

# OTMedia - The French Transmedia News Observatory

ina

Nicolas HERVÉ

FIAT/IFTA Media Management Seminar 2019

Stockholm – Garnisonen Conference Centre – May 23+24th

## Research on

- Signal processing
- Automatic content based indexing
- Data mining and Artificial Intelligence
- Semantic web

## ... applied to

- Preservation and restoration
- Documentation and valorisation
- Digital Humanities (history, economics, politics, media and gender studies, ...)

## Team

- Computer Science
- 15 researchers and engineers
- PhD students
  - > Face Recognition
  - > Speaker Recognition
  - > Analysis of Twitter
  - > Interactive image classification
  - > Graph visualization

**<http://research.ina.fr>**

# OTMedia main objectives

---

## Transdisciplinary Research

- Human Sciences
- Computer Sciences
- French News Ecosystem
- Quantitative studies on huge datasets

## Some questions we are trying to answer

- How does news spread today?
- What is the Internet's or Twitter's role compared with that of traditional players such as television, radio and the press?
- Who starts a media “buzz”?
- Who are the players involved?
- What place do images occupy in the media?

# OTMedia technical platform

---

## Software platform

- 24/24 - 7/7
- Collect
- Automatic analysis
- Indexing
- Monitoring

## Streams

- TV
- Radio
- News agencies
- Online press, news sites, blogs
- Twitter

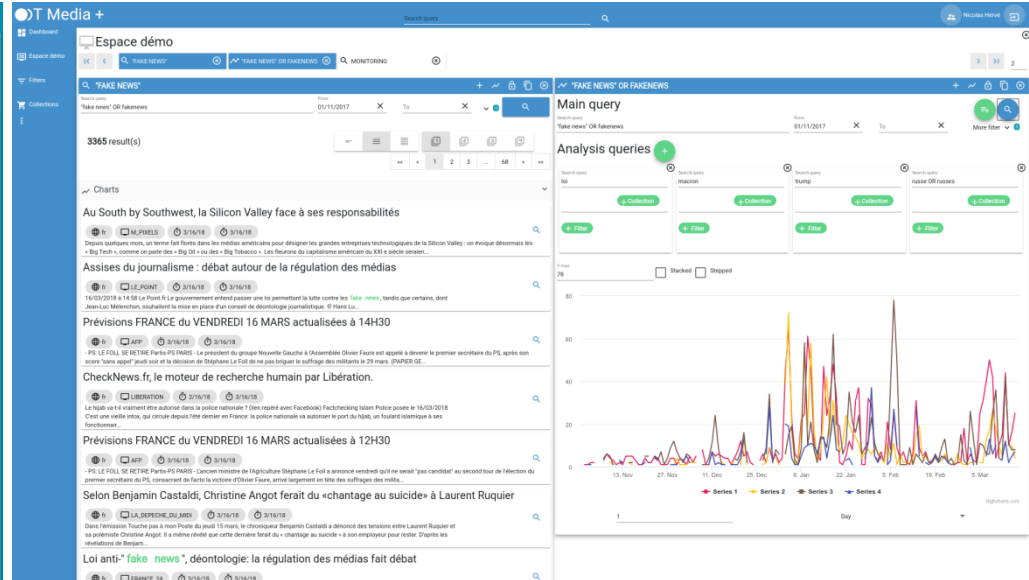
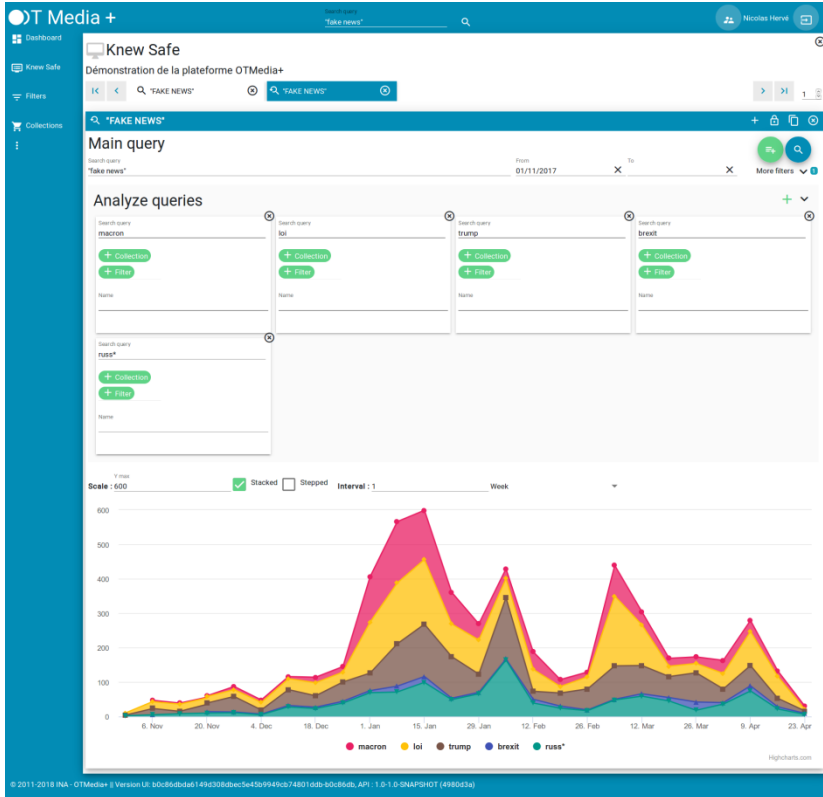
## Tools

- HTML parsing
- Named entities recognition
- Quote detection
- Reference detection
- Keyword extraction
- Categorization
- Plagiarism detection
- Event detection
- Speech-to-text
- Image indexing
- ...

# Towards interactive data mining – search UI

The screenshot displays the OT Media + search interface. At the top, there's a search bar with the query "FAKE NEWS" and a date range from 01/11/2017 to the present. The interface shows 3365 results. On the left, there are navigation options like Dashboard, Espace démo, Filters, and Collections. Below the search bar, there are several charts and filters. The "Publication date" chart shows a significant increase in activity starting in late January 2018. The "Document type (3)" chart shows that most results are "online". The "Issue name (10)" chart shows "NEXT\_INPACT" as a prominent source. The main content area displays search results for various news items, including articles about "Au South by Southwest, la Silicon V...", "Assises du journalisme : débat auto...", "Prévisions FRANCE du VENDREDI 1...", "CheckNews.fr, le moteur de recher...", "Selon Benjamin Castaldi, Christine...", "Loi anti- 'fake news', déontologie: ...", and "Mediamass, ce site où les faux arti...". Each result includes a snippet of the article and a small thumbnail image.

