

Auto-Detection of News on Twitter: Tuning and Testing the SocialSensor App

11th Dubrovnik Media Days Conference on Artificial Intelligence, Robots and Media,
30-31 October 2015

Neil Thurman

Neil.thurman@ifkw.lmu.de
@neilthurman

Stephen Hunt

Steve Schifferes

Nic Newman

Richard Fletcher

David Corney

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

DEPARTMENT
INSTITUT FÜR KOMMUNIKATIONSWISSENSCHAFT UND MEDIENFORSCHUNG



Centre for Research and Technology
Hellas (CERTH)



University of Koblenz-Landau (UKob)



Alcatel-Lucent Bell Labs (ALBLF)



Yahoo! Iberia S.L. (Yahoo)



The City University London (City)



Athens Technology Center SA (ATC)



Deutsche Welle (DW)



German Research Center for Artificial
Intelligence GmbH (DFKI)



IBM Israel - Science and Technology LTD
(IBM)



University of Klagenfurt (UNI-KLU)



JCP-Connect SAS (France)



The Robert Gordon University (RGU)





Jānis Krūms

@jkrums

Entrepreneur, Athlete, Latvian-American, CEO/Founder of @opportunity - Real-Time Professional Discovery - @YEC member Investor: @memsql @humandotco @plangrid



Boston Police Dept. ✓
@bostonpolice



Follow

Suspect in custody. Officers sweeping the area. Stand by for further info.

