Public crowd-sensing of heat-waves by social media data

Heatwaves' impact in Italy during summer 2015 by twitter social media audit.

Valentina Grasso (1,2), **Alfonso Crisci** (1), **Marco Morabito** (1), **Paolo Nesi** (3) and **Gianni Pantaleo** (3)

(1) CNR Ibimet, Italian National Research Council, Florence, Italy (v.grasso@ibimet.cnr.it)
(2) LaMMA Consortium, Italian National Research Council, Florence, Italy.
(3) DISIT Lab, Distributed [Systems and internet | Data Intelligence and] Technologies Lab, Dep. of Information Engineering (DINFO), University of Florence, Italy

16th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC) | 12–16 September 2016 | Trieste, Italy

Where atmosphere, sea and land meet: bridging between sciences, applications and stakeholders

Backgrounds

- Social media data become an unique realtime informative source to be used to monitor the impacts related of meteoclimatic events.
- Heat wave is the one of main issue in relation to the climatic hazard on population.
- The investigation on the reliability of social media content in relation to the crowd perception of natural risk as heat-wave is good topic.
- A methodological approach where data tools and framework give an effective exploitation of crowd-sensed data in regard to heat related hazard is a need today.

Aims of work

- Are the environmental hot conditions (heat-wave) associated to semantic related social media (SM) streams?
- Does SM activity "follow" the spatial and temporal patterns of heatwaves (HW) in Italy?
- Is a climatic daily classification (as HW days) able to discriminate different level of local SM activity?
- Do exist a social media sensitivity linked to local climatic parameters?

These are the basic questions to define the reliability of social media analytics to perform a real-time monitoring the impact of heatwave.

In the work will try to give some answers.

It is a prolonged period of excessive heat, often combined with excessive humidity lasting 2-+ days. Critical day definition for a generic location and season is needed.

Image from:
https://www.er24.co.za/blog/heat-wave-be-cautious/

- **EU-project EUROHEAT**(*) approach was adopted improving local daily climatologies (1981-2010) . It states that Critical day occurs when:

- $T_{max\ Apparent_{jd}} > T_{max\ Apparent_{jd}}\ 90^{th}$

- $T_{min_{jd}} > T_{min_{jd}}90^{th}$ & $T_{max\ Apparent_{jd}} > median\ T_{max\ Apparent_{jd}}$

* D'Ippoliti et(2010) Environ Health. "The impact of heat waves on mortality in 9 European cities: results from the EuroHEAT project."

- **Heat Wave** is when critical days spell exceed 2 days. The HW features are: duration (days) , intensity (T° > treshshold), seasonal precocity (jd timing) and spatial extent.

Heat wave 2015 in Italy

June 28-July 4, 2015

difference from average temperature (°C)
-9 -7 -5 -3 -1 1 3 5 7 9

Five main heatwaves episodes have occurred in Italy during summer 2015. The main one hits in July maintaining a persistent character for 1 months in reason to a atmospheric blocking ("omega block"). North Thyrrenic regions are the most impacted in July. The agust's episode have involved the southern eastern areas (mainly Apulia and Calabria).

June 28-July 4, 2015

upper-level (300 hPa) wind speed (kt)
10 20 30 40 50 60 70 80 90
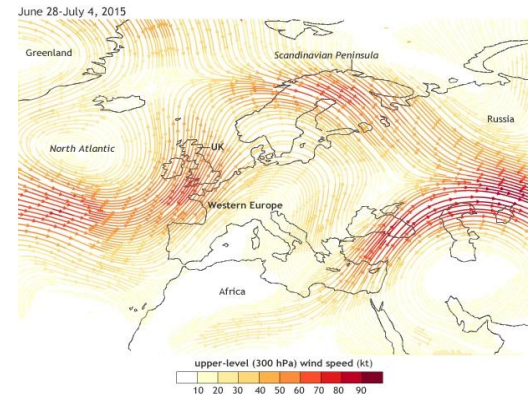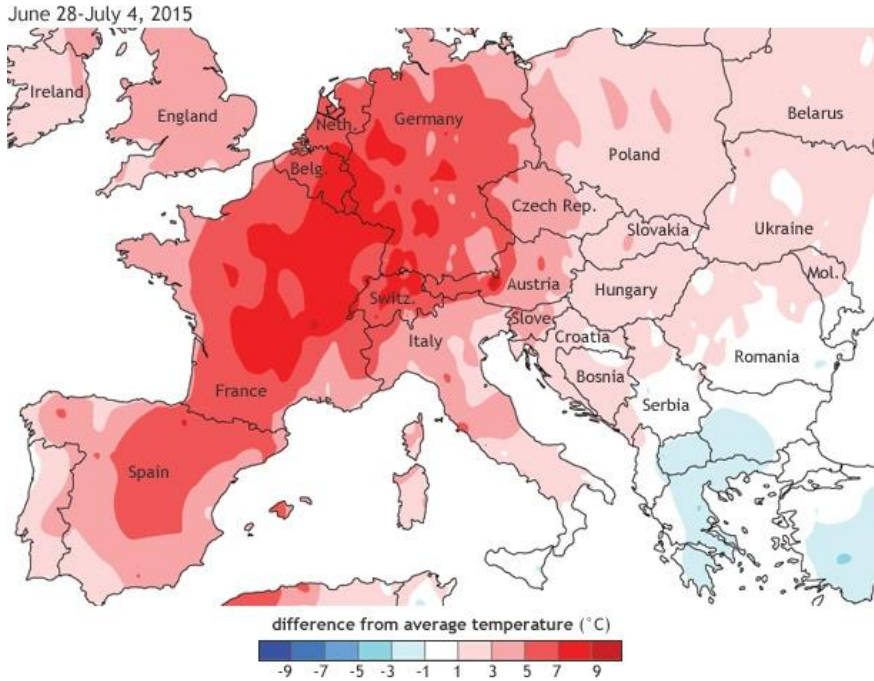
Performing crowd-sensing by Twitter data

On-line multilanguage platform for
social-networking and microblogging.
Twitter data perform significant crowd-sensing.

