

Much ado about one (single) thing ?

H2020 ODYCCEUS internal seminar

Claude Grasland, Jean-Marc Vincent

05/06/2020

Section 1

INTRODUCTION

Newspapers ...

Have you notice something strange in your favorite newspapers during the last months ?



An exceptional perturbation of media agenda ?

We have experimented an exceptional perturbation of media agenda by covid-19 crisis. Among many studies, we propose to introduce the debate by two recent works :

An analysis of non-trivial world in newspaper Le Monde (2000-2020)

- Title : Parle-t-on trop du coronavirus ?
- Author : Erwan X***, blogger
- Approach : IA & data mining
- URL : <https://www.iacomprise.net/?p=524>

A multimedia analysis of french classical and social media (Jan-May 2020)

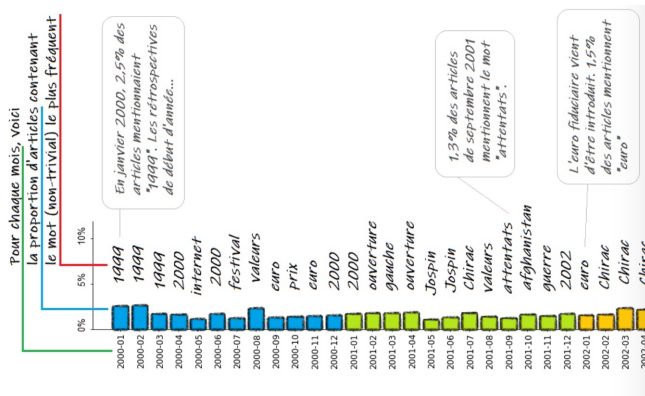
- Title : “Coronavirus - Étude de l'intensité médiatique”
- Author : Nicola Hervé, INA
- Approach : Media Studies & Political Science
- URL : <http://www.herve.name/pmwiki.php/Main/Etude-Coronavirus>

Non-trivial words in *Le Monde*

The author, specialized in IA and datascience, has measured the **top non-trivial word** in the newspaper *Le Monde* for each month between 2000 and 2020

Parle-t-on tant que cela du coronavirus ?

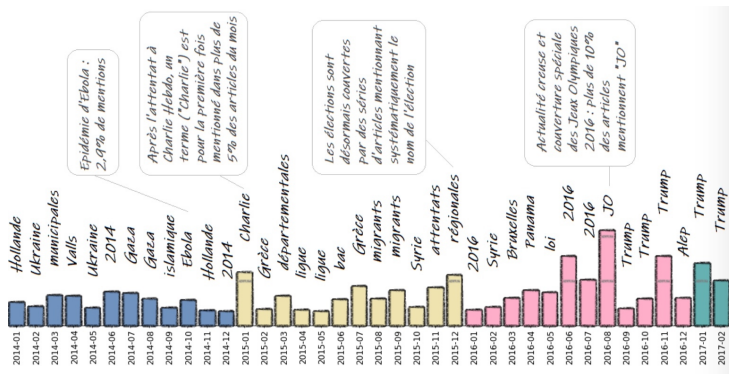
Analyse sur les titres d'articles sur [lemonde.fr](https://www.lemonde.fr)



Source : <https://www.iacomprise.net/?p=524>

Non-trivial words in *Le Monde*

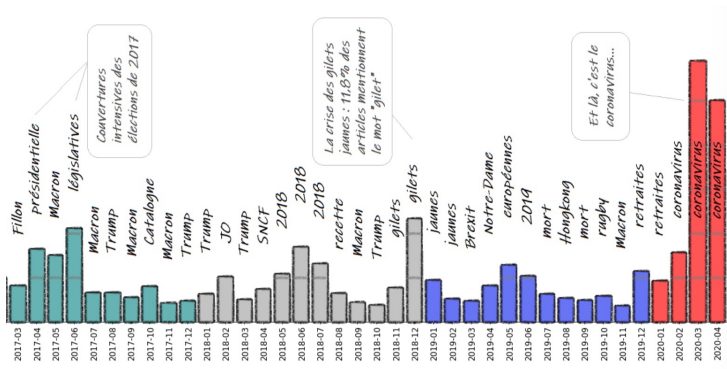
- The word *Ebola* was the top word in october 2014 (2.9% of news)
- The word *JO* (Olympic Games) reached 10% of news in August 2016



Source : <https://www.iacomprise.net/?p=524>

Non-trivial words in *Le Monde*

- The word *Gilets* reached 12% during the yellow jacket crisis.
- But the 20-years record was *coronavirus* in March 2020 (28 %)



Source : <https://www.iacomprise.net/?p=524>

Multimedia analysis of coronavirus in France

Nicolas Hervé (INA) has compared the timeline of various keywords associated to the crisis in the news published by AFP.

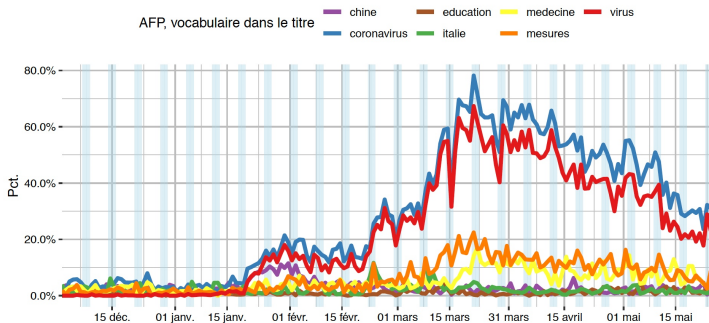


FIGURE 9 – Pourcentage de dépêches AFP pour lesquelles au moins un mot est trouvé dans le titre en fonction des groupes de vocabulaire.

Multimedia analysis of coronavirus in France

He has also measure the frequency of the same words in french tweets.

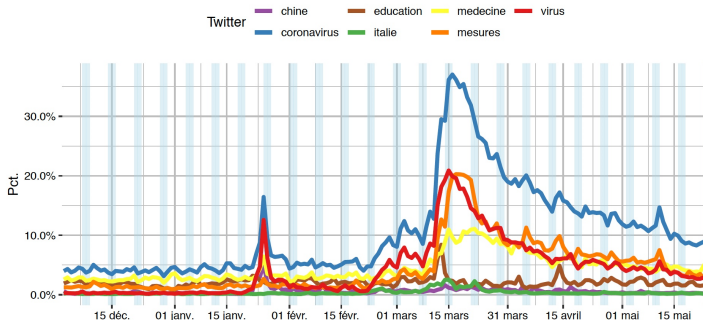
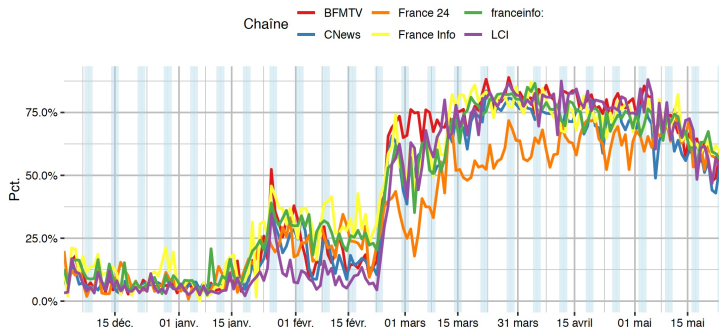


FIGURE 12 – Pourcentage de tweets en français captés pour lesquels au moins un mot du vocabulaire est trouvé en fonction des groupes de vocabulaire.

Multimedia analysis of coronavirus in France

And finally measure the share of time devoted to the crisis in the six most important TV channels of breaking news in France.



Multimedia analysis of coronavirus in France

Finally, he proposed a timeline related to specific explanatory events.

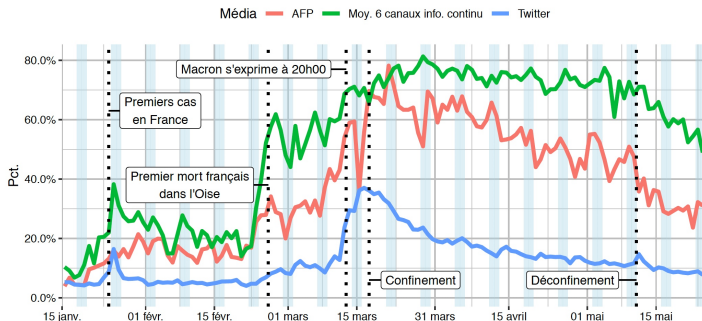


FIGURE 22 – Estimation du pourcentage des contenus diffusés qui sont liés aux principales thématiques du coronavirus : virus, médecine et mesures prises

Overarching hypothesis

We define our overarching question as follow:

Can we define the Covid media outbreak as a specific type of historical focusing event, a giant case of international spreading event, a combination or both or something completely different and never observed before?

To answer to this question, we propose to develop a space-time analysis of the topic based mainly on measures of salience over different time periods and at different scales of spatial and media analysis

- 1. Definition of the object :** we propose an objective method of measurement of the salience of the covid outbreak in daily newspapers
- 2. Data collection :** Collect the related data with a large geographical coverage in present time but also for historical comparison on a smaller subset
- 3. Geographical approach :** Comparison of the spread of media outbreak in various world countries in 2020
- 4. Historical approach :** Comparison between the covid outbreak of 2020 to Sept.11 terror attacks of 2001
- 5 Statistical modelling :** Propose a descriptive and explicative model of contemporary trends

Section 2

1. A MEDIA OUTBREAK

A media outbreak

- The word **coronavirus** can be considered as a signal that reveals the presence of a **media outbreak** that has been disseminated in the majority of stories published by newspapers during the first months of year 2020. The word **2019-nCoV** is a temporary name used by WHO at the beginning of the outbreak. The word **covid-19** is a mutation of the initial diseases but participate from the same outbreak.
- In China, for example, the outbreak was called (*Wuhan Pneumonia*) in the earlier stage of the diffusion. But in order to avoid the association of the disease with China, it was further renamed (*Novel Corona*) or (*Novel Corona Virus*) or (*Novel Corona Virus Pneumonia*).

N.B. (*italics*) refers to ideograms that are not reproduced here.

The concept of viral words

- **Coronavirus & covid-19** are twin viral words that can not be separated because they are generally combine in the same news. The mixture is clearly visible in 6934 news with more than 500 words published by *Le Figaro* from 1-1-2020 to 06-05-2020.