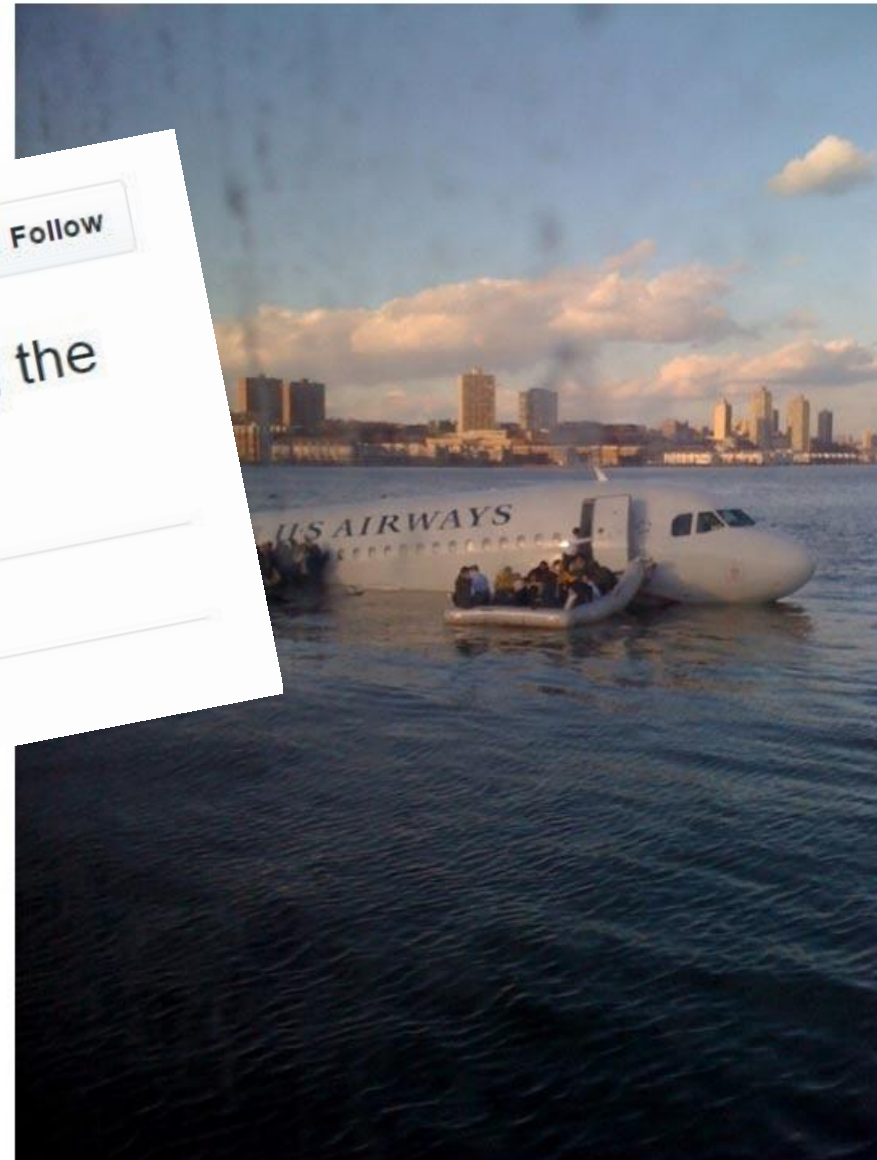


RETWEETS
71,380

FAVORITES
18,206



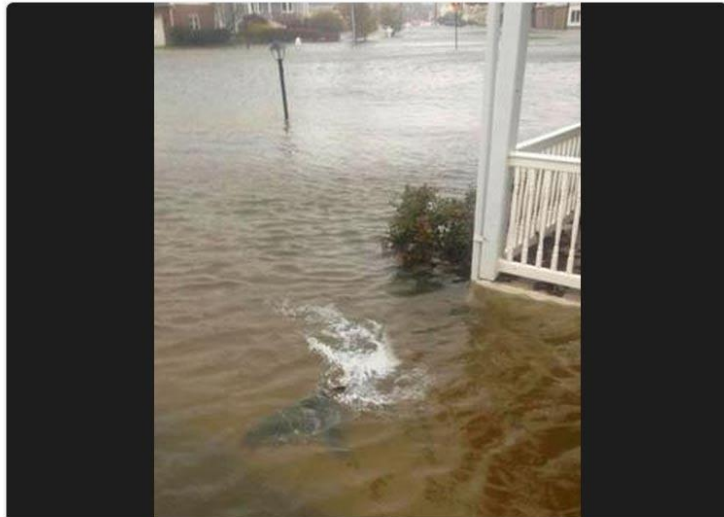
1:45 AM - 20 Apr 2013



There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy.

1,017,083 2181 days ago

- Volume
- Most not news
- How much can you believe



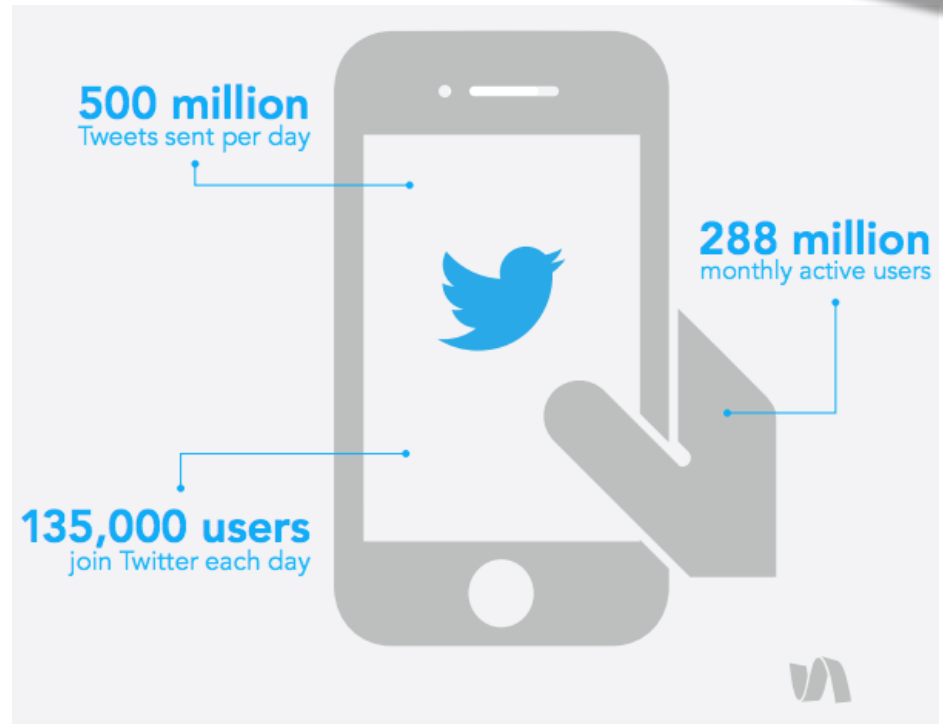
Mind Blowing
@TheMindBlowing



A shark was photographed swimming in the front yard of a flooded home in Brigantine Beach, New Jersey #sandy pic:

6:27 AM - 30 Oct 2012

8,117 RETWEETS 1,338 FAVORITES



“we need algorithms to take more onus off human being, to pick and understand the best elements” –
New York Times’ Social Media Team member (pers. Comm.)



