



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB



# *Twitter Motivation*

## *What we can do with Twitter data*

***Paolo Nesi, Gianni Pantaleo, Imad Zaza***

Distributed [Systems and internet | Data Intelligence and] Technologies Lab  
Dipartimento di Ingegneria dell'Informazione (DINFO),  
Universita' degli Studi di Firenze

Fax: 0039-055-2758570, tel: 0039-3355668674

<http://www.disit.dinfo.unifi.it> , <http://www.km4city.org>

<http://www.disit.org/tv>

<http://www.disit.org/rttv>

***Twitter Vigilance***



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

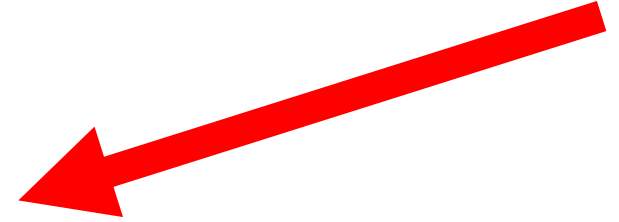
# Agenda

1. Introduction: What is Twitter
2. Overview on Twitter Data Capabilities
3. Twitter Vigilance how it is working
4. Exercitation Work Plan



# Agenda

**1. Introduction: What is Twitter**



**2. Overview on Twitter Data Capabilities**

**3. Twitter Vigilance how it is working**

**4. Exercitation Work Plan**



# Twitter Applications

- Social media analysis is becoming a very important instrument to monitor communities, users' preferences, and to make predictions
- Twitter is one of the most widespread microblogs allowing users to have a personal news feed and followers attached to it

# Introduction: What is Twitter ?

The evolution and growth of Social Networks and Media has recently brought new horizons and challenges for designing smart tools for event monitoring, providing also predictive capabilities

Microblogging services like **Twitter** have been proven to be an inestimable source of *continuously, instantly updated information*, useful for different purposes



Many different contexts and application areas:

- **Collect users' information about quality of services**
- **Event Monitoring** - crowd size estimation, voting results, predicting TV audience etc.
- **Early Warning** - monitoring critical situations for alerts providing (weather alerts, spread of contagious diseases, natural disasters etc.)



# Twitter in Numbers

- 350,000 Tweets are sent every minute
- In 2016, Twitter has 310 million monthly active users (*almost the same as the U.S. population*)
- A total of 1.3 billion accounts have been created
- Of those, 44% made an account and left before ever sending a Tweet

Followers receive some notification connected to the actions performed by the users they follow.



Hashtags # represent codified keywords

Twitter Users, Mentions and Citations @

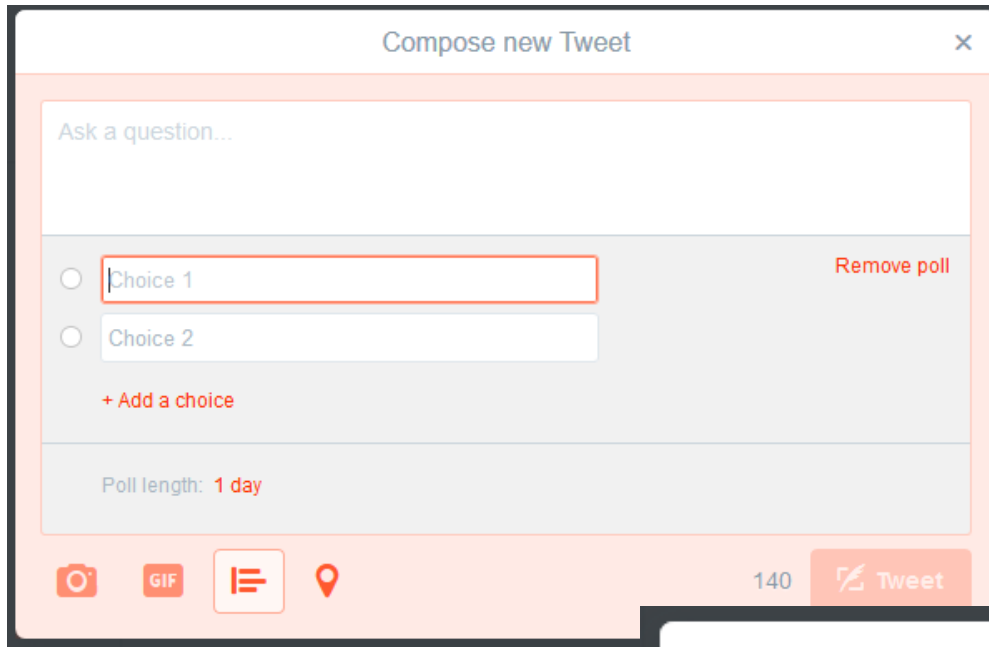
Typical actions:

- ❖ *Compose a new tweet*
- ❖ *Comment*
- ❖ *Like / Favourite*
- ❖ *Retweet*

Can provoke reactions, echoes and interest in viral processes that may lead to massive propagation in the user community, create influencers, pushers etc.

## SOCIAL ENGAGEMENT

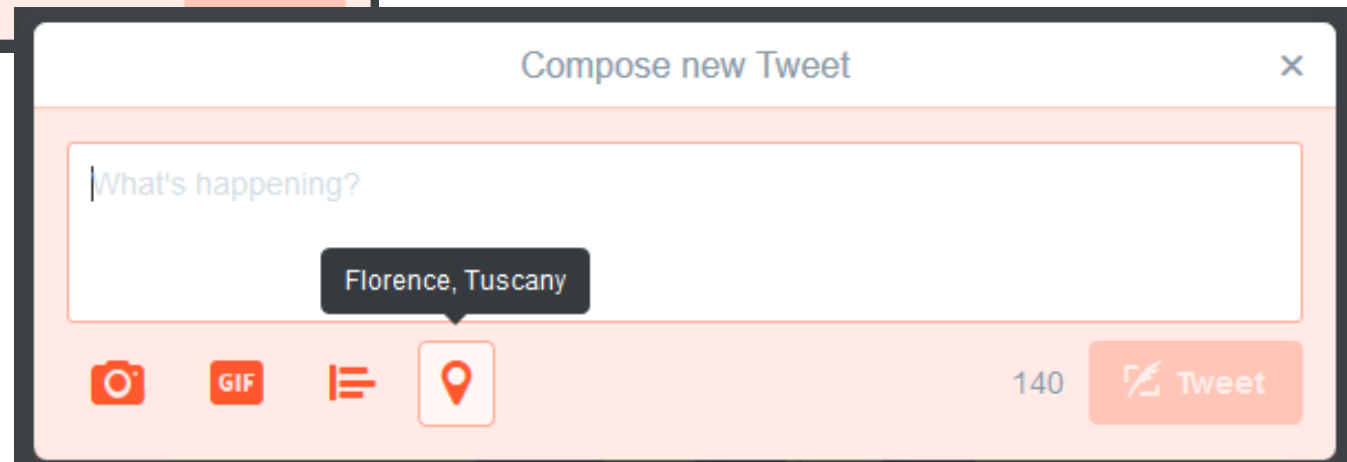
# Compose a New Tweet



- Compose a new text message (up to 140 characters)

- Add several kinds of metadata (pictures, Location...)

- Add polls for specific questions







# Twitter Users

- They are characterised by a number of features:
  - Following
  - Followers
  - Metadata information (age, location, sex, etc..)

# Different Kinds of Twitter Users

➤ Not only single users...

➤ but also News Agencies



➤ Public Institutions & Governments



➤ VIPs



➤ Teams, Schools, Municipalities,, etc.,

**Share, Promote and Communicate**

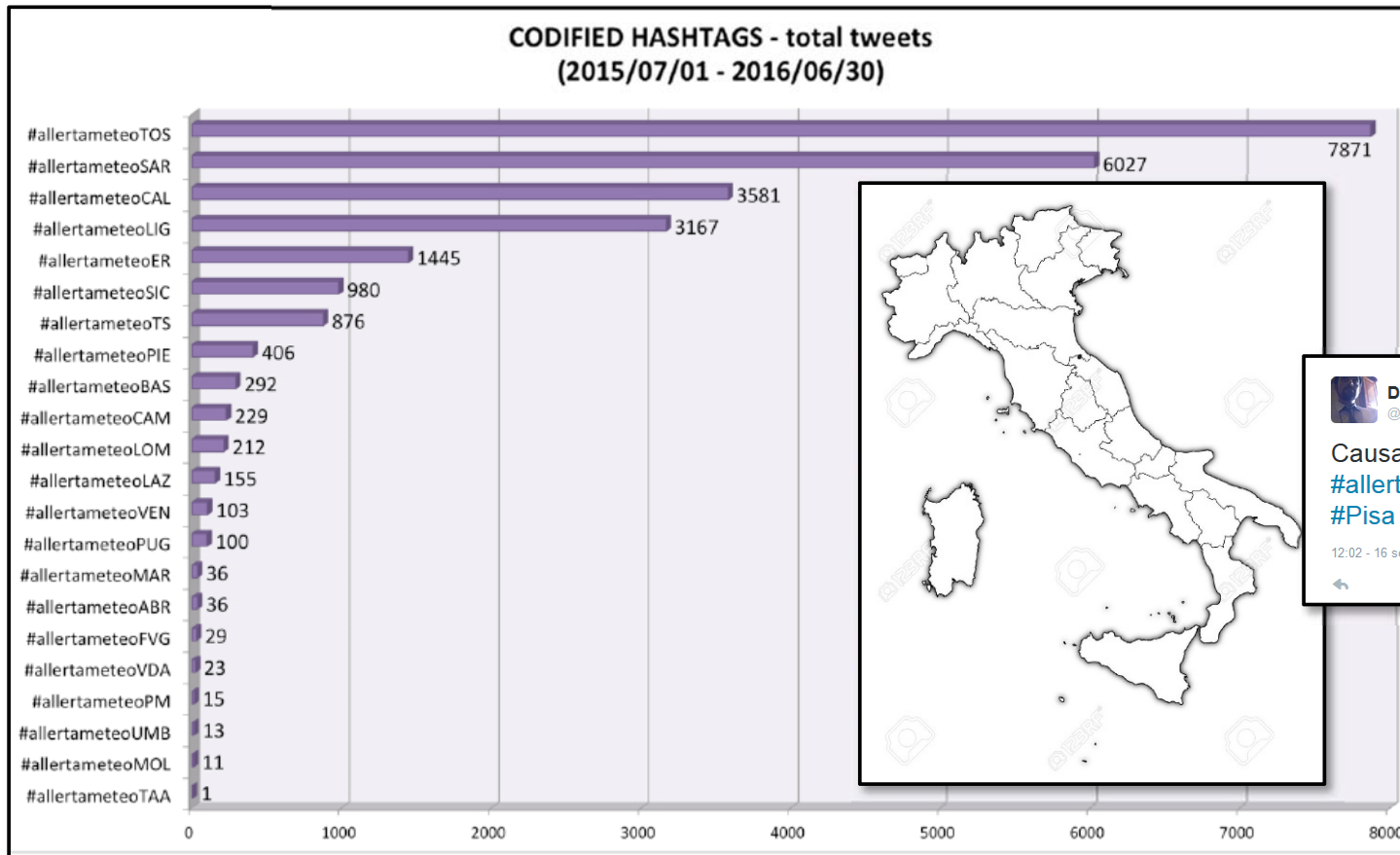


# Follower and Following of an user

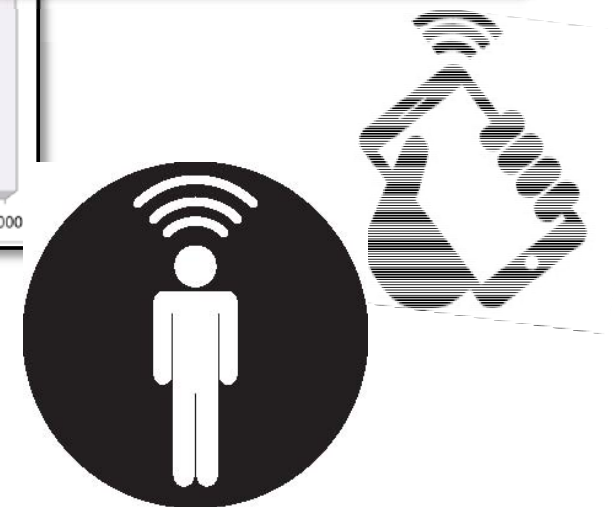
- Followers: are those that decided to be connected with the user receiving all updates from his activity and post.
  - A typical user has more following than followers
- Following: are the users followed by the identified user



# Twitter User as a Sensor



Codified Hashtags for Italian Regional  
Weather Report Services





# # Qualified VS Free Information ? #

- **Hashtags** “#<keyword>” can be invented and/or suggested to the community by the communication stakeholders (Public Administrations, companies and services, TV/web productions etc.), or created spontaneously by some users as well
- Qualified Hashtags VS Free Hashtags for Social Engagement ?
  - ❖ Codified hashtag are helpful as ***focused information channels***



## Citation (e.g., @paolonesi)

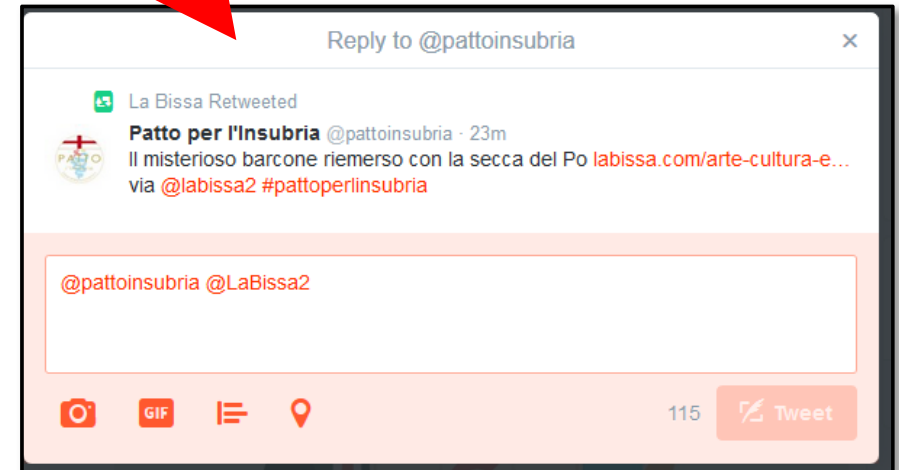
- When you can find a citation (e.g., @paolonesi) in the body of the tweet
  - Twitter sends to the CITED user (e.g., @paolonesi) a message pushing him/her to read and intervene
  - It is one of the main drivers of virality, connected with the number of Followers of the cited persons, since if the cited person reply, ALL his followers are going to received a notification of a new message

- Each Tweet can be *Favourite* by some user
- The number of *Favourite of a Tweet is an index of appreciation and create virality since the actions of Favourite are promoted on the network of Followers of*
  - *who has favorited and*
  - *who has produced the tweet favourite*



# Reply / Comment

- Each Tweet can be replied and the Reply is Connected with the former Tweet as in the figure.
- Twitter.com suggests you to put direct citations to former user and cited users





# ReTweet

- Given a Tweet, an user can do a ReTweet
- This implies that Twitter.com is promoting you action to:
  - your followers: @fgsfgklgh has retweeted.....
  - User of former tweet
  - the followers of the User of former Tweet





# Advertising ..... Also in Twitter

- **In general**
  - **Costs per Impressions, CPI (Costs per Thousands of Impressions, CPM, M is derived from M as 1000 in Roman numbers):**
    - a fee for each impression of the Ad
  - **Cost per Click, CPC:**
    - a fee for each click on the Ad
  - **Cost per Sale, CPS:**
    - a fee for each Sale of a given product/service



# Twitter Analytics

## 28 day summary with change over previous period



**Paolo Nesi** @paolonesi  
 13 Novembre 2015 <http://www.disit.org/smartcitybigdata2015...> #SmartCity e #BigData #km4city #LAMMA @UNI\_FIRENZE @gvannuccini  
<pic.twitter.com/4H1DdDlnzC>



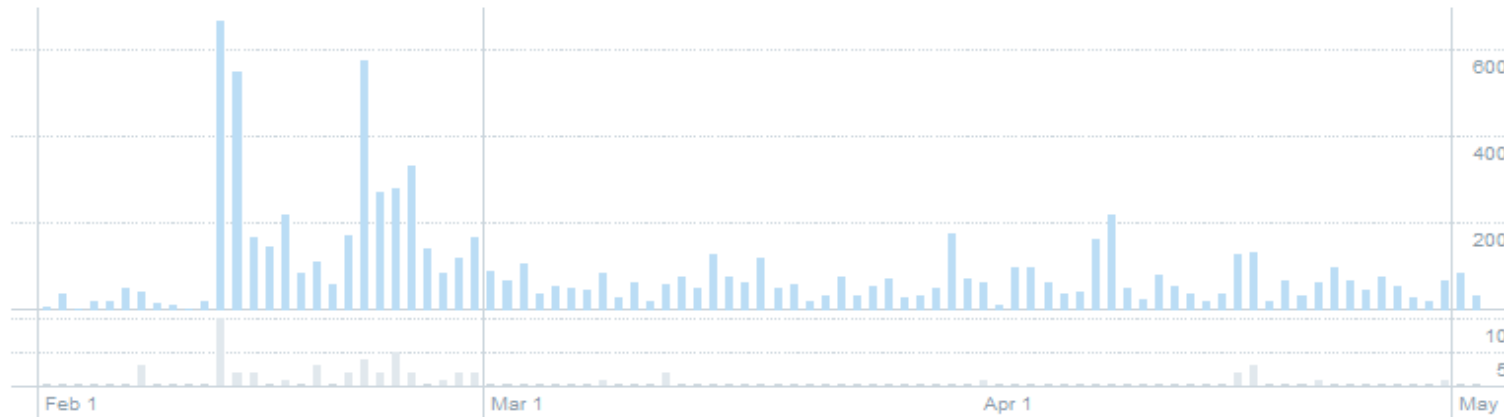
**Reach a bigger audience**  
 Get more engagements by promoting this Tweet!

Get started

Impressions	344
Total engagements	15
Retweets	7
Media engagements	2
Favorites	2
Follows	1
Link clicks	1
Detail expands	1
Profile clicks	1



Your Tweets earned **8.6K impressions** over this **91 day** period



**YOUR TWEETS**

During this 91 day period, you earned **95 impressions per day**.

Tweets **Top Tweets** Tweets and replies Promoted Impressions Engagements Engagement rate

	<b>Paolo Nesi</b> @paolonesi · Feb 12 FODD2015 - Firenze Open Data Day <a href="http://disit.org/fodd">disit.org/fodd</a> <a href="#">@CGiachi</a> <a href="#">@PoloNavacchio</a> <a href="#">@sclopit</a> <a href="#">View Tweet activity</a>	1,068	17	1.6%	<a href="#">Promote</a>
--	---	-------	----	------	-------------------------

	<b>Paolo Nesi</b> @paolonesi · Feb 22 #fodd slide e video <a href="http://disit.org/6593">disit.org/6593</a> #Firenze #OpenData #oddit15 <a href="#">@vbugli</a> <a href="#">@opentoscana</a> <a href="#">@gvannuccini</a> <a href="#">@comunefi</a> <a href="#">@anteos</a> <a href="#">@paolonesi</a> <a href="#">View Tweet activity</a>	757	23	3.0%	<a href="#">Promote</a>
--	---	-----	----	------	-------------------------

	<b>Paolo Nesi</b> @paolonesi · Feb 21 #fodd firenze open data day 2015 <a href="http://pic.twitter.com/QuwB3kpwj">pic.twitter.com/QuwB3kpwj</a> <a href="#">View Tweet activity</a>	663	13	2.0%	<a href="#">Promote</a>
--	--	-----	----	------	-------------------------

**Engagements**

Showing 91 days with daily frequency

**ENGAGEMENT RATE**

1.4%

May 2  
2.9% engagement rate



**LINK CLICKS**

57

May 2  
0 link clicks



On average, you earned **1 link clicks per day**



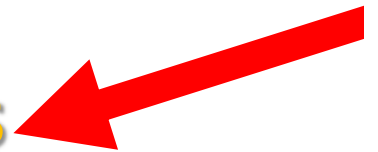
# Agenda

**1. Introduction: What is Twitter**

**2. Overview on Twitter Data Capabilities**

**3. Twitter Vigilance how it is working**

**4. Exercitation Work Plan**





# Twitter Vigilance: le analisi

- **Analisi e caratterizzazione della comunicazione**
- **Percezione sociale, eventi pubblici, naturali..**
- **Scoprire, identificare e calcolare**
  - **Nascita / crescita di nuove occorrenze** in tempo reale: eventi, fatti, meteo, condizioni critiche, etc.
    - *Supporto alla decisioni, ridurre i tempi di reazione, valutare la percezione, ridurre i costi, incrementare la resilienza come capacità di reagire, diagnosi precoce*
  - **Chi influenza** la comunicazione, le comunità e come: i pusher, gli attori, i *follower*, le sorgenti, etc.
  - **Predizione su eventi** periodici, per esempio presenze ad eventi, presenze sui canali televisivi, vendite aziende, etc.
  - **Misure indirette** basate sulla popolazione: rischio sicurezza, degrado, neve, grandine, vento, fallimenti, etc.



# Prediction/Assessment

- Football game results as related to the volume of Tweets
- Number of votes on political elections, via sentiment analysis, SA
- Size and inception of contagious diseases
- marketability of consumer goods
- public health seasonal flu
- box-office revenues for movies
- places to be visited, most visited
- number of people in locations like airports
- audience of TV programmes, political TV shows
- weather forecast information
- Appreciation of services

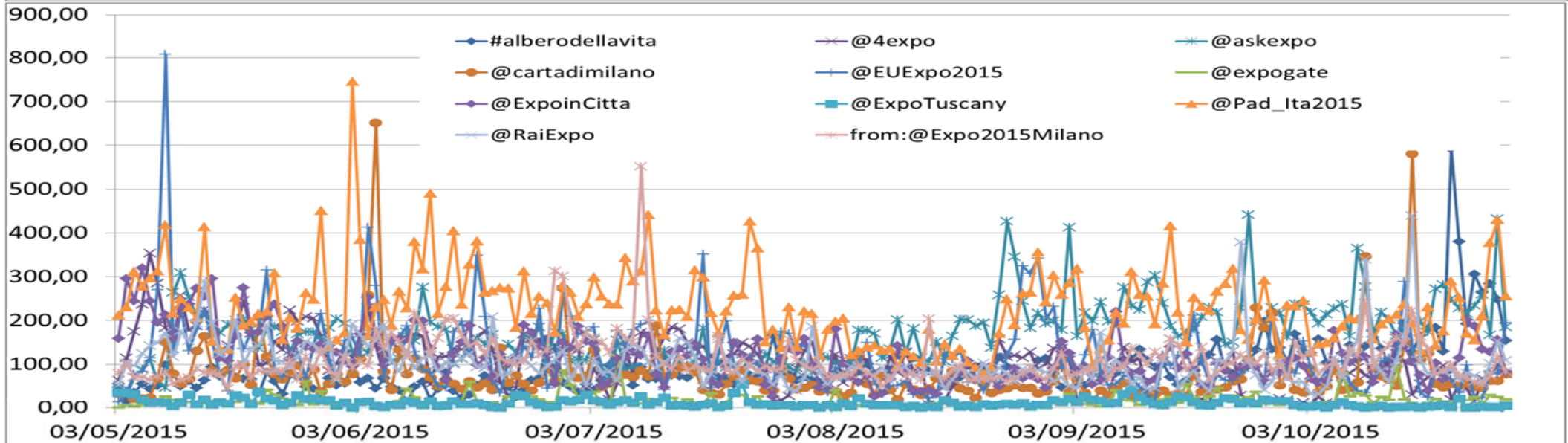
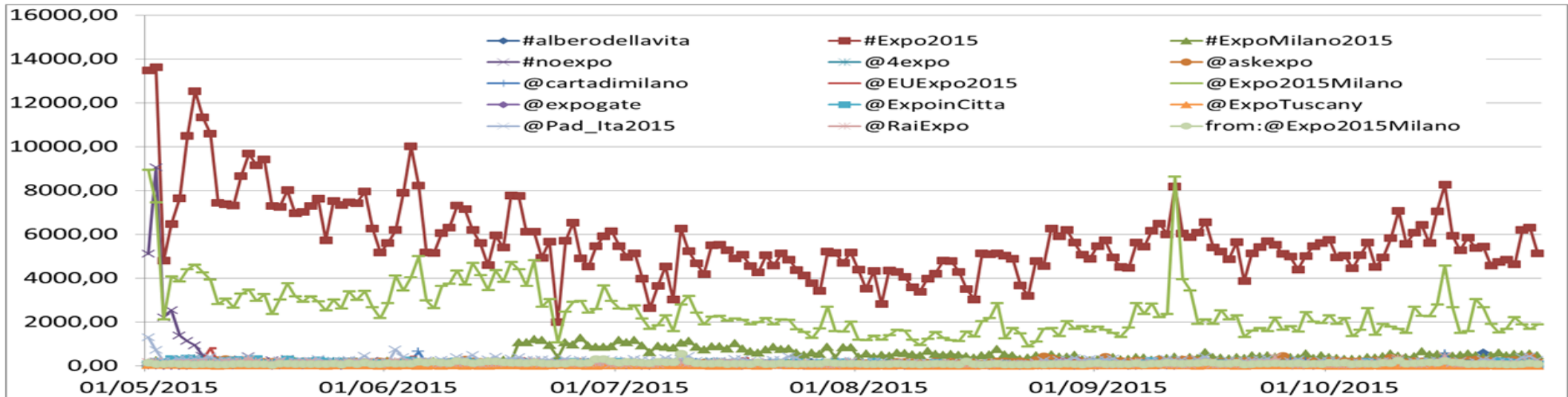


UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

# Twitter Vigilance



# Twitter Vigilance on EXPO2015 channel

**Twitter Vigilance**

MASTER: Big Data Analytics And  
Technologies For Management - MABIDA





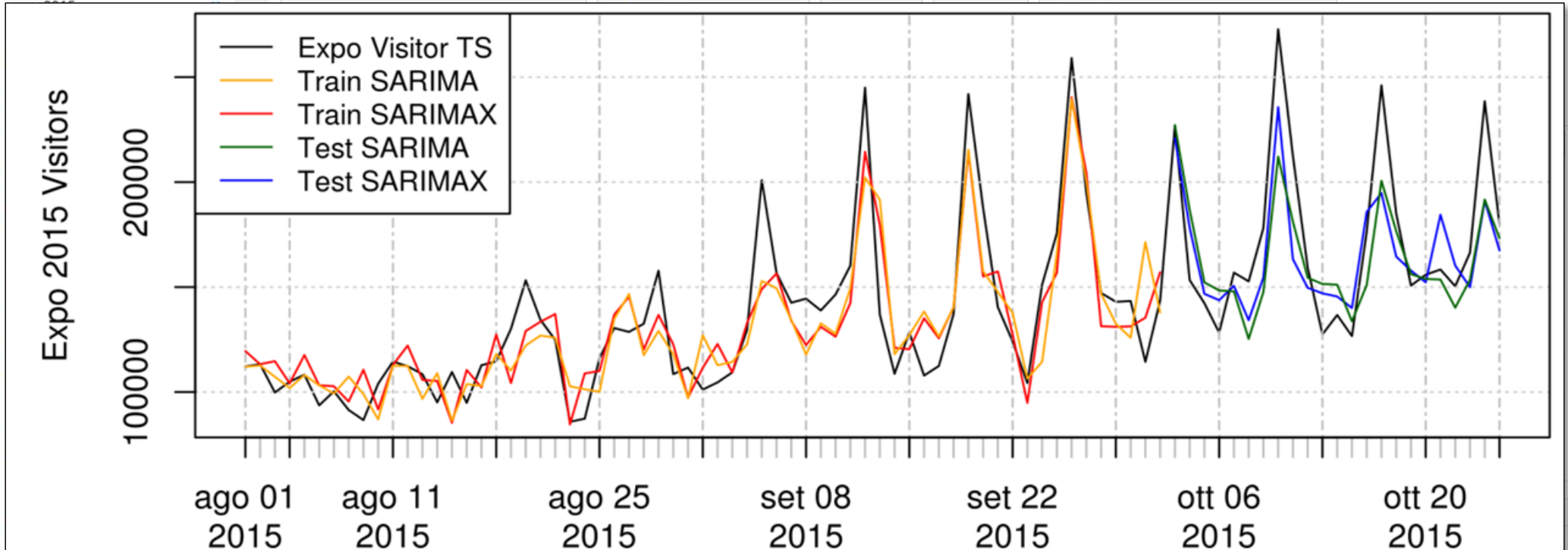
# Twitter Metrics

- TW: Number of Tweets per **Search/Channel** (as called Volume) , per day, per hour
- RTW: Number of ReTweets per **Search/Channel**, per day, per hour
- NRT/TW: ratio from ReTweets and Tweets per **Search/Channel**, per day, per hour
- NumSearch: number of Tweets including the Search per **Channel**, per day, per hour
- Sentiment Analysis Score per **Search/Channel**, per day, per hour
- Num of xxxxx



# Twitter Vigilance

## monitoring and predictions

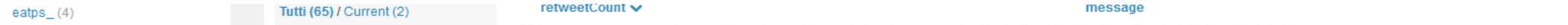


Training: 01-08-2015/02-10-2015 63 DD  
 Test period: 03-10-2015/25-10-2015 22 DD

Predizioni al 90%

# Twitter Vigilance on EXPO2015 channel

## Predicting volume of visitors for tuning the services



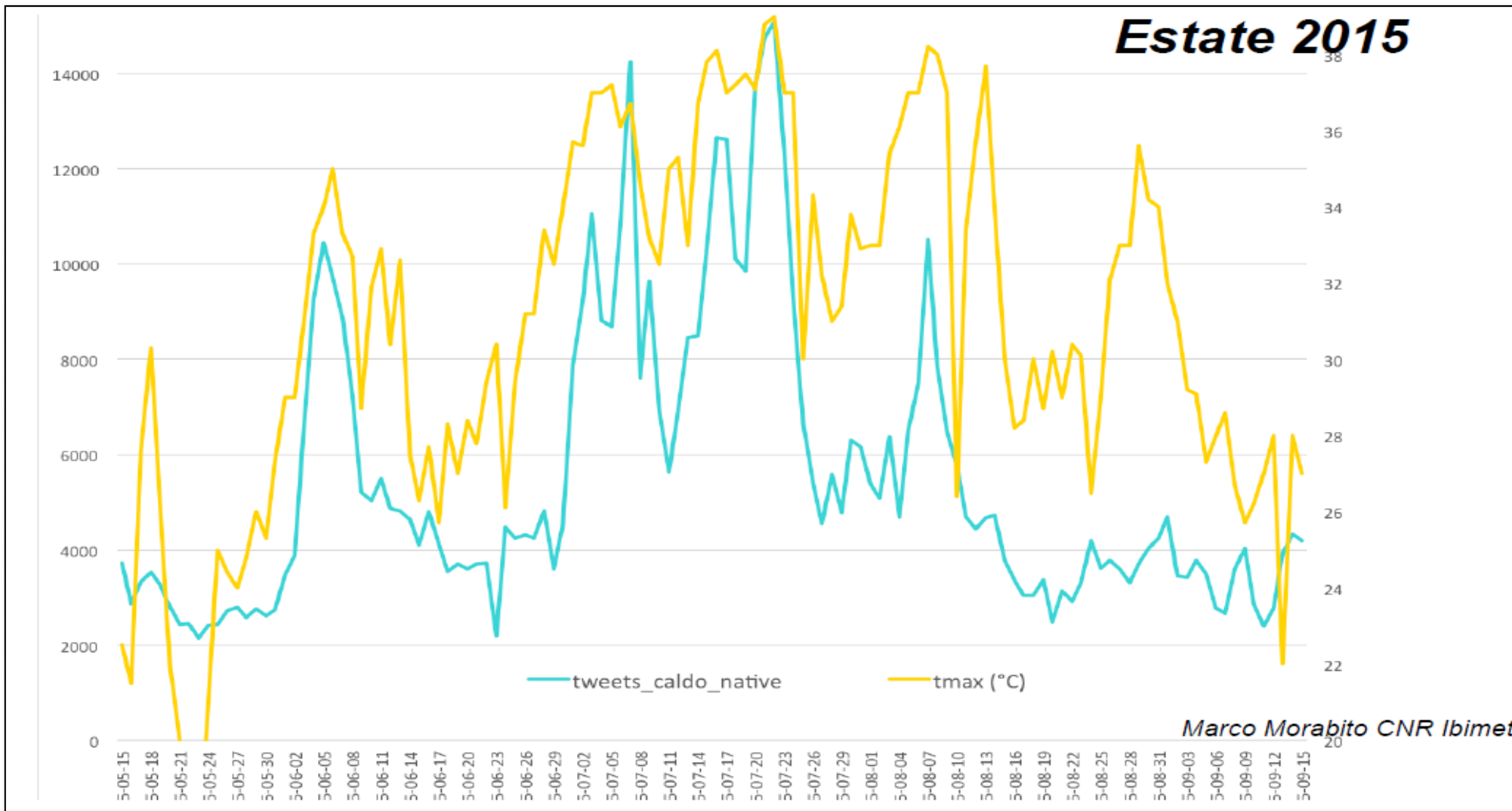


# Early Warning (diagnosi precoce)

- **Computing time**
  - Twitter Vigilance: 1 Day computing, 1 hour update, <http://www.disit.org/tv>
  - Real Time Twitter Vigilance: 5 minutes computing, <http://www.disit.org/rttv>
- **Alert of events inception**
  - Relatively Short Term prediction
  - E.g.: emerging drugs, disasters, emerging locations for social activities, new fashionable restaurants, etc.
- **Reaction time:**
  - Firing action on some metric: ReTweet, Alert, ....



# Twitter Vigilance and Hot Trends



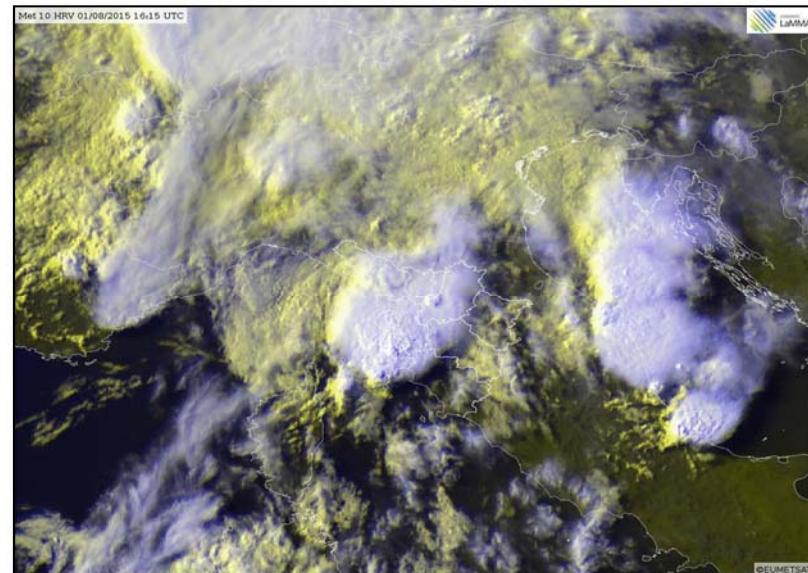
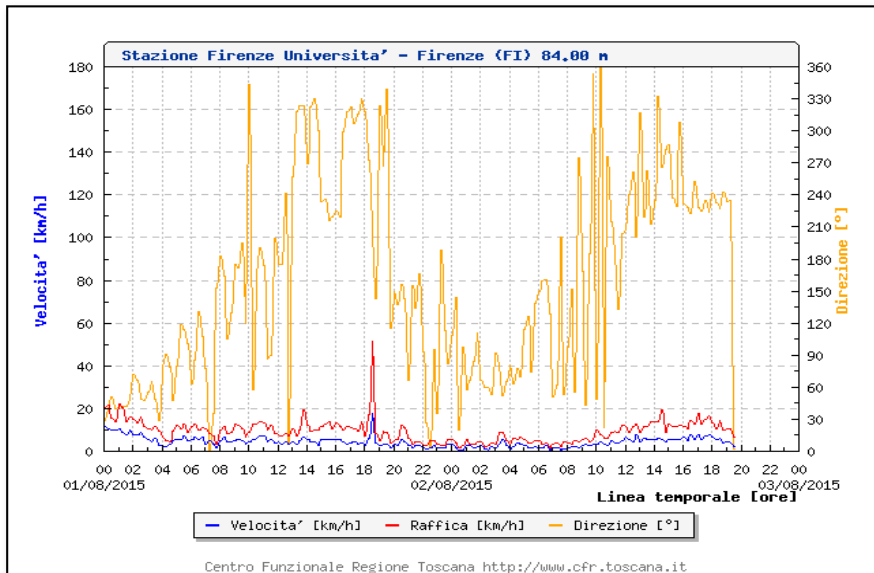
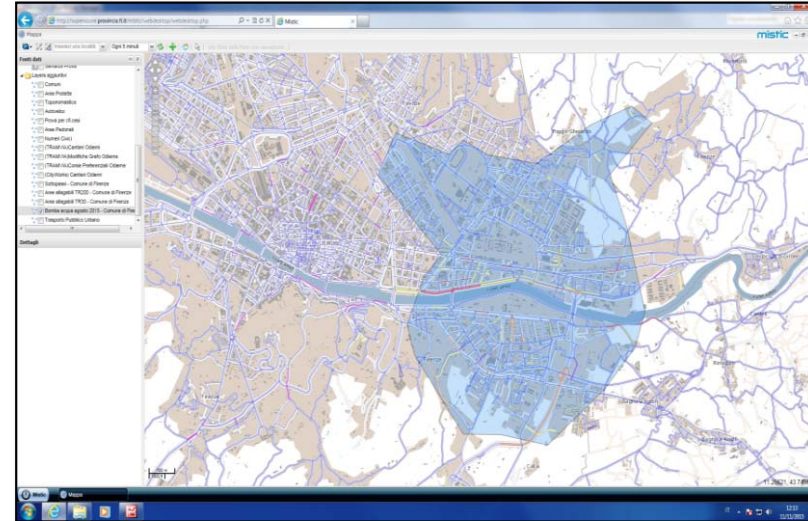
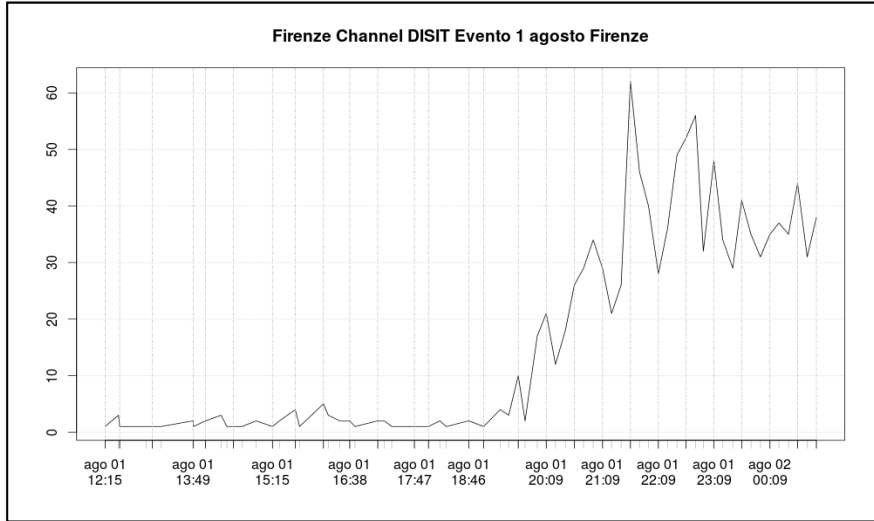


UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

# Twitter Vigilance and Water Bomb

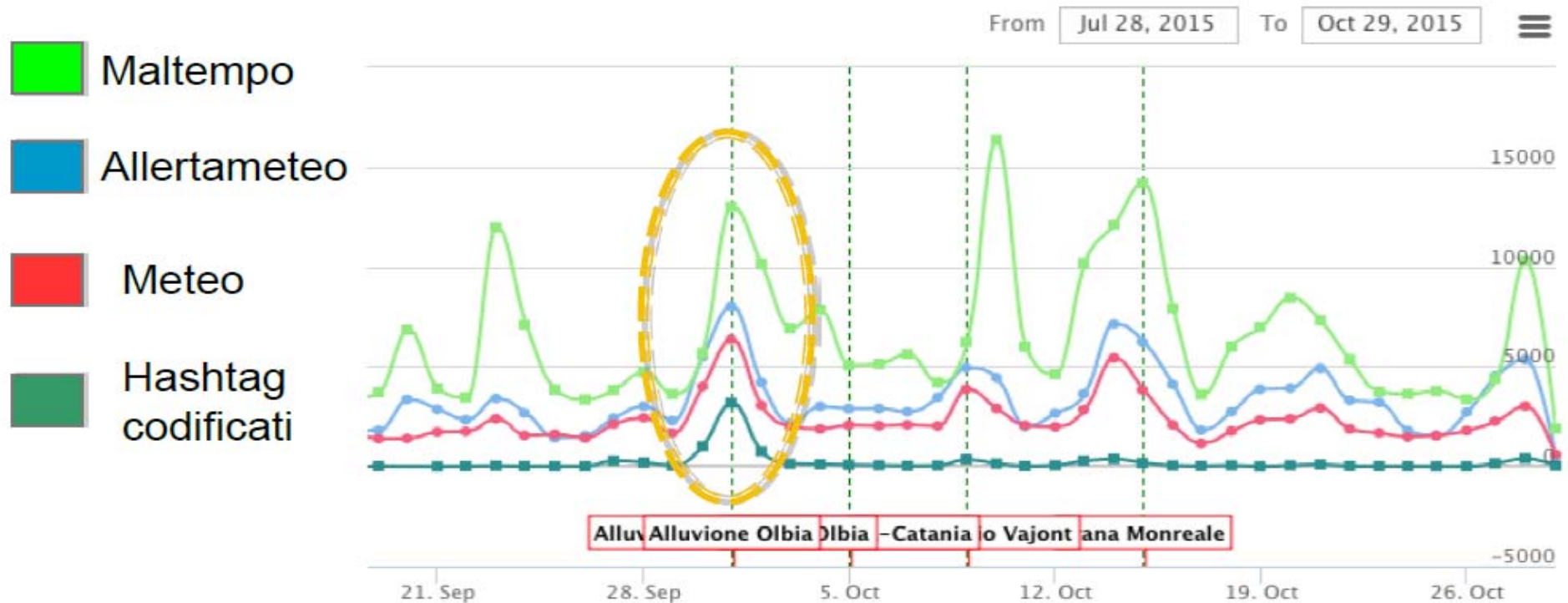


**Twitter Vigilance**

MASTER: Big Data Analytics And  
Technologies For Management - MABIDA

From IBIMET

# Strong correction with events different metrics remarking events



In collaboration with LAMMA



# NLP and SA analysis

- Analysis of Twitter text taking into account the specific grammatical constructs and lexicon of twitter
- Extraction of: verbs, adjectives, keywords, citations, hashtags, etc....
- Construction of metrics counting those elements
- Understanding the meaning of the text and taking into account the discourse, the context; disambiguating statements



UNIVERSITÀ  
DEGLI STUDI  
DI FIRENZE

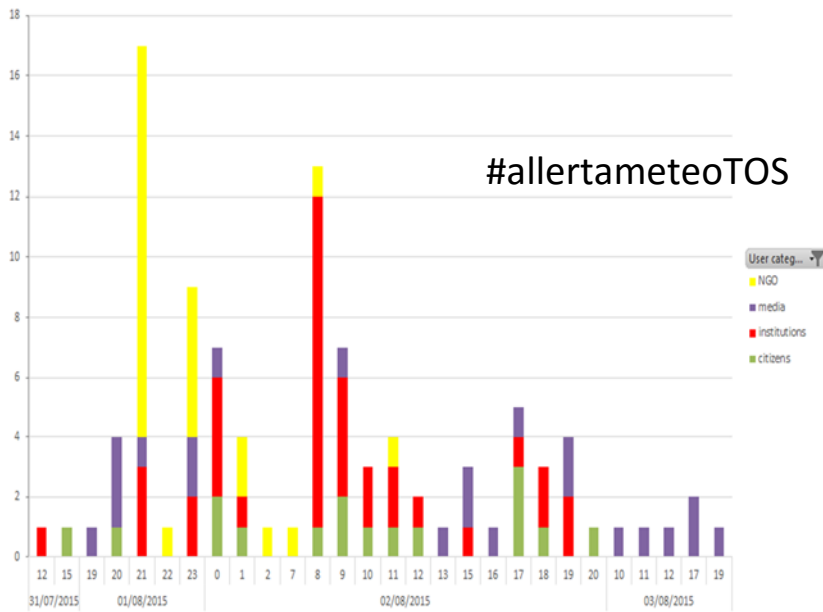
**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

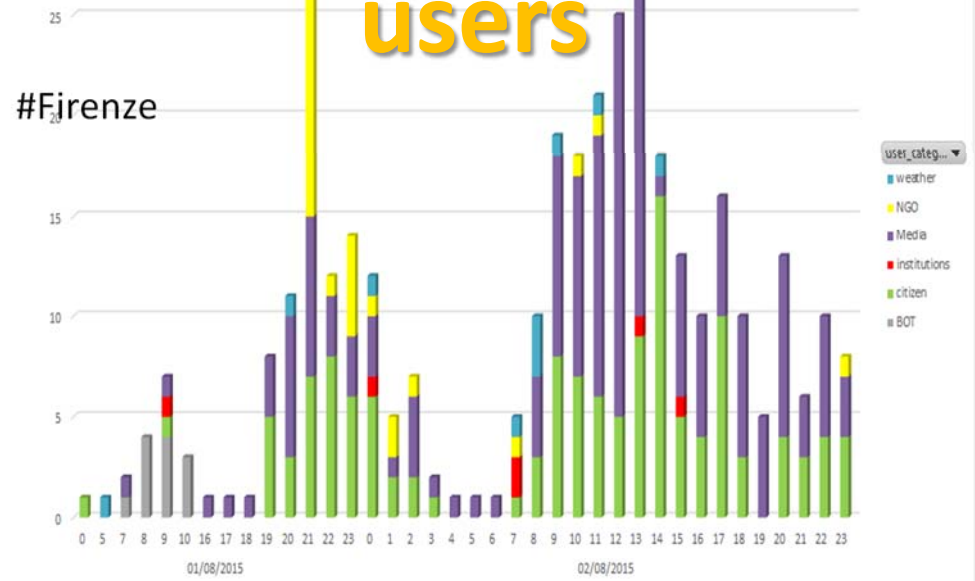
# Hashtags e comunità di users

#Firenze Tw nativi orari per categoria

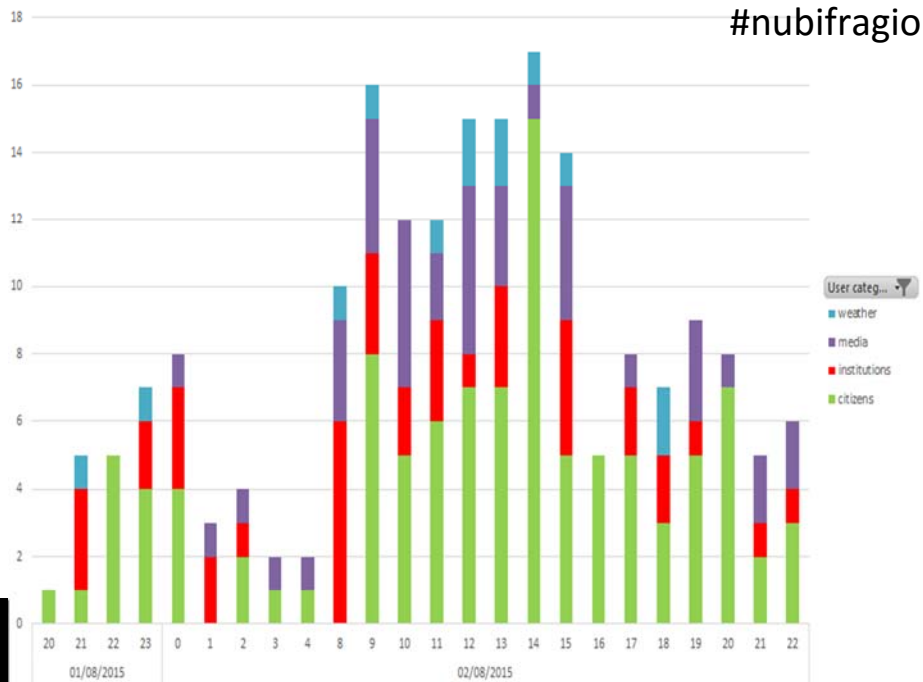
#allertameteoTOS native tweets per categoria utenti