# Towards interactive data mining – timeline UI



# Towards interactive data mining – Document UI

## Assises du journalisme : débat autour de la régulation des médias

© 2018-2019 OTMedia - 100 rue de la République - 75001 Paris - France

Entres

0 Tags 0 Catégories 1 Highlight 3 Organisations 1 Locations

oo 4 Lit



16/05/2018 à 16:58 Le Point.fr  
Le gouvernement entend passer une loi permettant la lutte contre les fake news, tandis que certains, dont Jean-Luc Mélenchon, souhaitent la mise en place d'un conseil de déontologie.  
Journaliste - @ Hans Lucas/RFI - Robin Jeffrey  
Abonnez-vous



Cette année, les 17es assises du journalisme à Tignes sont lieu de grands débats, sur la régulation des médias et depuis plusieurs mois, est au cœur de toutes les discussions et tensions. Une loi anti-fake news va-t-elle nuire à la liberté d'information ? Les dévies des journalistes doivent-elles être sanctionnées par un conseil de déontologie ? Les options divergent... Mais, la mission de la Culture est intervenue à Tignes afin de prendre la défense de la proposition de loi qui vise à régir les fake news en période d'élection, face aux critiques émanant de la sphère journalistique contre un projet de loi jugé sans responsabilité, sans transparence et « liberticide ».

Le chef français n'est plus adapté. Il ne permet pas d'être assés vite de façon assez systématique « contre la diffusion de fausses nouvelles via Facebook, YouTube ou Twitter », a ainsi plaidé François Nuyens, en assurant que « l'acte estival » « responsabilise », ces plateformes. Elle répondrait aux critiques d'organisations comme le Syndicat national des journalistes (SNJ), premier syndicat de la profession, qui a réclamé l'abandon d'un texte « potentiellement liberticide » qui pourrait « devenir un moyen d'entraver le travail des journalistes... » C'est difficile et journalistiquement dangereux, car on ne peut pas verser quelque chose qui pourrait nuire à la confiance », a affirmé lors des assises Vincent Lamine, premier secrétaire national du syndicat. Lui-même pour l'appeler les fake news le développement de l'information - une solution également défendue par le gouvernement - et un renforcement des moyens des rédactions.

« On veut toujours créer un problème de censure, et de toute façon on sera dépassé », a répondu Jean-Bernard Chevret, cofondateur du média numérique France 3. Mais il reconnaît une vertu à ce texte : « il pose le débat sur la place publique de manière très visible » et porte du « poids » la responsabilité des plateformes. Vers la création d'un conseil de déontologie ?

Le chef de file des insoumis, Jean-Luc Mélenchon, qui se dit régulateur la victime du journalisme travaillant à charge, et a tenu récemment des propos extrêmement virulents contre les médias, défend la mesure depuis des mois l'absence d'un conseil de déontologie des médias, un « tribunal professionnel » pour permettre aux citoyens de faire respecter le droit à une information objective et à une presse libre et indépendante.

Il a aussi dénoncé son positionnement en tant que « coup de force de 100 000 organisations. Références des profanes.

LOD défend, la lutte d'une instance indépendante du gouvernement comme le CSA, qui n'est pas de « pouvoir de sanctionner » ou jouer un rôle pédagogique, en portant au jour les délits et en encourageant les bonnes pratiques. Une forme d'autogestion que le SNJ voit d'un bon œil. « Il faut qu'on avance sur ce sujet dans la profession, sinon d'autres s'en chargeront à notre place. À nous de construire dans les années qui viennent ce qui pourrait être une instance de régulation, peut-être adossée à la CCJF (l'organisme professionnel qui délivre les cartes de presse) et que le public pourrait aller jusqu'à s'en servir ou une fausse déontologie », a estimé Laurent Lamine.

Revue nationale à l'initiative, l'agence LOD, les pouvoirs de presse qui se sont réunis pour le projet de loi des médias indépendants. Un rapport remis en 2014 au ministre de la Culture avait déjà donné du grain à moudre aux partisans d'une telle instance, défendue depuis 2004 par une association, l'APSP. De son côté, l'ONG Reporters sans frontières, qui regroupe aussi « journalistes » et la proposition de l'OCJ parlementaire le 3 avril au projet de « système de certification des médias » européens. Un dispositif qui désolait les médias respectant des normes rigoureuses.

Agrandir le texte

ISSUES

Context

- LE\_POINT\_BSL\_LE\_POINT\_POLITIQUE
- LE\_POINT\_BSL\_LE\_POINT\_ENLIGNEMENT1
- LE\_POINT\_BSL\_LE\_POINT\_ENLIGNEMENT2

Settings

Show document details

Debug

## Le pape en Birmanie et Bangladesh en plein drame des Rohingyas

24/11/2017 04:12:01 AFP 7892182-6967-4415-a090-8405963301e newsimmd.unnewsim.afp.com.20171124.doc.uk4g-1 1 1 1

Entres

1 People 4 Organizations 4 Locations

Ang les les les

8 Catégories 7 Tags 0 Uti 1 Highlight

Text Videos



Transcription

C'est la rencontre d'un pape catholique birman Dans quelques jours ils assisteront à une messe célébrée par le pape François en personne, une première historique dans ce pays à majorité bouddhiste où leur communauté ne représente qu'1 pourcent de la population C'est le chm et bien en actualité écrites, on a fait de la rencontre, on a vu qu'à la tête je vois TV5 les Verts il y a l Cette vidéo historique intervient alors que l'attention de la communauté internationale et bariégée sur la Birmanie et son traitement de la communauté agitée des Rohingyas pour échapper aux violences, ces musulmans sont depuis le mois d'août, plus de 100000 à avoir fui vers le Bangladesh voisin, les Nations-Unies parlent de nettoyage ethnique et l'opinion birmane a mal pris ces critiques Devant ce même genre j'ai pris pour La majorité des gens en Birmanie ne croit pas la version de la communauté internationale sur les abus contre les Rohingyas et le nombre de réfugiés au Bangladesh, cette version il est juste pas acceptée par les birmanes Je pense donc que si le pape vient et de des choses fortes sur le sujet, cela pourrait enflammer les tensions et le sentiment du public avec la Le souverain pontife doit rencontrer la dirigeante birmane Aung San Suu Kyi et le chef de l'armée avant de continuer son voyage vers le Bangladesh, là aussi, les catholiques sont une petite minorité et les attentats sont graves Il moi de chez Total, la Gaumont d' arrêter le moulin que à 4 On pense qu'il pourrait parler de la communauté des Rohingyas, on ne s'attend qu'à des mots de paix et d'harmonie Peut-être mourir à petit la chanelle que l'U2 le François a fait de l'harmonie entre les religions, une des priorités de son pontificat, mais la tâche est grande, un accord entre les 2 pays vient d'être signé, prévoyant un retour des Rohingyas en Birmanie d'ici 2 mois, mais les détails restent vagues et beaucoup ne quitteront pas le Bangladesh, sans garantie de sécurité