SAVE AS NEW PANEL

Search for a phrase, domain, person, ...

Search

Search full text of stories

Current Filters
Clear all

1m

Germany

PUBLISHED IN LAST

1h

3h

12h

24h

1w

1m

LOCATION

TOPICS

WORLD

NORTH AMERICA

EUROPE

GERMANY

UNITED KINGDOM

ALBANIA

AUSTRIA

BOSNIA AND

BELGIUM

BULGARIA

BELARUS

SWITZERLAND

CZECH REPUBLIC

DENMARK

ESTONIA

SPAIN

FINLAND

FRANCE

GREECE

CROATIA

HUNGARY

IRELAND

ICELAND

NEWS

POLITICS

ENTERTAINMENT

CULTURE

LIFE

FUN

OPINION

TECH

SCIENCE

BUSINESS

INDUSTRIES

SPORTS

Stories

Facebook Posts

Instagram Posts

Influencers

Creators

Watch List

Video only OFF

Language Any

Sort by Most Recent



1live's post

@lisakestel ist die nervigste Kollegin im Haus - sagt @schroederlich
#Essensdiebin #1live #mittagessen #nofoodporn

9 minutes ago by 1live

Video

More Info

Most Recent
5



herthabsc's post

Unser Kader für das Spiel Gelsenkirchen! #hahohe

11 minutes ago by herthabsc

Image

More Info

Most Recent
90



fcbayern's post

Last training ahead of #SVWFCB! #FCBayern #Bundesliga #MiaSanMia

18 minutes ago by fcbayern

Image

More Info

Most Recent
15,460



borisbeckerofficial's post

Another win in #Shanghai!!! Today against #Tomic#nofam #idemooo

35 minutes ago by borisbeckerofficial

Image

More Info

Most Recent
809

24,307 RESULTS

Refresh Options
+ Load More

Date Filter

From
mm/dd/yyyy hh:mm

To
mm/dd/yyyy hh:mm

Apply Clear

Networks

Twitter	9.4K
Instagram	13K
Facebook	100
YouTube	122
Flickr	217
Picasa	77
YikYak	70
Sina Weibo	100
VK	161