##		cor0	cor1	cor2	cor3	cor4	cor5	Sum
##	cov0	4624	557	213	107	42	42	5585
##	cov1	366	189	87	35	12	23	712
##	cov2	155	102	45	17	9	11	339
##	cov3	66	40	23	14	1	10	154
##	cov4	30	12	14	6	3	3	68
##	cov5	33	19	12	6	1	5	76
##	Sum	5274	919	394	185	68	94	6934

Source : FACTIVA

Detection of viral words by language

The twin viral words was easy to choose in latin's alphabet languages

code	langage	KW1	KW2
de	German	corona*	covid*
es	Spanish	coronav*	covid*
en	English	coronav*	covid*
fr	French	coronav*	covid*
it	Italian	coronav*	covid*
pt	Portuguese	coronav*	covid*
id	Indonesian	korona*	covid*

Detection of viral words by language

Otherwise, choice was made with support of native speakers :

code	langage	KW1	KW2
ru	Russian	корона*	covid*
ar	Arabic	كورونا	كوفيد
ir	Farsi	کورونا	کوفید
jp	Japanese	コロナウイルス -	
zh	Chinese	武汉肺炎	冠狀病毒
ko	Korean	코로나*	코로나19

Special thanks to Dr Kohei Watanabe, Dr Oul Han and Dr Chung-hong Chan, for their answers to my questions on the naming process of the Coronavirus/Covid-19 in Japan, China and Korea.

Which measure of news infection ?

We can imagine different solutions for the measure of the presence of viral words in news :

- presence of at least one viral words in the title of the news
- presence of at least viral words in the full text of the news
- presence of at least k_{min} viral words in the full text of the news
- density of viral words in the text greater than d_{min}
- presence of two different viral words in the text
- ...

Definition of the target phenomena

We have decided to adopt the simplest solution and to measure the **presence of at least one of the twin viral word coronavirus/covid-19 in the full text of news**, whatever the length of the story or the related density.

- A news N published in language L_N is composed of a set of words $W_N = \{W_1..W_k\}$.
- For each language L the infection of the news is defined by a set of keywords $KW_L = \{KW_{L,1}..KW_{L,p}\}$
- $f_{0,1}(N, L_N) \rightarrow (W_N \cap KW_{L_N}) \neq \emptyset$

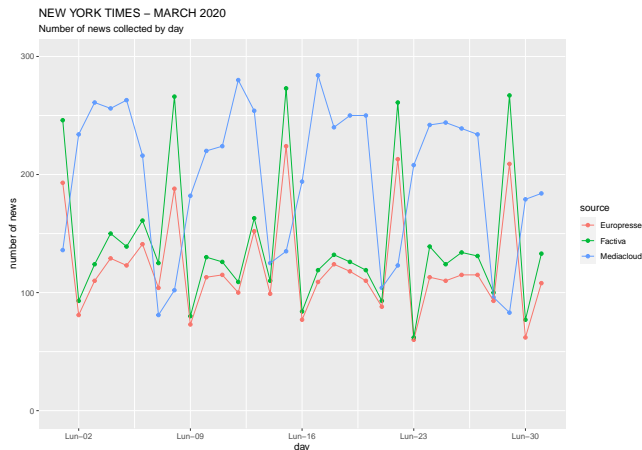
Section 3

2. DATA COLLECTION

Benchmarking of sources

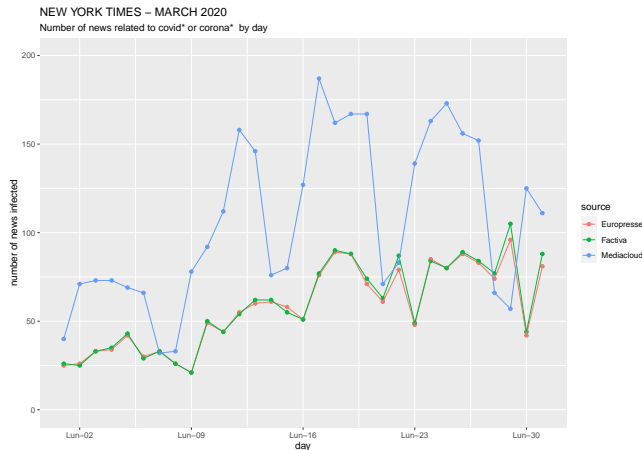
- 1 **Mediacloud** : collection of rss flows from newxspapers realized by MIT and Harvard, available for public research for free. Majority of media available only since March 2013
- 2 **Dow Jones Factiva** : private collection of newspapers available in some case since 1980's. World coverage.
- 3 **Europresse** : private collection of newspapers available in some case since 1980's. More focused on french and european newspapers.

Benchmarking of sources



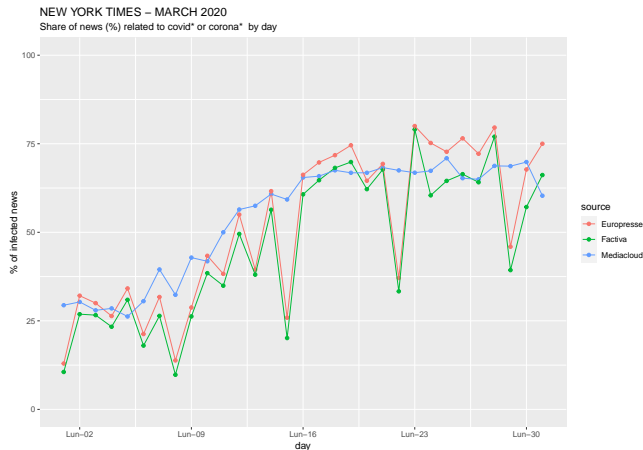
- Comment : the difference of production observed on sunday is associated to the supplement of the NYT available in Europresse and Factiva but not in Mediacloud

Benchmarking of sources



- Comment : the effect of the week-end is reduced because the supplement are mostly related to book review with few reference to covid/corona

Benchmarking of sources



- Comment : Mediacloud is more regular because it excludes the book review and other supplement of the week-end collected by Factiva and Europresse.

Benchmarking of sources

- 1 **Mediacloud** appears as the most interesting source for analysis related to the period 2013-present. It will be chosen for an extended **spatial coverage** (125 medias from 25 countries).
- 2 **Factiva or Europresse** are very similar in terms of content. They are usefull for the collection of full text and for the realisation from **historical comparisons**








Mediacloud collection

For each country, we have selected five representative daily newspapers in the Mediacloud database with the *source manager*.

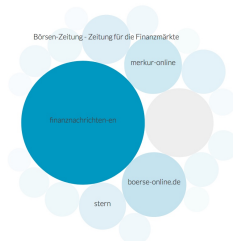
SOURCE LIST COLLECTION CONTENT

Sources

This collection includes 60 media sources.

Media Source	Stories per Day	First Story
 afp-deutsch	0	1/30/2012
 AktivOnline	2	8/15/2016
 ARD - Germany (German)	36	11/6/2017
 ard-text.de	0	6/11/2018
 berlinerumschau.com	0	12/29/2014
 bild.de ★	158	6/10/2013
 boerse-online.de	269	9/30/2013

Recent Source Representation ? DOWNLOAD OPTIONS...



<https://sources.mediacloud.org/>

Mediacloud collection

The regular expression and the time period are then introduced in the **news explorer** tool and applied to the media previously selected :

The screenshot shows the Mediacloud Explorer interface. At the top is a purple navigation bar with links: EXPLORER, TOPIC MAPPER, SOURCE MANAGER, SUPPORT, BLOG, ABOUT, and user icons. Below the navigation bar is a table with four media collections:

DEU_tazxxx	DEU_suddeu	DEU_frankf	DEU_tagspi
taz.de Jan 1, 2020 to May 20, 2020	suddeutsche Jan 1, 2020 to May 20, 2020	Frankfurter Allgemeine (FAZ) Jan 1, 2020 to May 20, 2020	Tagespiegel Jan 1, 2020 to May 20, 2020

Below the table is a search form with three main sections:

- 1 Enter search terms**: A text input field containing the query `corona* covid* ncov* korona*`. Below the input is a note: "Media Cloud will return stories that match your search query. We use standard boolean search syntax. Learn more about writing boolean search queries."
- 2 Select your media**: A dropdown menu showing `taz.de`. Below the dropdown is a note: "Choose individual sources or collections to be searched. Our system includes collections for a large range of countries, in multiple languages. Learn more about choosing media." and an **ADD MEDIA** button.
- 3 Enter dates**: Two date input fields showing `2020-01-01` and `2020-05-20`. Below the inputs is a note: "Enter your inclusive date range. Our database goes back to 2011, however the start date for different sources and collections can vary. Click on a source or collection to learn more about when we added it."

At the bottom of the search form are three buttons: **LOAD SAVED SEARCH...**, **SAVE SEARCH...**, and **SEARCH**.

source : <https://explorer.mediacloud.org/>

Mediacloud collection

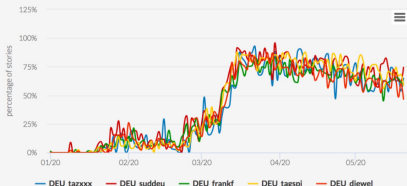
Data can be visualized in order to check for unexplained outliers or breaking time series before to download and store the results

Attention Over Time

Compare the attention paid to your queries over time to understand how they are covered. This chart shows the number of stories that match each of your queries. Spikes in attention can reveal key events. Plateaus can reveal stable, "normal", attention levels. Click a point to see words and headlines for those dates. Use the "view options" menu to switch between story counts and a percentage.

This chart includes one line for each query in your search. Each line charts the number of stories that matched your query per day in the sources and collections you have specified.

Roll over the line chart to see the stories per day in that period of time. Click the download button in the top right to download the raw counts in a CSV spreadsheet. Click the three lines in the top right of the chart to export the chart as an image file.