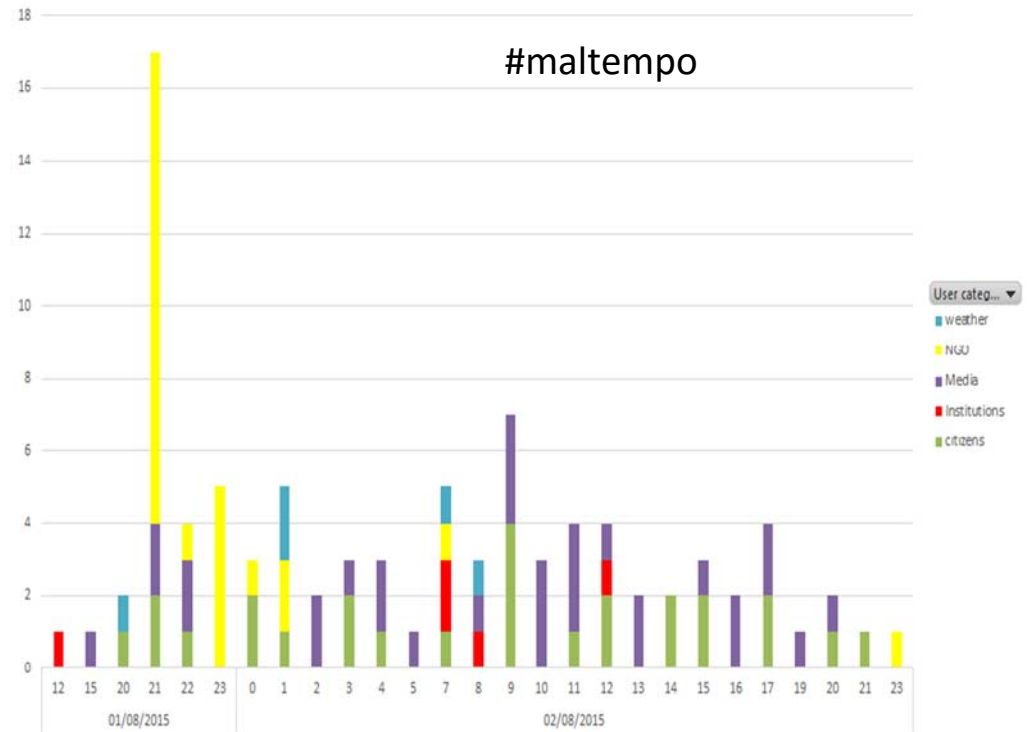
#Firenze



#nubifragio native tweets per categoria utenti



#Maltempo native tweets per categorie utenti



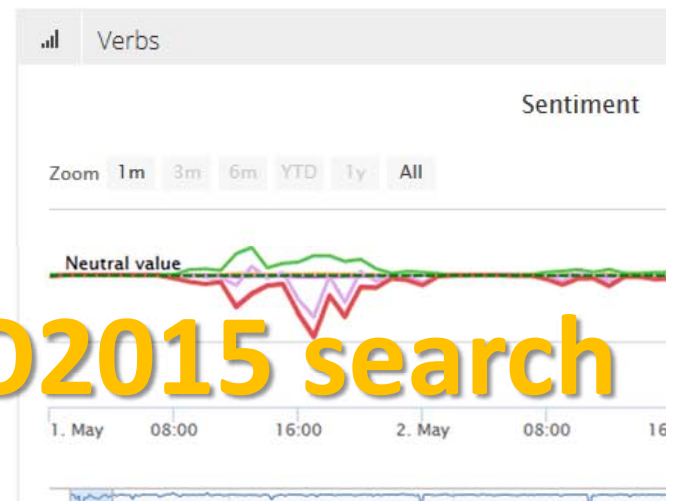
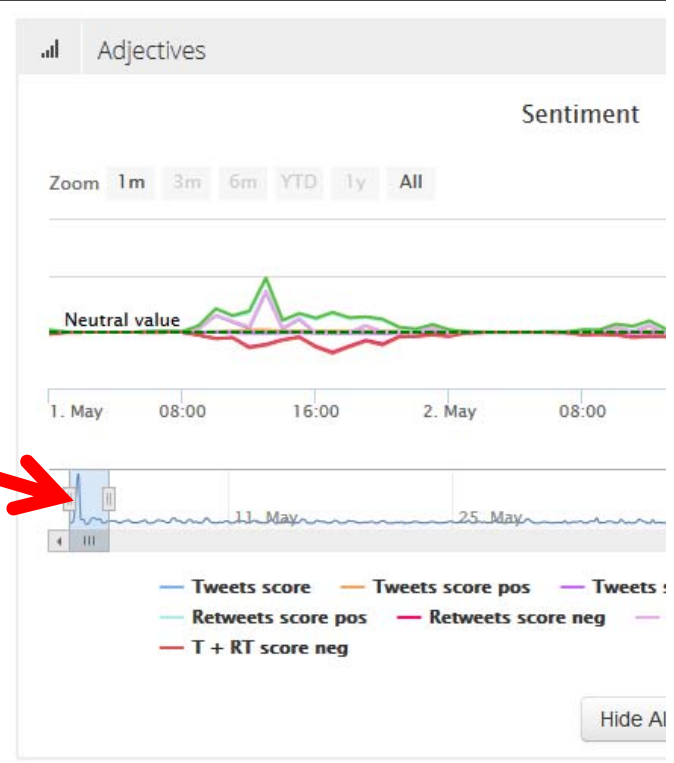
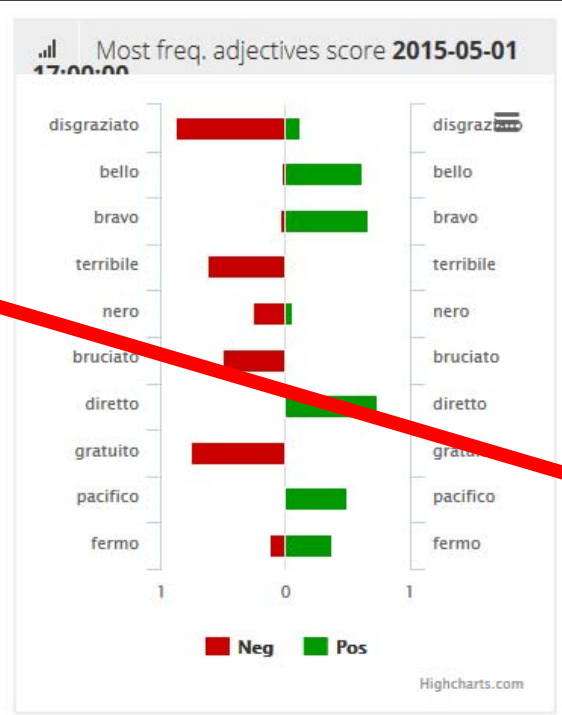
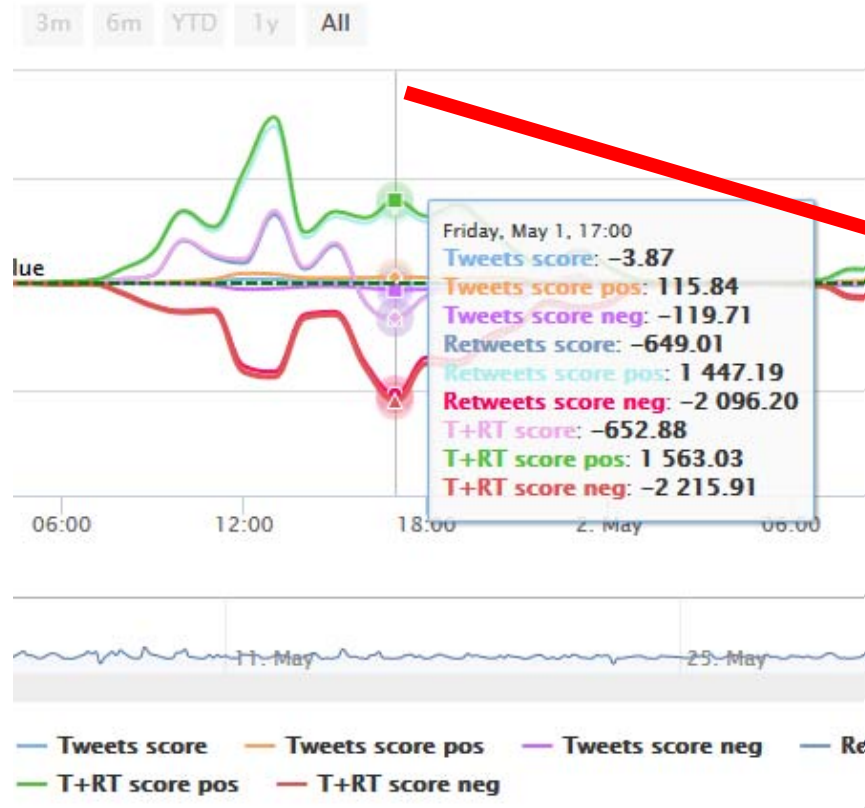




# *Twitter Vigilance: sentiment analysis*

- **Controlla e analizza in automatico il *livello di apprezzamento e/o dissenso* per:**
  - prodotti, servizi, promozioni, cambiamenti in città,
  - persone, azioni politiche, eventi,
  - programmi TV, attori, cantanti (surrogato di "auditel")
- **Permette di effettuare**
  - Valutazioni di **andamento** a breve e lungo termine
  - Valutazioni **comparative** a breve e lungo termine
  - **Predizioni** in certe situazioni
  - **Identificazioni in quasi real time** della nascita di eventi esplosivi, situazioni critiche, etc., uso di utenti come sensori diffusi

## sentiment analysis: #Expo2015



Some views on details of the SA of a Search

# Sentiment Analysis on #EXPO2015 search

# Twitter Vigilance

Most freq. adjectives score 2015-05-01

Adjectives Sentiment

Most freq. verbs score 2015-05-01

Verbs Sentiment

Example tweets

Results 1 - 1 of 1:

message	Fermate queste bestie. Animali, incivili, ignoranti, mostri. Fateschifo. Mi vergogno di essere italiana. #noexpo
twitterUser	SerenaSignorile
publicationTime	2015-05-01 18:29:06
locationUser	
retweet	0
twitterId	594176731541561344
originalTweet	
hashtagsOnTwitter	#noexpo #Expo2015
mentions	
links	
retweetCount	0
Total Sentiment Score	0.49

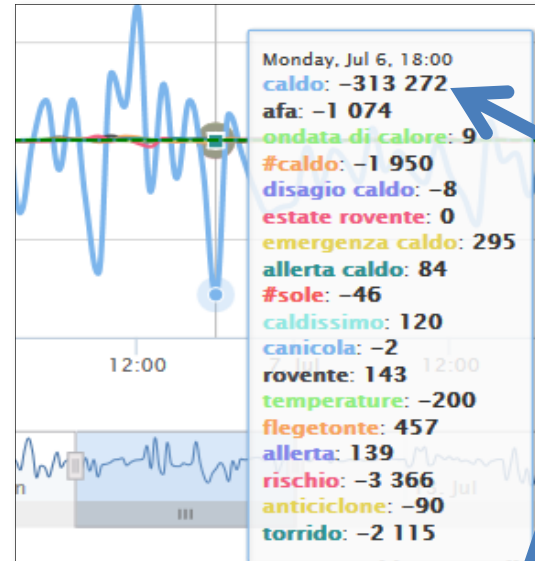
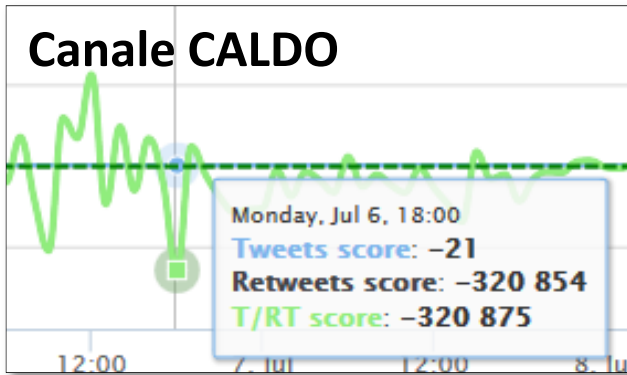
some views on details of the SA of a search

# Sentiment Analysis on #EXPO2015 search

## Twitter Vigilance

# Sentiment Analysis

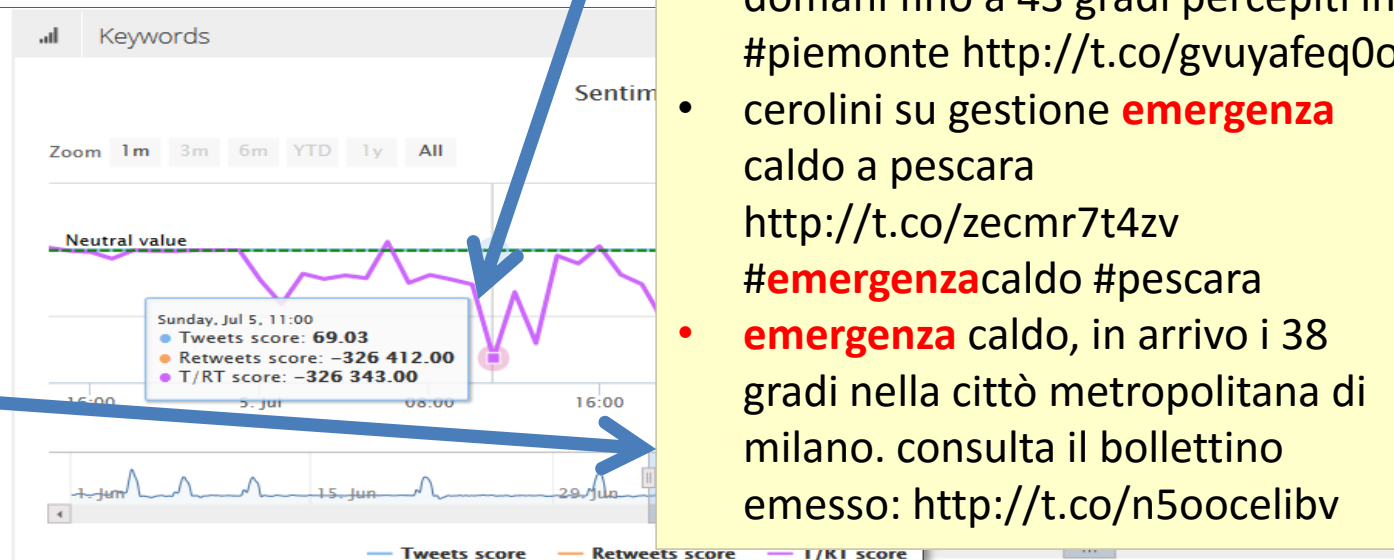
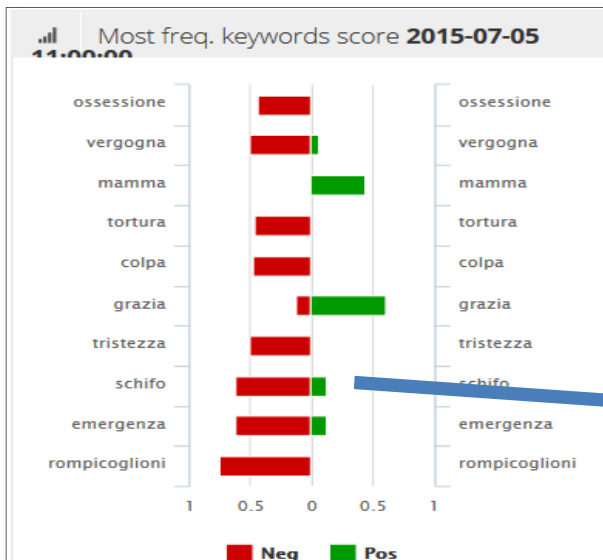
SA sul canale CALDO: la ricerca "caldo" ha dato un sentiment negativo



Un sentiment marcatamente negativo: -313272  
Anche su key multiple

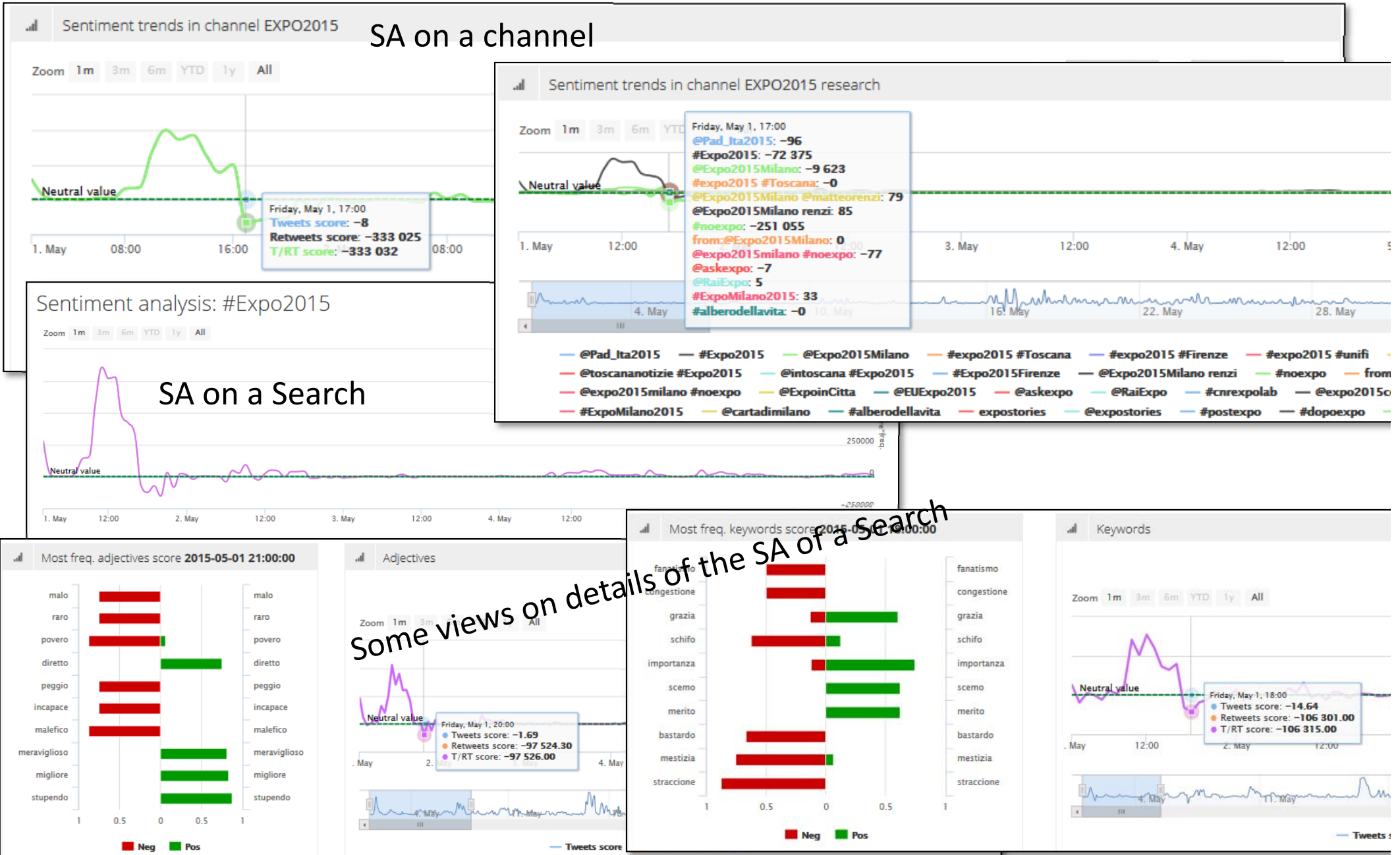
Pochi Tweet negativi hanno provocato un enorme numero di retweet

- **emergenza** caldo, una rete per i più deboli <http://t.co/x6e4dkwgk8>
- continua l'**emergenza** #caldo. domani fino a 43 gradi percepiti in #piemonte <http://t.co/gvuyafeq0o>
- cerolini su gestione **emergenza** caldo a pescara <http://t.co/zecmr7t4zv> #**emergenza**caldo #pescara
- **emergenza** caldo, in arrivo i 38 gradi nella città metropolitana di milano. consulta il bollettino emesso: <http://t.co/n5oocelibv>





# Sentiment Analysis

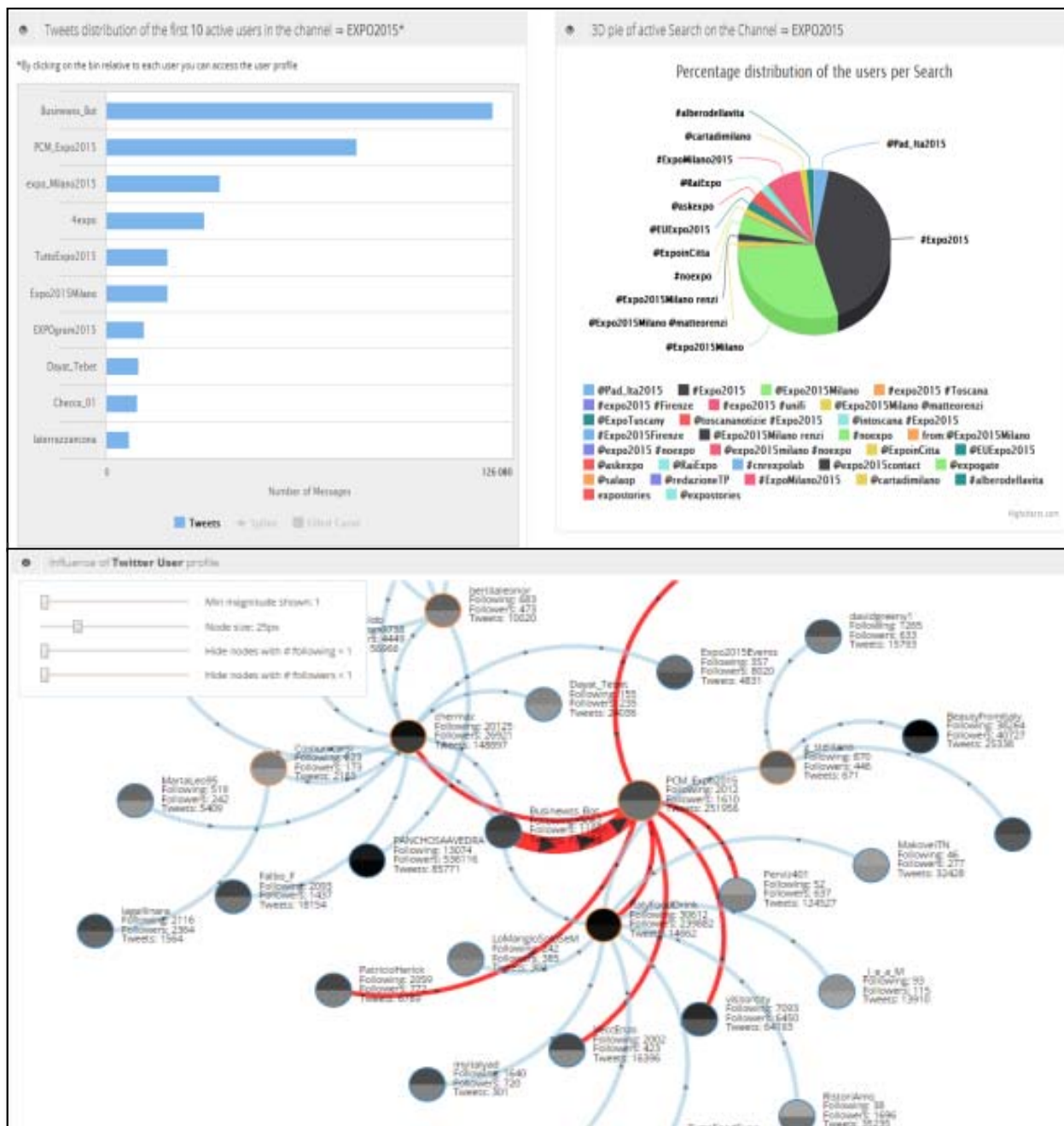




# Twitter user profile aspects

- **Twitter user:**
  - No much different from other social media users
  - User profile is not a confident and trusting info
  - Most of them does not provide info about their location, their language, etc.

- Identification of
  - Influencer
  - Chains of users
  - communities
  - pusher
  - BOT
  - Relationships
- Assessing and predicting Virality





# Metadata OF each Tweet

- Tweet itself, date and time, UserID
- Number of retweets
- Number of Favorites
- GPS coordinates
- Language
- Details about the UserID:
  - how many followers, following, lists, favorites, etc..
  - List of connected users...
- .....





# *Alcuni Canali di TwitterVigilance*

- <http://www.disit.org/tv> canali pubblicati
- **Esempi di Canali:**
  - EXPO 2015, CNR EXPO2015, Firenze, ApreToscana, maturità, ConsumoSuolo, meteo, allerta meteo toscana,
  - protezione civile, farmaco vigilanza, smart drug, ECLAP e Europeana, advertising TV, laudatesi, terrorismo, etc.
- **Aree tematiche:** meteo, ambiente, advertising, eventi pubblici, farmacovigilanza, smart city, politica, TPL, emergenza, Servizi, etc.