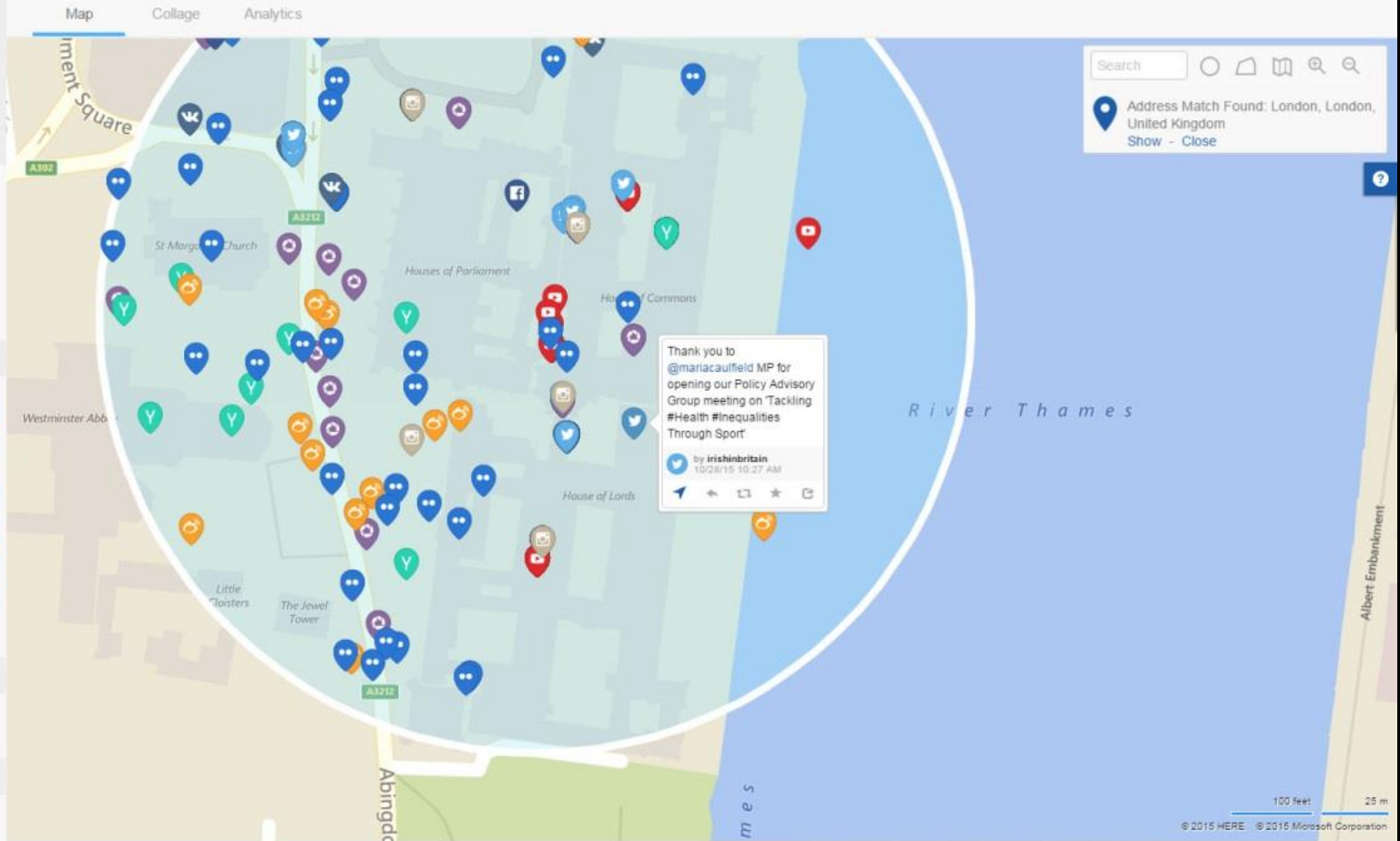
All On

Keyword Filters

Add Keyword Filter

Username Filters

Add Username Filter



social sensor






SocialSensor is a single tool that quickly surfaces trusted news stories from social media – with context.

- ❑ **A Single Tool**: one platform, one interface
- ❑ **Quickly**: in real time
- ❑ **Surfaces**: automatically discovers and clusters
- ❑ **Trusted**: automatic support in verification process
- ❑ **Material**: any material (text, image, audio, video = multimedia)
- ❑ **Social Media**: across relevant social media platforms
- ❑ **With Context**: location, time, sentiment, influence

Q: Which social networks should you monitor to detect news?

- Flickr
- Instagram
- Facebook
- Twitter
- Google+
- Tumblr
- YouTube
- Picasa
- Yammer
- etc

Broke
1st or
1st=

				
Event		Twitter	Facebook	Google Plus
Uruguay legalizes cannabis		11:55 Dec 11	01:21 Dec 11	02:35 Dec 11
Pope person of year		12:48 Dec 11	12:55 Dec 11	16:03 Dec 11
Bubonic plague outbreak		19:30 Dec 11	18:00 Dec 12	00:22 Dec 11
Golden Globe nominations named		13:22 Dec 12	13:23 Dec 12	14:32 Dec 12
Jang Sung-taek executed		21:35 Dec 12	23:00 Dec 12	22:19 Dec 12
Google removes privacy feature		09:48 Dec 12	07:25 Dec 14	09:50 Dec 12
Car bomb in Mali		09:48 Dec 14	19:18 Dec 14	14:04 Dec 14
Peter O'Toole dies		18:21 Dec 15	21:20 Dec 15	01:18 Dec 16
Chinese on moon		13:38 Dec 14	16:39 Dec 14	17:42 Dec 14
Jameis Winston wins Heisman		03:19 Dec 15	03:59 Dec 15	02:37 Dec 15
Jane Fontaine dies		01:05 Dec 16	02:29 Dec 16	06:01 Dec 16
Google buys Boston Dynamics		08:40 Dec 14	05:04 Dec 15	08:10 Dec 14
Michelle Bachelet elected		03:56 Dec 16	05:38 Dec 16	22:28 Dec 15
RnR Hall of Fame inductions		04:37 Dec 17	07:43 Dec 17	13:07 Dec 17
Santiago wins Miss World		12:55 Dec 17	12:45 Dec 17	19:41 Dec 17
Two sentenced for drug smuggling		17:46 Dec 17	16:00 Dec 17	19:18 Dec 17
Angela Merkel reelected		09:24 Dec 17	13:44 Dec 17	07:21 Dec 18
Australia wins Ashes		06:14 Dec 17	06:24 Dec 17	06:05 Dec 17
Gonzalo Inzunza Inzunza killed		20:26 Dec 18	00:54 Dec 19	00:36 Dec 20
Pussy Riot given amnesty		13:36 Dec 9	18:12 Dec 11	12:43 Dec 18
Ronnie Biggs dies		07:04 Dec 18	12:45 Dec 18	10:33 Dec 18
Target credit cards compromised		22:07 Dec 18	13:01 Dec 19	22:39 Dec 18
Apollo Theatre ceiling collapses		20:31 Dec 19	21:08 Dec 19	09:40 Dec 20
Canada overturns prostitution laws		17:20 Dec 19	09:15 Dec 20	07:38 Dec 20
Alan Turing pardoned		21:19 Dec 23	22:36 Dec 23	23:24 Dec 23
Shinzo Abe visits Yasukuni shrine		01:52 Dec 26	05:15 Dec 26	03:22 Dec 26
Joaquin Guzman heart attack		11:22 Dec 26	00:10 Dec 27	17:41 Dec 28
MS King Seaways catches fire		22:58 Dec 28	23:43 Dec 28	06:27 Dec 29