VIEW OPTIONS... DOWNLOAD OPTIONS...

source : <https://explorer.mediacloud.org/>

Mediacloud collection

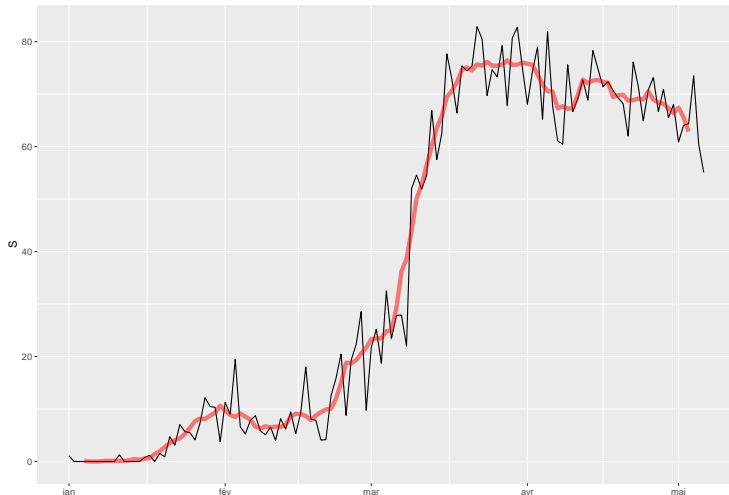
The 125 media selected in 25 countries are stored in a single data table with following format :

country	media	day	news_tot	news_covid
DEU	frankf	2020-01-01	92	1
DEU	frankf	2020-01-02	115	0
DEU	frankf	2020-01-03	131	0

country	media	day	news_tot	news_covid
DEU	frankf	2020-05-04	132	97
DEU	frankf	2020-05-05	124	75
DEU	frankf	2020-05-06	236	130

Visualization at daily or weekly level

For a better visualization of trends, we will use for some analysis a rollmean with a window of 7 days (which balance the effects of specific perturbation of news during the week-end).



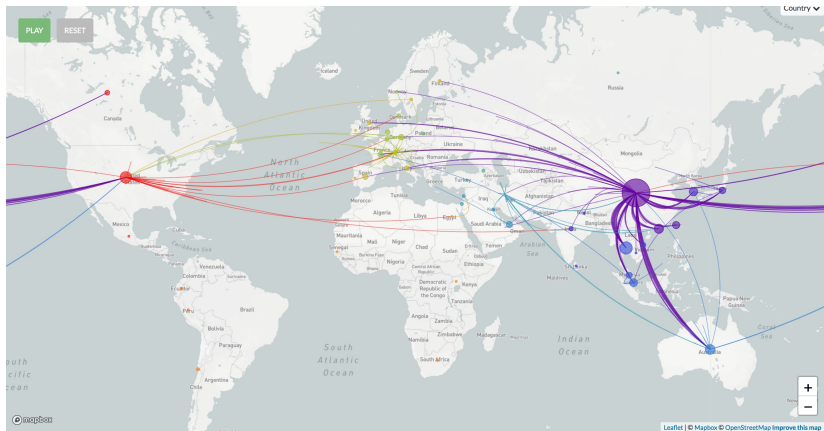
Section 4

2.SPACE-TIME DIFFUSION

Chronology of disease diffusion (until 31 January)

Before to analyze media outbreak, we propose a brief reminder of the spread of disease provided by the website

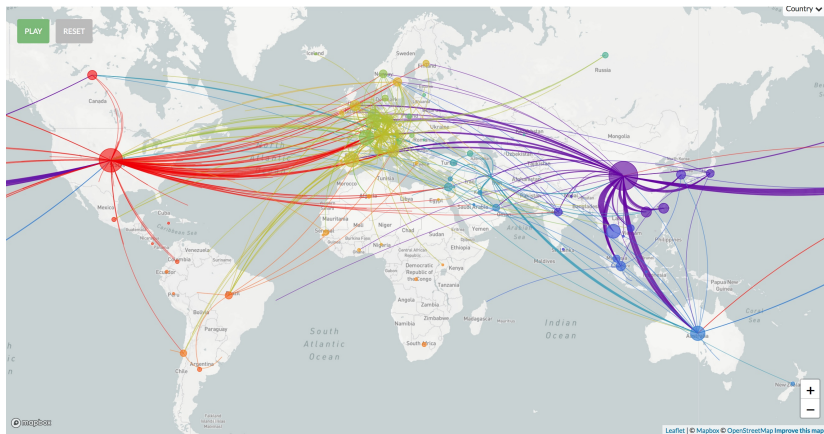
<https://nextstrain.org/ncov/global>



Chronology of disease diffusion (until 29 February)

Before to analyze media outbreak, we propose a brief reminder of the spread of disease provided by the website

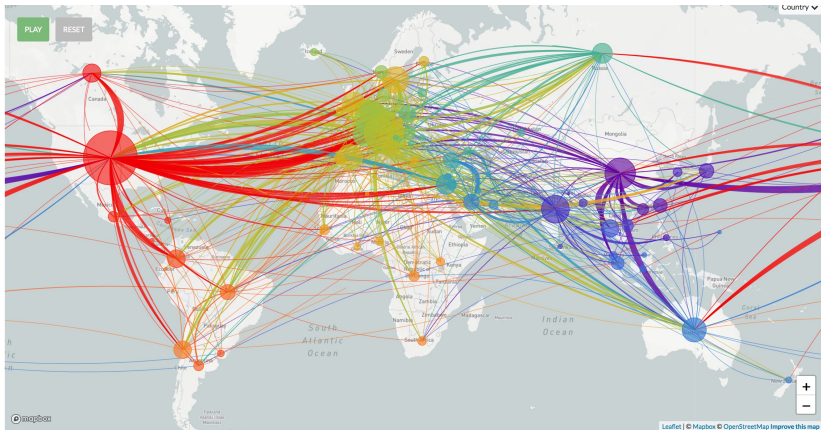
<https://nextstrain.org/ncov/global>



Chronology of disease diffusion (until 31 March)

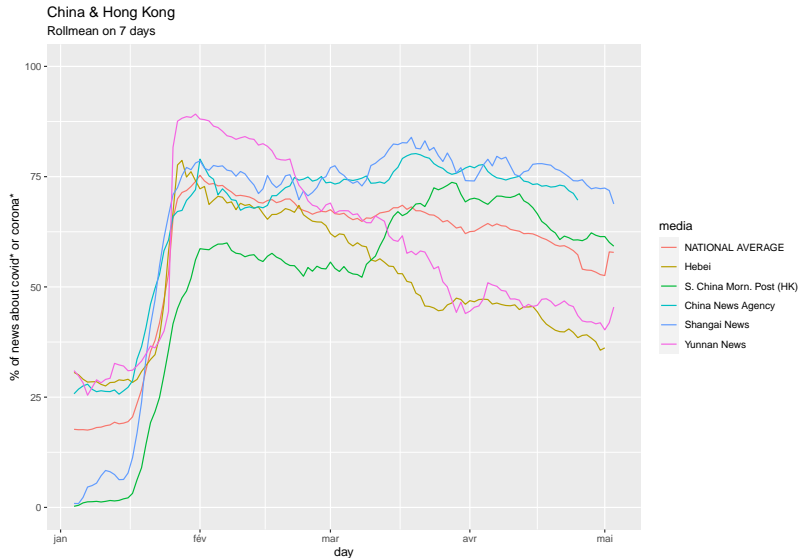
Before to analyze media outbreak, we propose a brief reminder of the spread of disease provided by the website

<https://nextstrain.org/ncov/global>



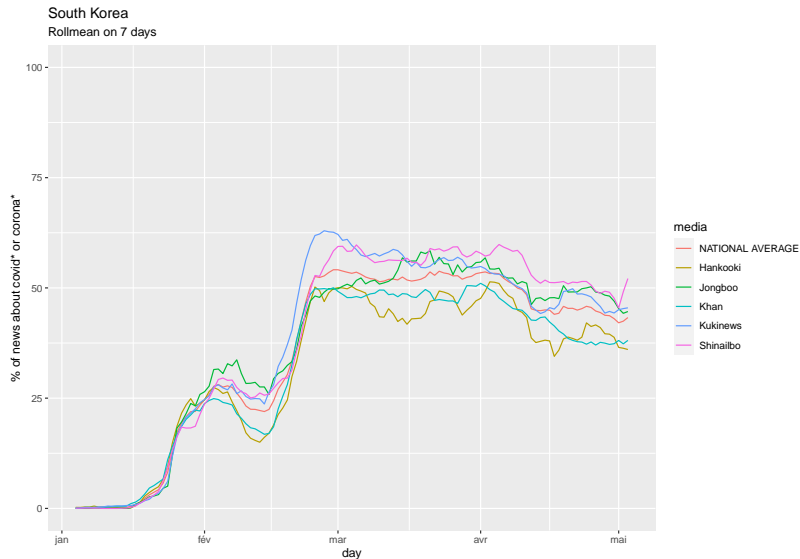
China : > 50% the 25th January

The country of origin of the disease, first concerned by media outbreak.



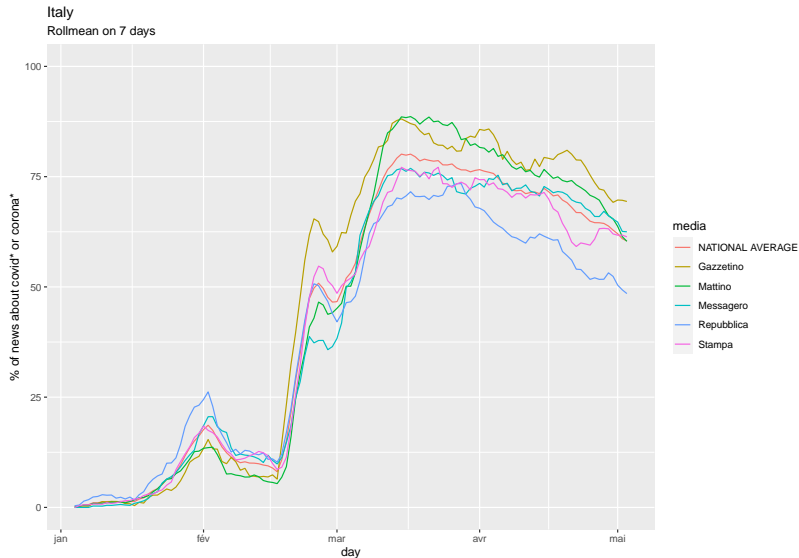
South Korea : $> 50\%$ the 24th february

Early dissemination but low level of saturation.