# Twitter Vigilance Tools

- Twitter Vigilance:
  - 1 Day computing, 1 hour update
  - <http://www.disit.org/tv>
- Real Time Twitter Vigilance:
  - 5 minutes computing
  - <http://www.disit.org/rttv>



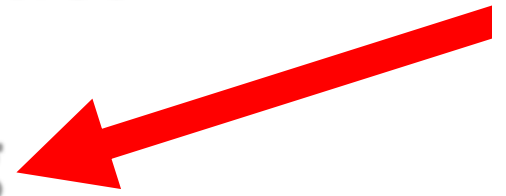
# Agenda

**1. Introduction: What is Twitter**

**2. Overview on Twitter Data Capabilities**

**3. Twitter Vigilance how it is working**

**4. Exercitation Work Plan**

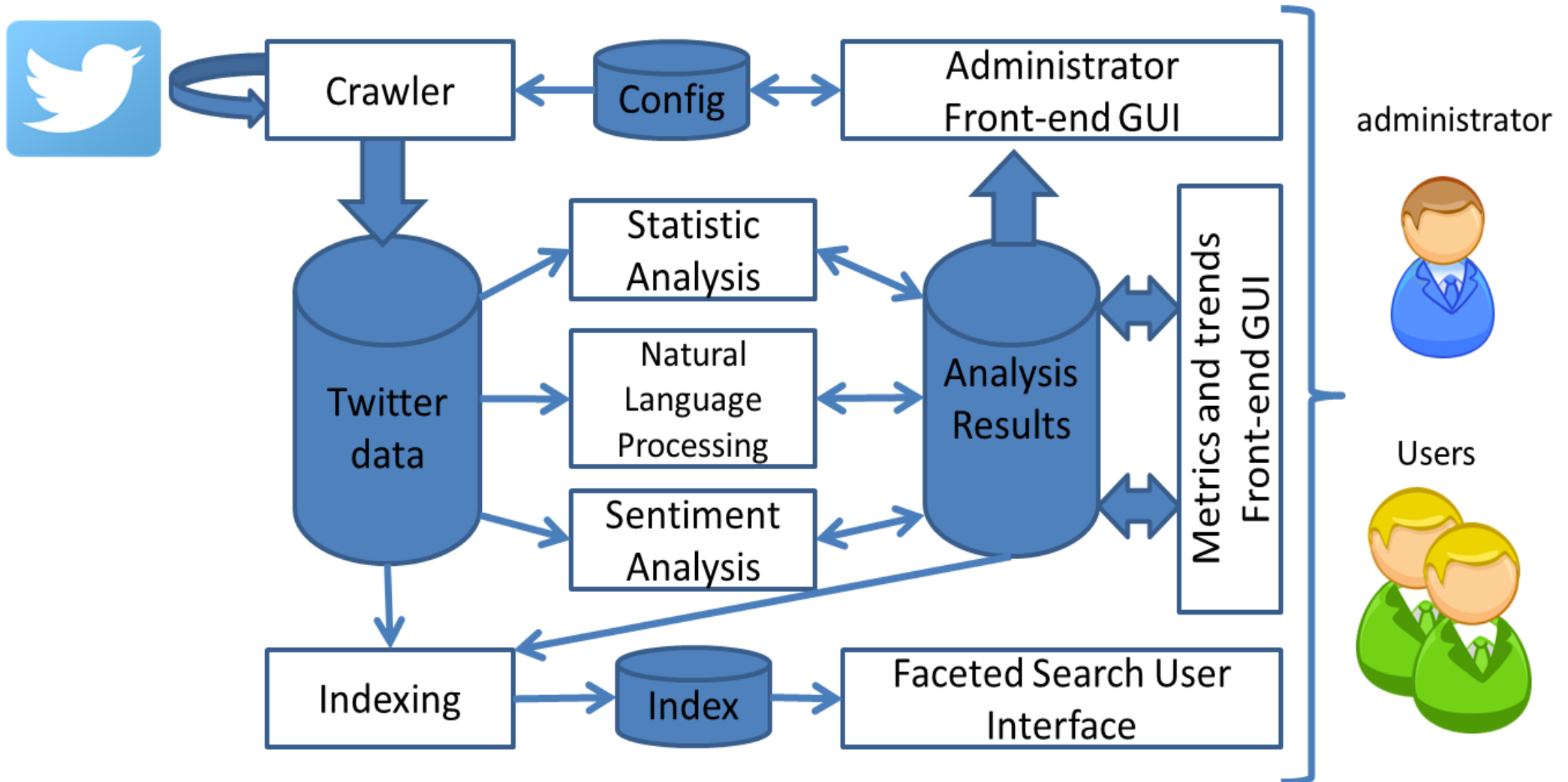




# Obiettivi: Twitter Vigilance

- **Monitorare** Canali Twitter con
  - **Elevata affidabilità** e precisione, seguendo eventi lenti, veloci ed esplosivi
  - **Gestione multiutente** per canali multipli: pubblici e privati
  - **Data analytics e sentiment analysis** in modo assistito e diretto
- **Canale di Twitter Vigilance:** un insieme di ricerche attive e adattive su social media Twitter.
  - Ogni ricerca può essere semplice o complessa
  - Ogni utente qualificato può gestire più canali e ricerche
- **Attivo**
  - da Aprile 2015 per il collezionamento dati,
  - da settembre per il calcolo automatico big data analytics,
  - da ottobre per la Sentiment Analysis automatizzata.
- **Gestisce** con disinvoltura
  - Oltre 1.3-2.5 Milioni di Tweet per giorno per server
  - Oltre 190 Milioni di Tweet
  - Lavora anche Real Time con RealTimeTwitterVigilance

# Twitter Vigilance





UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

# Http://www.disit.org/tv



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE  
DINFO  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

## Twitter Vigilance Dashboard

admin ▾

Last crawling: 2016-06-23 00:13:01

Search parameters

Crawler Statistics

Config Twitter API

Data analysis 3

Channel Statistics

Search statistics

Retweet Statistics

Twitter Users  
statistics

Sentiment analysis 1

LOGS 9

Processes

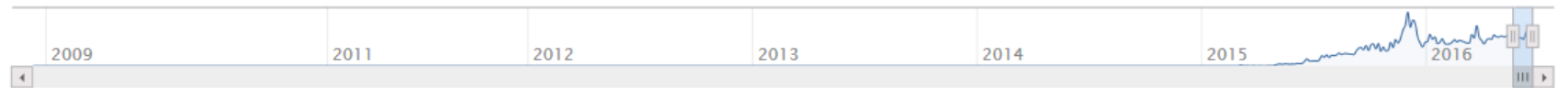
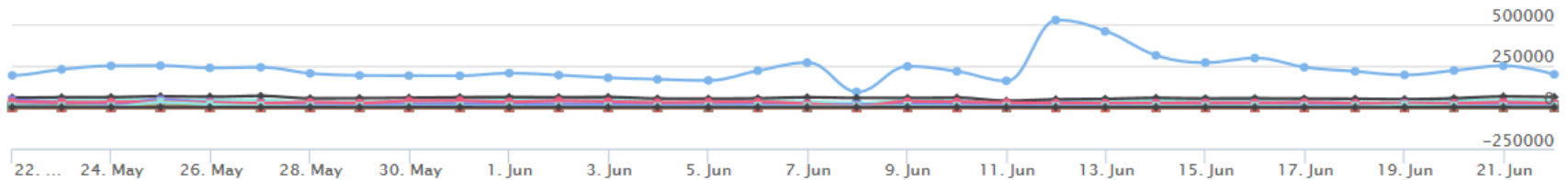
INFO

Home > Channel statistics

### Global view of user channels

Zoom 1m 3m 6m YTD 1y All

From May 22, 2016 To Jun 22, 2016



- ads
- aeroporto firenze
- Agenzie-ambientali
- Allertameteo
- Allertameteo TOSCANA
- Alluvione
- apretoscana
- ARPAT
- CALDO
- cambiamenti climatici
- cardboard
- chi parla di meteo
- Clima
- CNRxEXPO
- Codified Hashtags Allerta
- comfortUS
- ConsumoSuolo
- eclap
- Emergenza\_acqua
- estra
- Europeana
- EXPO2015
- EXPO2015Toscana
- fallbegins
- Farmacovigilanza
- Firenze
- firenzefuochi
- FirenzeICT
- geocomuniTOS
- Giubileo
- idrogeologico
- iene
- influenza
- iononrischio
- LaMMA
- laudatoSi
- laudatoSi\_1
- majorcities2016
- Maltempo
- Maltempo Firenze
- Maltempo Grosseto
- maturità 2015
- MeteoAlert
- MeteoUSER
- MissioneRosetta
- monitAllergie
- mrm
- mugnone2016
- mymeteo
- NASA New Horizons
- nutella
- PAAlert
- PAEnvironment
- PAMeteoNews
- PAMeteoNewsStud
- paolonesi
- papafrancesco
- PAProtCivile
- PA\_social\_PA
- protezione civile toscana
- reflusso
- resilienza
- rettore
- rossano
- siena
- Smart Drugs
- smartcity
- smartcitybigdata2015
- SOLO allertameteoTOSCANA
- tech
- terrorismo
- TPL
- turismo Firenze
- turismo Firenze\_1
- turismo Firenze\_2
- uber
- Univ-Firenze
- vargroup
- Vda
- VoragineFi
- Xfactor9
- Xylella
- zanzara
- Events

Hide All



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

# Real time Twitter Vigilance

[Http://www.disit.org/rttv](http://www.disit.org/rttv)

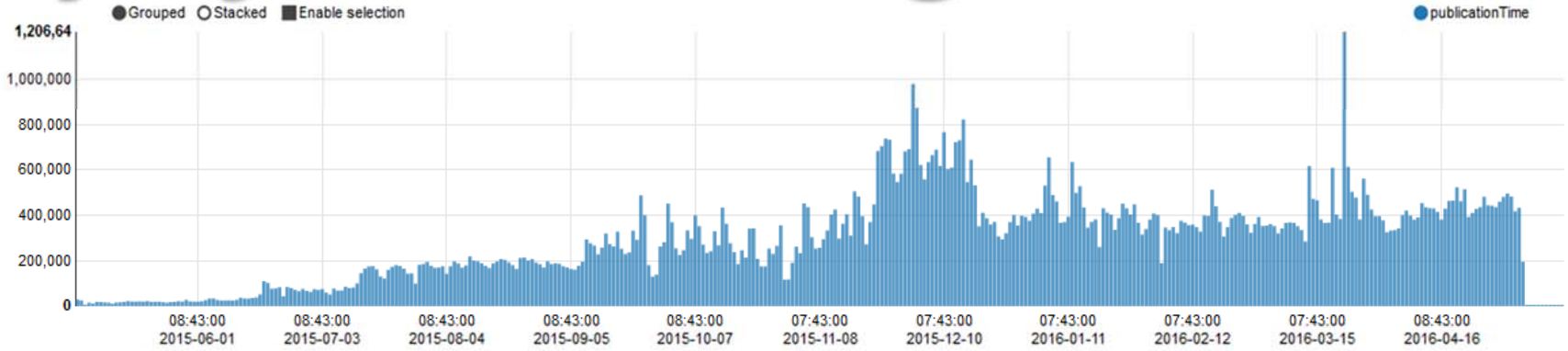


**Twitter Vigilance**

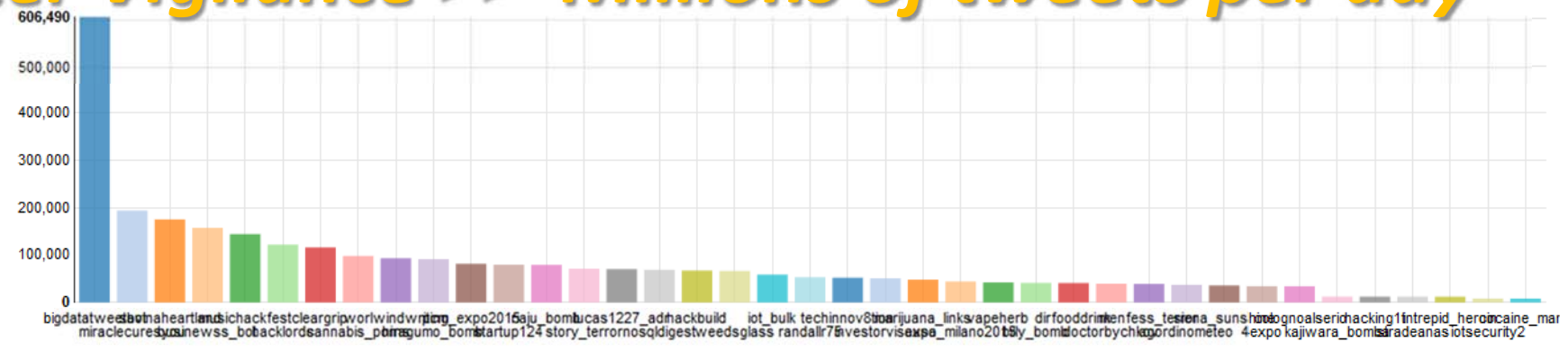
MASTER: Big Data Analytics And  
Technologies For Management - MABIDA

# Full text and Faceted Search Facility of Twitter Vigilance db

- terrorismo (342208)
- farmacovigilanza (9746639)
- tech (8495537)
- uber (7337300)
- smart drugs (7062871)
- ads (91055)
- firer (9100)
- carlameridiana (3198)
- reflusso (3039363)
- clima (2958966)
- xfactor9 (2672493)
- turismo firenze\_2 (2451517)
- caldo (2395023)
- influenza (2374427)
- tpl (2051480)
- expo2015 (1643031)
- maltempo (1624816)
- allertameteo (1623214)
- cnrxexpo (1541848)
- mymeteo (1275964)
- xytella (984595)
- allertameteo toscana (901664)
- turismo firenze (759910)
- turismo firenze\_1 (759910)
- chi parla di meteo (699119)
- smartcity (646207)
- resilienza (636137)
- siena (496935)
- papafrancesco (485570)
- giubileo (413690)
- geocomunitas (399976)
- nasa new horizons (336838)
- laudatosi (321243)
- laudatosi\_1 (321243)
- cardboard (294675)
- meteocaler (283981)
- nutella (252075)
- emergenza\_acqua (245792)
- monitallergie (193804)
- mugnone2016 (175584)
- Show more...



Twitter Vigilance >> Millions of tweets per day



- twitterUser
- bigdatatweetbot (606490)
- miraclecuresyou (194035)
- savinaheartland (175355)
- business\_bot (157586)
- musichackfest (144278)
- hacklords (122107)
- cleargrip (116207)
- ...

Grid Results

Filter fields

Tutti (65) / Current (2)

Field Name	retweetCount	message
<input type="checkbox"/>	3338556	rt @theellenshow: if only bradley's arm was longer. best photo ever. #oscars http://t.co/c9u5notgap
<input type="checkbox"/>	3337499	if only bradley's arm was longer. best photo ever. #oscars http://t.co/c9u5notgap
<input type="checkbox"/>	3336849	rt @theellenshow: if only bradley's arm was longer. best photo ever. #oscars http://t.co/c9u5notgap
<input type="checkbox"/>	3336475	rt @theellenshow: if only bradley's arm was longer. best photo ever. #oscars http://t.co/c9u5notgap
<input type="checkbox"/>	3335875	rt @theellenshow: if only bradley's arm was longer. best photo ever. #oscars http://t.co/c9u5notgap

[Http://tvsolr.disit.org](http://tvsolr.disit.org)

Twitter Vigilance

MASTER: Big Data Analytics And Technologies For Management - MABIDA



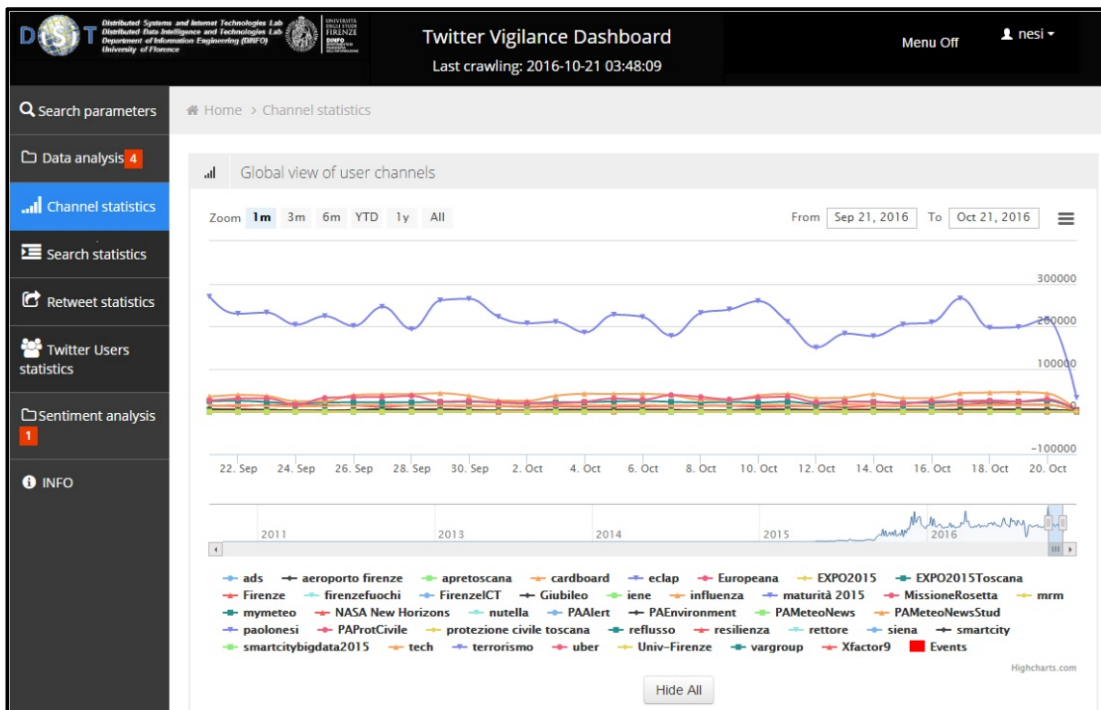


# Definitions

- Channel := { Search }
- Search := <request> AND <request>
- Request := keyword | from:<username> |  
#<hashtag> | @<username> | .....

# Twitter Vigilance (TV) – Channel

<http://disit.org/tv/>

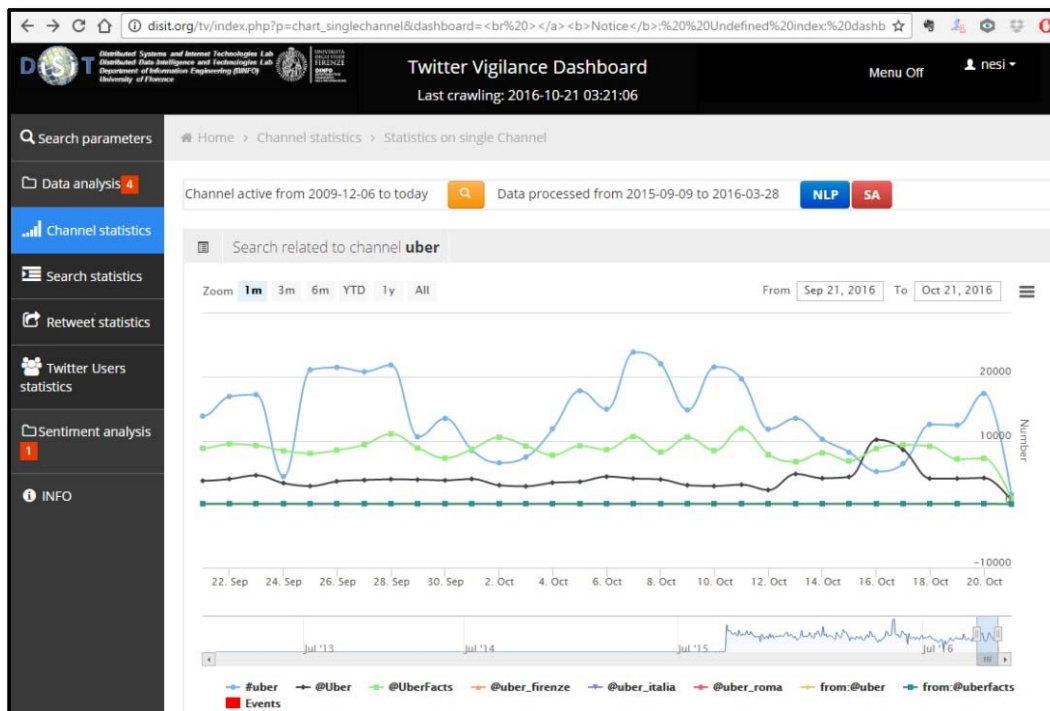


➤ Data Acquisition is based on the concept of **Channel**

➤ A **Channel** is a thematic set of simple and complex **Searches**

# Twitter Vigilance (TV) – Search

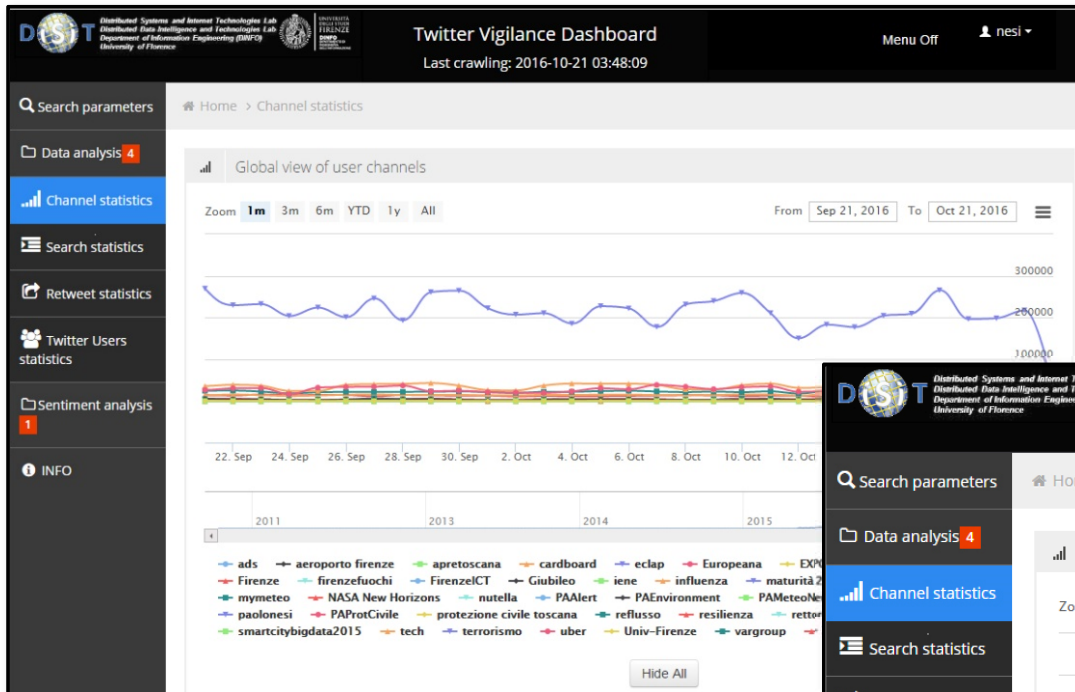
<http://disit.org/tv/>



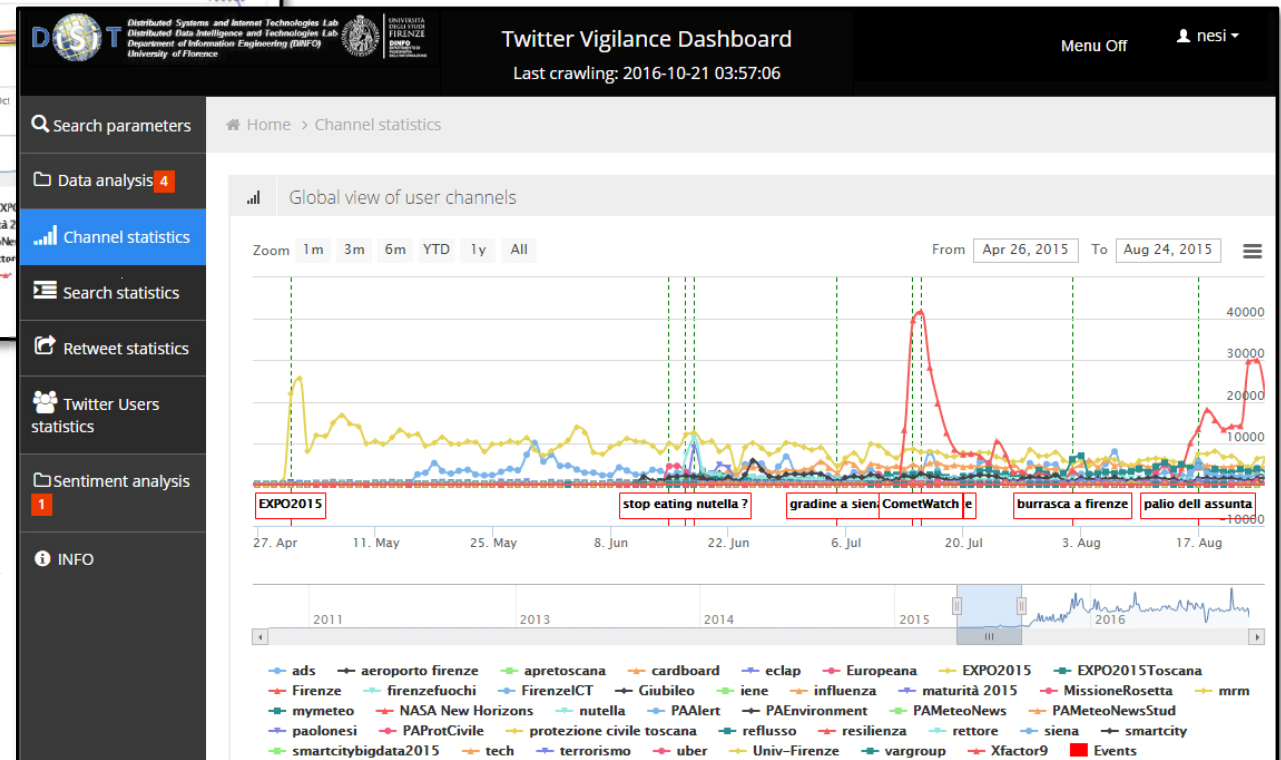
➤ **Searches** can be single or groups of keywords used to monitor and extract Tweets information through Twitter APIs

➤ A complex **Search** can combine several keywords such as hashtags, user IDs, mentions, citations with logical operators (AND, OR...)