6.4 million of active users in Italy (2015)
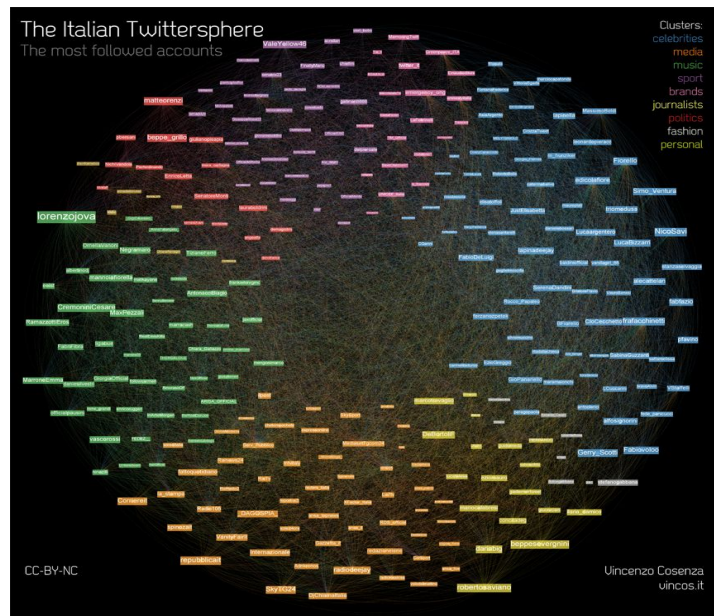
Who feed information in 🐦 ?

Many and various

Citizen, Institution, Institutional Public Services, Business
Companies, Community - NGO organization,
Media,Conversational Bots and Sensors Bots.

Daily activity common metrics

RTW_TW: N° of tweets & retweets
RTW: N° of tweets & retweets
TW: N of native tweets & retweets
U_native_users: N° of native TW authors
U_full_user: N° authors of TW & RTW
U_unique_hashtag: N° of hash-TAG

Twitter Vigilance DISIT Platform

http://www.disit.org/tv/
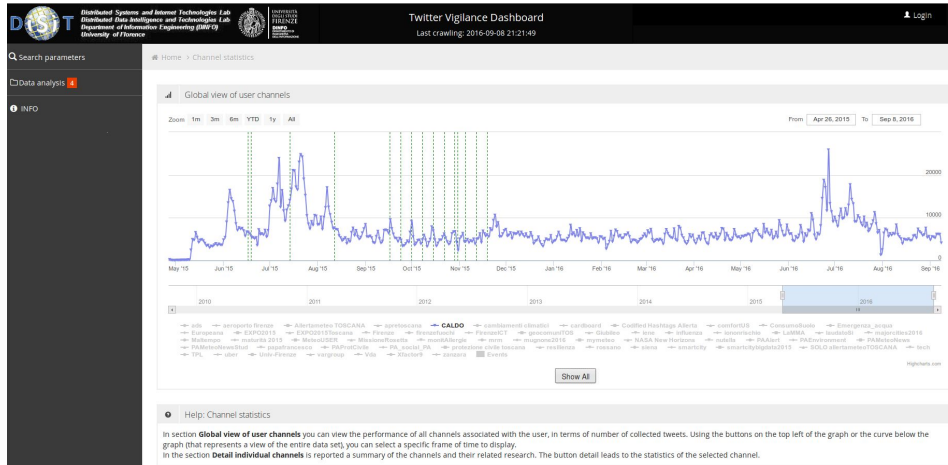
Twitter VIgilance platform is an environment developed by DISIT University of Florence that:

- Manage multiple queries in twitter API
- Store the data of messages collected by user defined queries -> channel
- It is a dashboard able to visualize data collecting process & analytics of twitter metrics of channel

Social media data: Channel "CALDO"

http://www.disit.org/tv/index.php?p=chart_singlechannel&canale=CALDO

Channel	Related research	Total ▼	N° tweets	N° tweets(%)	N° retweets	N° retweets(%)	Details	Analysis
CALDO	#caldo #sole afa allerta allerta caldo anticiclone caldissimo caldo canicola disagio caldo emergenza caldo estate rovente flegetonte ondata di calore rischio rovente temperature torrido	3344111	1851650	55.37%	1492461	44.63%	From 2009-10-23 To today	From 2015-05-15 To2015-09-15 NLP SA

- GSOD [1] daily weather data for 21 italian location are acquired corresponding to main city for each region.
- Daily Max T° Apparent are assessed by using Steadman [2] approach.
- Daily normals & quantile (1981-2010) are performed for Tmin and Tmax by using 31-day window for JD.
- Critical days and Heatwave days are calculated for the period (2015-05-15 / 2015-05-15). N=124.
- Statistics for each region are calculated.

Climatic analysis are carried out to link outcomes with social media data for the period investigated (15 May to 15 September).
Main city are considered as representative of own region.

[1] https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod
[2] Steadman, R.G. 1984. A universal expression of apparent temperature. J. Appl. Meteor. **23**, 1674-87

Channel "caldo" data and geo-extraction of local twitter streams.