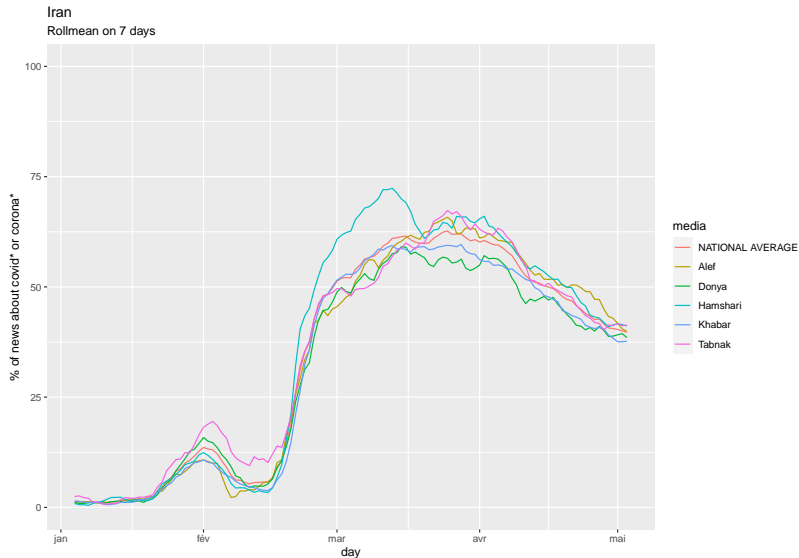
Italy : > 50% the 26th February

First country heavily concerned in European Union



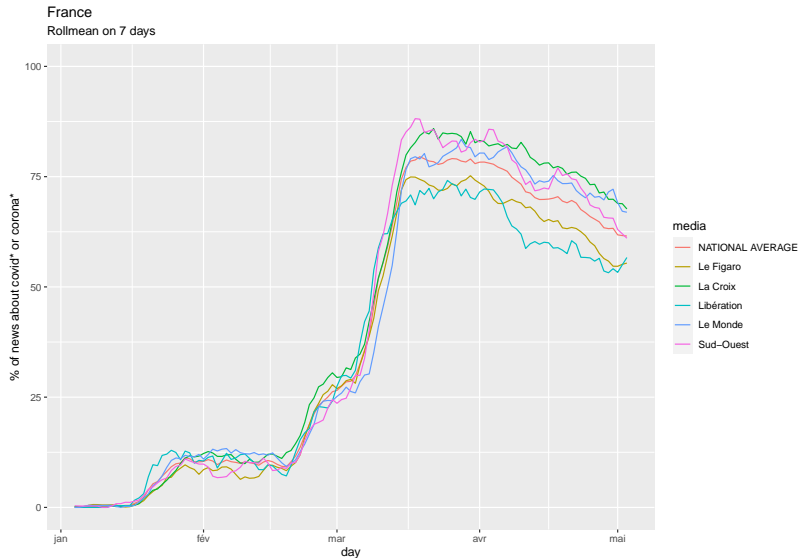
Iran : > 50% the 2nd March

Similar to Italy but with lower peak and earlier decline.



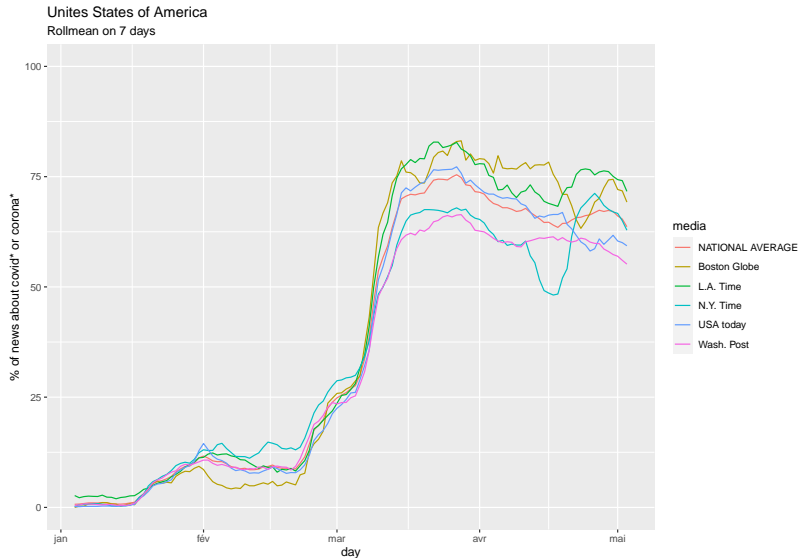
France : > 50% the 10th March

The second country concerned in European Union, 10 days after Italy.



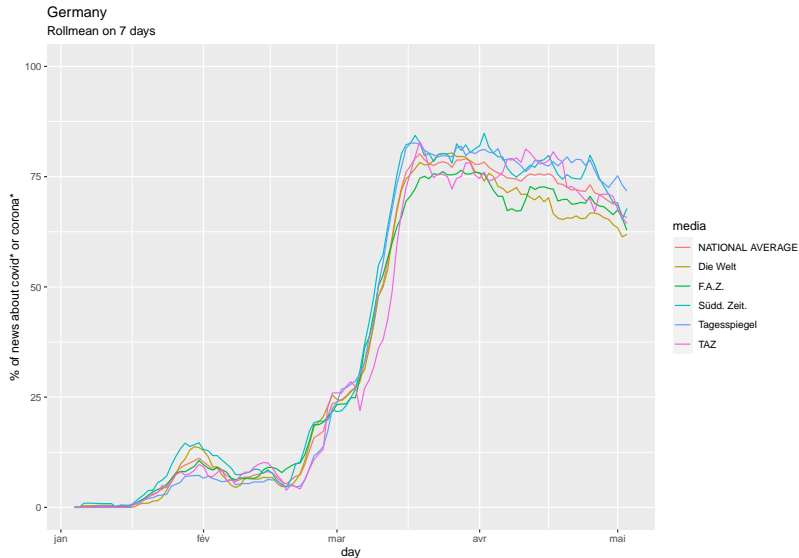
USA : > 50% the 10th March

Similar to France but with longer maximum plateau.



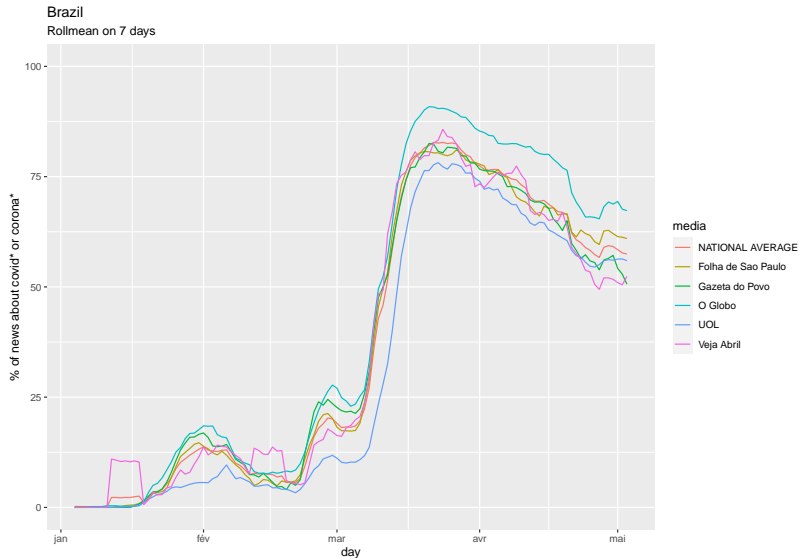
Germany : > 50% the 11th March

Nearly equivalent to France, with one or two days of delay



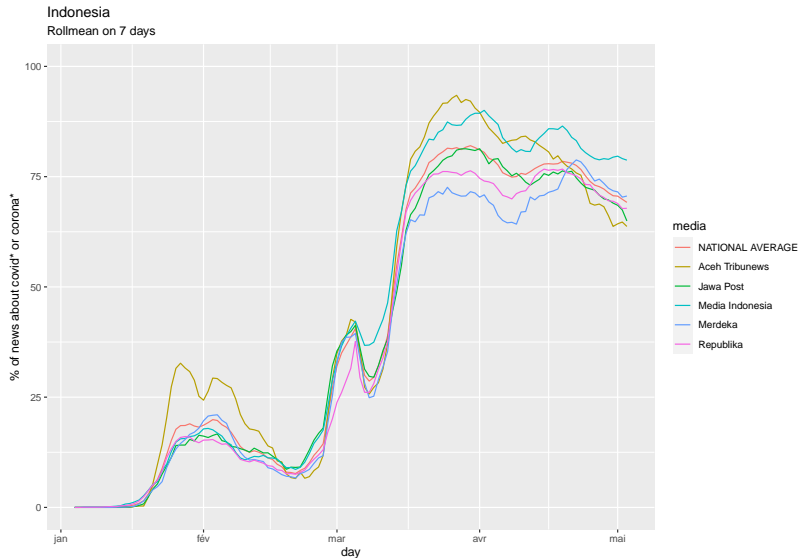
Brazil : $> 50\%$ the 12th March

Similar to France. Perturbations of *Veja Abril* curve are aretfacts.



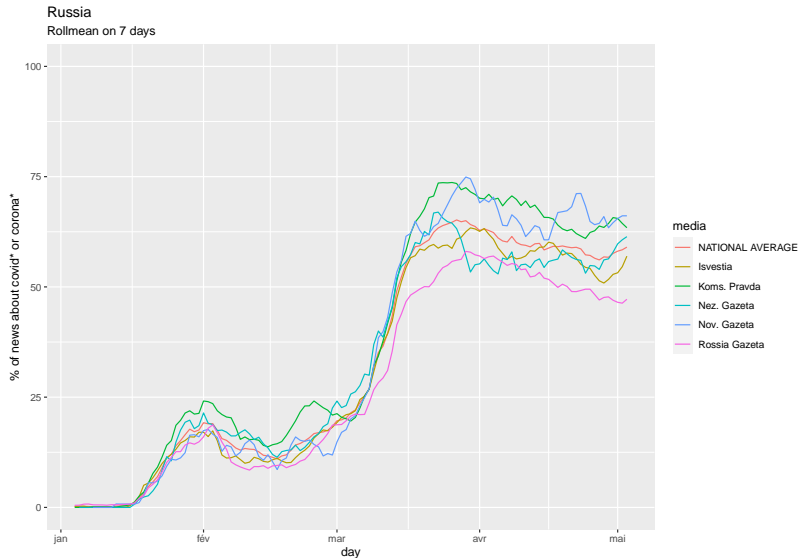
Indonesia : > 50% the 14th March

General profile but with stronger oscillations at the beginning.