# TV Graphs: Channel View

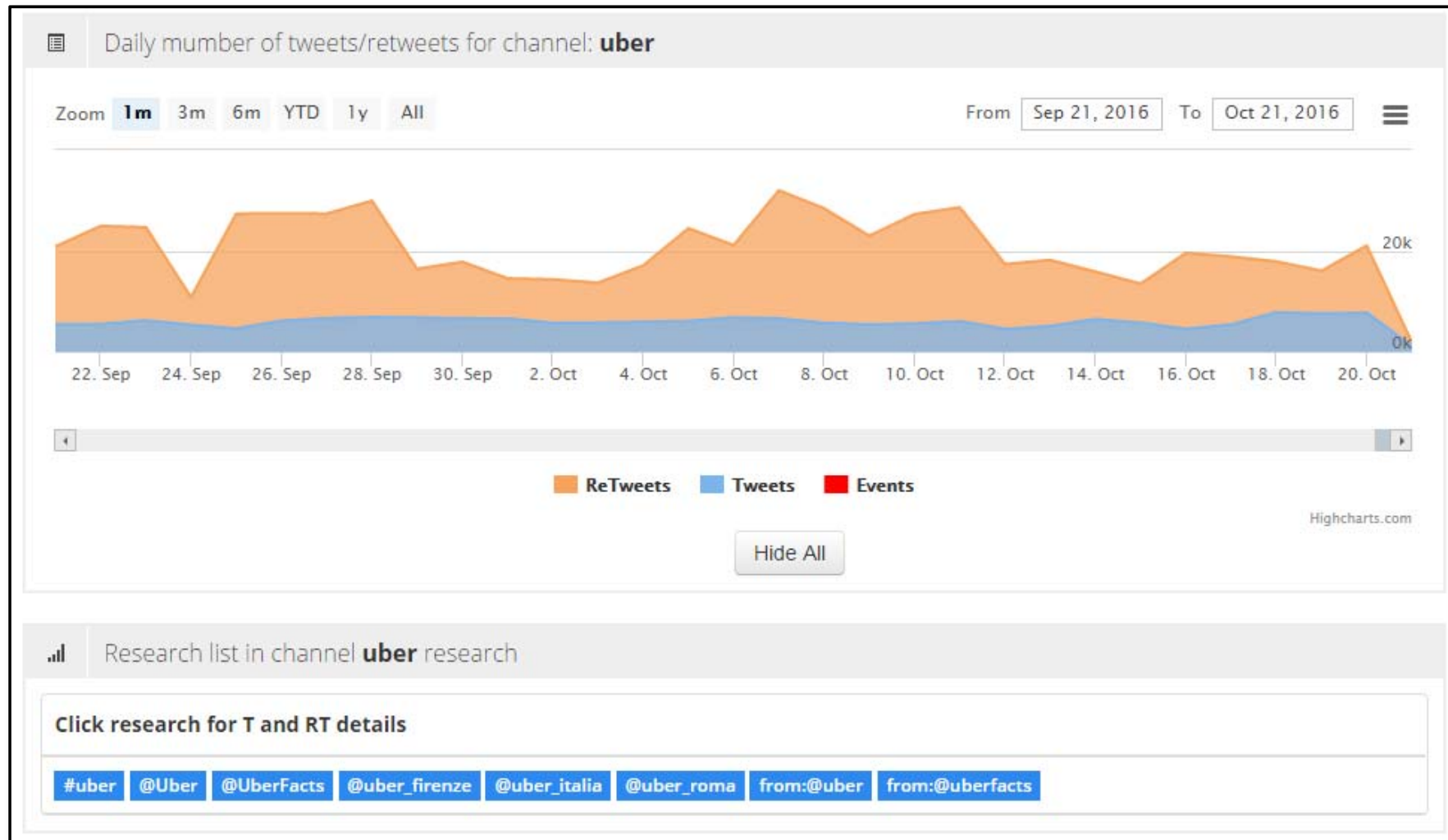


➤ Display / Hide single Channels or all Channels

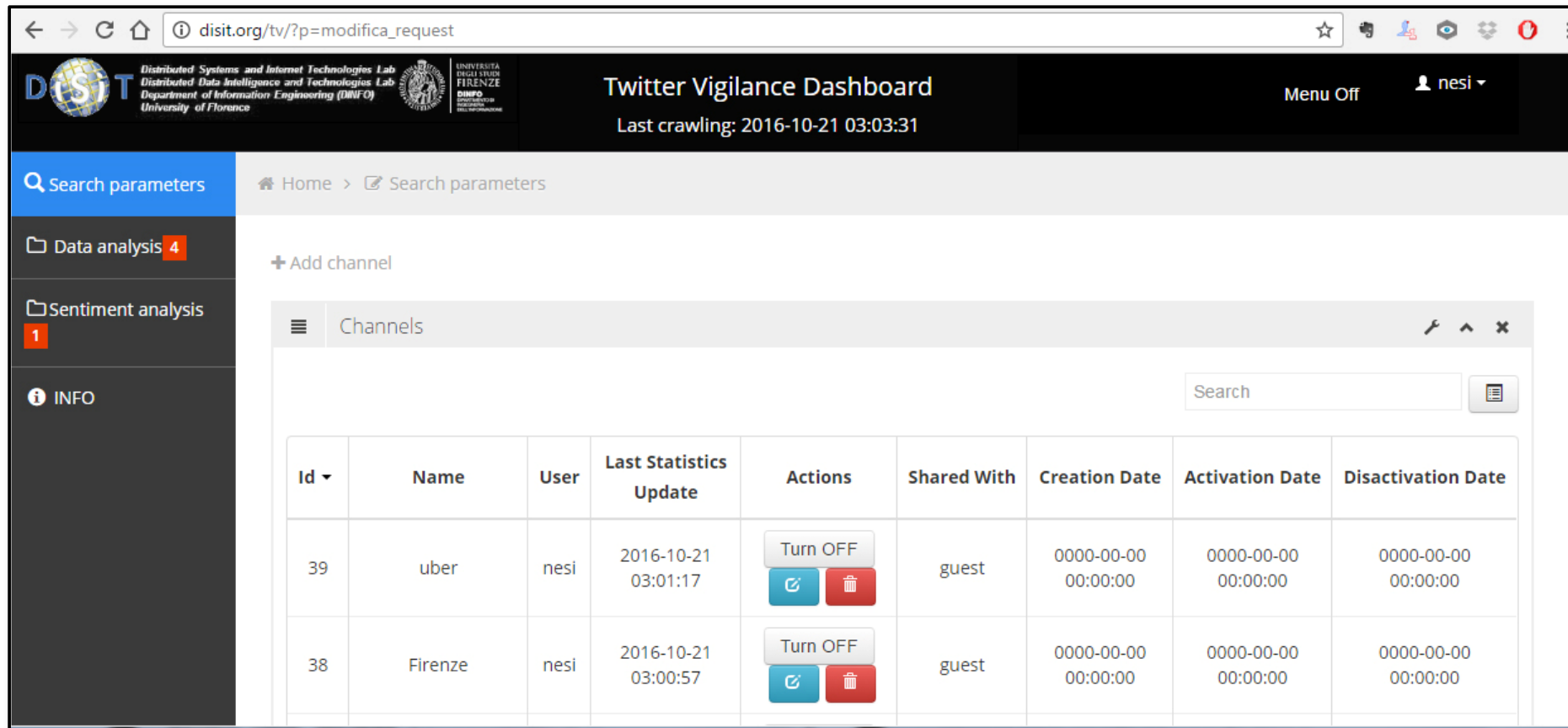


➤ Define and Mark events to improve readability

# TV Graphs – Channel View



# TV Graphs – Channel View



The screenshot shows the Twitter Vigilance Dashboard interface. The top navigation bar includes the DISIT logo, the text 'Distributed Systems and Internet Technologies Lab', 'Distributed Data Intelligence and Technologies Lab', 'Department of Information Engineering (DINFO)', and 'University of Florence'. The main title is 'Twitter Vigilance Dashboard' with a subtitle 'Last crawling: 2016-10-21 03:03:31'. A left sidebar contains navigation options: 'Search parameters', 'Data analysis 4', 'Sentiment analysis 1', and 'INFO'. The main content area shows a '+ Add channel' button and a 'Channels' table with a search input field.

Id	Name	User	Last Statistics Update	Actions	Shared With	Creation Date	Activation Date	Disactivation Date
39	uber	nesi	2016-10-21 03:01:17	Turn OFF [Share] [Delete]	guest	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00
38	Firenze	nesi	2016-10-21 03:00:57	Turn OFF [Share] [Delete]	guest	0000-00-00 00:00:00	0000-00-00 00:00:00	0000-00-00 00:00:00

- Add, Edit and Delete Channels
- Share Channels with Other Users
- Activate / Deactivate Channels
- See Channel Statistics

# TV Graphs – Channel Edit

**Twitter Vigilance Dashboard**  
Last crawling: 2016-10-21 03:03:31

Menu Off | nesi

Search parameters | Home > Search parameters

Channel name:   
Version:   
Share with:

ID	Status	Text	Language (ISO 639-1)
1	Active	#meteo	it
2	Disactive	#previsionimeteo #Firenze	it
3	Disactive	#meteo #neve #Firenze	it
6	Active	#ODDIT15 #Firenze	it
7	Active	#fodd	it
8	Active	#OpenDataDay #Firenze	
9	Active	@flash_meteo	it
10	Active	@firenzeditale	it
11	Active	@UNI_FIRENZE	it

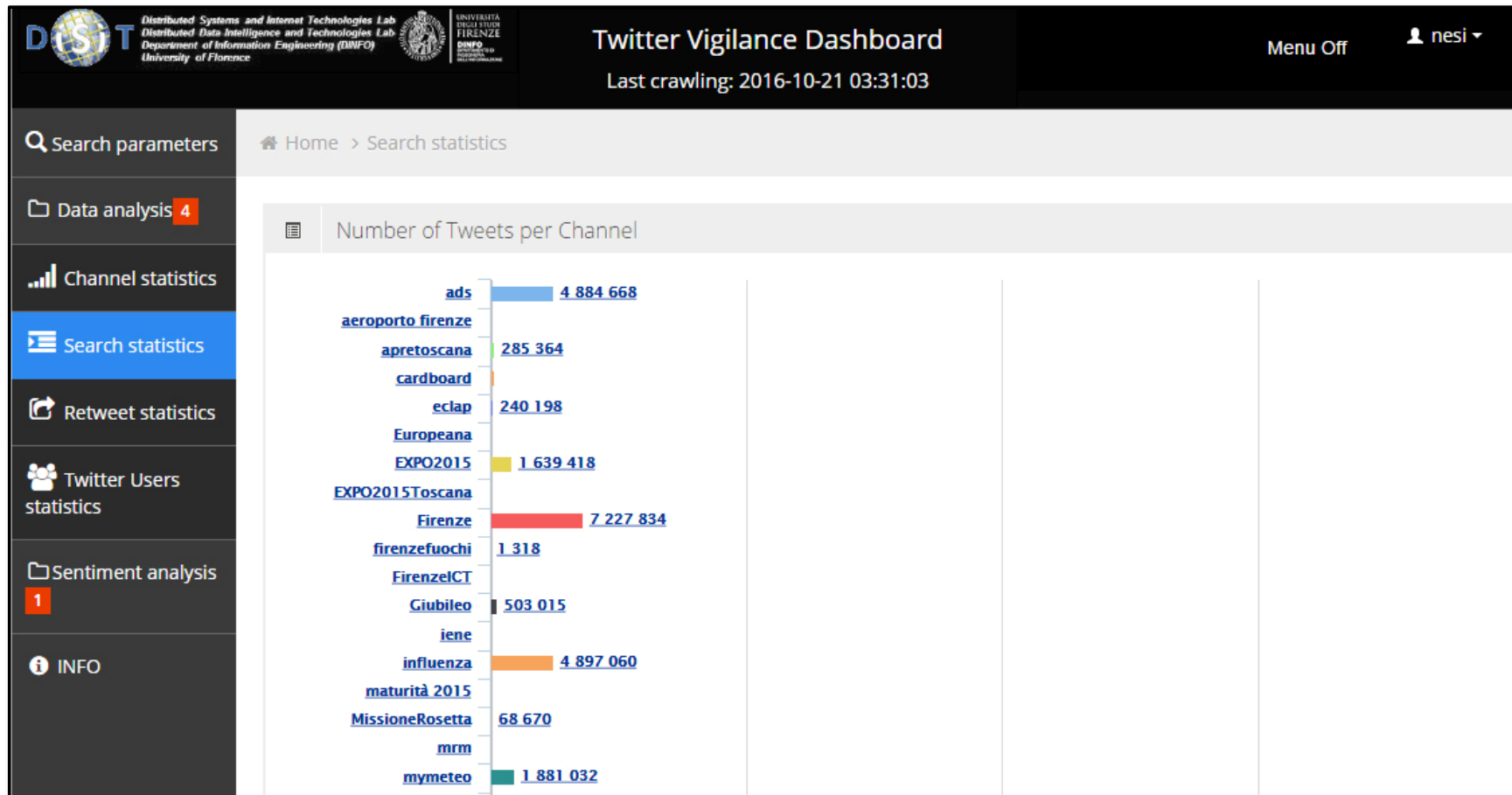
ID	Status	Text	Language (ISO 639-1)
490	Active	@Uber	
491	Active	@UberFacts	
492	Active	@uber_roma	
493	Active	@uber_firenze	
494	Active	@uber_italia	
495	Active	from:@uber	
496	Active	from:@uberfacts	
497	Active	#uber	

Showing 1 to 8 of 8 rows

➤ Add and /or Remove Searches for a Channel



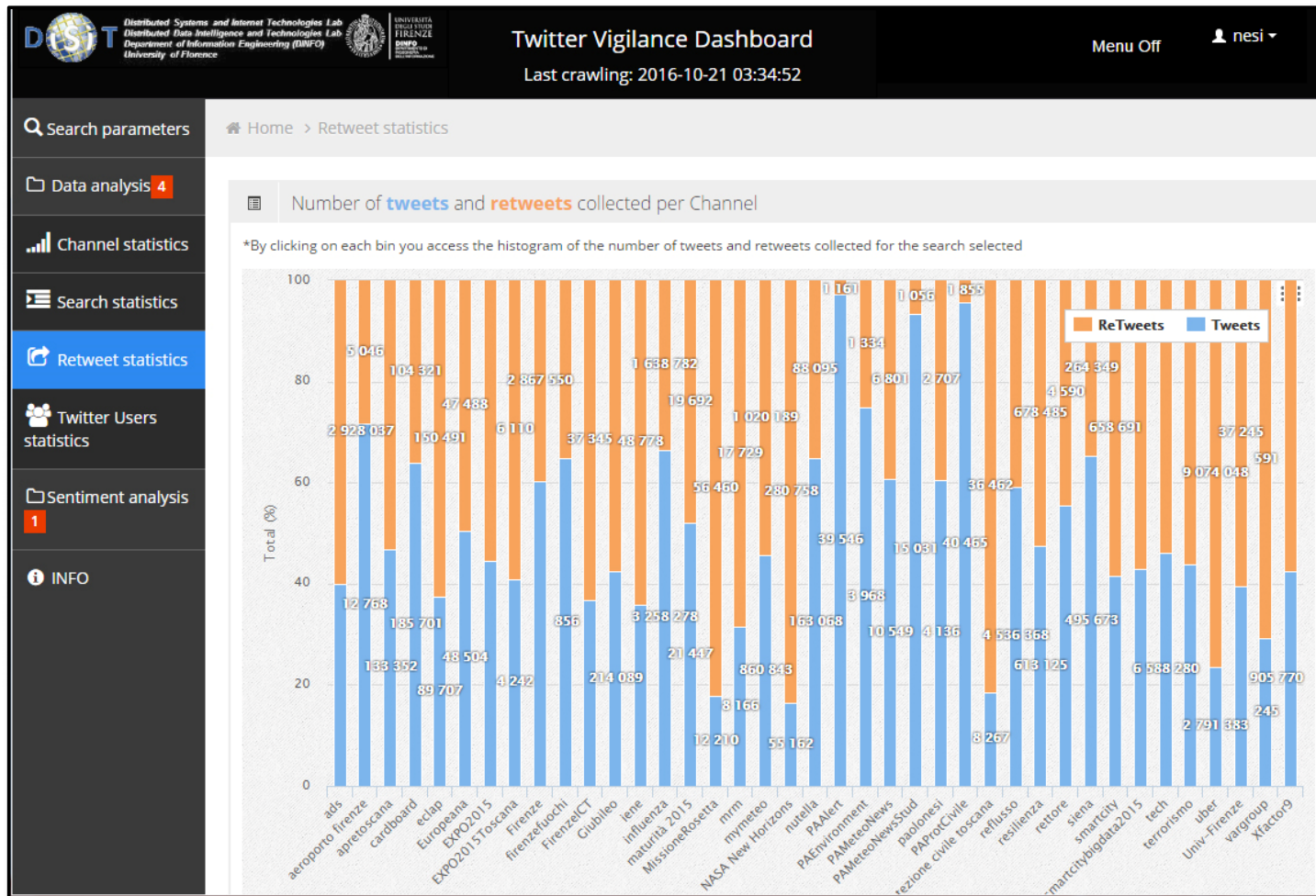
# TV Graphs – Channel Statistics



➤ Statistics Resume Charts for all Channels



# TV Graphs – Channel Statistics



➤ Number of tweets and retweets collected per Channel

# TV Graphs – Channel Statistics

Details active channels

uber

Channel ▾	Related research	Total	N° tweets	N° tweets(%)	N° retweets	N° retweets(%)	Details	Analysis
uber	<a href="#">#uber</a> <a href="#">@Uber</a> <a href="#">@UberFacts</a> <a href="#">@uber_firenze</a> <a href="#">@uber_italia</a> <a href="#">@uber_roma</a> <a href="#">from:@uber</a> <a href="#">from:@uberfacts</a>	11865431	2791383	23.53%	9074048	76.47%	From 2009-12-06 To today 	From 2015-09-09 To 2016-03-28 

➤ Channel Statistics: Total Number of Posts (Tweets + Retweets)

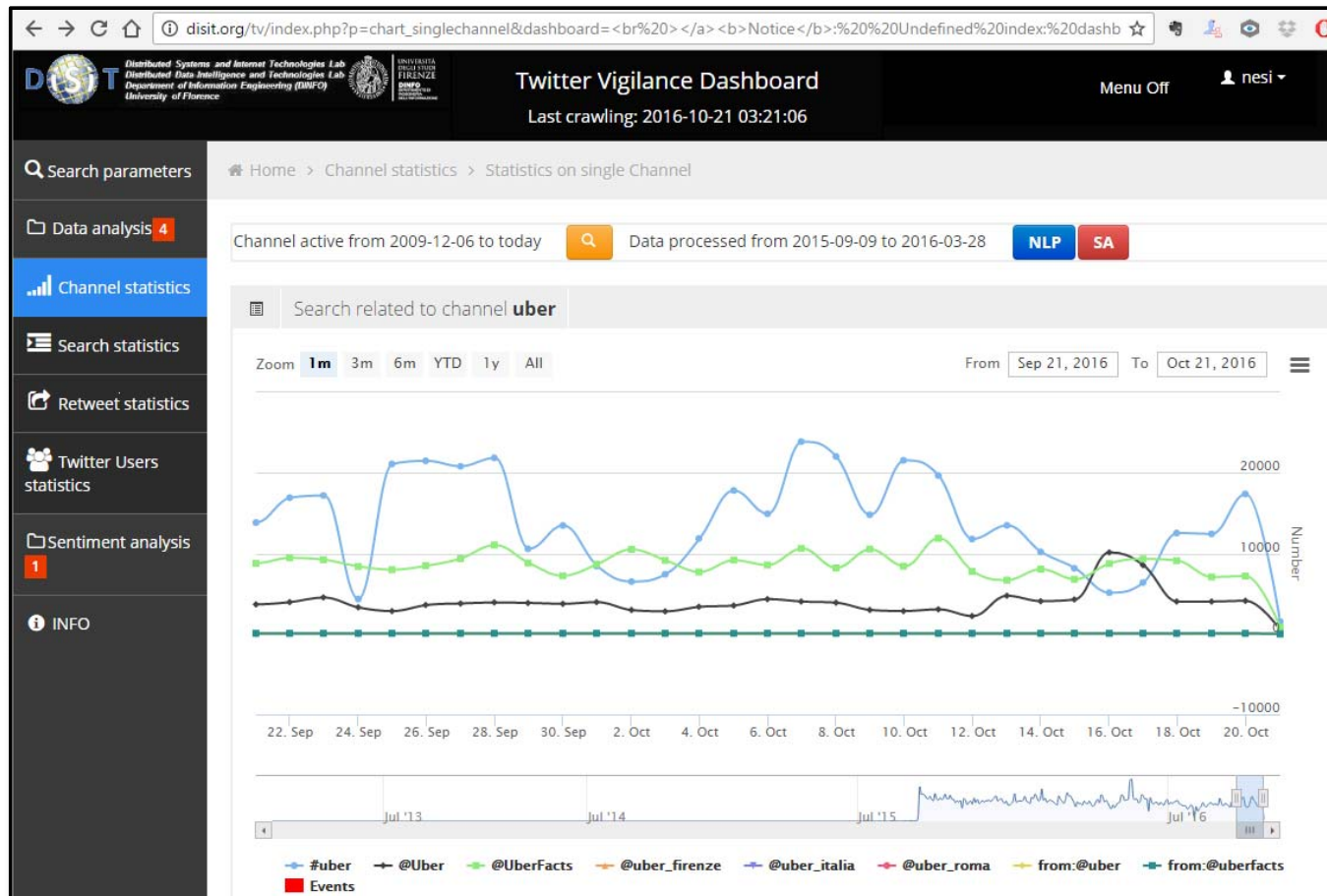


➤ Graphical details: charts, temporal trends of Tweets and Retweets for each Search of selected Channel...



➤ Natural Language Processing (NLP) and Sentiment Analysis (SA) of collected Tweets

# TV Graphs: Search View



- Display / Hide single Searches in a specific Channel
- Temporal Trends
- Zoom on Timeline