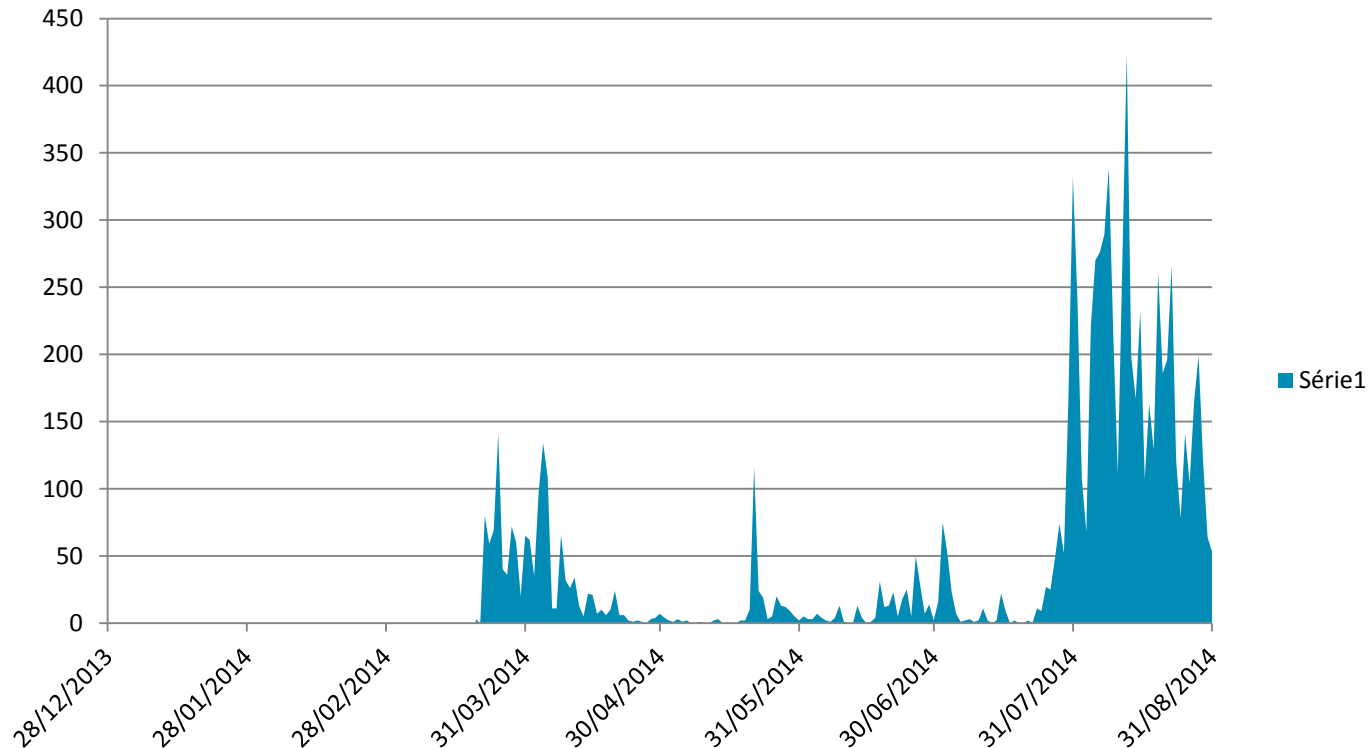


# Towards interactive data mining – Visual query

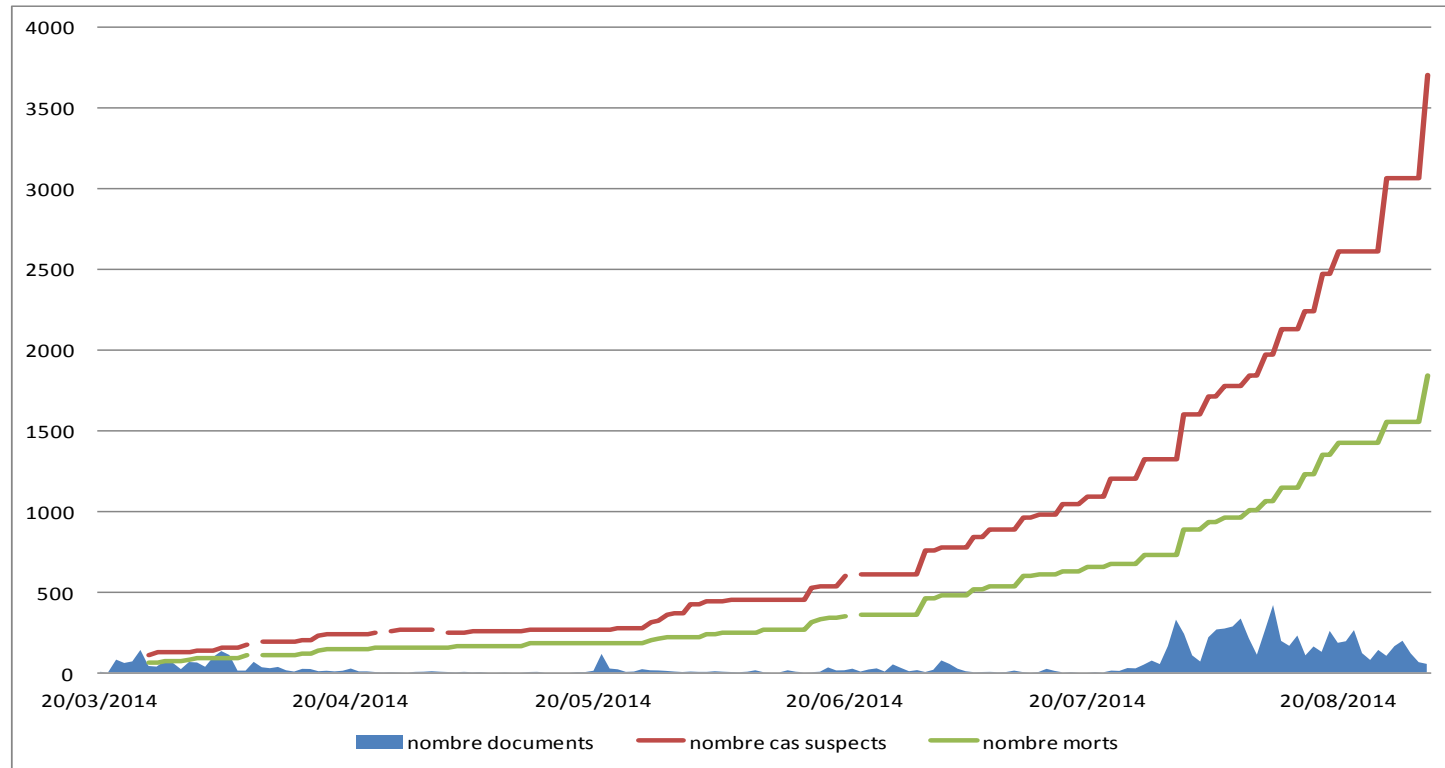




# Towards interactive data mining – Ebola example



# Towards interactive data mining – Ebola example



## Object emergence: media events

---

One of the fundamental tasks of data mining is to group and label sets of "close" items in order to bring out "higher level" objects.

OTMedia : detect, circumscribe, monitor, measure, analyze and study the spread of large-scale media events.

# Object emergence: detection of visual events

2 to 2011-09-19 to 2011-09-26 to **2011-10-03 to** 2011-10-10 to 2011-10-17 to 201  
-19 2011-09-26 2011-10-03 **2011-10-10** 2011-10-17 2011-10-24 20

**2011-10-03 to 2011-10-10**

Mort de Steve Jobs : les fans d'Apple en deuil.

keywords : [ Décès ] [ Steve Jobs ] [ Apple ] [ USA ]

Steve Jobs  
1955-2011

16 to 2011-10-03 to 2011-10-10 to **2011-10-17 to** 2011-10-24 to 2011-10-31 to 201  
03 2011-10-10 2011-10-17 **2011-10-24** 2011-10-31 2011-11-07 20