Area	RTW_TW	TW	RTW	retweet	N_users	F_users	N_hashtag	Most_retweeted
marche	106	77	29	152	51	73	55	ITnewsAN
puglia	934	473	461	25581	306	716	256	99Ludo MNW_Puglia
emilia	606	461	145	850	313	433	220	LucaLombroso
trentino	221	171	50	162	121	157	93	alto_adige
sardegna	648	457	191	1279	295	452	195	paoloigna1
molise	61	53	8	28	29	37	24	mangiareinmolis
calabria	916	228	688	507633	180	859	137	paynehoran16
toscana	1662	1098	564	3133	487	760	397	flash_meteo
liguria	1320	910	410	1896	303	519	314	Ossmeteobargone
lombardia	2944	1949	995	50893	1391	2152	1075	besidenjall
sicilia	1785	874	911	222818	622	1449	412	mariamdj202
umbria	353	247	106	495	96	144	90	ItnewsPG
abruzzo	240	181	59	324	98	138	111	DonatellaPetre1
lazio	2705	1834	871	11851	1308	2060	868	TrastevereRM
piemonte	1145	698	447	3480	418	758	313	ArpaPiemonte chermaz
friuli	218	134	84	411	85	140	97	
veneto	589	463	126	577	283	392	224	Gazzettino
Totals	16453	10308	6145	831563	6386	11239	4881	
Italia	940123	585286	354837	49113594	159917	233553	61385	VoiceItalia

- Original Twitter Channel "Caldo" are filtered matching message containing the heat related words ["caldo","afa","canicola","sudore","caldissimo","torrido","record","allarme","emergenza","bollino","bere","anziani","sete","umidità","sudore","anticiclone","disagio","canicola","caronte","umido","flegetonte","sudo","sudato"] . FIltered Channel "Caldo" is indicated in table as "Italia".
- Local streams are originated by a further geographical key-term filtering by the using terms linked to location considered ["city name","region name"].
- For each tweets collection the daily metrics of activity is calculated.
- Daily data are merged with the respective local weather data.

Channel "Caldo Italia" and Heatwave periods 2015 (15 May to 15 September)

Filtered Channel CALDO vs mean TMaxApp in Italy
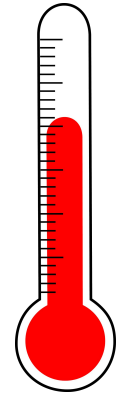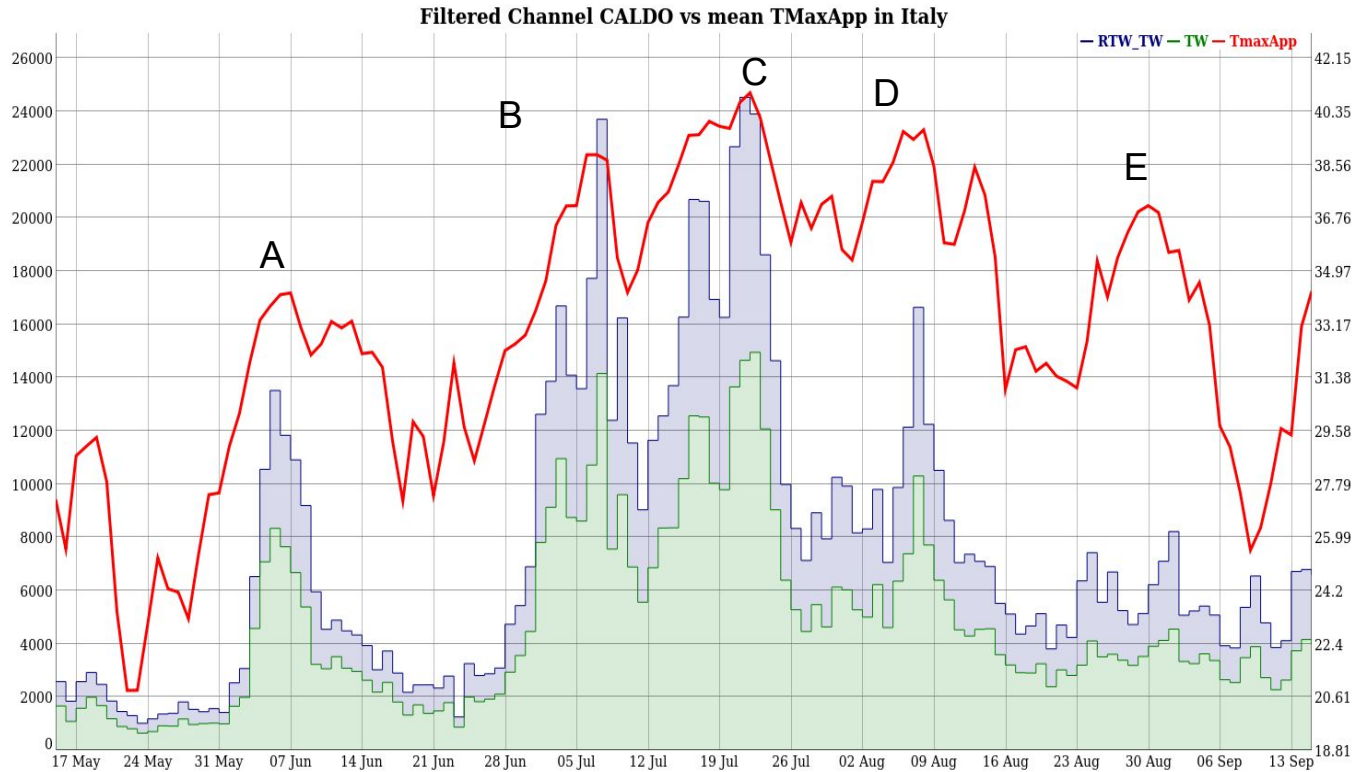
The long and intense HeatWaves 2015: a very deep impact in Italy.
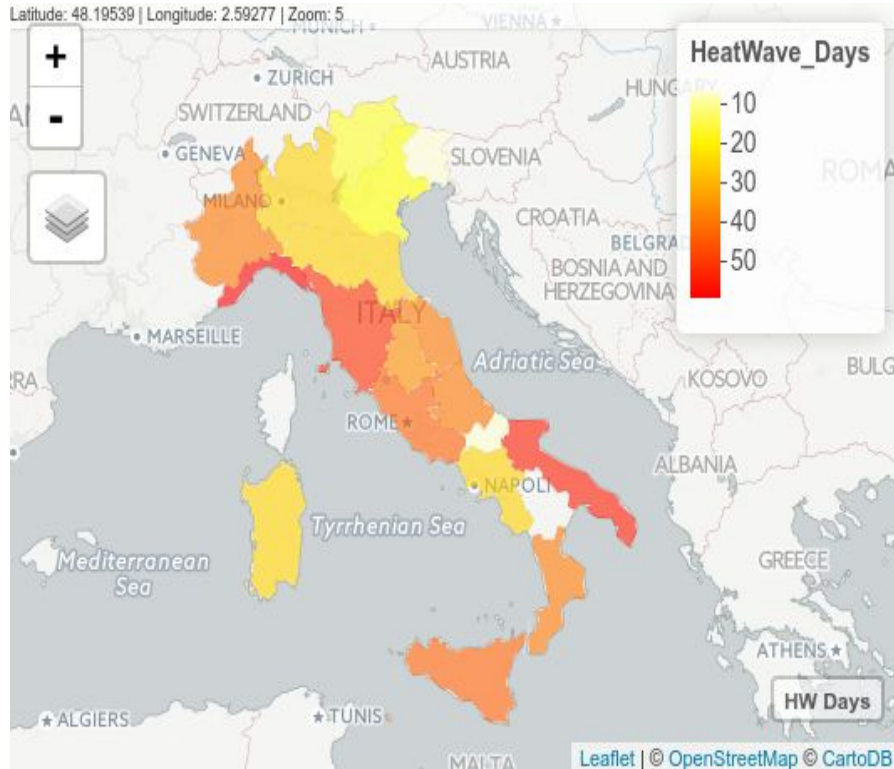