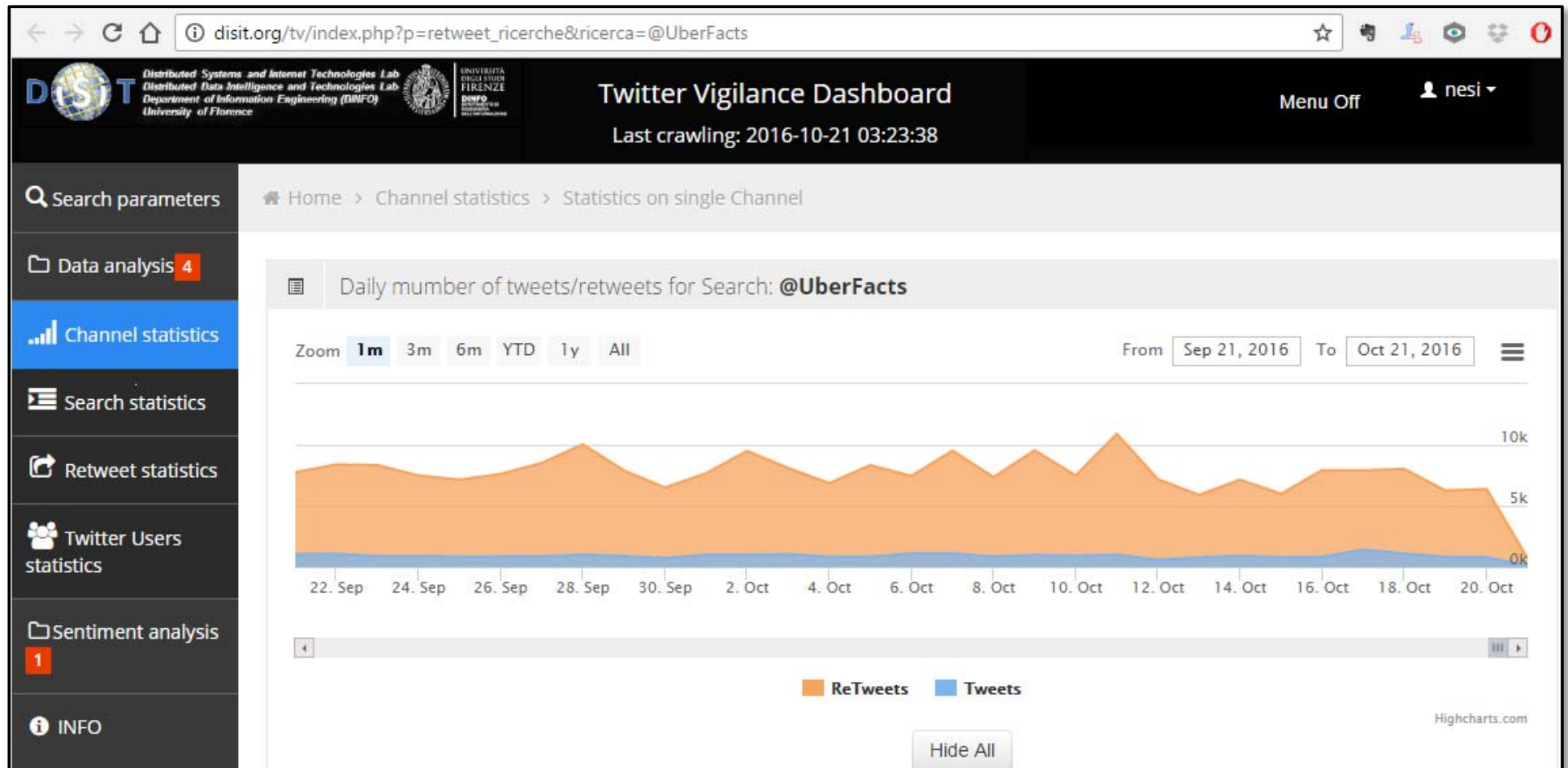


UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

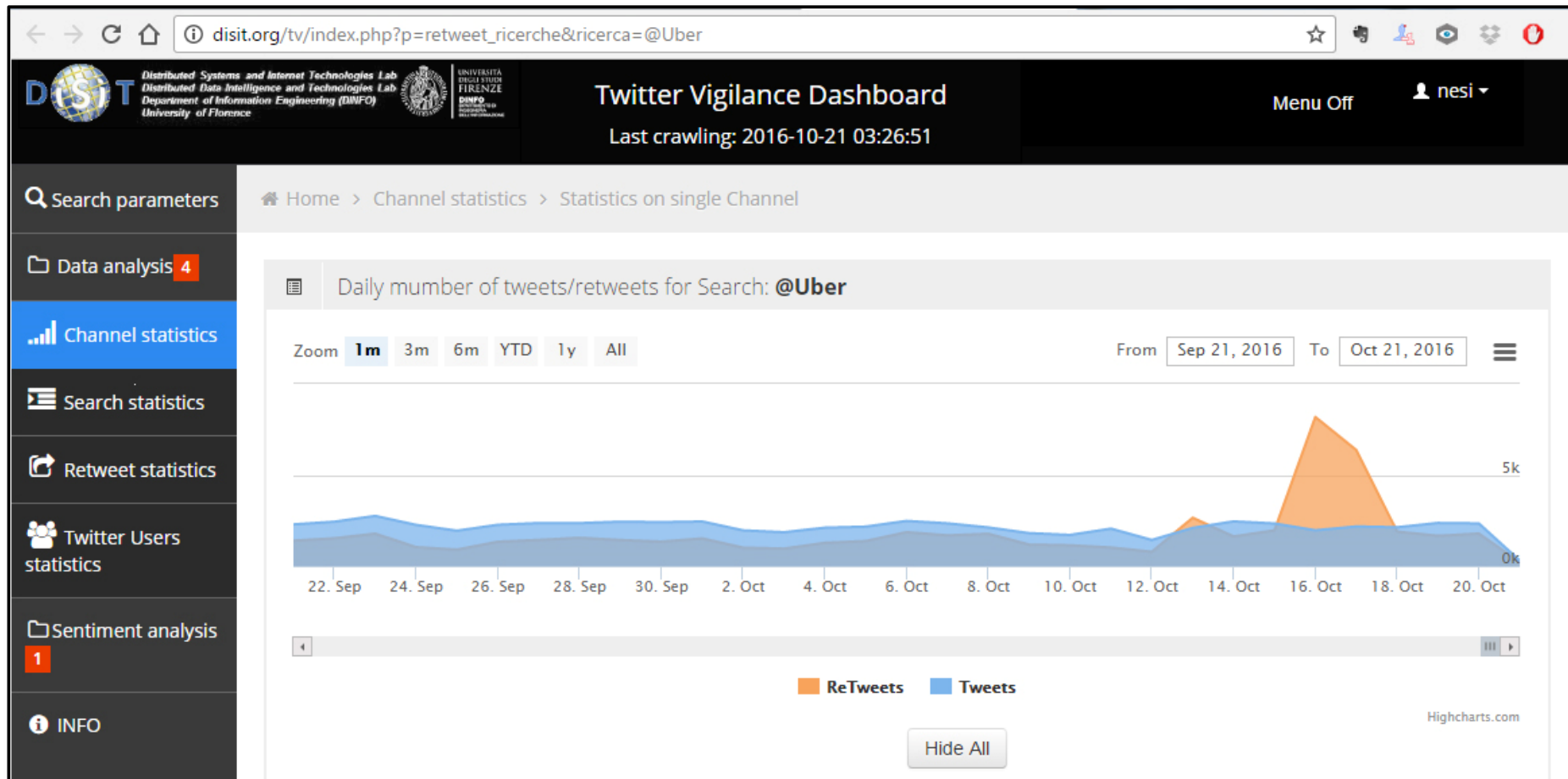
# TV Graphs – Single Search Details



**Twitter Vigilance**

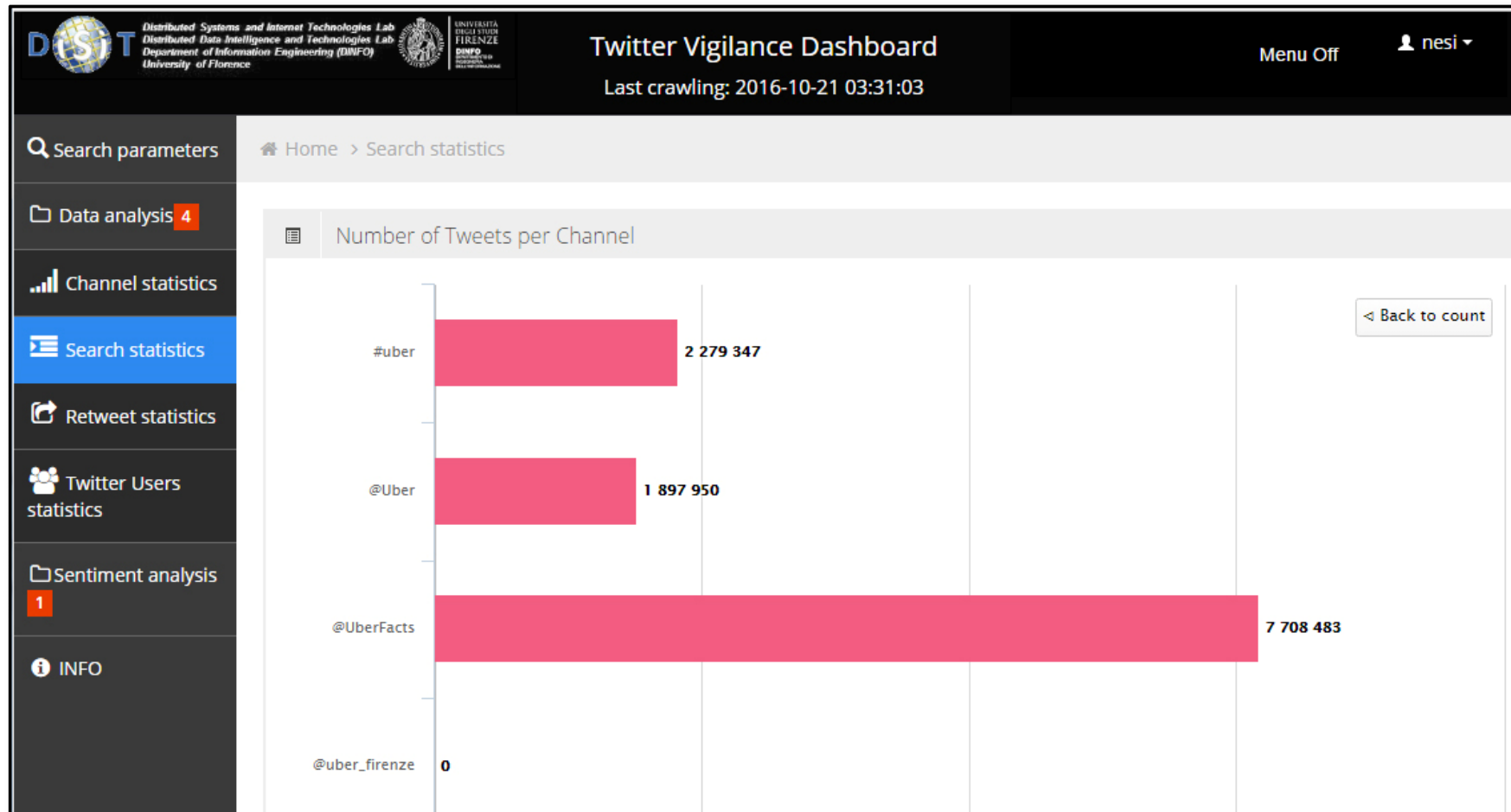
MASTER: Big Data Analytics And  
Technologies For Management - MABIDA

# Twitter Vigilance – Single Search Details



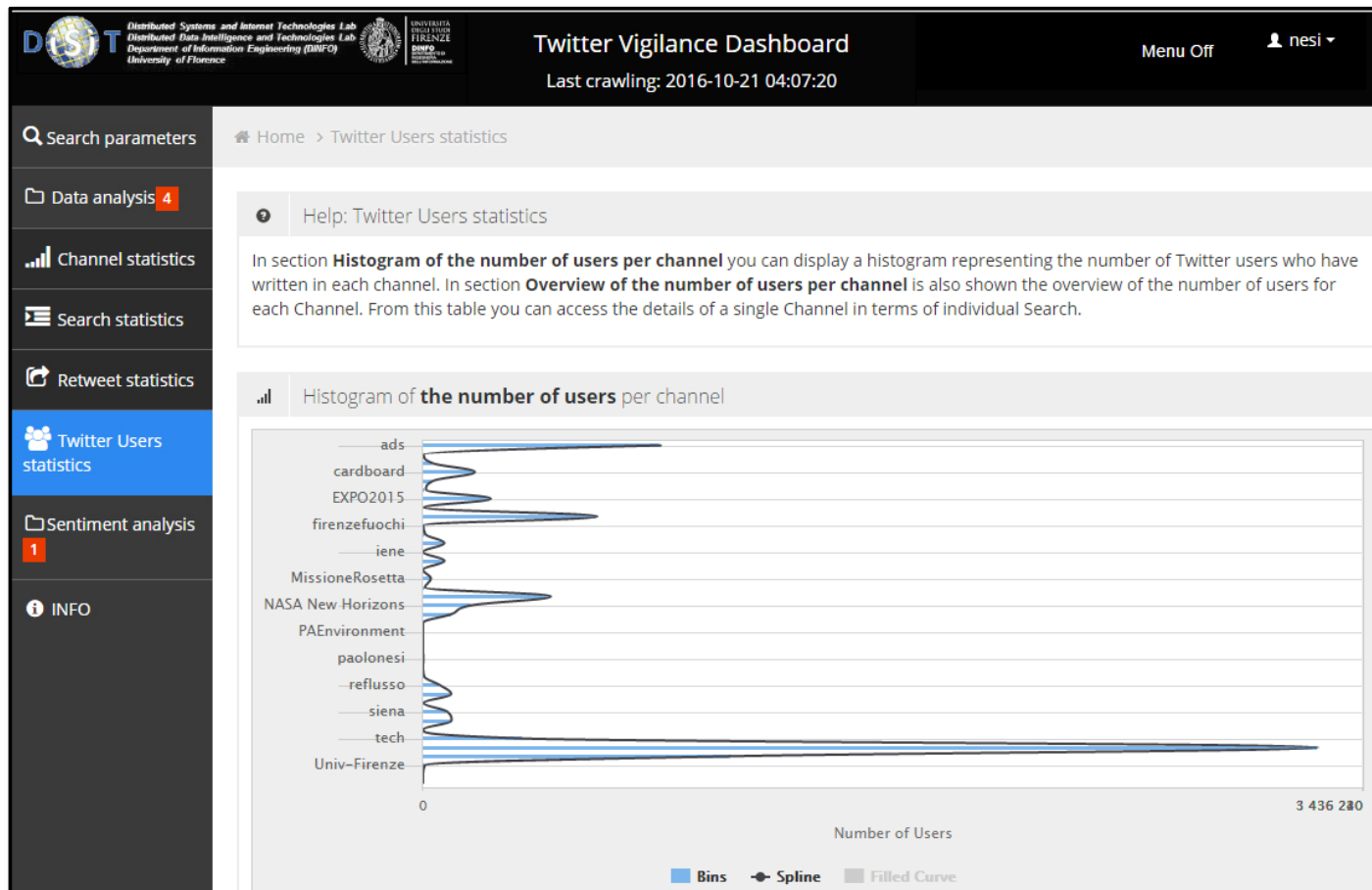
➤ Find and Retrieve Communication Peaks

# Twitter Vigilance – Search Statistics



- Statistics Resume charts for all the Searches of a single Channel

# Twitter Vigilance – User Statistics



- Statistics Resume charts for all the Users posting tweets / retweets for each Channel

# Twitter Vigilance – User Statistics

Overview of the number of users per channel

Channel	Number of Users	Details
terrorismo	3272573	[Details]
uber	1125597	[Details]
ads	868645	[Details]
Firenze	637456	[Details]
mymeteo	467649	[Details]
tech	365296	[Details]
EXPO2015	247650	[Details]
cardboard	189274	[Details]
NASA New Horizons	177771	[Details]
smartcity	104386	[Details]

Showing 1 to 10 of 39 rows | 10 records per page

Twitter Vigilance Dashboard  
Last crawling: 2016-10-21 04:12:06

Home > Twitter Users statistics > Details of the number of users grouped by search

Help: Details of the number of users grouped by search

In section **Number of users grouped by search** you can see in tabular form the distribution of twitter users within each Search associated with the Channel. From this table you can access the details of a single Search. In section **Tweets distribution of the first 10 active users in the channel** From this table you can access the names of the corresponding Twitter users. In section **3D pie of active Search on the Channel** ⇒ **Search** is shown in the form of a circular chart the percentage of distribution of the number of users for each searches for that channel.

Number of users grouped by search

Search	Number of Users	Details
@UberFacts	728394	[Details]
@Uber	250891	[Details]
#uber	144566	[Details]
@uber_italia	712	[Details]
@uber_roma	29	[Details]
from:@uber	1	[Details]

➤ Users' Details  
for each Channel



# Twitter Vigilance – User Statistics

Search parameters

Home > Twitter Users statistics > Details of the number of users grouped by search > Userlist per search

Help: Userlist per search

In section **Userlist channel** ⇒ **search** you can see in tabular form a list of twitter users regarding a single search and the number of tweets. From this table is also possible to access the *User profile*, the total number of tweets that specific user for each search / channel.

Userlist Channel **uber** ⇒ Search **@UberFacts**

User	Total Number of Tweets	Number of Tweets	Number of ReTweets	Profile
SteveStevens47	6817	15	6802	
LanaShyTV	6726	0	6726	

## ➤ Details for single users

Home > Twitter Users statistics > Details of the number of users grouped by search > Twitter User Profile

Definition of **Twitter User** profile

Is defined as user profile, the association of a twitter user with the number of tweets collected for each individual search. For each search are indicated channels to which it is associated.

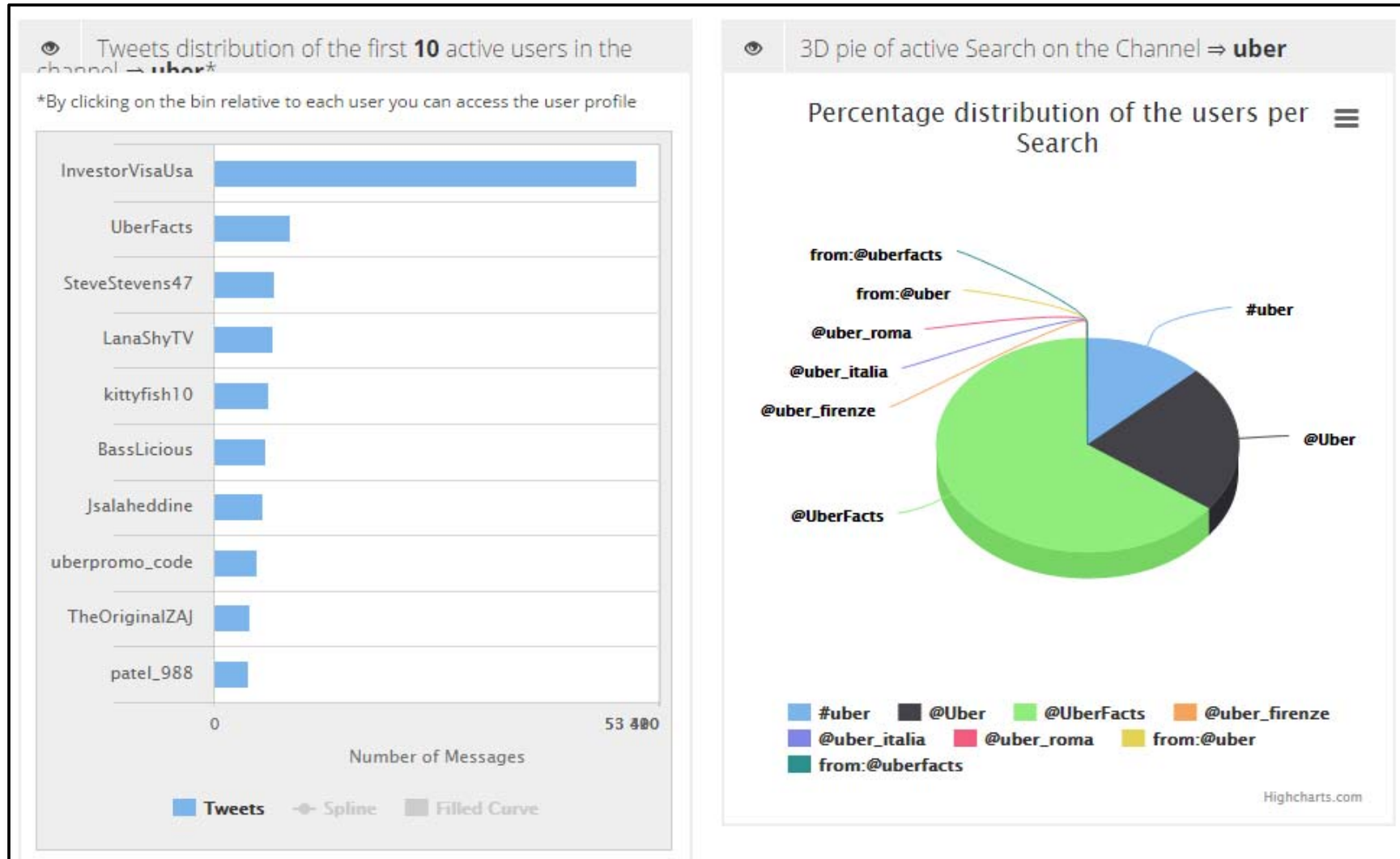
User profile: **SteveStevens47**

Profile Creation	Number of Favourites Tweets	Number of Followers	Number of Following	Number of List	Location	Numt
2013-02-13 20:36:52	230	1802	10	802	Bristol, England	

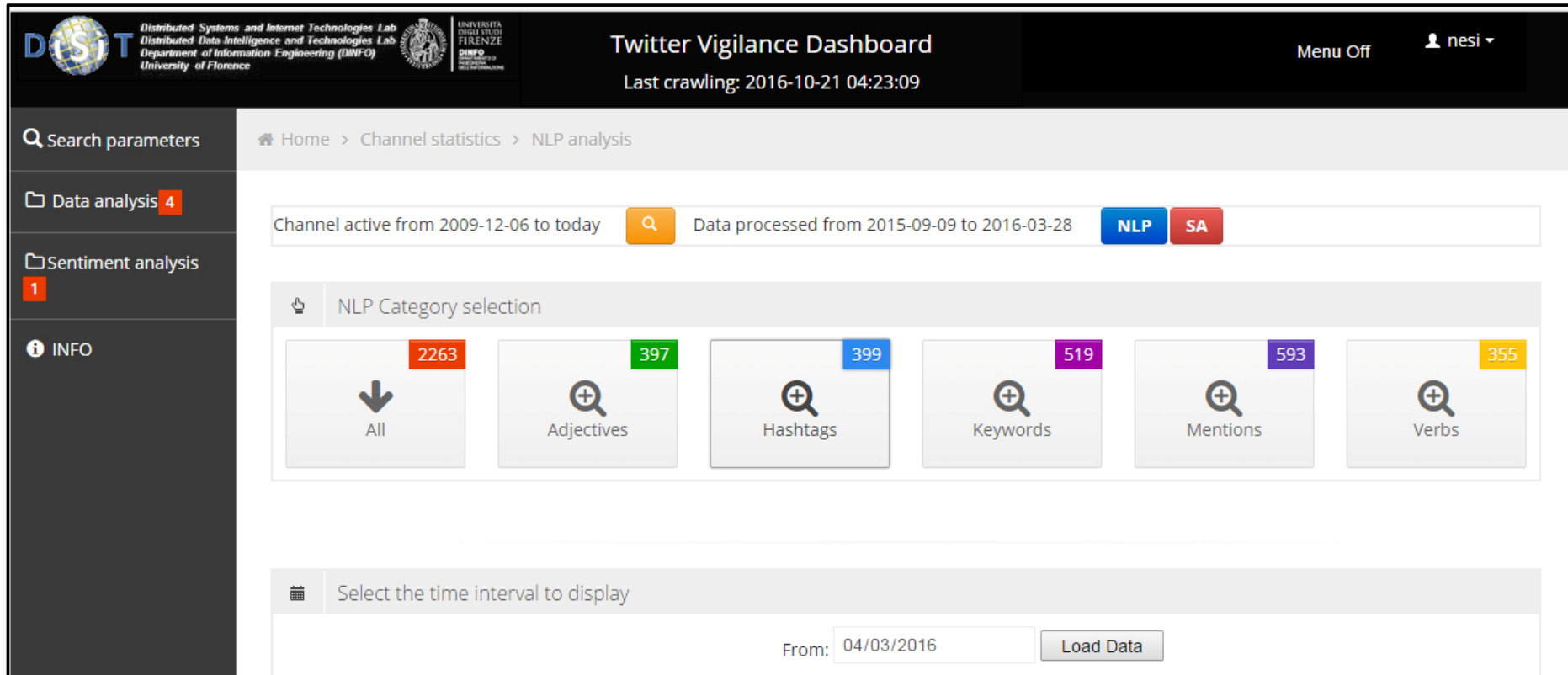
Search

Search	Associated Channels	Number of Tweets+ReTweets	Number of Tweets	Number of ReTweets
@UberFacts	uber	6817	15	6802
neve	mymeteo	73	0	73
bomb	terrorismo	35	0	35

# Twitter Vigilance – User Statistics



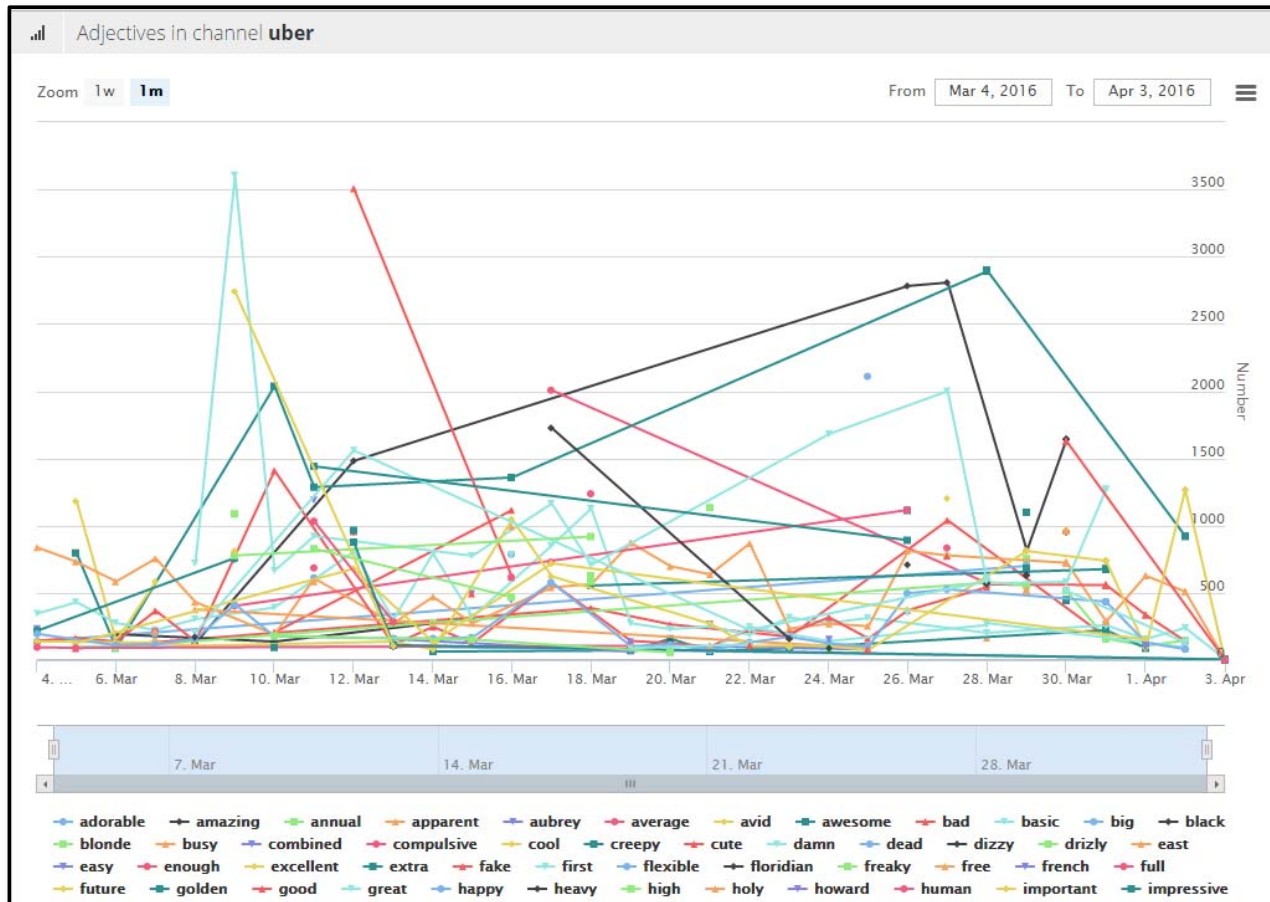
# Twitter Vigilance – NLP & SA



The screenshot displays the Twitter Vigilance Dashboard interface. The top navigation bar includes the DISIT logo, the lab name 'Distributed Systems and Internet Technologies Lab', the department 'Distributed Data Intelligence and Technologies Lab', and the university 'University of Florence'. The dashboard title is 'Twitter Vigilance Dashboard' with a user profile 'nesi' and a 'Menu Off' option. The last crawling time is '2016-10-21 04:23:09'. The main content area shows a breadcrumb trail: 'Home > Channel statistics > NLP analysis'. Below this, there are search parameters for 'Channel active from 2009-12-06 to today' and 'Data processed from 2015-09-09 to 2016-03-28', with buttons for 'NLP' and 'SA'. The 'NLP Category selection' section features six buttons: 'All' (2263), 'Adjectives' (397), 'Hashtags' (399), 'Keywords' (519), 'Mentions' (593), and 'Verbs' (355). At the bottom, there is a 'Select the time interval to display' section with a 'From: 04/03/2016' input and a 'Load Data' button.

- Natural Language Processing (NLP) of Tweets content.