**2011-10-17 to 2011-10-24**

Le régime de Mouammar Khadafi aura duré 42 ans. Un régime inauguré à la fin des années 60 par un jeune officier qui deviendra un dictateur redouté dans son pays et au delà des frontières libyennes, notamment en raison de son soutien au terrorisme.

keywords : [ Libye ] [ Mouammar Khadafi ] [ OTAN ] [ FRA ]

201


26 2011-10-03 2011-10-10 **2011-10-17** 2011-10-24 2011-10-31 20

**2011-10-10 to 2011-10-17**

Le débat d'entre-deux tours a lieu mercredi entre les finalistes de la primaire. » VIDÉO - Hollande et Aubry vont s'affronter au second tour de la primaire PS.

keywords : [ Présidentielles ] [ François Hollande ] [ PS ] [ FRA ]

Cluster - 427



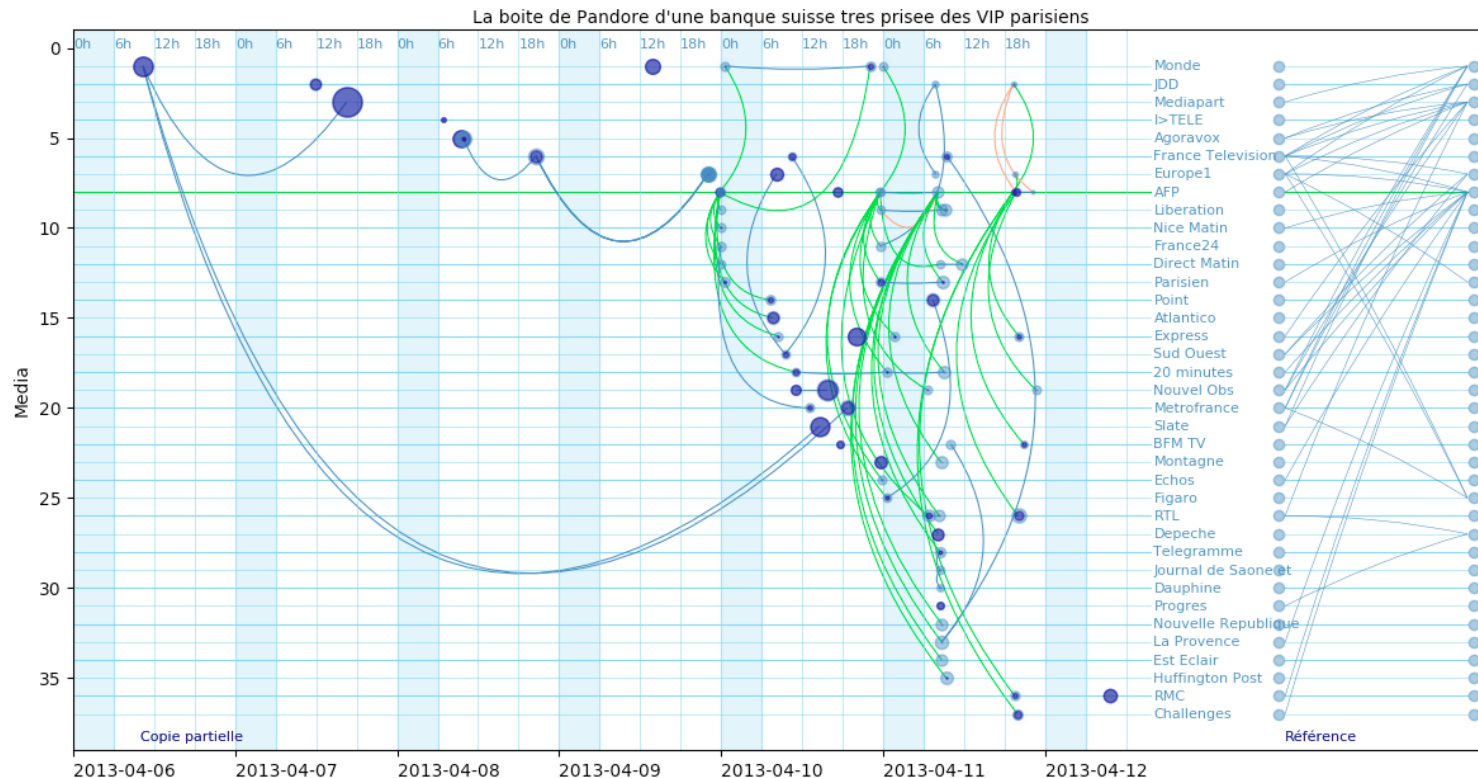
Zoom 36 / 47 files

AAA  AAB  AB  BB  Composite  Demo  Graphic  Illustration  Logo  News

Person  Pub  Texte

Web UI - 1 Web UI - 2

# Object emergence: detection of textual events



# Economic study of the French news ecosystem



## The Production of Information in an Online World: Is Copy Right?\*

Julia Cagé<sup>1</sup>, Nicolas Hervé<sup>2</sup>, and Marie-Luce Viaud<sup>2</sup>