Source: Osborn, M. and Dredze, M (2014) Facebook, Twitter and Google Plus for Breaking News: Is there a winner? Proceedings of the Eight International AAAI Conference on Weblogs and Social Media.

Q: Should your software tool try to listen to all accounts or just some?

TWITTER USAGE / COMPANY FACTS

An infographic with three circular data points. Each point consists of a light blue circle containing a value, with a corresponding label below it. The first circle contains '316M' and is labeled 'Monthly active users'. The second circle contains '500M' and is labeled 'Tweets sent per day'. The third circle contains '80%' and is labeled 'Active users on mobile'.

316M

Monthly active
users

500M

Tweets sent per
day

80%

Active users on
mobile

Twitter Search API

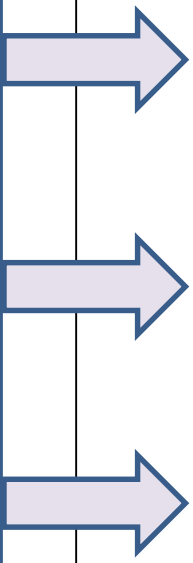
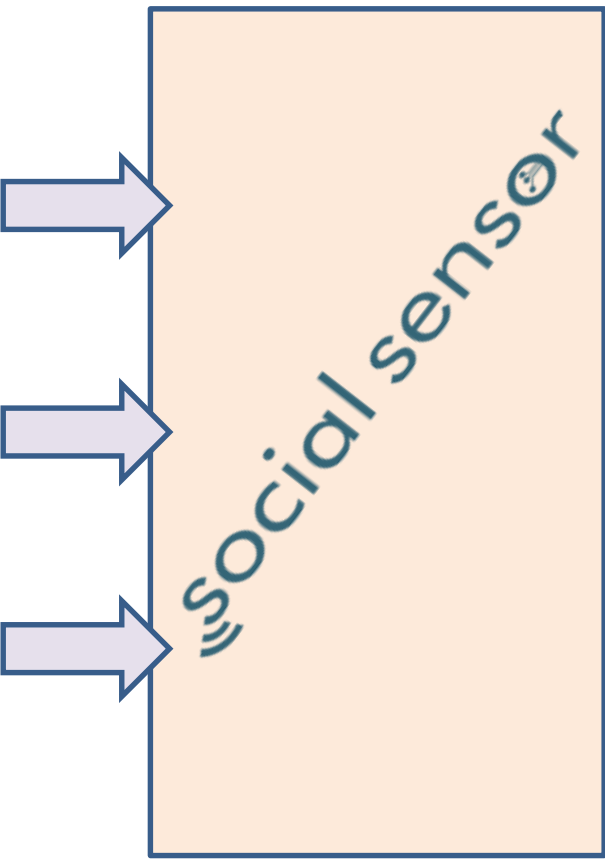
It will return a maximum of 3,200 results with a limit of 180 searches every 15 minutes.

