apps that leverage users' fears about Covid-19. The prevailing purpose of these threats seems to be economic, in addition to stealing credentials for future attacks

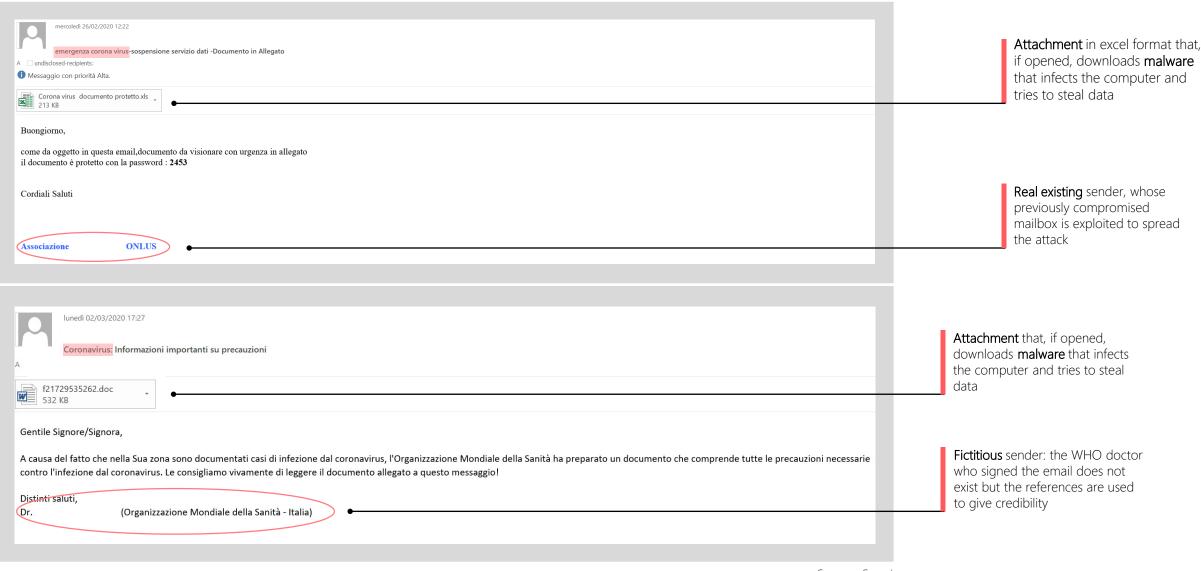
Malicious campaigns spread in certain parts of the world can often resemble and potentially be **replicated** on a **global scale** 

Source: Sogei

# Cyber attacks related to the coronavirus in Italy



#### Examples of phishing campaigns carried out through malicious coronavirus emails





#### COMPOSITION OF THE DOCUMENT DATABASE

If not differently stated, the figures reported in Section 1 of this Report, are the result of AGCOM elaborations carried out on a database built from data extrapolated through the platform developed by *Volocom Technology*.

More specifically, the analyzes were conducted on the entire textual content extrapolated from around 5 million documents created in Italy (from January 1 to April 20, 2020) by more than 2,000 information sources (national television and radio channels, newspapers, news agencies, websites of traditional publishers, native online news outlets, and related pages and accounts of social networks), and sources of disinformation (websites and social pages/accounts) identified as such by external subjects specialized in debunking activities. The amount of online disinformation produced in Italy was therefore estimated using a subjective methodology, i.e. considering the total number of documents created monthly by the aforementioned sources of disinformation.

The reference set of sources of information and disinformation is constantly updated. As a result, adjustments may need to be made to the estimates for previous periods, so that, for some indicators, there may be slight deviations in the values reported in different issues of the Report.

The database is composed by the entire textual content of all documents produced during a day by every information and disinformation source. For document, we mean the entire article, in the case of newspapers and websites; the transcription of a transmission segment; in the case of Tv and radio; all tweets/posts in the case of online platforms.

#### **TOPIC MODELING**

For the study of the main topics of coronavirus news (Section 1), analyses were conducted on the entire content disseminated by information and disinformation sites from January to April 2020, adopting a methodological approach known as **topic modeling**.

A topic model is a statistic model for the automatic individuation of topics 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 co-occurrence of the terms used in the collection of documents.

The same is at the basis of LDAvis (Sievert e Shirley, 2014), the interactive visualization system that 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 terms**, associated to every one of them.

In detail, 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 that explain the variability of the data.

#### SUPPLEMENTARY DOCUMENTS

Further studies and updates on the AGCOM initiatives regarding online disinformation and the Covid-19 epidemic are available in a dedicated area of the AGCOM webiste: Tavolo piattaforme digitali e Big data – Emergenza Covid-19, and in the related pages Covid-19 per gli utenti and Fact-checking

A more detailed description of the methodology is presented in the AGCOM Report (2018),

# News vs. fake in the information system

More in-depth analysis on the definitions and technical dimensions of online disinformation are in the AGCOM Report (2018), <u>Online disinformation strategies and the fake content supply chain</u>, conducted within the <u>Roundtable on pluralism and online platforms</u>

The AGCOM policy recommendations and action plan on the online news system have been recently published in the <u>document</u> closing the <u>Sector Inquiry on digital platforms and the online information system</u> (2020)

www.agcom.it/osservatorio-sulla-disinformazione-online

Department of Economics and Statistics

# CONTRIBUTIONS AND ACKNOWLEDGEMENTS

The realization of the Report on online disinformation – Special issue on Coronavirus s the result of monitoring activities based on the integration of multiple sources and the analysis, among other things, of a large amount of data made available by the participants of the **Roundtable on Digital Platforms and Big Data – Covid-19 Emergency**.

For the precious and active collaboration, we would like to especially thank:

- Auditel
- Comscore
- Newsguard
- Sensemakers-Shareablee
- Sogei

The Report and, more generally, the Roundtable on Digital Platforms and Big Data can also benefit from the research results of the **Data Science Task Force** activated by AGCOM - Economic-Statistics Department on the theme of online disinformation during the Covid-19 emergency.

For the scientific cooperation and the continuous comparison on the most advanced study techniques and methodologies, we thank the research centres and academic institutions involved in the Task Force:

- Sapienza University of Rome, Physics Department
- Università Ca' Foscari di Venezia Research Institute for Complexity
- SONY Computer Science Lab Paris
- CNR Institute for Complexity Systems
- Enrico Fermi Study and Research Center





www.agcom.it



ses@agcom.it

DEPARTMENT OF ECONOMICS AND STATISTICS