<sup>1</sup>Sciences Po Paris and CEPR

<sup>2</sup>Institut National de l'Audiovisuel

December 2017

# Economic study of the French news ecosystem

→ Figure 4

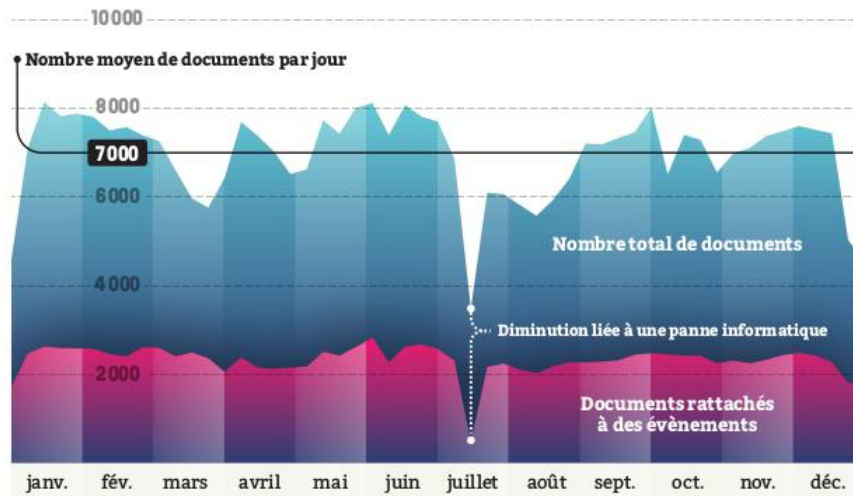
## UN CORPUS COMPOSÉ DE 86 MÉDIAS D'ACTUALITÉ



→ Figure 5

## UNE BASE DE DONNÉES DE PLUS DE 2,5 MILLIONS DE DOCUMENTS DONT 1/3 CLASSÉS DANS DES ÉVÉNEMENTS

Nombre de documents par jour



Un événement ici est un événement médiatique. Les événements sont définis à partir de notre algorithme de détection d'évènements, décrit en détails dans le texte, qui nous a permis de mettre ensemble tous les articles se rapportant au même « fait ».



# Economic study of the French news ecosystem

FRANCE 13

## Le Monde

### COMMENT LA NSA ESPIONNE LA FRANCE

Publication : 21/10/2013 06:03:41 - Taille : 1371 - Copie : 1,97%

Le Monde 13 21 10 2013 06:03:41 Des Jacques Follorou et Glenn Greenwald (Journaliste)

L'avenir dira peut-être, un jour, pourquoi Paris est resté si discret, par rapport à Berlin ou Rio après les révélations sur les programmes d'espionnage électronique américain dans le monde. Car la France a été tout autant ciblée et dispose aujourd'hui de preuves tangibles que ses intérêts sont quotidiennement visés.

## Comment la NSA espionne la France

Le Monde 9 | 21 10 2013 | 08:01 • Mis à jour le 27 05 2014 à 11:57 |

Par Jacques Follorou et Glenn Greenwald (Journaliste)



### LA NSA A RECULÉ DES MILLIONS DE DONNÉES EN FRANCE ( PRESSE )

Publication : 21/10/2013 06:03:41 - Taille : 1371 - Copie : 08,85%

## AFP

### LA NSA A RECULÉ DES MILLIONS DE DONNÉES EN FRANCE ( PRESSE )

Publication : 21/10/2013 06:03:41 - Taille : 1371 - Copie : 40,48%

L'agence américaine de renseignement, NSA, a intercepté de façon massive des communications téléphoniques des citoyens français, révèle-t-il dans le quotidien Le Monde, citant des documents de l'ancien membre de la NSA, Edward Snowden.

## Le Monde

### COMMENT LA NSA ESPIONNE LA FRANCE

Publication : 21 09 2013 06:03:41 - Taille : 1371 - Copie : 46,34%

Le Monde 13 21 09 2013 06:03:41 Des Jacques Follorou et Glenn Greenwald (Journaliste)

L'avenir dira peut-être, un jour, pourquoi Paris est resté si discret, par rapport à Berlin ou Rio après les révélations sur les programmes d'espionnage électronique américain dans le monde. Car la France a été tout autant ciblée et dispose aujourd'hui de preuves tangibles que ses intérêts sont quotidiennement visés.

## AFP

### LA NSA A RECULÉ DES MILLIONS DE DONNÉES EN FRANCE ( PRESSE )

Publication : 21/10/2013 06:03:41 - Taille : 1371 - Copie : 74,36%

L'agence américaine de renseignement, NSA, a intercepté de façon massive des communications téléphoniques des citoyens français, révèle-t-il dans le quotidien Le Monde, citant des documents de l'ancien membre de la NSA, Edward Snowden.

L'agence américaine de renseignement, NSA, a intercepté de façon massive des communications téléphoniques des citoyens français, révèle-t-il dans le quotidien Le Monde, citant des documents de l'ancien membre de la NSA, Edward Snowden.

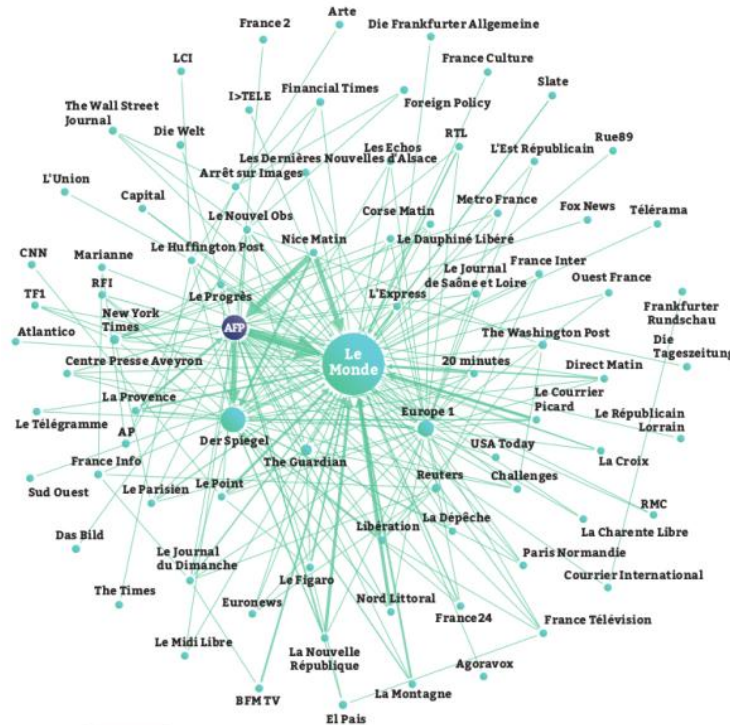
L'agence américaine de renseignement, NSA, a intercepté de façon massive des communications téléphoniques des citoyens français, révèle-t-il dans le quotidien Le Monde, citant des documents de l'ancien membre de la NSA, Edward Snowden.