[source](#)

- On Twitter and Facebook **7% of users generate more than 50% of the content**
- These **top users** are quite **stable over time**
- (includes non news content)

Baeza-Yates and Saez-Trumper (2015)

Wisdom of the Crowd or Wisdom of a Few? An Analysis of Users' Content Generation



Social Sensor BETA

Trending

Topics

Algeria launches hostage

Supermarkets remove prod

Guardian under fire for Julie

Pep Guardiola joins Bayern

Elephant and Castle helicop

Police arrest second man in

Wallington evacuated as flo



Example list used to 'seed' News Hound database



Lyse Doucet, Chief International Correspondent: [@BBCLyseDoucet](#)

Gavin Hewitt, Europe Editor: [@BBCGavinHewitt](#)

Lucy Williamson, Paris Correspondent: [@LucyWilliamson](#)

Fergal Keane, World Affairs Correspondent: [@FergalKeane47](#)

Chris Morris, Correspondent: [@BBCChrisMorris](#)

Christian Fraser, Correspondent: [@ChristianFraser](#)

Damian Grammaticas, Correspondent: [@DNGBBC](#)

Simon Wilson, Europe Bureau Editor: [@Siwilso](#)

Piers Schofield, Senior Europe Producer: [@Inglesi](#)

Natalie Morton, Senior Producer: [@NatalieMortonTV](#)

Imelda Flattery, Senior Producer: [@ImeldaFlattery](#)

Frank Gardner, Security Correspondent: [@FrankRGardner](#)

Gordon Corera, Security Correspondent: [@GordonCorera](#)



lyse doucet ✓

@bbclysedoucet

BBC Presenter and Chief International Correspondent. Canada is my country, London is my city, but am at home in many places.

bbc.co.uk/lysedoucet

Joined January 2010

Tweet to lyse doucet

11 Followers you know



Who to follow · Refresh · View all



Oli Reed @olireedoutdoors

Follow



Jo Burt @VecchioJo

Follow



Charlie @charlbubblehead

Follow

TWEETS
16.9K

FOLLOWING
2,068

FOLLOWERS
75.2K

FAVORITES
30

LISTS
3



Follow



Magnus Shaw

@TheMagnusShaw

Magnus Shaw - staff writer at GroupBrand and editor of The Rocking Vicar. All opinions my own.



Follow



King Abdullah

@KAbdullahKSA

H.M. KING ABDULLAH BIN ABDULAZIZ AL SAUD, Custodian of the Two Holy Mosques, King of Saudi Arabia. Official Twitter Account



Follow



Misha Glenny

@MishaGlenny

Author, ex-BBC Correspondent, books, politics, history, org crime, hi-tech crime and security



Follow



Lisa Daftari

@LisaDaftari

@FoxNews analyst | #MidEast #CounterTerrorism #HumanRights | Founder @MidEastFaces TV/Radio Personality | 4-lingual, ♥ Piano,...



Follow



colleen cleveland

@colcleve

Humanitarian / Global HR



Follow



Osman Sert

@osmansrt

Başbakan Basın Başmüsaviri, Senior Press Adv to Turkish Prime Minister, tweeting Turkish/English



Follow



Scoring newshounds

Criteria	Score
On 'seed' list?	150
Each seed that follows them	5
Each seed they follow	2
Send at least 10 tweets per day	50
Verified with Twitter's blue tick	25
Presence on at least 50 Twitter lists	25

- **Highly active** users better?
- **Highly connected** users better?
- **More** newshounds better?
- Geographical focus of **newshounds'** packs?
- **How fast** is Social Sensor vs **mainstream media channels**?

- **Better to replace inactive users with highly active ones?**
- No – didn't change performance of tool. Even with inactive users included could still capture 83-90% of 'news.

- **Better to include users who are highly connected?**
- No. Indeed those newshounds with fewer followers posted more news-related tweets.
- (note: even our less well connected newshounds were relatively well connected)

- **Better to have a larger group of newshounds?**
- 500 vs. 1000 vs. 5000
- Bigger was better, but only up to a point.

- What effect does sourcing the ‘seeds’ from different geographical contexts have (e.g. UK, US)?
- News hounds tend to find news from their geographical domain.