Fig. 1 - Italia: Frequenza mensile di morti. Anni 2014 e 2015
(Fonte www.demo.istat.it)

http://www.neodemos.info/68-mila-morti-in-piu-nel-2015/
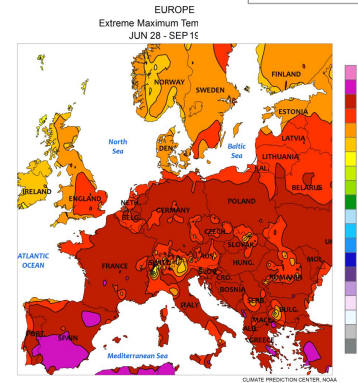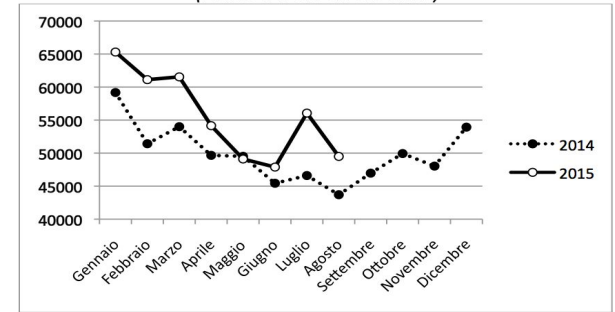
Image from Wiki Commons 2015 European Heat wave.

https://commons.wikimedia.org/wiki/Category:2015_european_heat_wave?uselang=de

Social Response Pattern during peak days

Twitter social response are influenced not only by heatwave conditions but it is also strongly dependent by population density.

A Jun

B July

C July

D August

E August

Association between daily Twitter Metrics (RTW_TW, TW, RTW, native_users, full_users, N hashtag) and main HW parameters by cities.