# Twitter Vigilance – NLP & SA



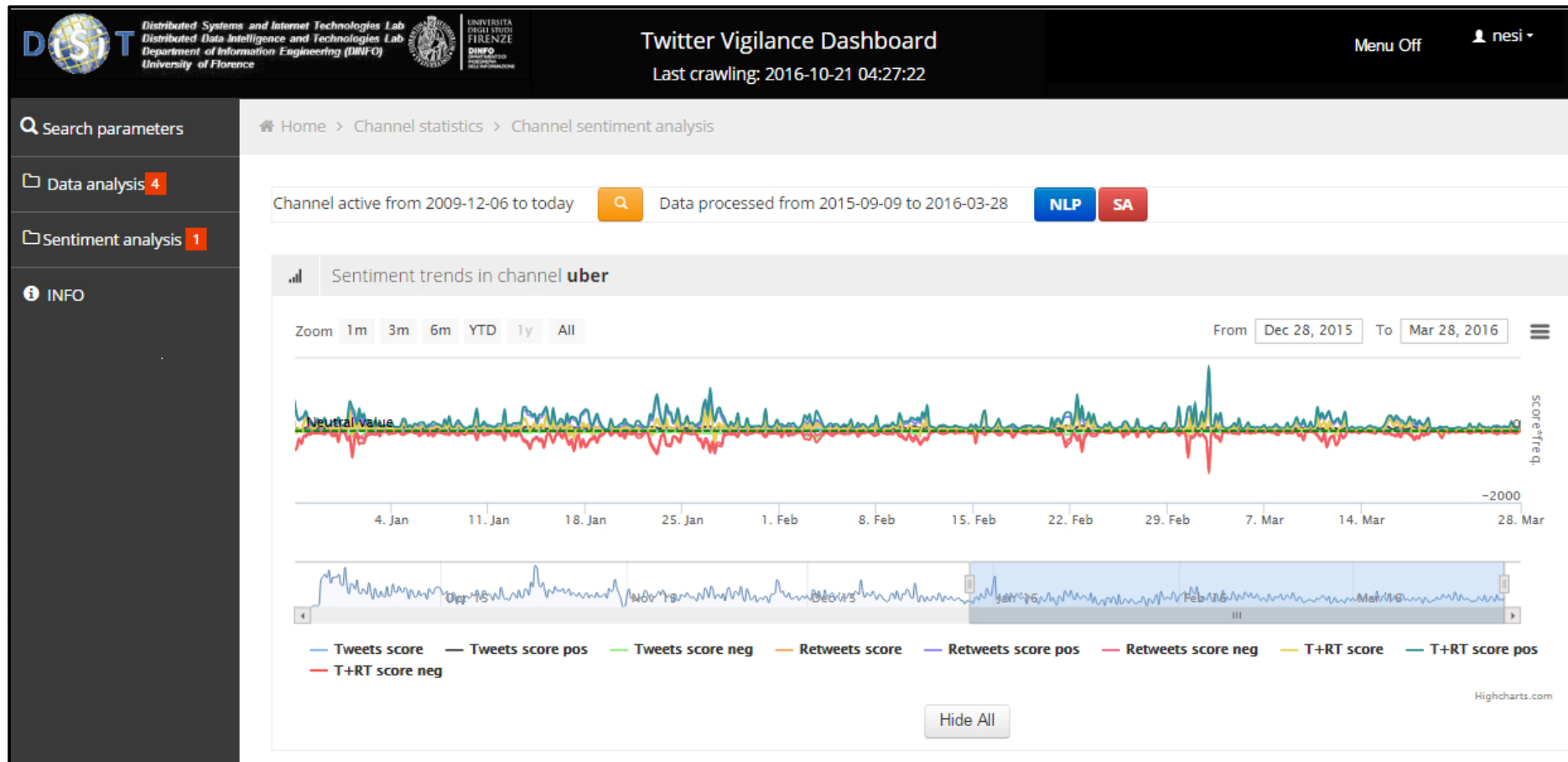
Different NLP categories:

➤ **Keywords (nouns)**

➤ **Adjectives**

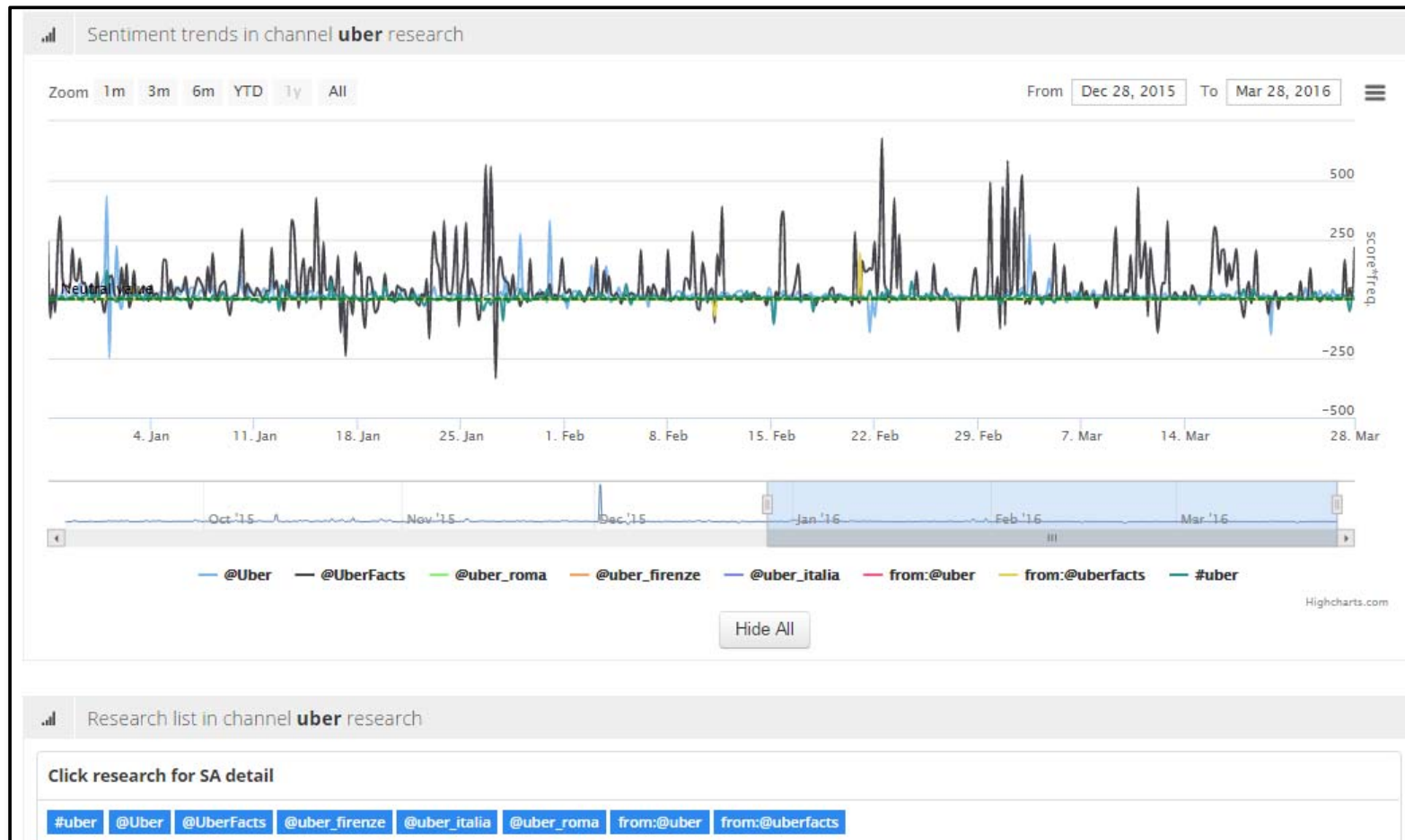
➤ **Verbs**

# Twitter Vigilance – NLP & SA



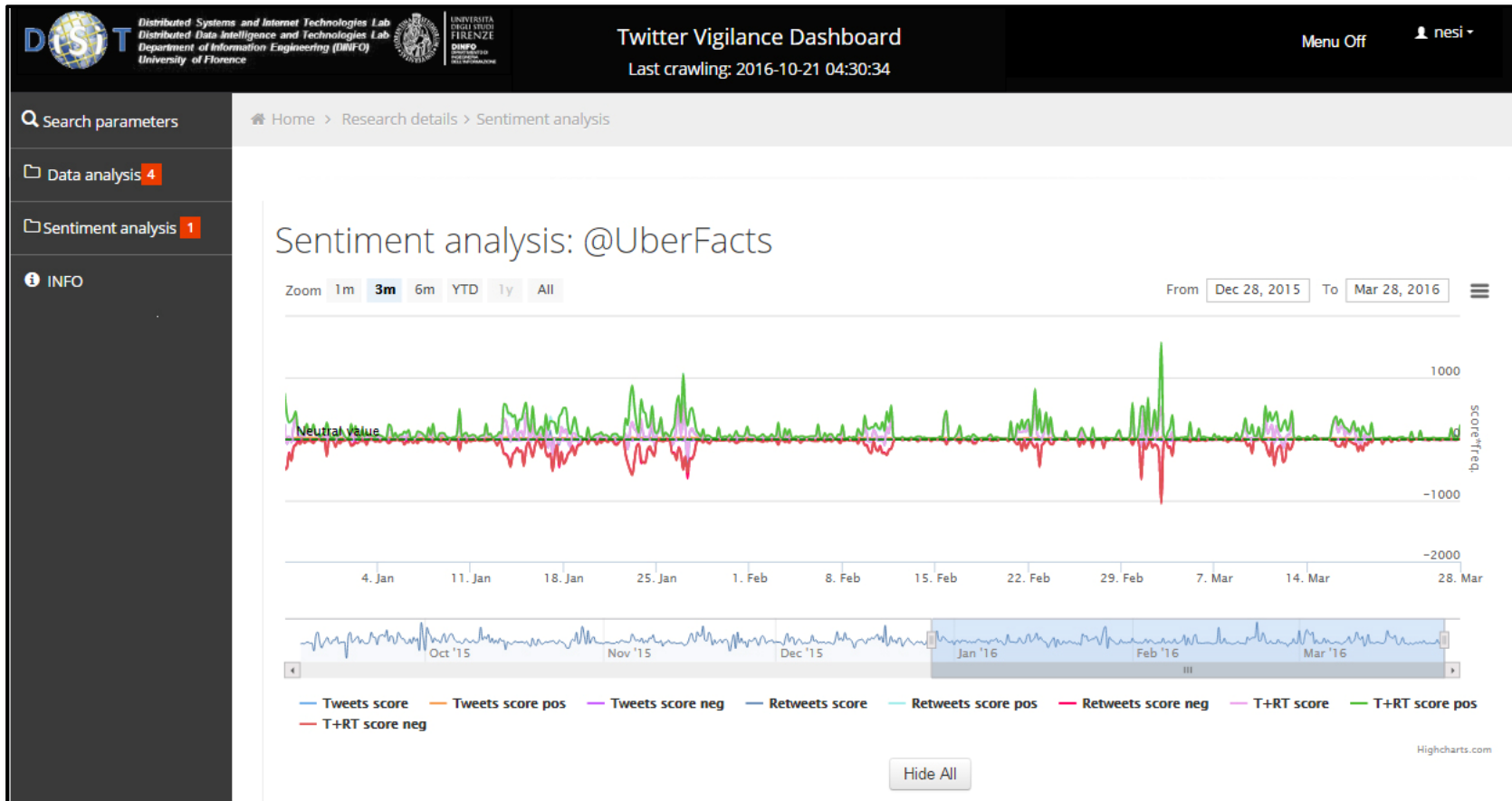
➤ Sentiment Analysis (SA) of Tweets content at Channel Level.

# Twitter Vigilance – NLP & SA



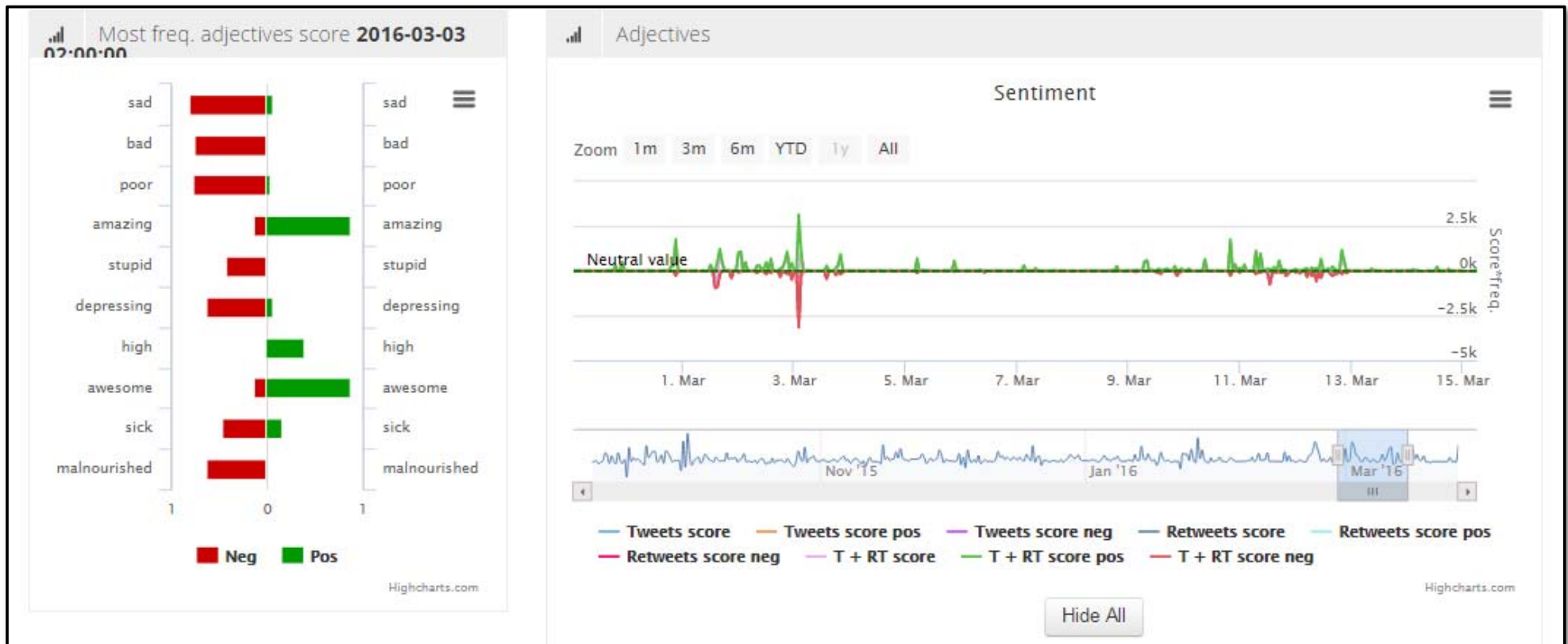
- Sentiment Analysis (SA) of Tweets content for each Search of a specific Channel.

# Twitter Vigilance – NLP & SA



- Sentiment Analysis (SA) of Tweets content at Search Level.

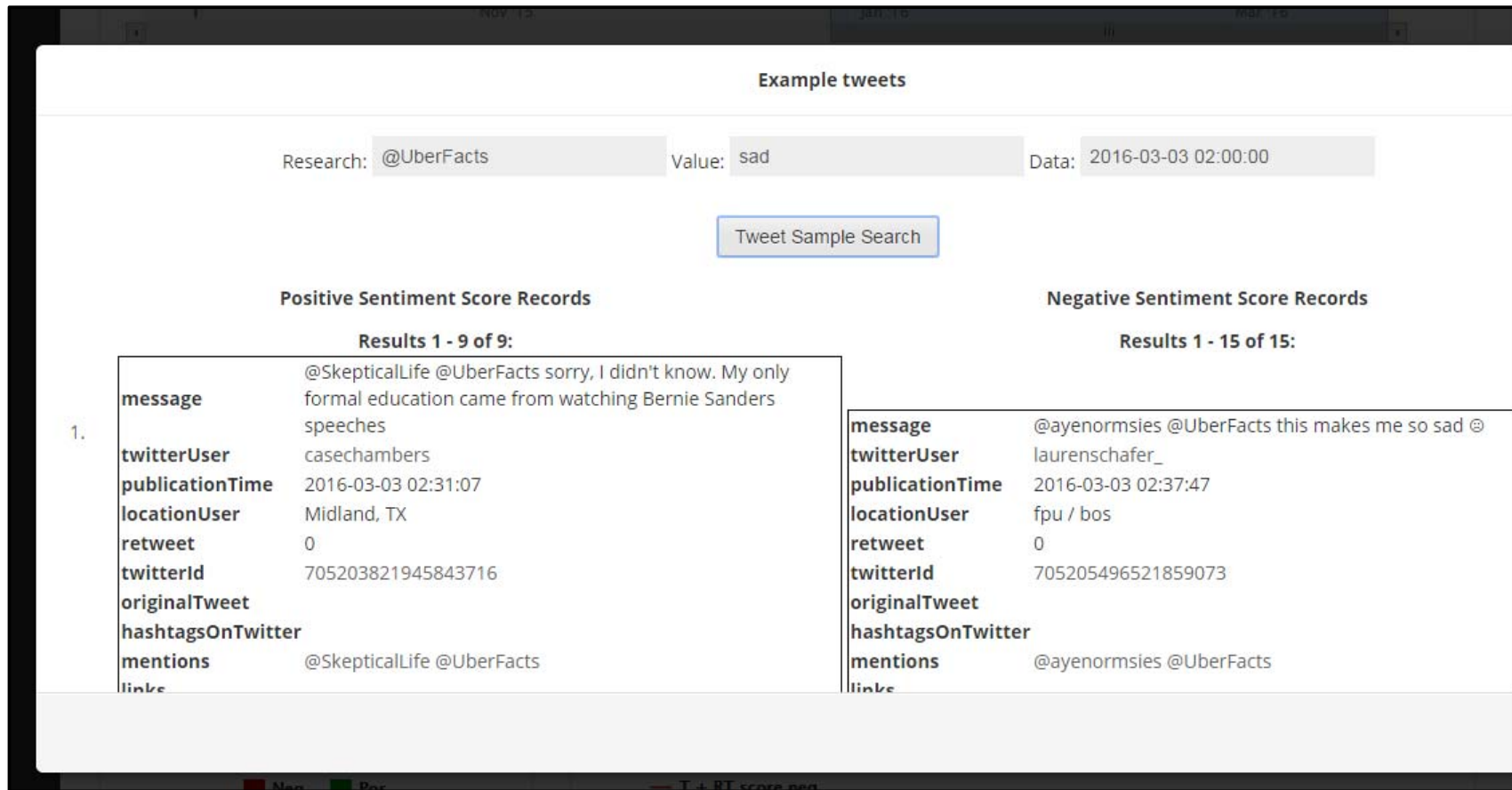
# Twitter Vigilance – NLP & SA



- Sentiment Analysis (SA) at Search level specifically for Keywords, Adjectives and Verbs.



# Twitter Vigilance – NLP & SA



Example tweets

Research: @UberFacts Value: sad Data: 2016-03-03 02:00:00

Tweet Sample Search

**Positive Sentiment Score Records**

Results 1 - 9 of 9:

1.	message	@SkepticalLife @UberFacts sorry, I didn't know. My only formal education came from watching Bernie Sanders speeches
	twitterUser	casechambers
	publicationTime	2016-03-03 02:31:07
	locationUser	Midland, TX
	retweet	0
	twitterId	705203821945843716
	originalTweet	
	hashtagsOnTwitter	
	mentions	@SkepticalLife @UberFacts
	links	

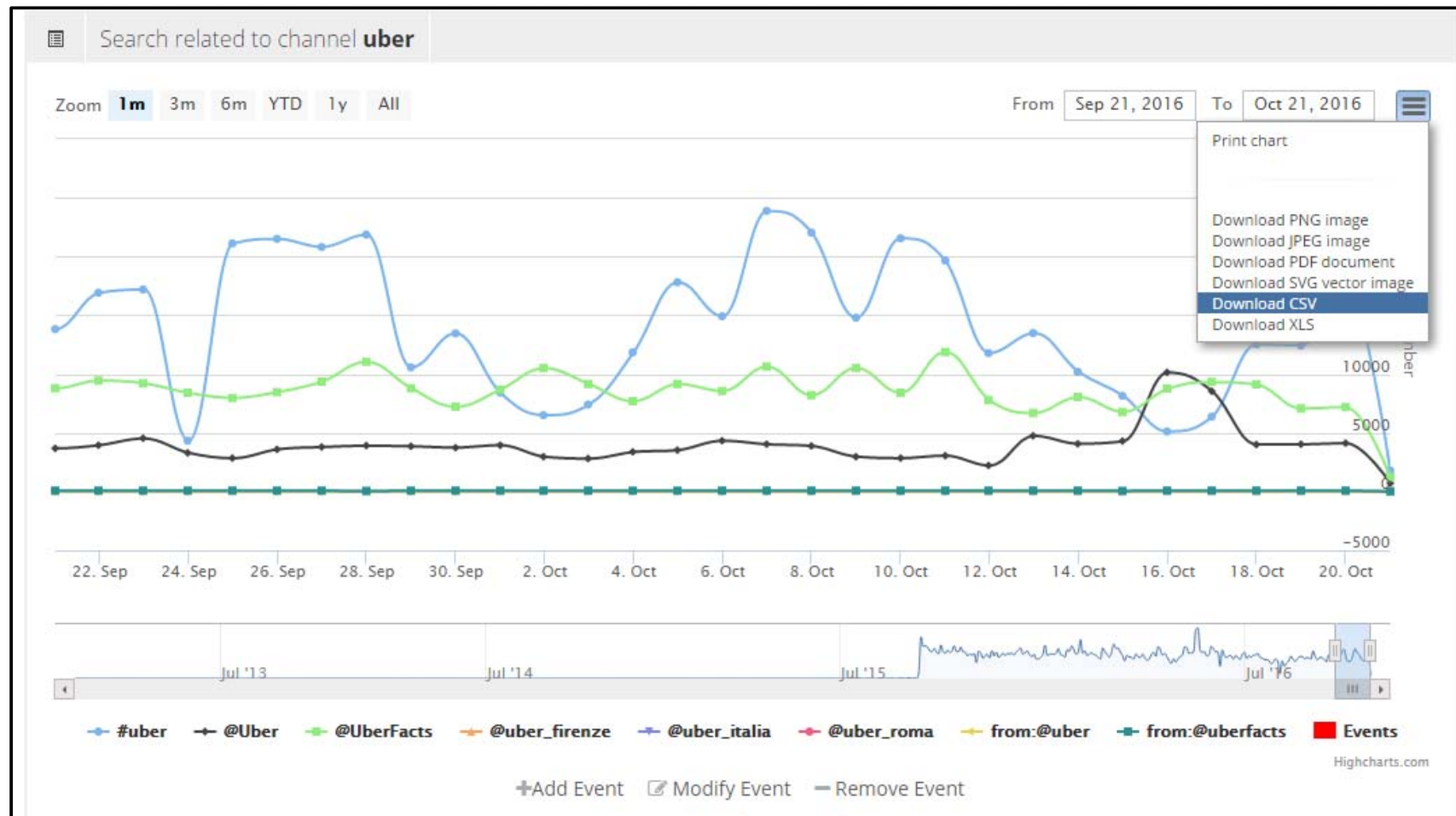
**Negative Sentiment Score Records**

Results 1 - 15 of 15:

	message	@ayenormsies @UberFacts this makes me so sad ☹
	twitterUser	laurenschafer_
	publicationTime	2016-03-03 02:37:47
	locationUser	fpu / bos
	retweet	0
	twitterId	705205496521859073
	originalTweet	
	hashtagsOnTwitter	
	mentions	@ayenormsies @UberFacts
	links	

- Show samples and details of those Tweets which most have effects upon the Sentiment Analysis final computation results.


# Twitter Vigilance – Data Download



- It is possible to download data for further processing and analysis
- It is possible to specify the date time interval



# Agenda

1. Introduction: What is Twitter
  2. Overview on Twitter Data Capabilities
  3. Twitter Vigilance how it is working
  4. Exercitation Work Plan
- 
- A thick red arrow originates from the right side of the slide and points diagonally down and left towards the text '4. Exercitation Work Plan'.



# Facciamo le Esercitazioni

- 5 Gruppi
- 3-4 Persone per gruppo con Skill Diverso
  - noi facciamo I gruppi
- Compiti diversi:
  - Uber:.....
  - TPL:.....
  - Firenze “Turismo”:.....
  - Xfactor:.....
  - Firenze “Meteo”:.....