Social Sensor^{BETA}

Refresh


Trending in social media

Filtered by the last **30 minutes** based on **UK Trends**

Trending topics

- UK Trends
- US Trends

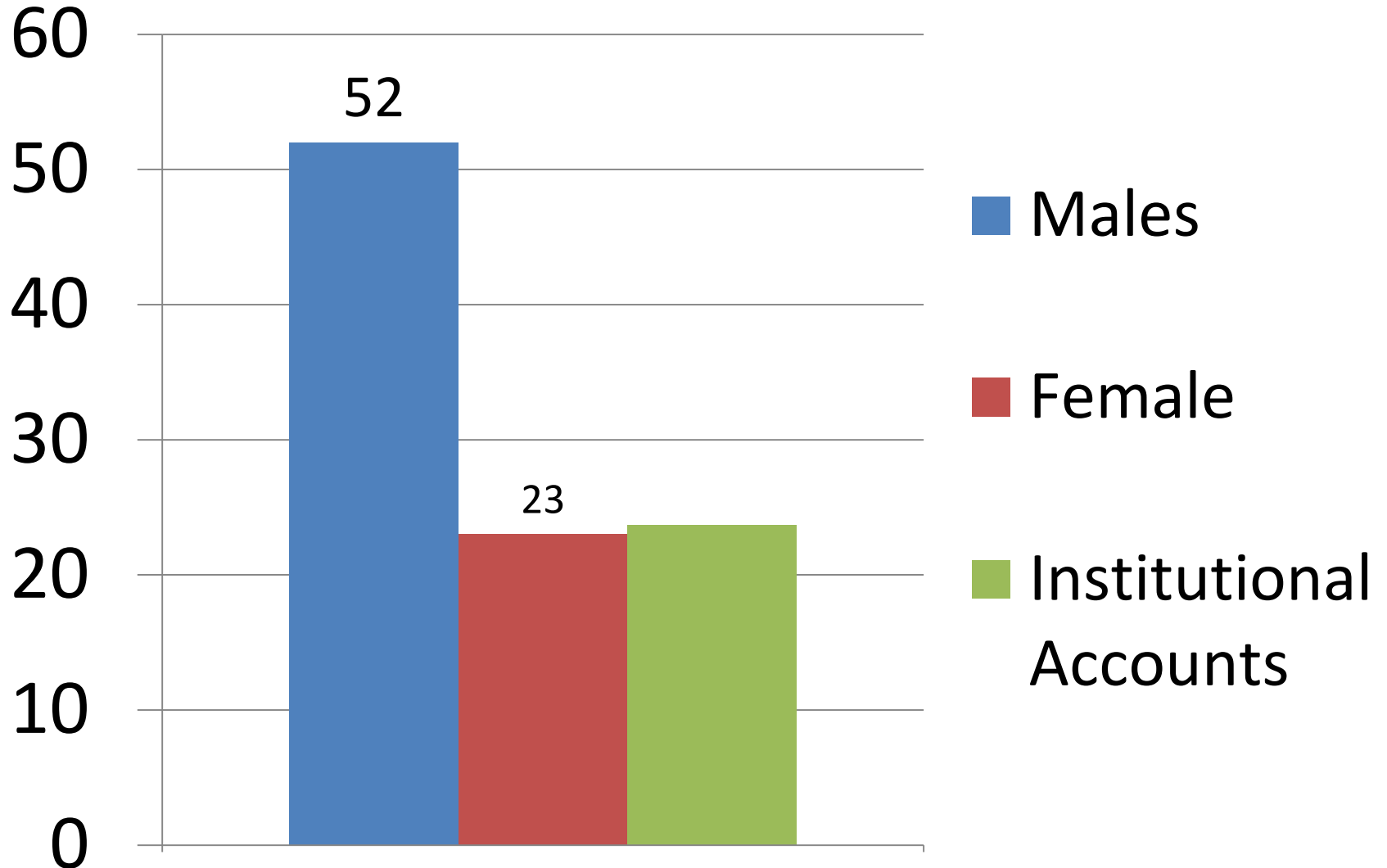
Massive inflatable jack-o-lantern terrorizes Arizona town