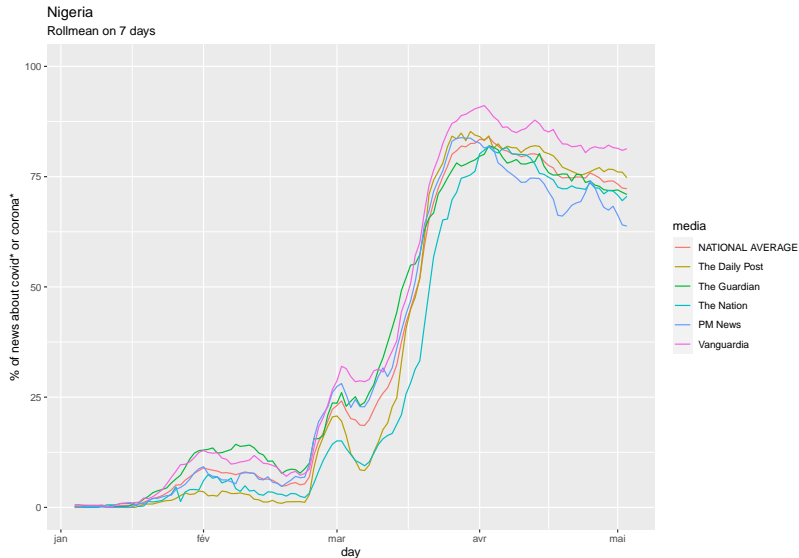
Russia : > 50% the 15th March

Similar to Iran with later diffusion and relatively low maximum level

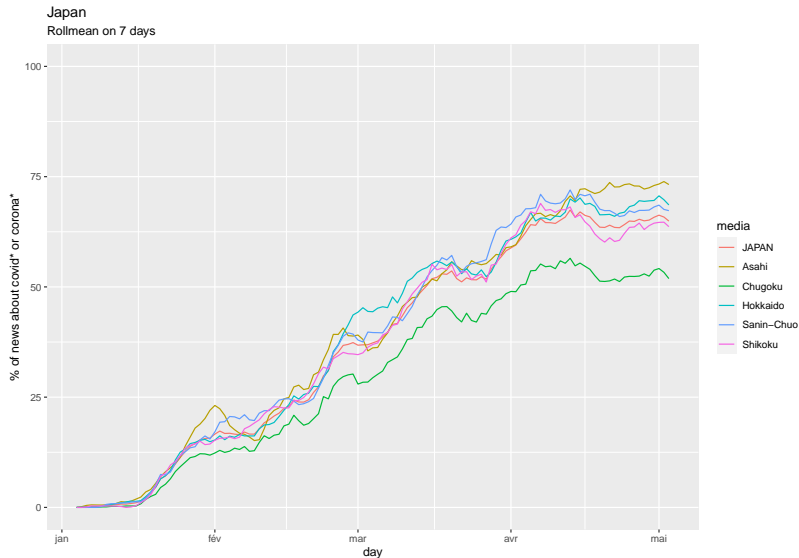


Nigeria : > 50% the 19th March

Similar to european curves but delayed by one week.



The strangest curve ... Linear with small cycles.

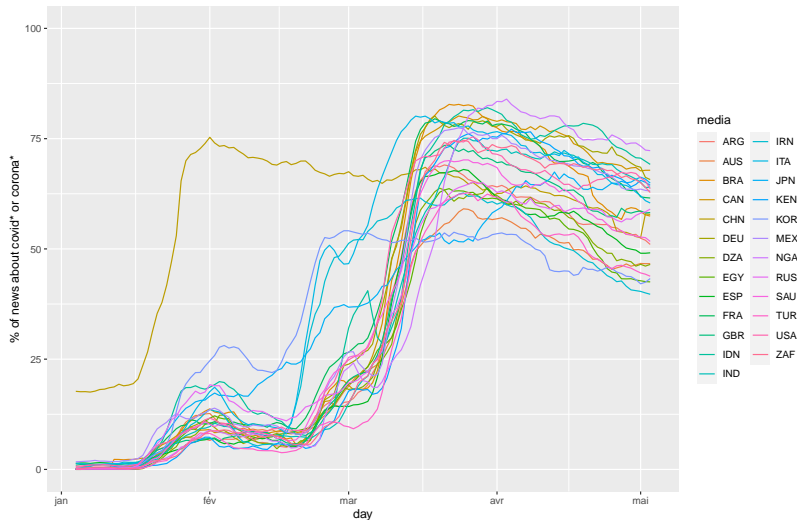


Comparison of the 25 countries

Comparison of average national trends in 25 countries.

25 top world countries

Rollmean on 7 days

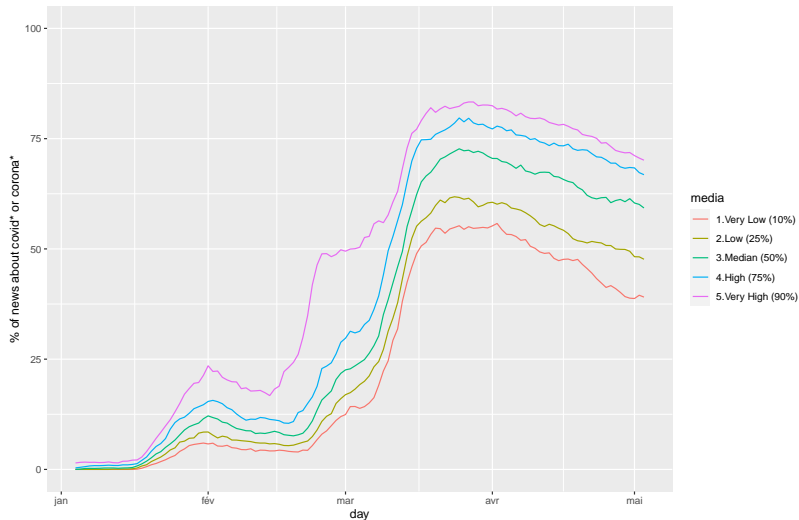


Global trends

Summary of trajectories of the 125 media through quantiles.

25 top world countries

Rollmean on 7 days



Section 5

4.A GIANT FOCUSING EVENT ?

The concept of “focusing event”

‘A focusing event is an event that is sudden; relatively uncommon; can be reasonably defined as harmful or revealing the possibility of potentially greater future harms; has harms that are concentrated in a particular geographical area or community of interest; and that is known to policy makers and the public simultaneously.’

Source : Birkland, Thomas A. 1997. After Disaster: Agenda Setting, Public Policy, and Focusing Events. Georgetown University Press.

Sept.11 terror attacks as “focusing event”

‘The news coverage of the September 11 events follows the typical trend, in which the news media aggressively cover an issue for a short time, and then coverage fades as the event recedes into the past and as political institutions decide whether or not to act in substantive or symbolic ways’

Source : Birkland, Thomas A. 2004. “ ‘The World Changed Today’: Agenda-Setting and Policy Change in the Wake of the September 11 Terrorist Attacks.” *Review of Policy Research* 21(2): 179–200.

H1. The level of salience reached by the Covid-19 is exceptional, compared to the reference event of Sept. 11

- H1.1 The peak of coverage of covid-19 is **higher** than the one observed in Sept. 11
- H1.2 The duration of the peak of high coverage is **longer** for Covid-19 than for Sept.11
- H1.3 The world coverage is **more homogeneous** for Covid-19 than for Sept. 11

Experimental protocol

We propose to compare the salience of viral words :

- the salience of *terror-* in news published by media from August to Decembre 2001
- the salience of *coronav-* or *covid-* in news published by media from january to June 2020

We have chose a set of media available at each period

- 3 daily newspapers from USA
- 3 daily newspapers from European countries
- 3 daily newspapers from Devlopping countries
- 3 press agencies

The historical corpus (source : **FACTIVA**)

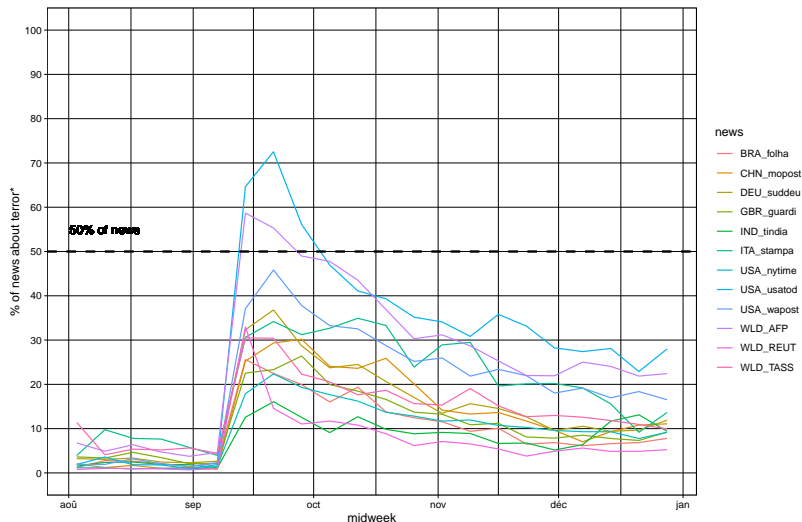
Table 3: Table 1 : selection of media from Factiva database

code	space	description
en_GBR_guardian	NORTH	UK national newspaper covering world affairs,
de_DEU_sueddeutsche	NORTH	Munich-based daily national German newspaper
it_ITA_stampat	NORTH	Leading Italian daily newspaper covering polit
en_IND_tindian	SOUTH	The Times of India covers business, economic
en_CHN_mingpao	SOUTH	South China Morning Post covers general and
pt_BRA_folha	SOUTH	Sao Paulo-based daily newspaper including a
en_USA_nytimes	USA	A major daily known world-wide for its in-dep
en_USA_usatoday	USA	Daily national newspaper known for its concis
en_USA_washingtonpost	USA	Coverage of national, international, and local
en_WLD_reuters	WORLD	Global news from Reuters covering all leading
en_WLD_afp	WORLD	AFP general news wire in English from Agenc
ru_WLD_tass	WORLD	Comprehensive news from Russia, the CIS reg