L'agence américaine de renseignement, NSA, a intercepté de façon massive des communications téléphoniques des citoyens français, révèle-t-il dans le quotidien Le Monde, citant des documents de l'ancien membre de la NSA, Edward Snowden.

L'avenir dira peut-être, un jour, pourquoi Paris est resté si discret, par rapport à Berlin ou Rio après les révélations sur les programmes d'espionnage électronique américain dans le monde. Car la France a été tout autant ciblée et dispose aujourd'hui de preuves tangibles que ses intérêts sont quotidiennement visés.

Selon les documents de l'Agence nationale de sécurité (CNSA) obtenus par Le Monde, les communications téléphoniques des citoyens français sont, en effet, interceptées de façon massive. Ces pièces, dévoilées en juin par l'ex-consultant de l'agence, Edward Snowden, décrivent les techniques utilisées pour capter illégalement les secrets ou la simple vie privée des Français. Certains éléments ont été évoqués par l'hebdomadaire allemand Der Spiegel et le quotidien britannique The Guardian. D'autres sont inédits.

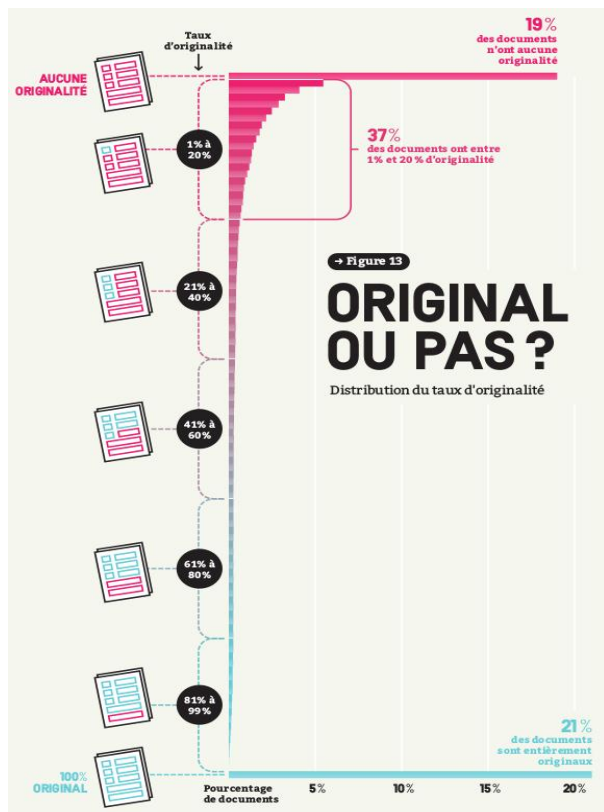
# Economic study of the French news ecosystem



→ Figure 16

LE MONDE, SNOWDEN ET LES ÉCOUTES DE LA NSA

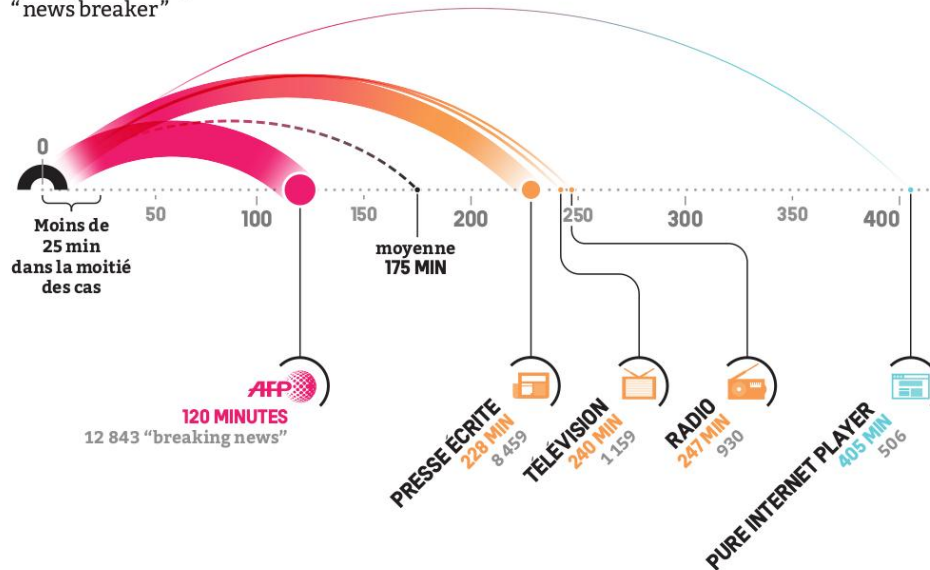
# Economic study of the French news ecosystem



→ Figure 11

## MOINS DE 3H EN MOYENNE POUR QU'UNE INFORMATION SORTIE PAR UN MÉDIA SOIT REPRIS SUR LE SITE INTERNET D'UN CONCURRENT

Temps de réaction en minutes selon le support hors ligne du média  
"news breaker"