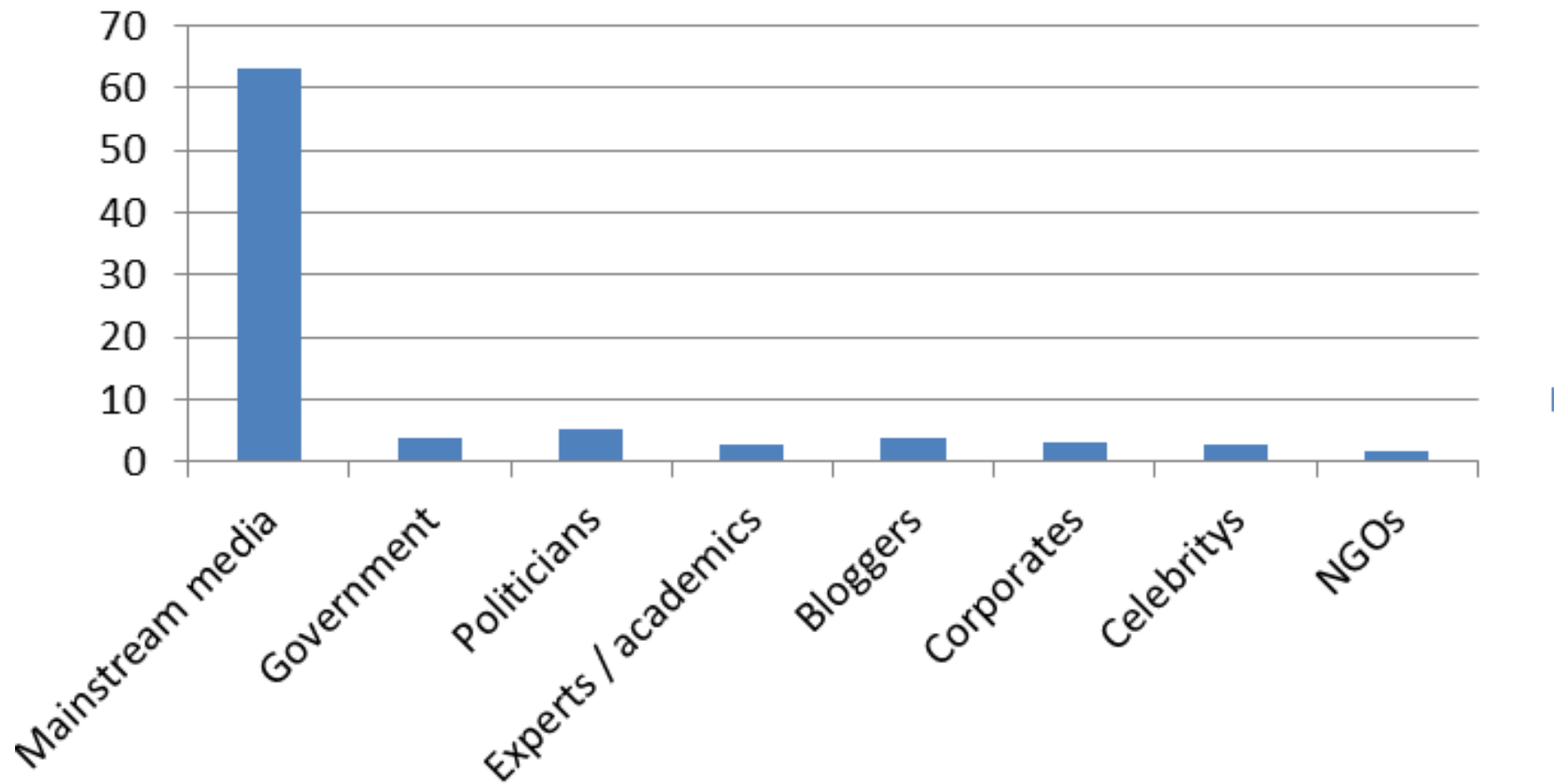
Diego Ramirez

- **How fast can Social Sensor find news on social media compared with mainstream media channels?**
- 75% of the time news breaks first in a range of mainstream media. Other 25% Social Sensor could get to first, e.g. outcomes of trials.
- However against a single competitor (@BBCbreaking) Social Sensor was 1st 88% of the time.

Who are the 'news hounds'?



Who are the 'news hounds'?



News hounds scoring system

Criteria	Score
On 'seed' list?	150
Each seed that follows them	5
Each seed they follow	2
Send at least 10 tweets per day	50
Verified with Twitter's blue tick	25
Presence on at least 50 Twitter lists	25

Computational journalism tools



```
graph TD; A[Computational journalism tools] --- B[Tuned to: Most mentioned, Most followed & Most Vocal...?or]; A --- C[Agents for change?]
```

Tuned to: Most mentioned, Most followed & Most Vocal...?or

Agents for change?

Digital 'Nose for News'

- Rely on journalistic input
- Success measured against journalistic 'ground truth'
- Created in our own image

Digital 'Nose for News'

- **Its biases are ours:**
 - short-termism
 - 'Personalization'
 - Demography

Thank you.

Prof Neil Thurman

Neil.thurman@ifkw.lmu.de

@neilthurman

Stephen Hunt

Steve Schifferes

Nic Newman

Richard Fletcher

David Corney