The salience of *terror* in 2001

The coverage of Sept.11 attacks in selected newspapers and press agencies

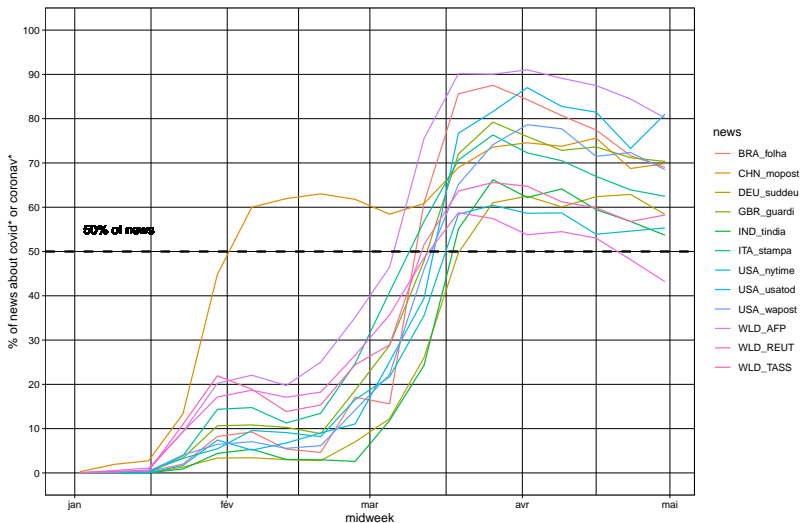
Source : Factiva



The salience of *covid/corona* in 2020

The coverage of covid outbreak in selected newspapers and press agencies

Source : Factiva



Benchmarking of min, max and range

	min_t	max_t	range_t	min_c	max_c	range_c
BRA_folha	1.22	25.58	24.36	0.00	87.54	87.54
CHN_mopost	0.85	30.19	29.34	0.24	75.61	75.37
DEU_suddeu	2.29	36.80	34.51	0.00	62.89	62.89
GBR_guardia	2.09	26.40	24.31	0.00	79.21	79.21
IND_tindia	1.13	16.08	14.95	0.00	66.21	66.21
ITA_stampa	3.92	34.92	31.00	0.00	76.33	76.33
USA_nytime	0.83	22.36	21.54	0.00	60.43	60.43
USA_usatod	0.98	72.53	71.55	0.00	87.01	87.01
USA_wapost	1.47	45.83	44.36	0.00	78.63	78.63
WLD_AFP	3.76	58.68	54.92	0.07	91.03	90.96
WLD_REUT	0.70	32.97	32.27	0.00	58.76	58.76
WLD_TASS	4.10	30.51	26.42	0.04	65.58	65.54
Mean	1.94	36.07	34.13	0.03	74.10	74.07
Std. Dev.	1.29	15.96	15.82	0.07	11.16	11.15
Coeff. Var.	0.66	0.44	0.46	2.42	0.15	0.15

Discussion

- **H1.1 is true** : Whatever the media, the peak of salience is always higher for covid outbreak of 2020 than for terror attack in 2001
- **H1.2 is true** : The decline of salience started one or two weeks maximum after sept. 11 terror attacks in 2001. In the case of covid/corona, we observe rather a plateau with very slow decline (several weeks above 50% of salience).
- **H1.3 is true** : The homogeneity of maxima or range is more important for covid outbreak (Coeff. of variation = 15%) than for sept. 11 attacks (Coeff. of variation = 44%)

Therefore :

- **H1 is validated** : The level of salience reached by the Covid-19 is *exceptional*, compared to the reference event of Sept. 11 ...



Discussion

... but we can really ask if it belongs to the same category of **focusing events** ?

Pros

- Both events has been associated to the narrative "*The world will be changed ...*"
- Both events has been associated to a crisis of international relations and a redistribution of world power.

Cons

- Covid-19 can not be assigned to an event occuring a single day, in a single place with a single cause.
- Covid-19 is characterized by a very different timeline with much more breakpoints than sept. 11 attacks.

Sept.11 attacks as schock model

We use a bayesian estimation for thr research of breakpoint based on the package **mcp**

```
Family: binomial(link = 'logit')
```

```
Iterations: 9000 from 3 chains.
```

```
Segments:
```

```
1: X | trials(N) ~ 1
```

```
2: X | trials(N) ~ 1 ~ 1 + t
```

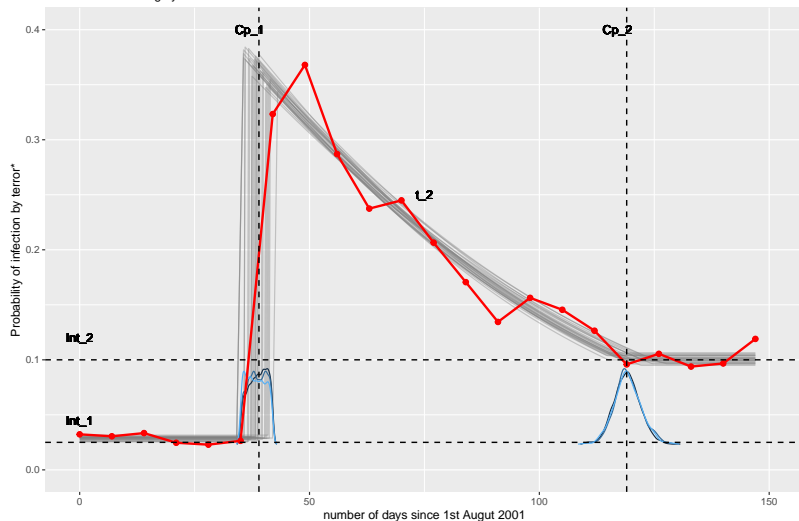
```
3: X | trials(N) ~ 1 ~ 0
```

```
Population-level parameters:
```

name	mean	lower	upper	Rhat	n.eff
cp_1	38.60	35.396	41.983	1	477
cp_2	119.20	113.284	125.338	1	1181
int_1	-3.54	-3.666	-3.396	1	5709
int_2	-0.55	-0.659	-0.449	1	363
t_2	-0.02	-0.022	-0.018	1	554

Sept.11 attacks as shock model

Segmented model of transition with shock
Süddeutsche Zeitung by week



Covid outbreak as logistic model

In the case of covid outbreak, a logistic model seems better adapted and can be fit with only three parameters as the initial level is equal to zero.

Formula: `probability ~ SSlogis(time, Asym, xmid, scal)`

Parameters:

	Estimate	Std. Error	t value	Pr(> t)
Asym	0.59797	0.01092	54.756	< 2e-16 ***
xmid	70.13203	0.77018	91.059	< 2e-16 ***
scal	4.70056	0.67306	6.984	1.6e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

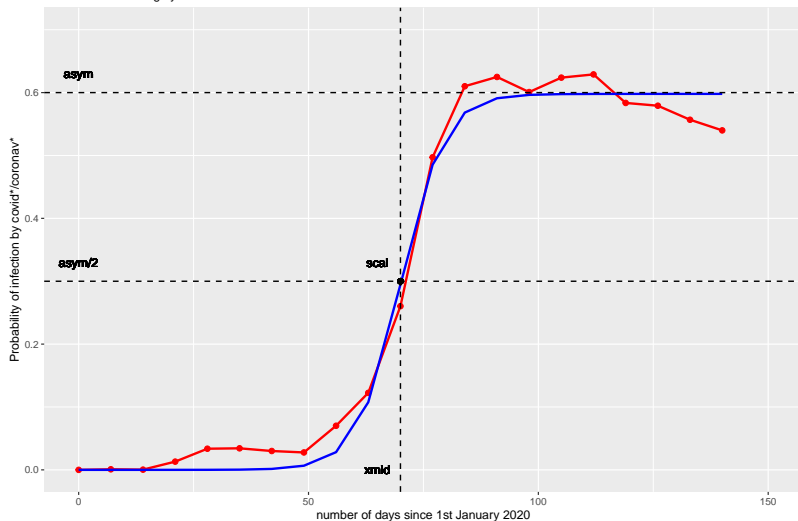
Residual standard error: 0.03092 on 18 degrees of freedom

Number of iterations to convergence: 0



Covid outbreak as logistic model

Logistic model of increase with saturation
Süddeutsche Zeitung by week



Section 6

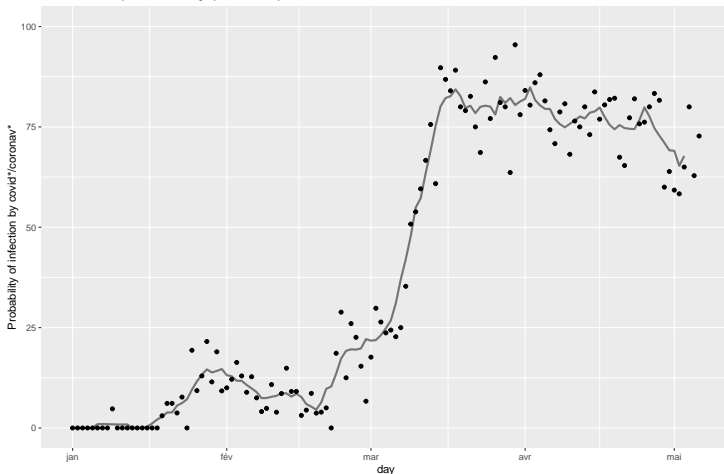
5. A COMPLEX SPREADING EVENT WITH MULTIPLE CAUSALITIES

Looking for discontinuities

- Are they **breakpoints** in this time serie ? How many ? When ?

Süddeutsche Zeitung

black dots = daily observation / gray line = 7-days rollmean



Breakpoints (1): intercept models

The intercept model assume the existence of **levels of media attention** that are constant a period of time but subject to brutal changes after **breakpoints** localised in $(t_1 \dots t_k)$.

It can be formulated as

$$\begin{cases} p_{t_0, t_1} = \alpha_0 \\ p_{t_1, t_2} = \alpha_1 \\ \dots \\ p_{t_{k-1}, t_k} = \alpha_{k-1} \end{cases}$$

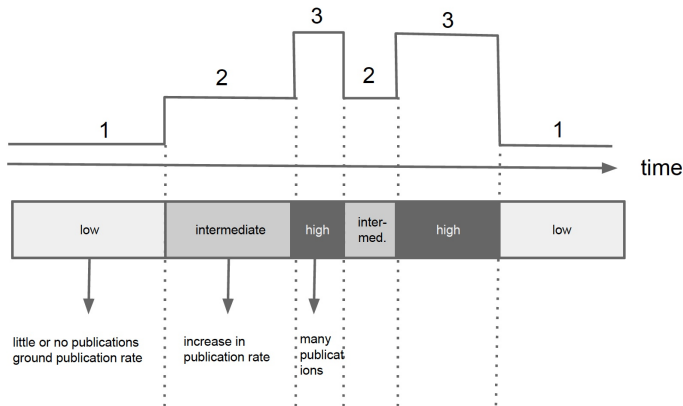
Or as

$$p_t = \alpha_0 + (I_{t > t_1} * \beta_1) + (I_{t > t_2} * \beta_2) + \dots + (I_{t > t_k} * \beta_k)$$

where β_i indicates the changes of the intercept at point i

Breakpoints (1): intercept models

A specific case of intercept models is represented by the family of *Hidden Markov Models (HMM)* where we assume that the number of levels of attention is finite and can be lower than the number of breakpoints.