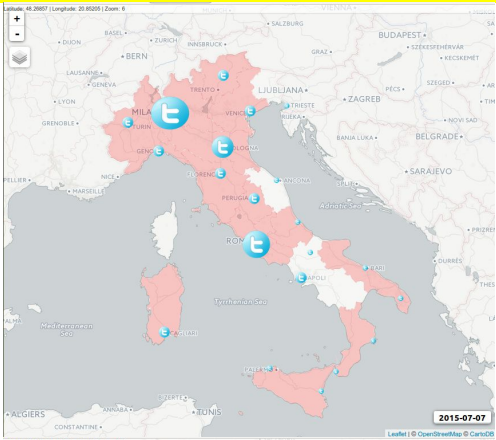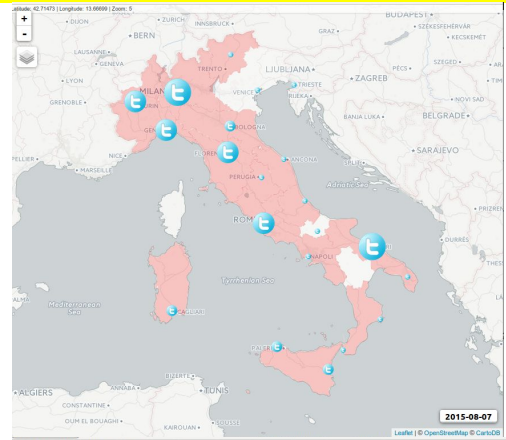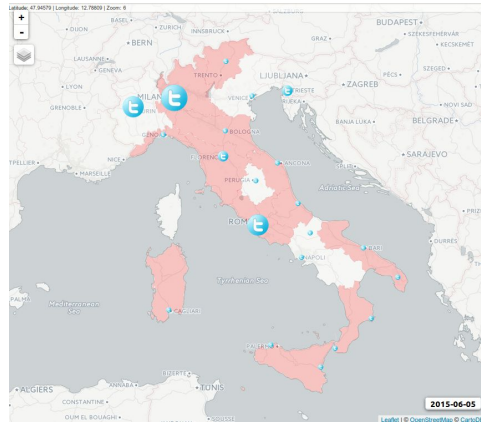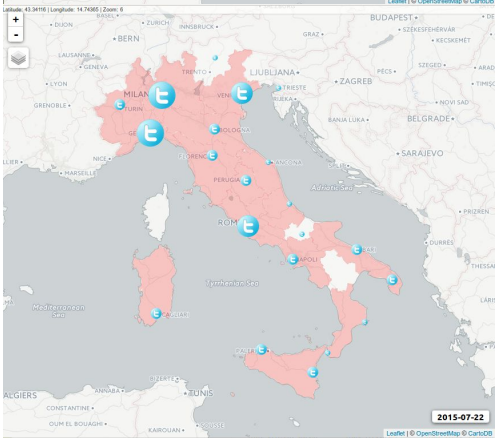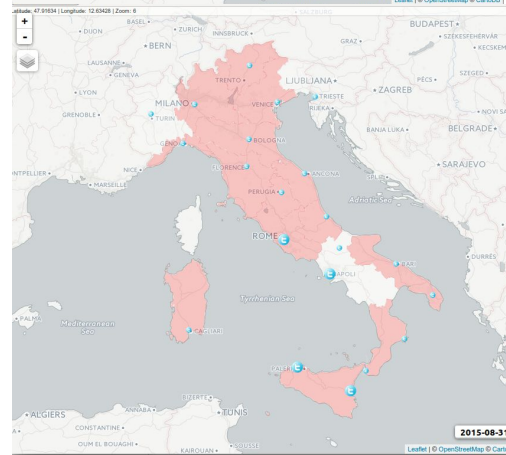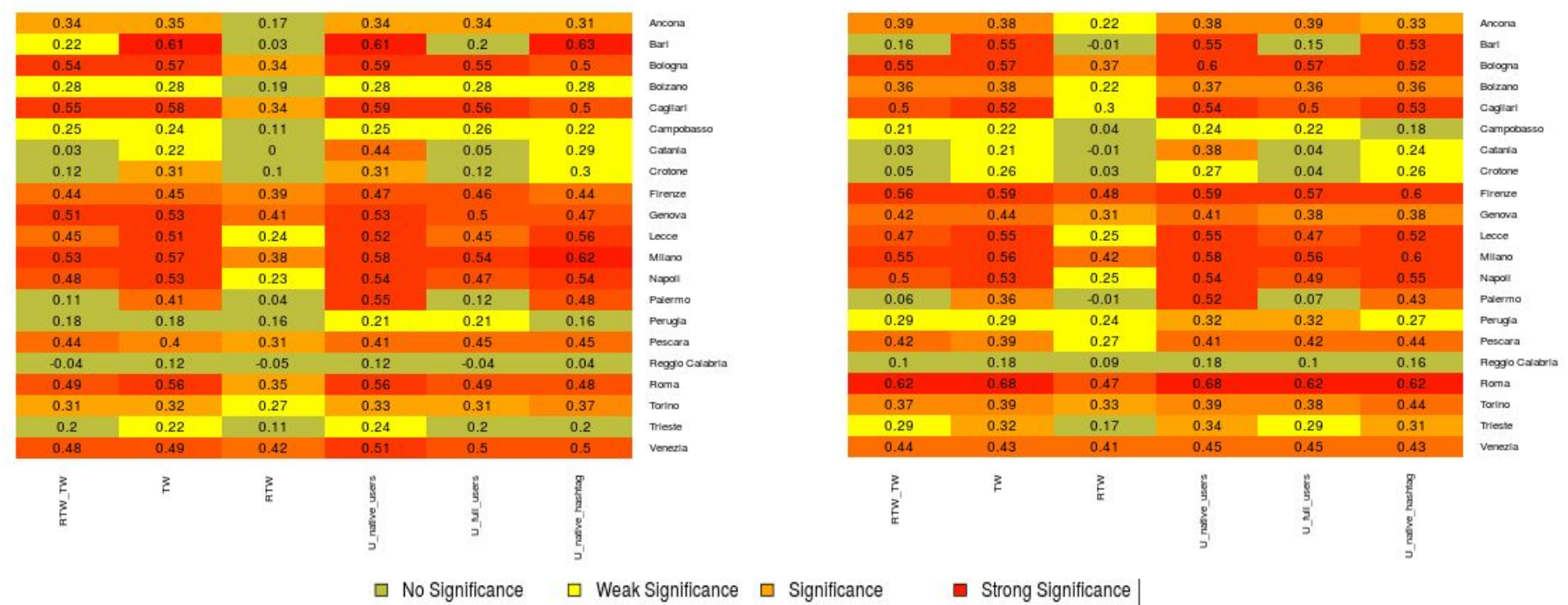Linear Correlation between daily values in 15 may-15 september. N=124 . Minimum Level of significance r(0.05) ~ 0.15