# Economic study of the French news ecosystem

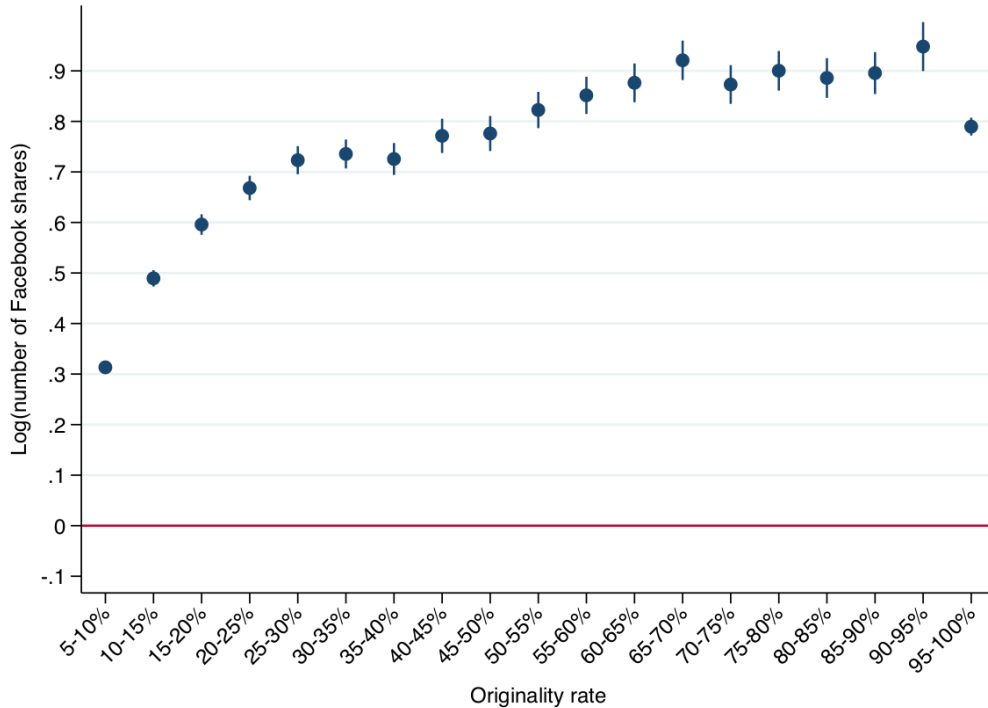
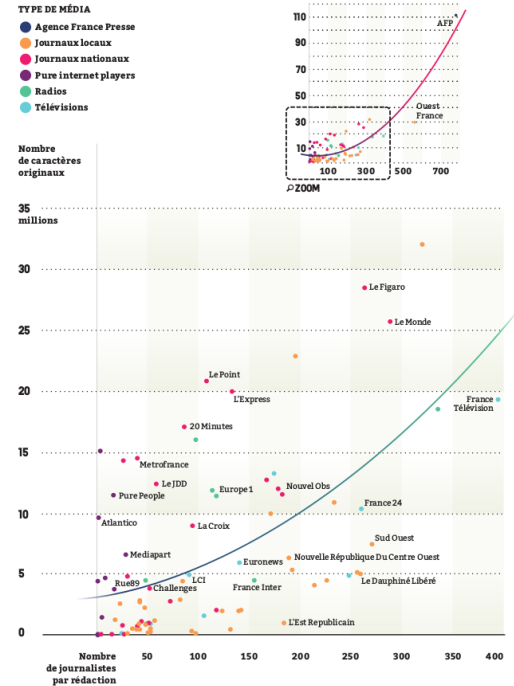


Figure 19 E

## PRODUCTION D'INFORMATION ET TAILLE DES RÉDACTIONS

Contenu original et taille des rédactions





# Preliminary study – TV and radio vs. Twitter

17:59 • DIRECT  
AVENUE DES CHAMPS-ÉLYSÉES (PARIS)  
7 JOURS BFM

PORTES OUVERTES  
HIGH RISE

BFM TV  
LES CHAMPS-ÉLYSÉES SACCAGÉS.

00:00:02.92 / 02:00:00.00

Marche pour le climat : 45 000 personnes ont manifesté à Paris, selon d'Occurrence pour plusieurs médias dont BFM TV

VOCAPIA

HISTO-Gilets jaunes

HISTO-Marche climat

SEGMA-Marche climat

SEGMA-Gilets jaunes

5 éléments.

## Transcription



Et Stanislas Gaudon, secrétaire national syndicat Alliance police nationale, j'espère que je n'ai oublié personne, merci à tous, on va tout de suite contre la direction

00:00:00

00:00:09



Des Champs-Élysées rejoindre Fanny Régnaud

00:00:09

00:00:16



Effectivement, vous le voyez cette avenue des Champs-Élysées et c'est très nombreuses enseignes pieds ici, nous sommes devant le magasin d'une célèbre marque de jeans, vous le voyez, les pillards allez les Black Blocks ceux notamment qui ont, qui se sont engouffrés et qui ont ravagé cette boutique, vous voyez les mannequins par terre, des mannequins qui se trouve initialement dans les magasins, il y en a un petit peu tout le long de cette avenue des Champs-Élysées, parce que c'est ça devient difficile, finalement, de trouver des boutiques qui n'ont pas été abîmées ici sur les Champs, des dizaines et des dizaines d'enseignes de vitrine complètement ravagé, hein, vous le voyez, là c'est une autre enseigne de cosmétiques qui est en train de filmer Alexia à faire un petit peu plus loin, une boulangerie ça continue comme ça, sur quasiment totalité de cette avenue des Champs-Élysées, cette avenue des choses Élysée où le calme est revenu depuis plusieurs minutes, maintenant il y a un petit peu tout

00:00:16

00:01:13



Il y a des gilets jaunes, il y a aussi beaucoup de passants tout simplement queer emprunte à nouveau cette avenue des Champs-Élysées, les forces de l'ordre ont copieusement ratissé tout le bas de l'avenue, les casseurs se sont retranchés dans les petites rues adjacentes, il y a aussi la des dégâts des voitures notamment qui ont été incendiés par endroits, donc le calme qui est momentanément revenu en tout cas on espère que cela va durer, désormais comme ça, mais c'est difficile de se rendre compte de l'évolution de la situation en tout cas, regardez ces dégâts extrêmement importants ici sur cette avenue des Champs-Élysées, finalement, toutes les boutiques qui n'avaient pas protégé leurs vitrines à l'aide de planches bien ont été ravagés et donc partiellement pillé la vous le voyez encore une qui est bien, par endroits, a été complètement vidée de sa marchandise ici donc une marque de vêtements et ça continue comme ça, sur l'essentiel de cette avenue des Champs-Élysées