- Accedendo ai dati e all'account di TV per il vostro elaborato accettate le seguenti condizioni:
  - Non siete autorizzati a
    - comunicare a terzi i dati che ottenete da Twitter Vigilance in nessuna forma, anche se elaborati (non potete passarli neanche ai docenti degli altri corsi)
    - Passare a terzi la vostra password
    - Pubblicare i risultati delle vostre elaborazioni, se non in agreement con il DISIT lab.
  - Siete responsabili di cosa accade dentro il vostro account che avrà una durata limitata al corso
  - Ci autorizzate a pubblicare i risultati che ottenete in ogni forma ci possa essere utile, riconoscendovi un acknowledgement (ringraziamento)



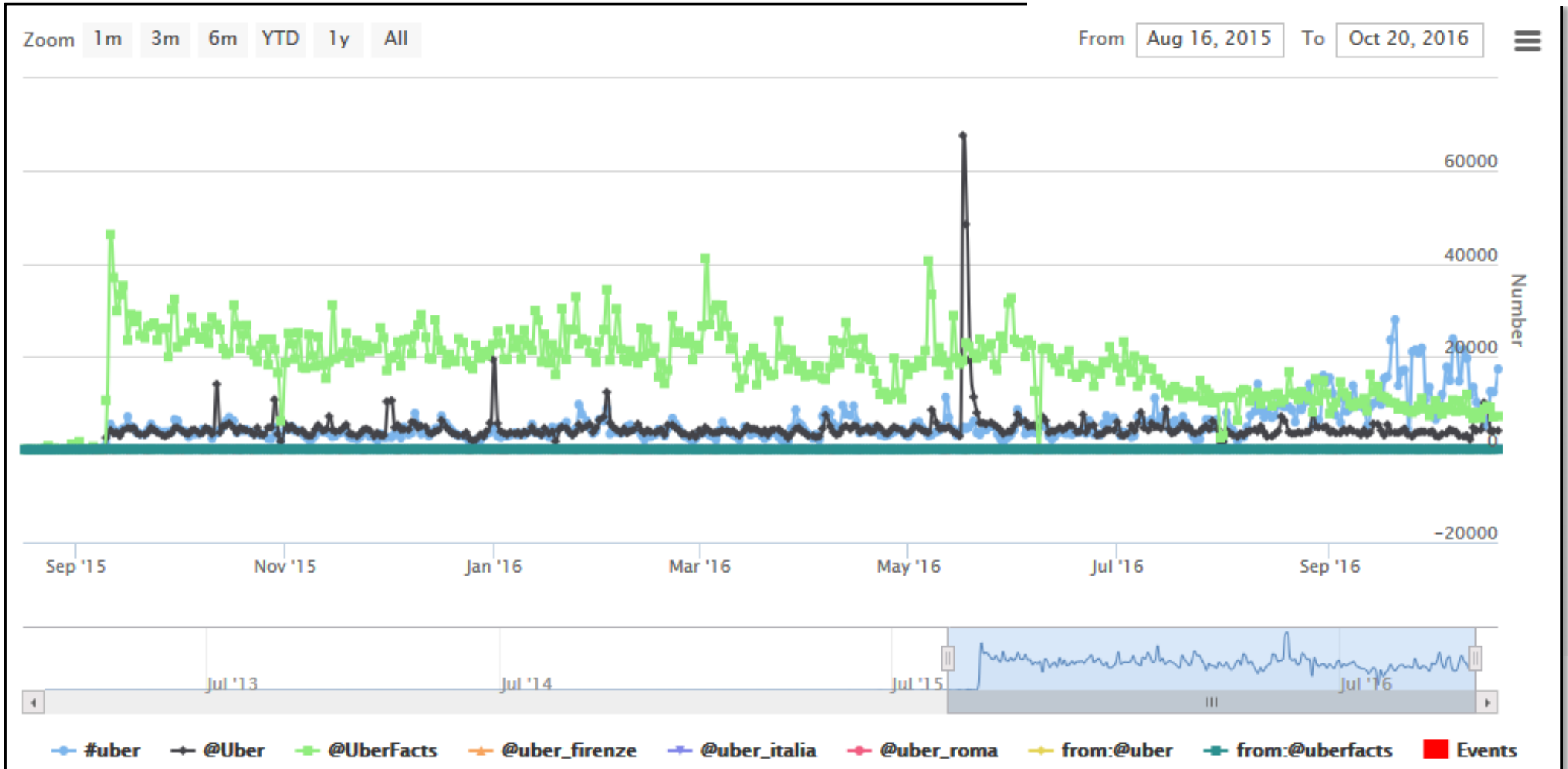
UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

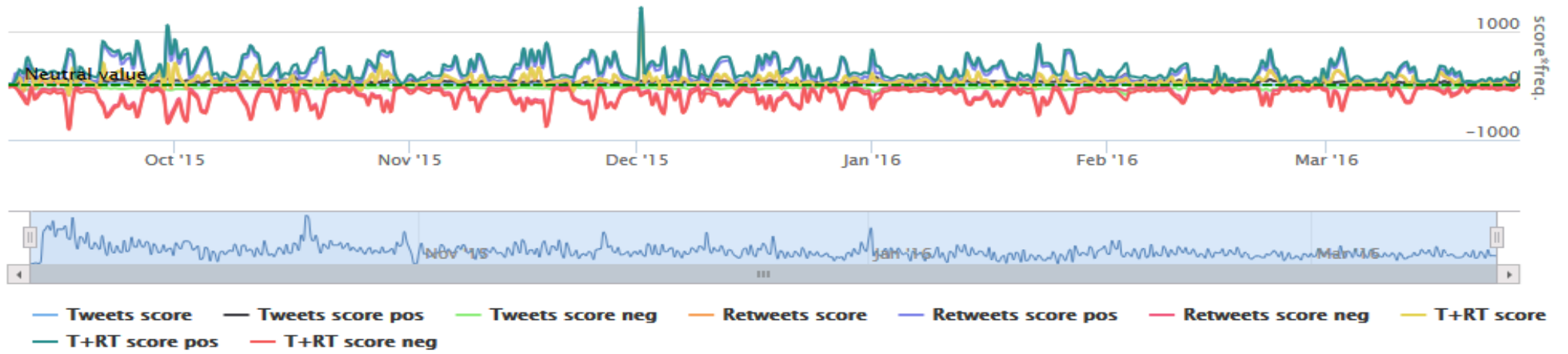
#uber | @Uber | @UberFacts  
 @uber\_firenze | @uber\_italia  
 @uber\_roma | from:@uber  
 from:@uberfacts

# Canale "Uber"



Zoom 1m 3m 6m YTD 1y All

From Sep 9, 2015 To Mar 28, 2016



Highcharts.com

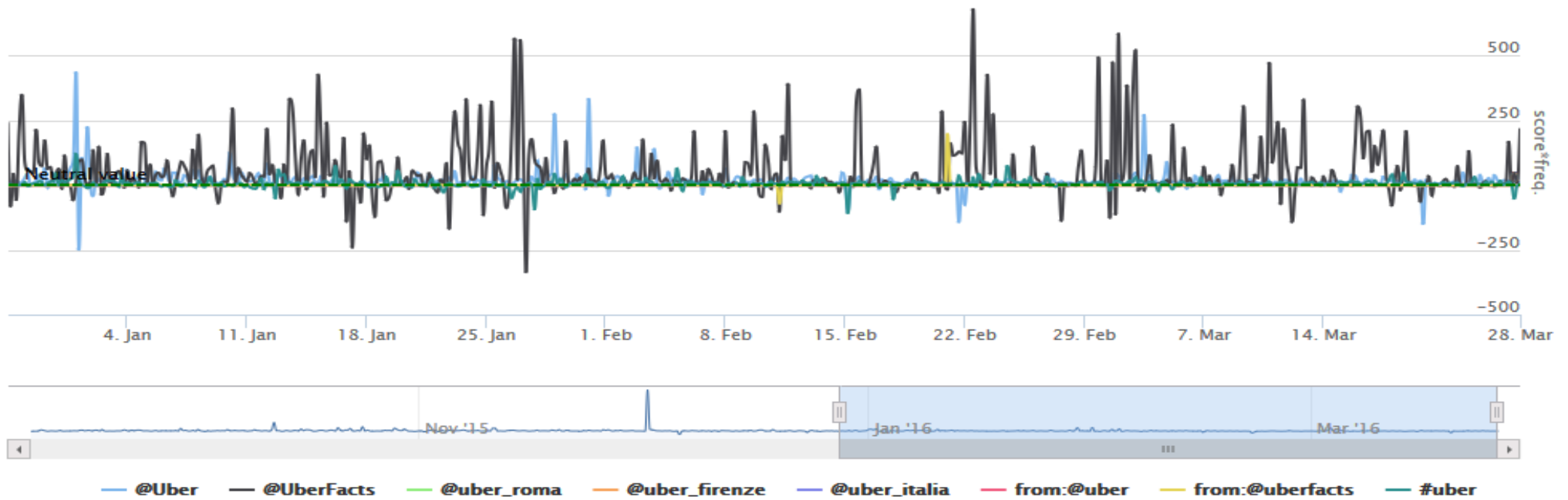
Hide All

# Uber Sentiment Analysis

Sentiment trends in channel uber research

Zoom 1m 3m 6m YTD 1y All

From Dec 28, 2015 To Mar 28, 2016





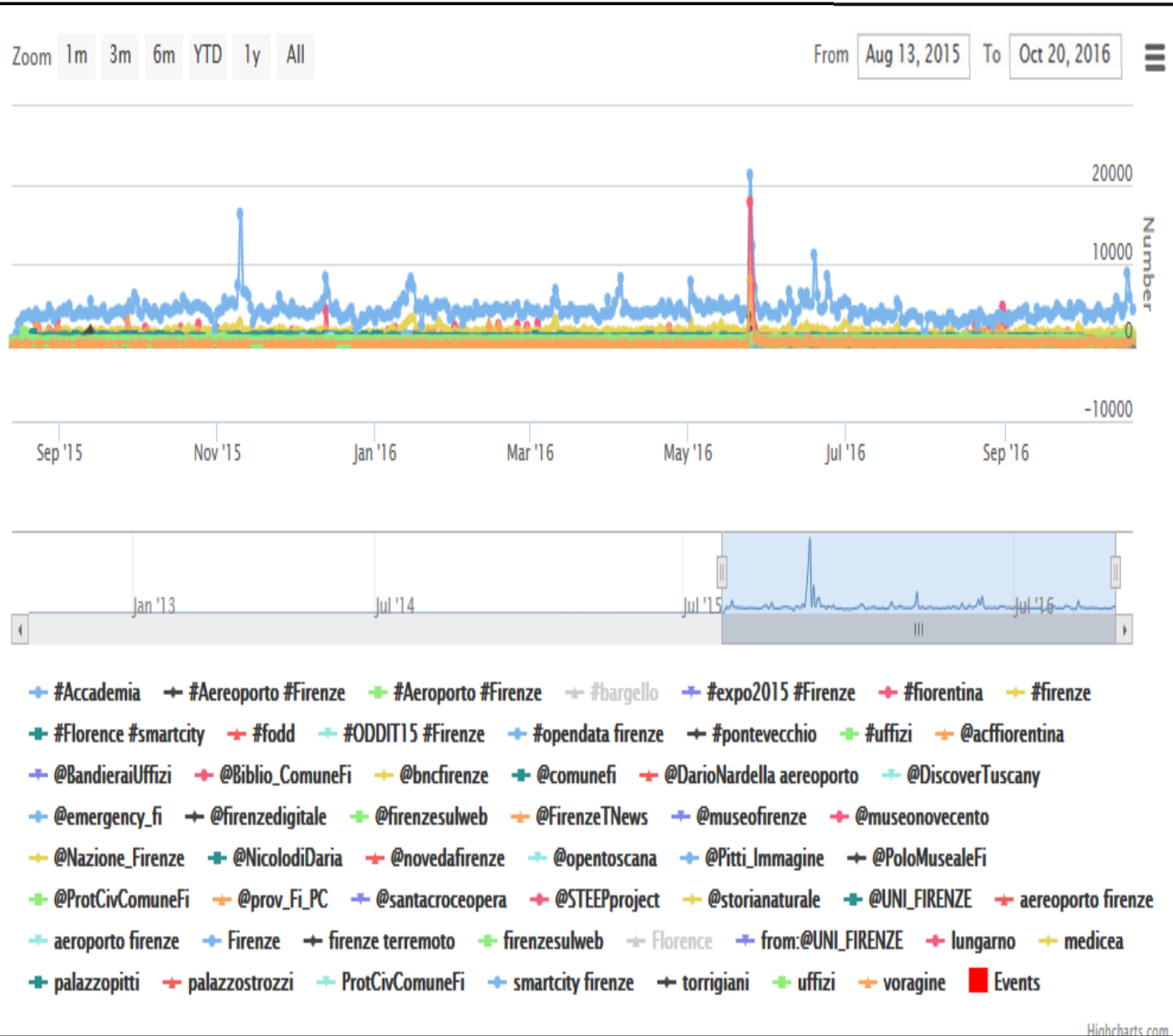
# Canale “Uber”

- 11 Milioni di Tweet
- L'azienda e/o il servizio stanno perdendo l'attenzione del mercato
- Cosa ha determinato alcuni dei picchi ?
  - Potevano essere previsti ?
- Identificazione delle Sorgenti di rumore
- Identificazione di nuove Keywords/Search





# Canale "Firenze"



- #Accademia #Aeroporto #Firenze
- #Aeroporto #Firenze #bargello
- #expo2015 #Firenze #fiorentina #firenze
- #Florence #smartcity #fodd
- #ODDIT15 #Firenze #opendata firenze
- #pontevecchio #uffizi @acffiorentina
- @BandieraiUffizi @Biblio\_ComuneFi
- @bncfirenze @comunefi
- @DarioNardella aeroporto
- @DiscoverTuscany @emergency\_fi
- @firenzedigitale @firenzesulweb
- @FirenzeTNews @museofirenze
- @museonovecento @Nazione\_Firenze
- @NicolodiDaria @novedafirenze
- @opentoscana @Pitti\_Immagine
- @PoloMusealeFi @ProtCivComuneFi
- @prov\_Fi\_PC @santacroceopera
- @STEEPproject @storianaturale
- @UNI\_FIRENZE aeroporto firenze
- aeroporto firenze Firenze
- firenze terremoto firenzesulweb Florence
- from:@UNI\_FIRENZE lungarno medicea
- palazzopitti palazzostrozzi
- ProtCivComuneFi smartcity firenze
- torrigiani uffizi voragine

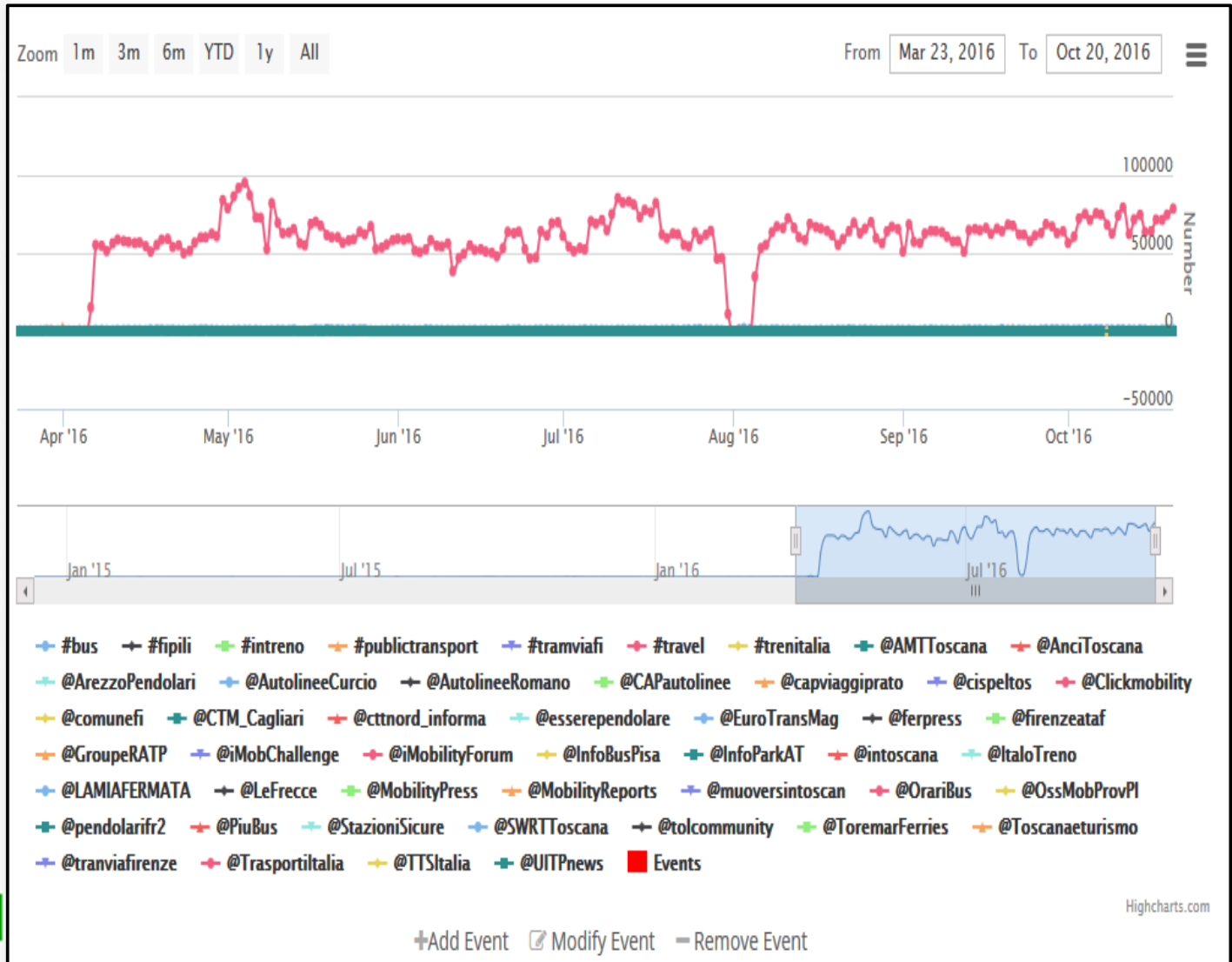


- 7 Milioni di Tweets
- Identificazione delle Sorgenti di rumore
- Identificazione di Keywords/Search giuste, e non inserire le altre
- Quali sono le metriche correlate con i servizi piu' ricercati
  - Cosa cercano I turisti, dove vanno, cosa fanno
  - Cosa non gli e' piaciuto ?, Cosa gli e' piaciuto di piu' ?
  - Perche tornano ?
- I cittadini cosa dicono?



- Canale “TPL”, 12 M tweet

- #bus #fipili #intreno
- #publictransport #tramviafi
- #travel #trenitalia @AMTToscana
- @AnciToscana @ArezzoPendolari
- @AutolineeCurcio
- @AutolineeRomano @CAPautolinee
- @capviaggioprato @cispeltos
- @Clickmobility @comunefi
- @CTM\_Cagliari @cttnord\_informa
- @esserependolare @EuroTransMag
- @ferpress @firenzeataf
- @GroupeRATP @iMobChallenge
- @iMobilityForum @InfoBusPisa
- @InfoParkAT @intoscana
- @ItaloTreno @LAMIAFERMATA
- @LeFrecece @MobilityPress
- @MobilityReports
- @muoversintoscan @OrariBus
- @OssMobProvPI @pendolarifr2
- @PiuBus @StazioniSicure
- @SWRTToscana @tolcommunity
- @ToremFerries @Toscanaeturismo
- @tranviafirenze @Trasportitalia
- @TTSItalia @UITPnews





- Identificazione delle Sorgenti di rumore
- Identificazione di Keywords/Search giuste, e non inserire le altre
- Quali metriche sono correlate con l'apprezzamento del servizio
  - Confronto fra i vari operatori di TPL sulla base della qualità/apprezzamento
- Tipologia delle lamentele degli utenti



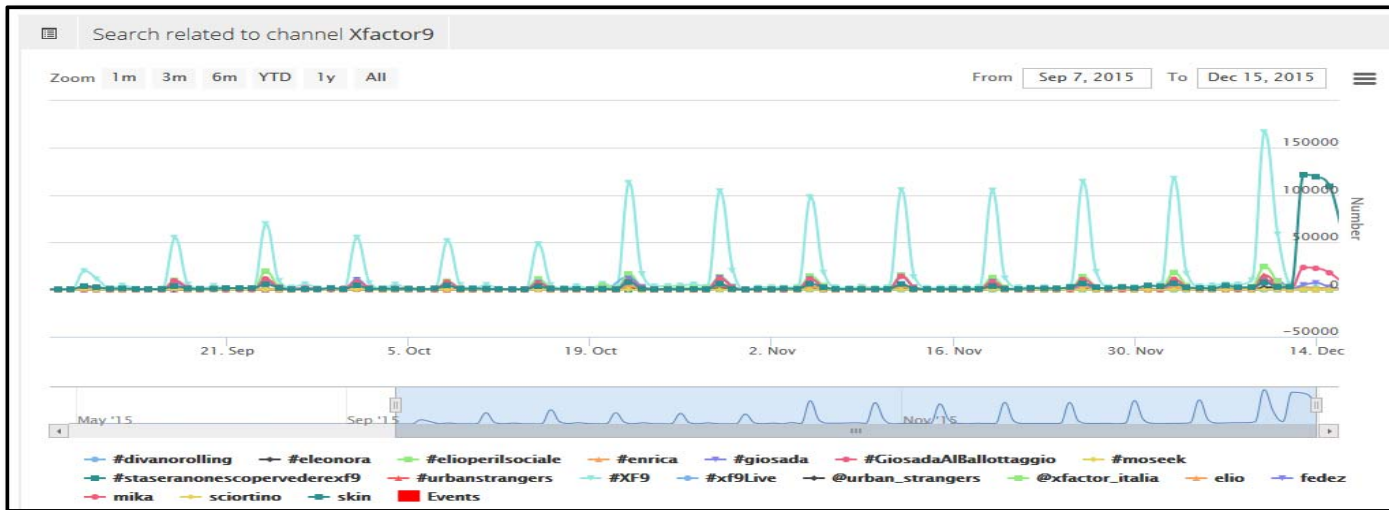
UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

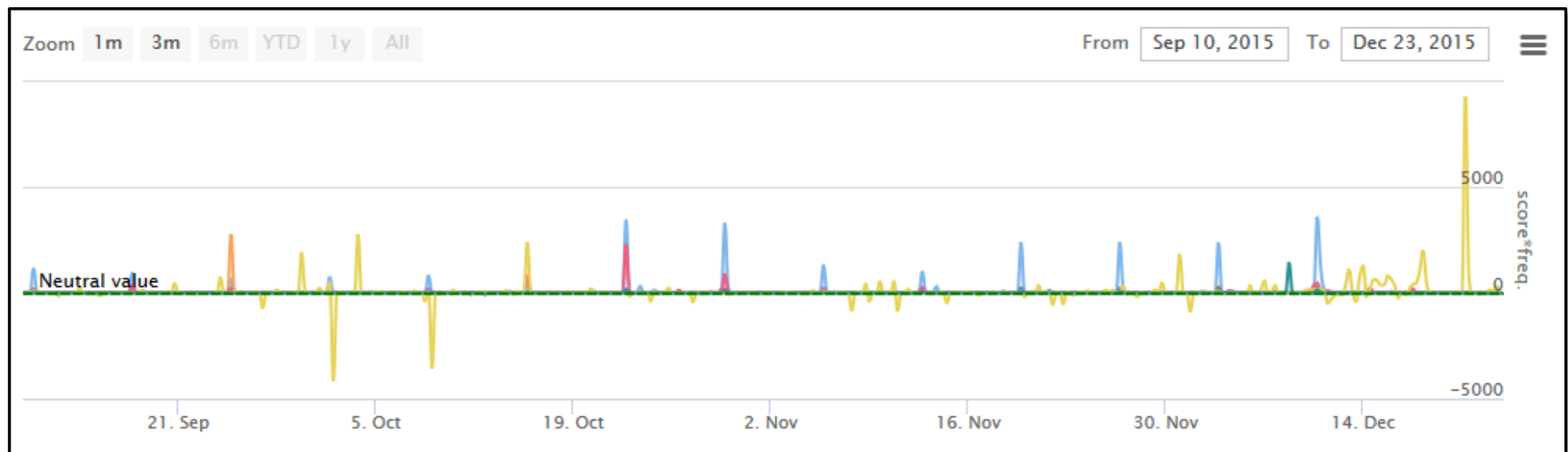
**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

# Canale "XFactor9", "xf10"

- #divanorolling
- #eleonora
- #elioperilsociale
- #enrica
- #giosada
- #GiosadaAlBallottaggio
- #moseek
- #staseranonescopervederexf9
- #urbanstrangers
- #XF9
- #xf9Live
- @urban\_strangers
- @xfactor\_italia
- elio
- fedez
- mika
- sciortino
- skin



- 2 Milioni di Tweet





## Canale “XFactor9”, “xf10”

- Identificazione delle Sorgenti di rumore
- Identificazione di Keywords/Search giuste, e non inserire le altre
- Quali metriche sono correlate con:
  - il vincitore di Xfactor9
  - il numero di telespettatori
  - casi di abbandono, come l’anno presa I concorrenti ?
- Che tipo di risposta hanno avuto I giudici ?



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

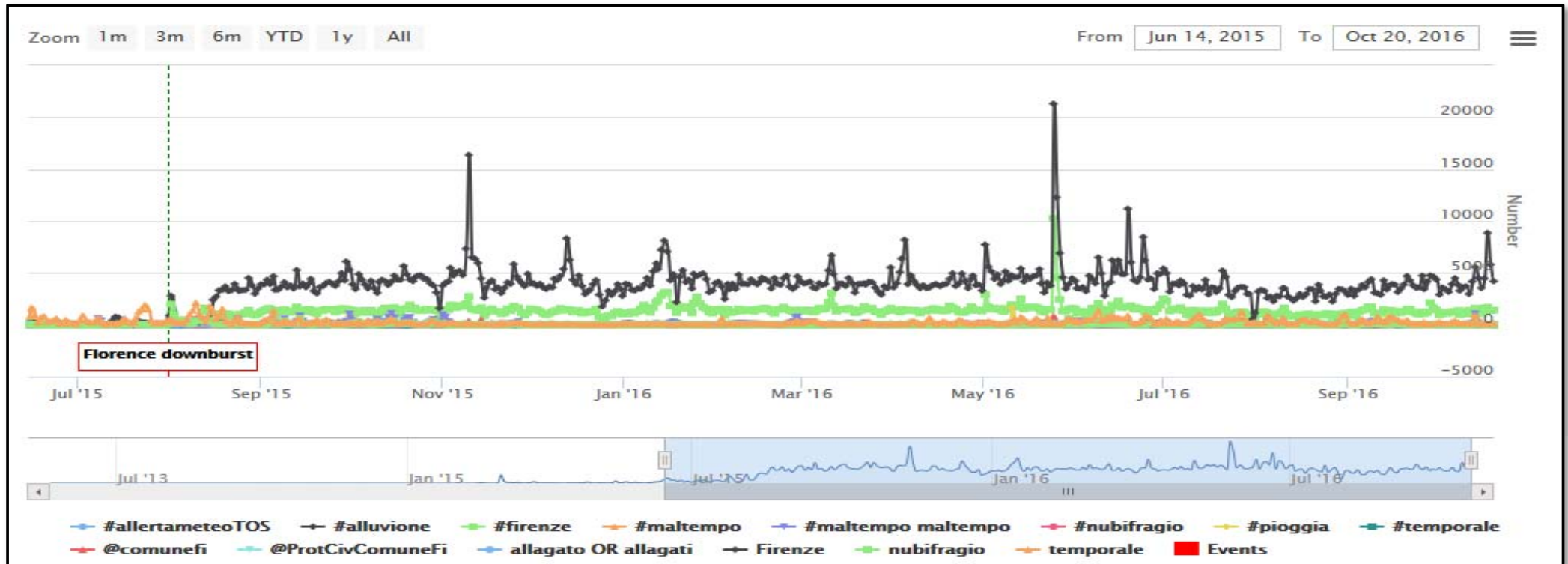
**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

# Meteo

#allertameteoTOS  
 #alluvione #fiorenze  
 #maltempo  
 #maltempo maltempo  
 #nubifragio #pioggia  
 #temporale @comunefi  
 @ProtCivComuneFi  
 allagato OR allagati Firenze  
 nubifragio temporale

- Canale “Maltempo Firenze”
- 2 Milioni di Tweet





- Identificazione delle Sorgenti di rumore
- Identificazione di Keywords/Search giuste, e non inserire le altre
- Identificare altre keyword ?
- Che percentuale di questi tweet riguarda veramente firenze
- Che correlazione c'è fra eventi meteo gravi e il numero di Tweet, Retweet, etc.
- Che anticipo medio si ha fra I primi tweet virali su Twitter riguardo ad eventi meteo gravi e le notifiche della protezione civile?
- Quali metriche sono correlate con il maltempo a Firenze





- Analizzate I dati prima!!!!!! Da utenti guest
- Vi vengono dati degli account vuoti
- Dovete creare il vostro/i canale/i: **XXXGruppoXXX**
- Rimettono dentro le Search (gia presenti) che ritenete opportune per ogni canale
  - Aggiungono nuove search → aspettare un paio di giorni poiché i dati sono molti e anche voi siete in molti: non esagerate con il numero delle search nuove !!!!! Max 5 nuove search per ogni gruppo
- Analizzate I nuovi dati, scaricando il data set specifico nell'intervallo specifico
- Utilizzo di Excel o R
- Estrazione dei risultati e realizzazione di una micro relazione: descrizione, cose buone, cose cattive, risultati...



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB

• **END**