RTW_TW	TW	RTW	U_native_users	U_full_users	U_native_hashtag		RTW_TW	TW	RTW	U_native_users	U_full_users	U_native_hashtag	
0.34	0.35	0.17	0.34	0.34	0.31	Ancona	0.39	0.38	0.22	0.38	0.39	0.33	Ancona
0.22	0.61	0.03	0.61	0.2	0.63	Bari	0.16	0.55	-0.01	0.55	0.15	0.53	Bari
0.54	0.57	0.34	0.59	0.55	0.5	Bologna	0.55	0.57	0.37	0.6	0.57	0.52	Bologna
0.28	0.28	0.19	0.28	0.28	0.28	Bolzano	0.36	0.38	0.22	0.37	0.36	0.36	Bolzano
0.55	0.58	0.34	0.59	0.56	0.5	Cagliari	0.5	0.52	0.3	0.54	0.5	0.53	Cagliari
0.25	0.24	0.11	0.25	0.26	0.22	Campobasso	0.21	0.22	0.04	0.24	0.22	0.18	Campobasso
0.03	0.22	0	0.44	0.05	0.29	Catania	0.03	0.21	-0.01	0.38	0.04	0.24	Catania
0.12	0.31	0.1	0.31	0.12	0.3	Crotone	0.05	0.26	0.03	0.27	0.04	0.26	Crotone
0.44	0.45	0.39	0.47	0.46	0.44	Firenze	0.56	0.59	0.48	0.59	0.57	0.6	Firenze
0.51	0.53	0.41	0.53	0.5	0.47	Genova	0.42	0.44	0.31	0.41	0.38	0.38	Genova
0.45	0.51	0.24	0.52	0.45	0.56	Lecce	0.47	0.55	0.25	0.55	0.47	0.52	Lecce
0.53	0.57	0.38	0.58	0.54	0.62	Milano	0.55	0.56	0.42	0.58	0.56	0.6	Milano
0.48	0.53	0.23	0.54	0.47	0.54	Napoli	0.5	0.53	0.25	0.54	0.49	0.55	Napoli
0.11	0.41	0.04	0.55	0.12	0.48	Palermo	0.06	0.36	-0.01	0.52	0.07	0.43	Palermo
0.18	0.18	0.16	0.21	0.21	0.16	Perugia	0.29	0.29	0.24	0.32	0.32	0.27	Perugia
0.44	0.4	0.31	0.41	0.45	0.45	Pescara	0.42	0.39	0.27	0.41	0.42	0.44	Pescara
-0.04	0.12	-0.05	0.12	-0.04	0.04	Reggio Calabria	0.1	0.18	0.09	0.18	0.1	0.16	Reggio Calabria
0.49	0.56	0.35	0.56	0.49	0.48	Roma	0.62	0.68	0.47	0.68	0.62	0.62	Roma
0.31	0.32	0.27	0.33	0.31	0.37	Torino	0.37	0.39	0.33	0.39	0.38	0.44	Torino
0.2	0.22	0.11	0.24	0.2	0.2	Trieste	0.29	0.32	0.17	0.34	0.29	0.31	Trieste
0.48	0.49	0.42	0.51	0.5	0.5	Venezia	0.44	0.43	0.41	0.45	0.45	0.43	Venezia

□ No Significance   □ Weak Significance   □ Significance   ■ Strong Significance |

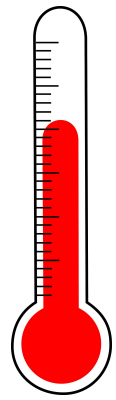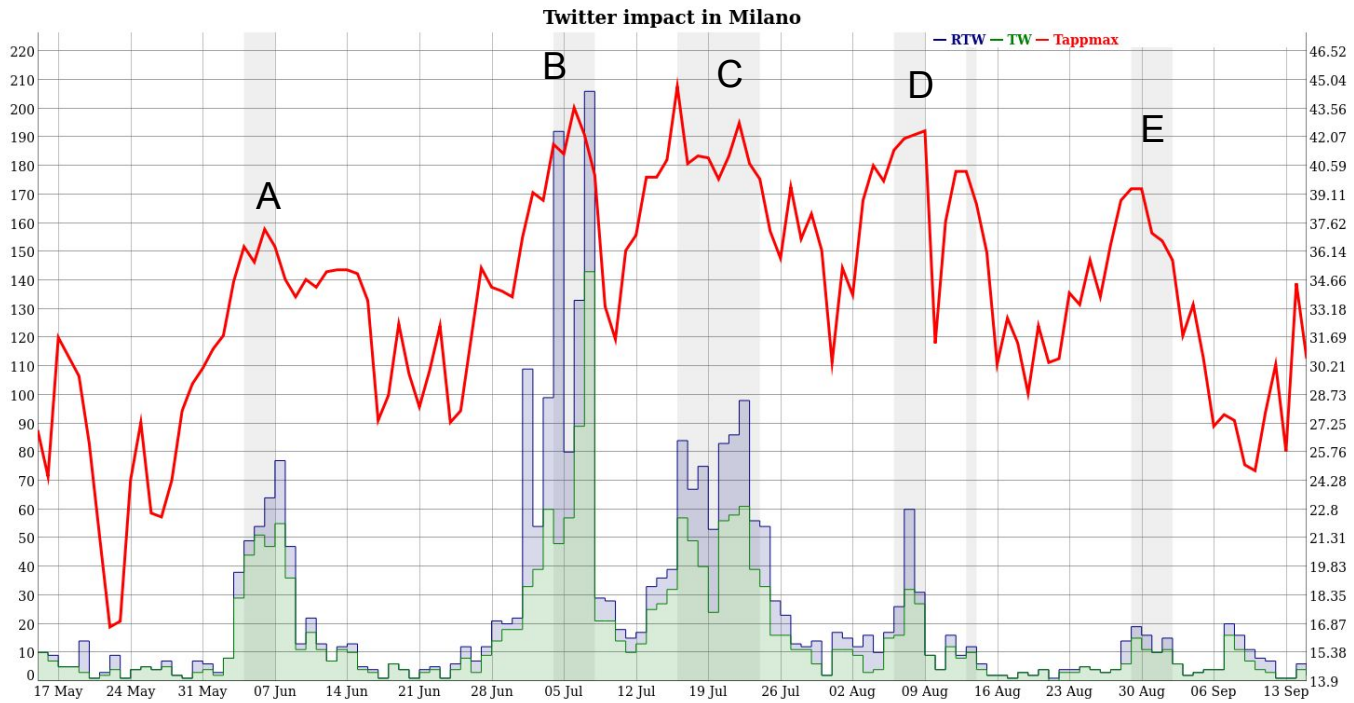Daily Minimum temperature   Daily apparent maximum temperature