Source :  ODYCCEUS

Studeny A., Lamarche-Perrin R., Vincent J.M., 2015, "Studying Media

Claude Grasland, Jean-Marc Vincent

Much ado about one (single) thing ?

05/06/2020

68 / 88

Breakpoints (1): intercept models

Hidden Markov Models (HMM) has been applied in particular in ANR Geomedia for the modelisation of ebola outbreak of 2014

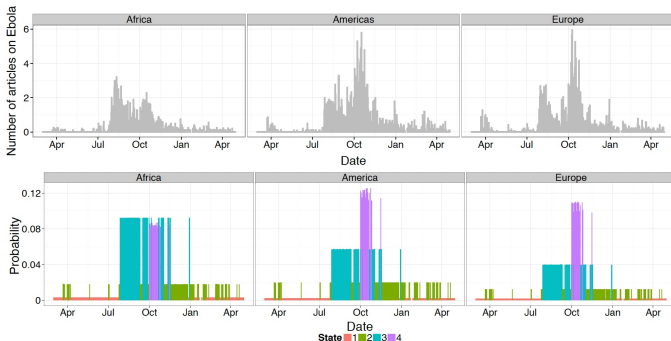
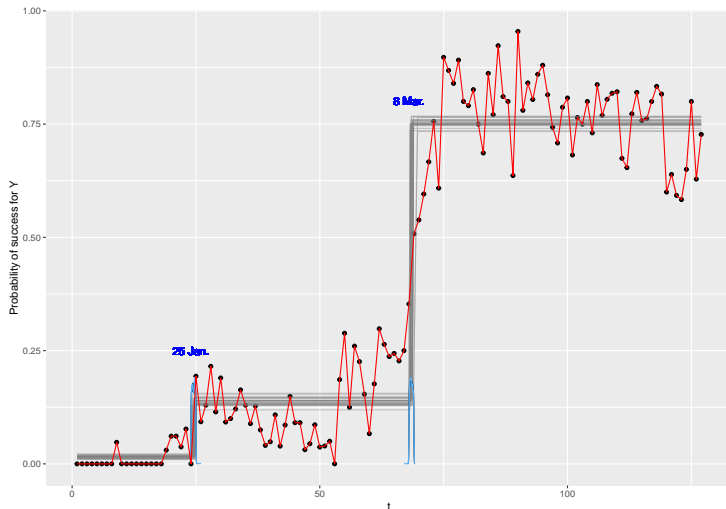


Figure 5.6: The mean number number of articles published daily on Ebola by African, American and European RSS news feeds, respectively (upper panel) and the probabilities of publishing an article on Ebola as predicted by Model II (lower panel), for each of the three possible geographical locations and based on the common sequence of most probable states.

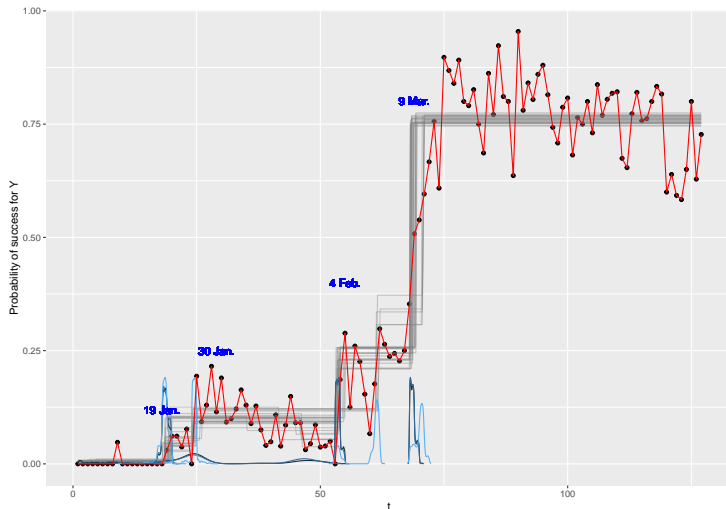
Breakpoints (1): intercept models

Application to *Süddeutsche Zeitung* with $k=3$



Breakpoints (1): intercept models

Application to *Süddeutsche Zeitung* with $k=5$



Breakpoints (2): the slope models

The slope assume the existence of **variations of rates of variation of media attention** after **breakpoints** localised in $(t_1 \dots t_k)$.

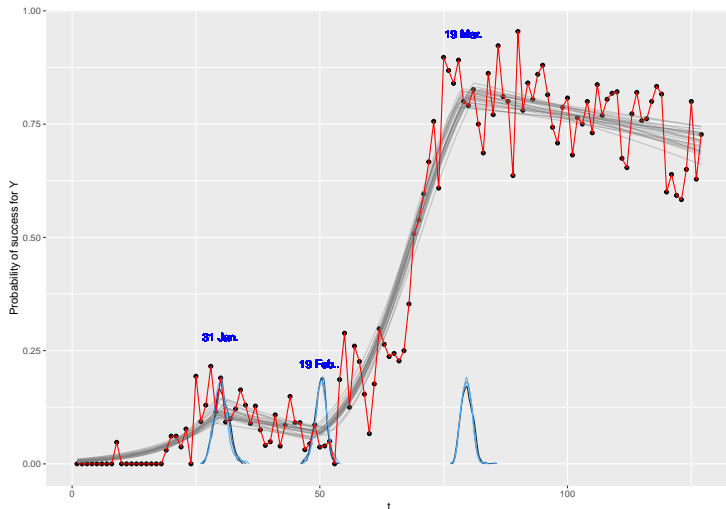
It can be formulated as

$$\begin{cases} p_{t_0, t_1} = \exp(\alpha_0 + \beta_0 \cdot t) \\ p_{t_1, t_2} = \exp(\alpha_1 + \beta_1 \cdot t) \\ \dots \\ p_{t_{k-1}, t_k} = \exp(\alpha_{k-1} + \beta_{k-1} \cdot t) \end{cases}$$

If we assume the hypothesis of continuity, it is possible to replace the parameters $(\alpha_0 \dots \alpha_k)$ by constraints of continuity between segments at the level of breakpoints (**pure slope model**) or to allow brutal changes (**slope+intercept models**)

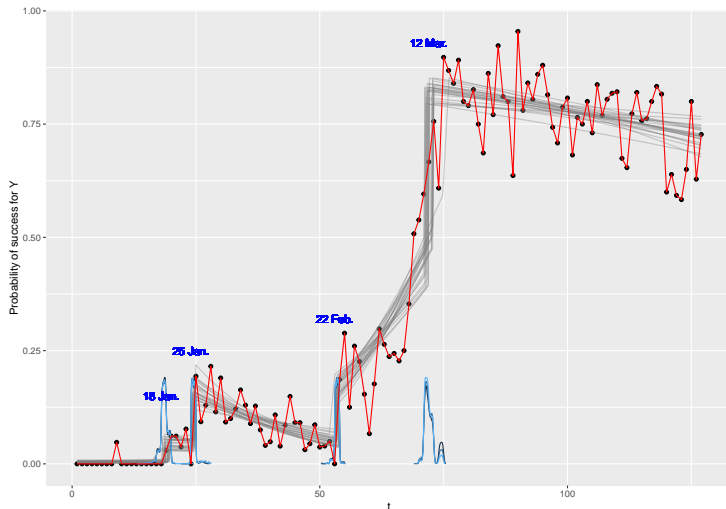
Breakpoints (2): slope

Application of **pure slope model** to *Süddeutsche Zeitung* with $k=4$



Breakpoints (2): slope

Application of **slope+intercept model** to *Süddeutsche Zeitung* with $k=5$




Interpretation of breakpoints

The **inductive approach** has suggested with a high level of probability the existence of potential breakpoints associated to brutal changes in the behaviour of media outlets regarding the covid-19 outbreak. We can now turn to a more **deductive approach** and explore what can be the explanatory variables of the breakpoints revealed by the analysis. We have many potential candidates for such explanation:

- ① influence of global events like declarations of WHO ?
- ② co-variation between number of cases/deaths and number of news ?
- ③ influence of national events like beginning and end of lockdown ?
- ④ spatial diffusion of news between neighbouring countries ?
- ⑤ other ...

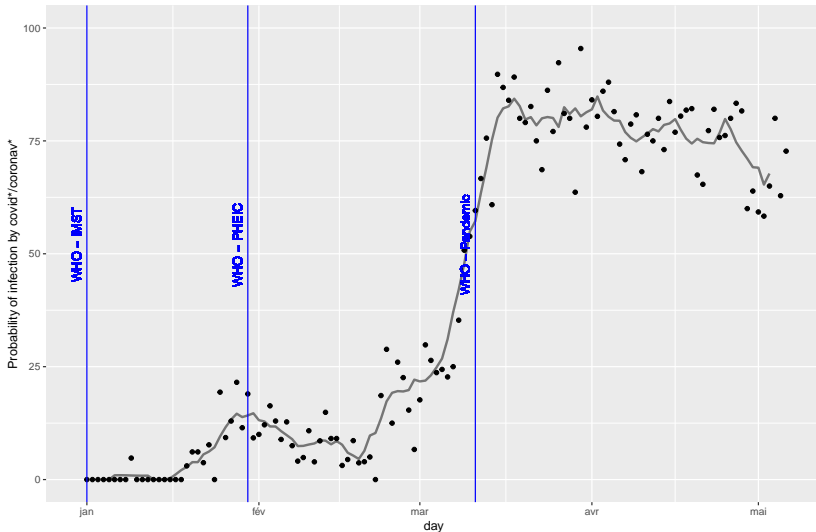
Global events : The declarations of WHO

- **1 January 2020** : Wuhan Municipal Health Commission, China, reported the 31 Dec 2019 a cluster of cases of pneumonia in Wuhan, Hubei Province. A novel coronavirus was eventually identified. As a result, the WHO decided the 1st Jan 2020 to set up the **IMST (Incident Management Support Team)** across the three levels of the organization
- **30 January 2020** : The WHO Director-General reconvened the Emergency Committee (EC). This was earlier than the 10-day period and only two days after the first reports of limited human-to-human transmission were reported outside China. This time, the EC reached consensus and advised the Director-General that the outbreak constituted a **Public Health Emergency of International Concern (PHEIC)**. .
- **11 March 2020** : Deeply concerned both by the alarming levels of spread and severity, and by the alarming levels of inaction, WHO made the assessment that COVID-19 can be characterized as a  **ODYCEUS pandemic**.