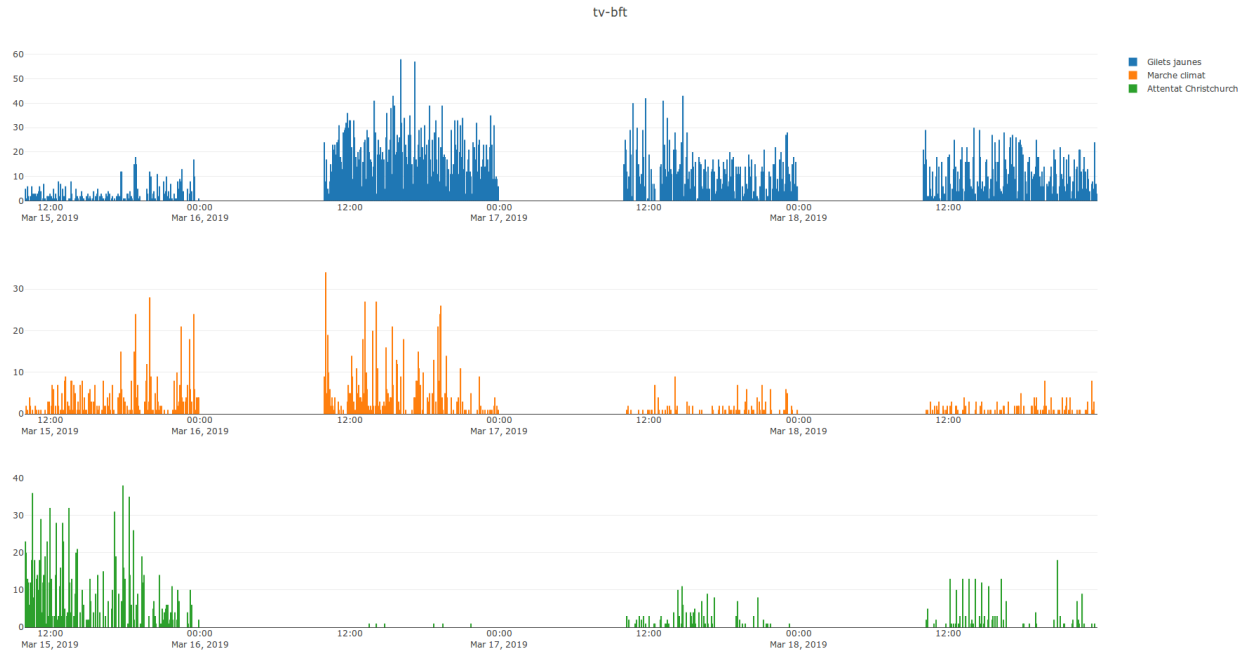
00:01:13

00:02:14

# Preliminary study – TV and radio vs. Twitter

[Channels - Terms](#)

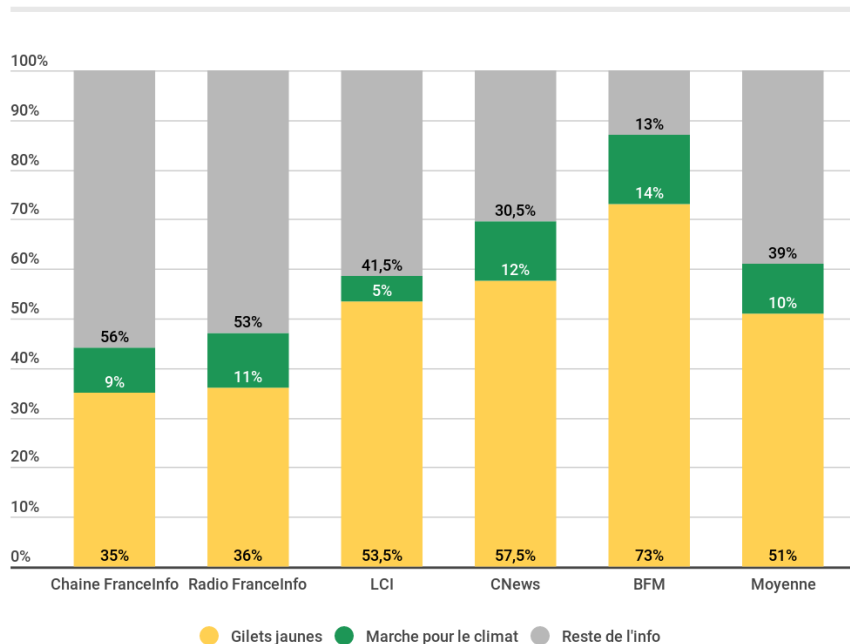
| [Gilets jaunes](#) | [Marche climat](#) | [Attentat Christchurch](#) | [tv-bft](#) | [tscie](#) | [tsc28a](#) | [tsc3a](#) | [radiofr](#)



Preliminary results, figures are not yet fully validated

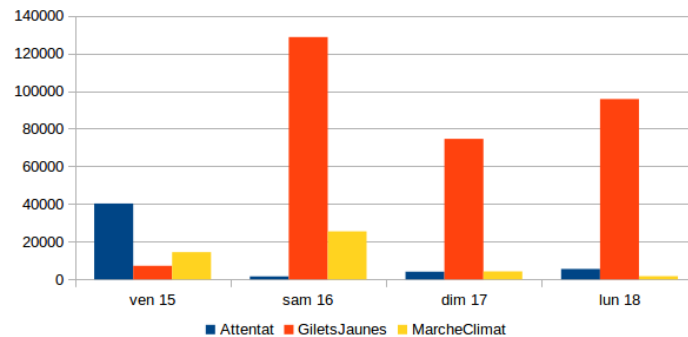
# Preliminary study – TV and radio vs. Twitter

## Samedi 16 mars, les manifs "gilets jaunes" monopolisent l'antenne

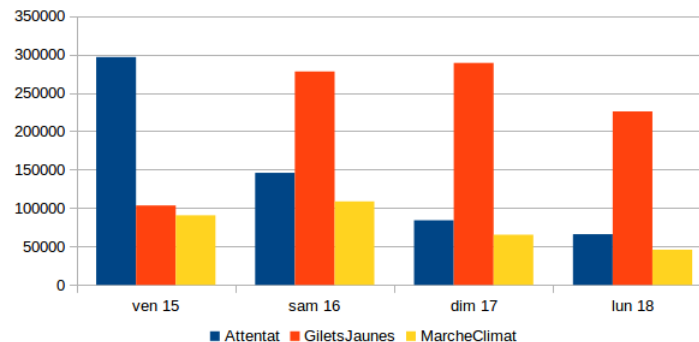


Preliminary results, figures are not yet fully validated

## Chaînes d'info



## Twitter





# Ongoing and future work

---

## News propagation

- Link between TV, radio, Press and Twitter
- Image propagation
- Fake News propagation

## Technically

- Stream automatic segmentation and categorization
- Graphical models features extraction
- Semantic models improvement

# Thanks

---

[nherve@ina.fr](mailto:nherve@ina.fr)