HeatWaves and daily social reliability (how & where: t-student comparison & map)

City (*Significant)	All tweets NO HW	HeatWave	Native tweets NO HW	HeatWave	Native Users NO HW	HeatWave	Population
Ancona *	0.5	1.6	0.4	1.2	0.4	1.2	100861
Bari *	5.1	10.7	2	6	1.9	5.6	1263820
Bologna *	2.9	10.9	2.1	8.6	2	7.8	1005831
Bolzano	1.6	1.9	1.2	1.6	1.1	1.5	106441
Cagliari *	3.3	11.4	2.3	8	2.2	7.6	154460
Campobasso	0.5	0.8	0.4	0.8	0.4	0.8	225622
Catania (*)	13.9	10	5.2	6.7	4.6	6.6 *	1115535
Crotone	9.7	4.8	1.6	2.2	1.5	2.1	62178
Firenze	8.5	20.8	5.7	13.6	4.7	11.5	1013348
Genova *	6.5	14.8	4.6	10.1	3.7	7.4	854099
Lecce *	1.8	5.8	1.4	3.7	1.3	3.6	94773
Milano *	11.9	60.3	8.2	38.7	7.8	35.3	3208509
Napoli *	5.8	14.3	4.5	11	3.9	9	3113898
Palermo *	14.2	14.8	5.7	9.5	5	9.2	1271406
Perugia	2.6	3.5	1.8	2.6	1.5	2.2	662110
Pescara *	1.1	3.8	0.9	2.8	0.8	2.4	121014
Reggio Calabria	8.7	4.3	1.5	2.1	1.4	2	555836
Roma*	11.5	41.3	8.2	27.3	7.5	25.1	4340474
Torino	7.8	13.2	4.9	7.6	4.6	7.1	2282197
Trieste	1.8	1.4	1.1	1	1	1	234874
Venezia *	2.7	14.5	2.2	11.1	2.1	9.9	855696
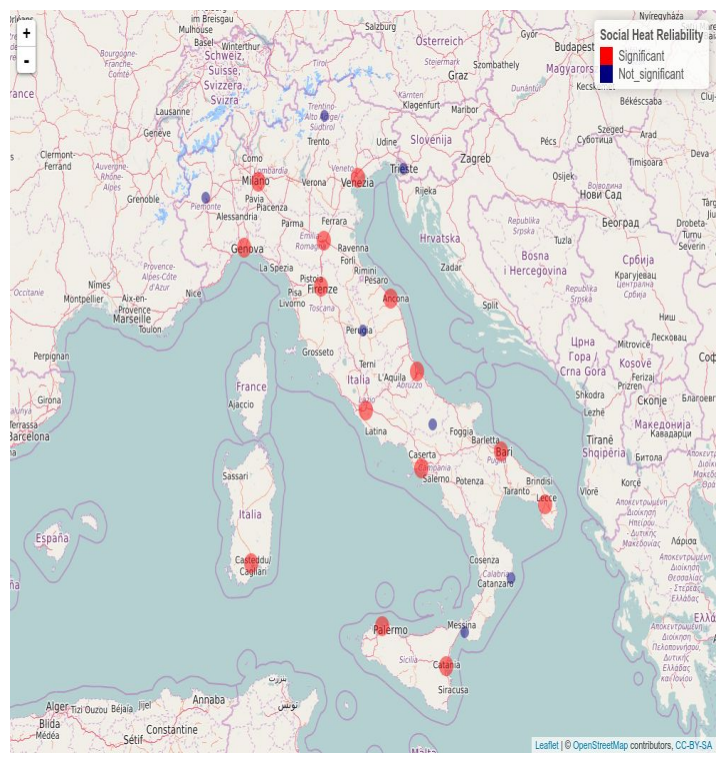
HeatWaves and daily twitter activity Milano

Parametric sensitivity of local daily twitter flows (GAM models Threshold is X the when Y>=0)