Global events : The declarations of WHO

Süddeutsche Zeitung

black dots = daily observation / gray line = 7-days rollmean



Co-variation : number of cases or death reported

Previous observations made on Ebola outbreak suggest that the co-variation between number of cases or death reported and number of news published is not obvious. As in the case of covid-19, the reaction of media from rich countries was very late, and certainly related to the arrival of outbreak in the “North”

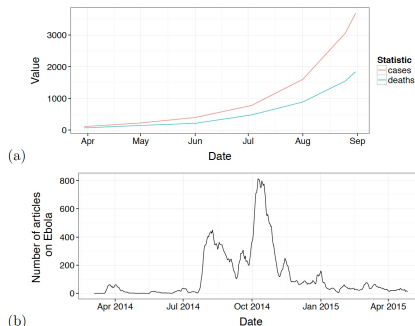
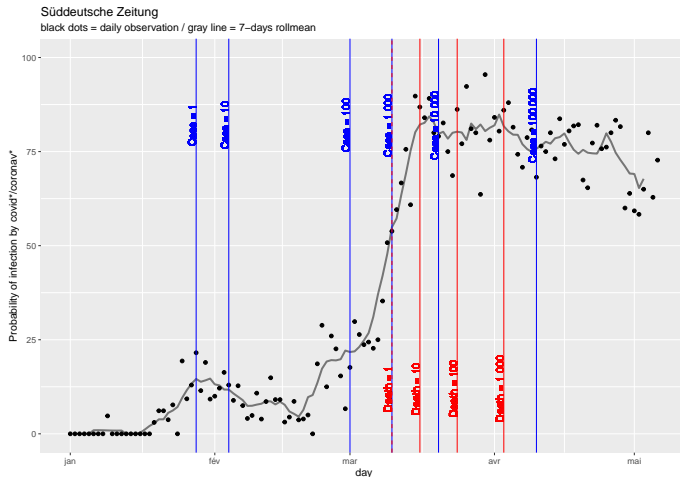


Figure 5.1: (a) Number of Ebola cases and deaths across Guinea, Sierra Leone and Liberia between the 30th of March and the 31st of Aug 2014 (compiled from WHO statistics) and (b) rolling mean across 7 days of the proportion of Ebola related articles from all articles pooled for 39 international RSS news feeds from March 2014 until April 2015.

Co-variation : number of cases or death reported

Number of **cases reported at national level** (Germany in the case of *Süddeutsche Zeitung*) can be a good predictor. We use here voluntary **symbolic threshold numbers** (1, 10, 100, ...)



National political responses : the stringency index

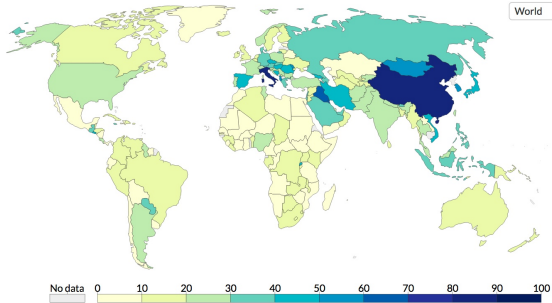
Data related to these topics are not easily to collect and harmonize. Therefore we are particularly grateful to the **Blatavenik School of Government at University of Oxford** that has spend time to collect a world database of particular interest on national political responses from which they have derived a “*Stringency Index*”

COVID-19: Government Response Stringency Index, Mar 11, 2020

The Government Response Stringency Index is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response).

Our World
in Data

World



Source: Hale, Webster, Petherick, Phillips, and Kira (2020). Oxford COVID-19 Government Response Tracker – Last Updated 3rd June.
Note: This index simply records the number and strictness of government policies, and should not be interpreted as ‘scoring’ the appropriateness or effectiveness of a country’s response.
OurWorldInData.org/coronavirus • CC BY

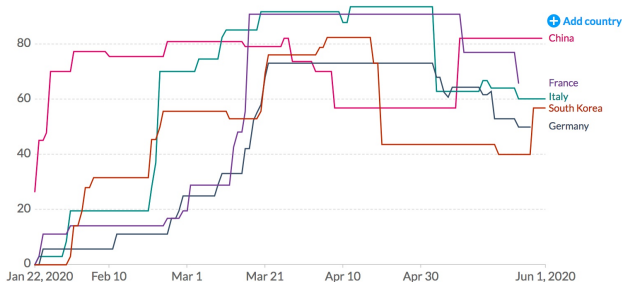
National political responses : the stringency index

The Oxford's index of stringency (OIS) is available for all countries of the world on a daily basis.

COVID-19: Government Response Stringency Index, Jan 22, 2020 to Jun 1, 2020

Our World
in Data

The Government Response Stringency Index is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response).



Source: Hale, Webster, Petherick, Phillips, and Kira (2020). Oxford COVID-19 Government Response Tracker – Last Updated 3rd June.

Note: This index simply records the number and strictness of government policies, and should not be interpreted as 'scoring' the appropriateness or effectiveness of a country's response.

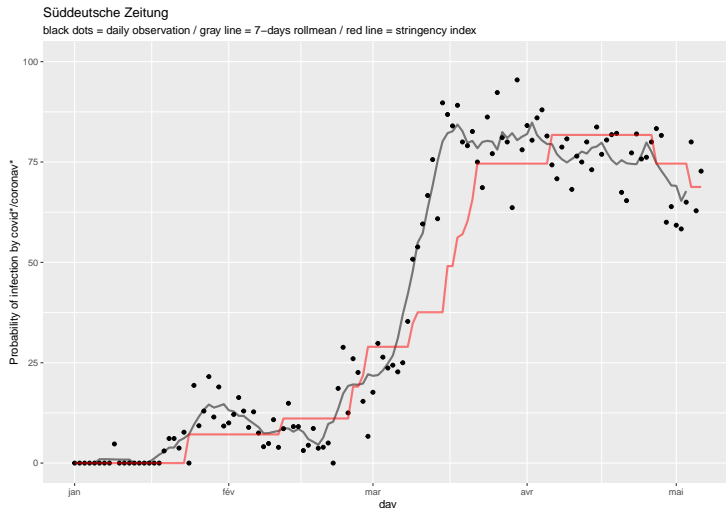
OurWorldInData.org/coronavirus • CC BY

► Jan 21, 2020 Jun 1, 2020



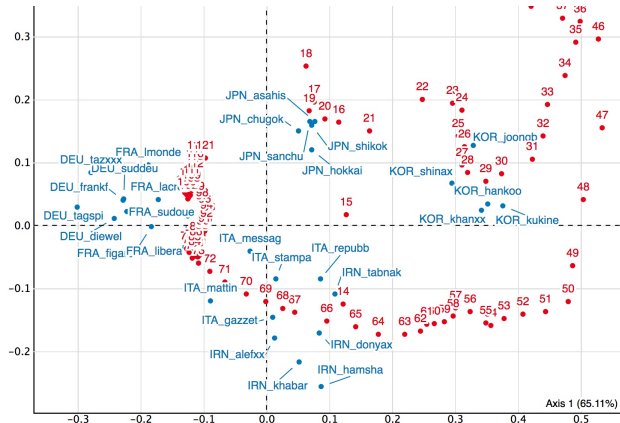
National political responses : the stringency index

Nice covariations can be observed between the OIS and the media outbreak in Germany as in other countries



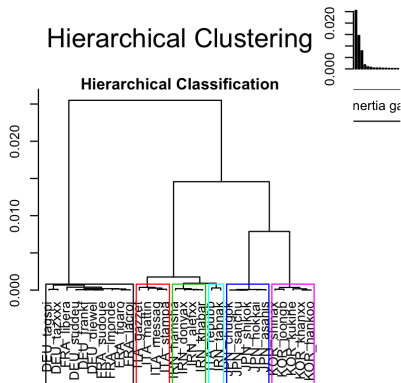
Neighbourhood and network effects

We can finally imagine processus of diffusion of news between media of the same country or media of different countries. Onepossible solution is to realise a **correspondande analysis** assoaited to a hierarchical clustering of timelines of media.



Neighbourhood and network effects

The correspondence analysis confirms the strong coherency of national trends but also the proximity between France and Germany.



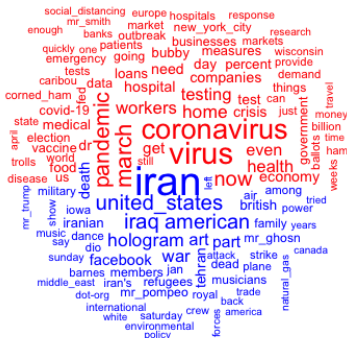
Section 7

CONCLUSION

Media agenda and discursive shift

The semantic content of media has been completely modified by the perturbation. As an example, we have collected a sample of news published the first weeks of January and April in the *New York Times* and compare the specific vocabulary with **quanteda**.

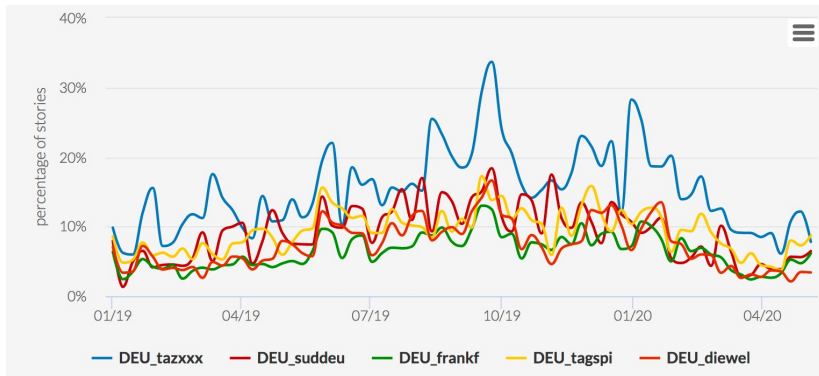
April



Jan

Competition between stories

The coronavirus has destroyed the majority of topics during three months. A good example is the topic of *klima* in German newspapers from 1-1-2019 to present. After a major peak in 2019, the topic collapsed in march-april 2020. What's next ?



Thank you for your attention !