Milano - All tweets vs TmaxApp

Roma - All tweets vs TmaxApp

Milano - All tweets vs Tmin

Roma - All tweets vs Tmin

Genova - All tweets vs Tmin

Venezia - All tweets vs Tmin
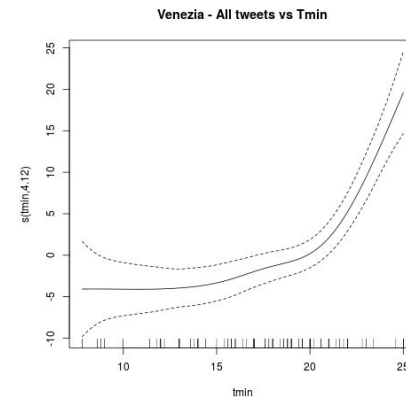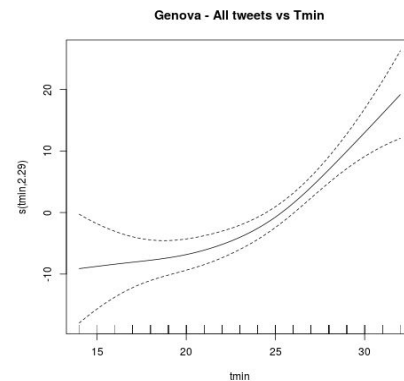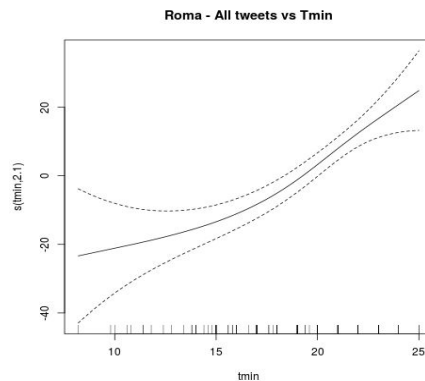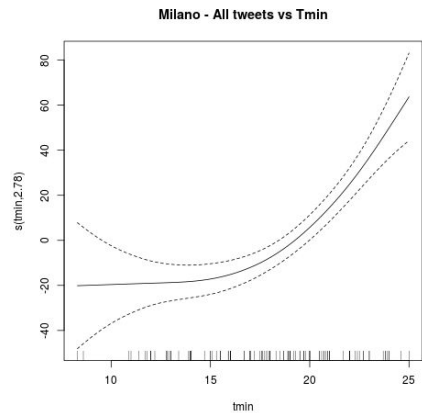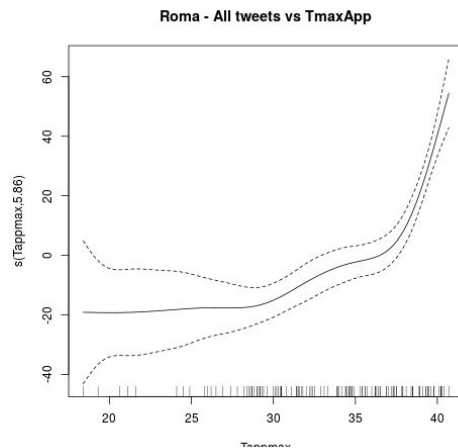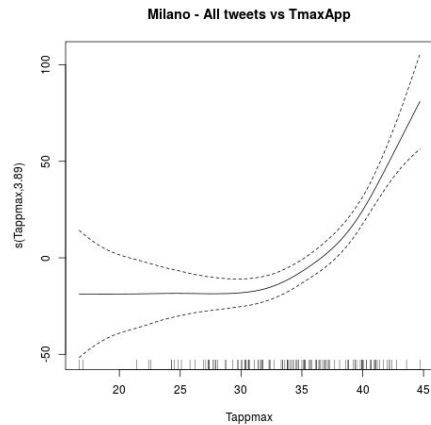
Inlands cities show threshold around 35°C in TMax App.

Analysis suggest that coastal areas have more higher threshold in Tmin (25°C and more).

These considerations are valid where association between HeatWave and tweets is significant. Further investigations are needed.

Seven points remarks

- Social media (SM) data collection require right search-terms to obtain suitable data. the semantic tuning of TW channel is ever required. Platforms as TwitterVigilance are useful.
- Channel as "Caldo" have followed well the italian summer thermal profile in 2015. HW events are easily recognizable following channel activity.
- Simple local SM stream obtained by geo-term filtering, could be reliable only in a medium-large city context (> 800.000 ab) to reach an adequate SM data numerosity.
- SM data suffers a strong dependence with population density. This is a considerable limit when SM data are used outside of big metropolitan areas.
- When data numerosity is adequate the association SM activity vs HW parameters is detected as significant. Moreover HW status (with some exception) implies ever higher level of SM activity.
- SM flows exhibits often a spatial pattern following the timing and the extent of HW phenomenon.
- SM sensitivity activity vs HW parameters is statistically detectable. In large urban areas SM data could be an help to evaluate quantitatively the impact of phenomenon as HW and disaster response.

Contacts

Communication analyst

Valentina Grasso
grasso@lamma.rete.toscana.it
v.grasso@ibimet.cnr.it

http://www.lamma.rete.toscana.it
http://www.ibimet.cnr.it/

Biometeorology
Alfonso Crisci
a.crisci@ibimet.cnr.it
alfcrisci@gmail.com
@alf_crisci

Marco Morabito
marco.morabito@unifi.it

http://www.ibimet.cnr.it/

Informatics Science
Paolo Nesi
paolo.nesi@unifi.it
Gianni Pantaleo
gianni.pantaleo@unifi.it

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

Data & Code: https://github.com/alfcrisci/EMS_heat_twitter

Image from :https://commons.wikimedia.org/wiki/File:Thank-you-word-cloud.jpg