



**CITY UNIVERSITY  
LONDON**

# **A Digital Nose For News: Automating Sourcing?**

**Neil Thurman and Steve Schifferes**

Future of Journalism Conference, Cardiff University, 10-11  
September 2015





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

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



RETWEETS  
71,380

FAVORITES  
18,206



1:45 AM - 20 Apr 2013



Follow



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

“the biggest problem is how to exploit the vast amount of content in social media with a small team” – **MSN journalist** (pers. Comm.)

“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.)

“Current tools aren’t powerful enough” – **CNN social media expert** (pers. Comm.)



WE ARE HIRING!

[Home](#)

[About](#)

[Case Studies](#)

[Blog](#)

[Spike](#)

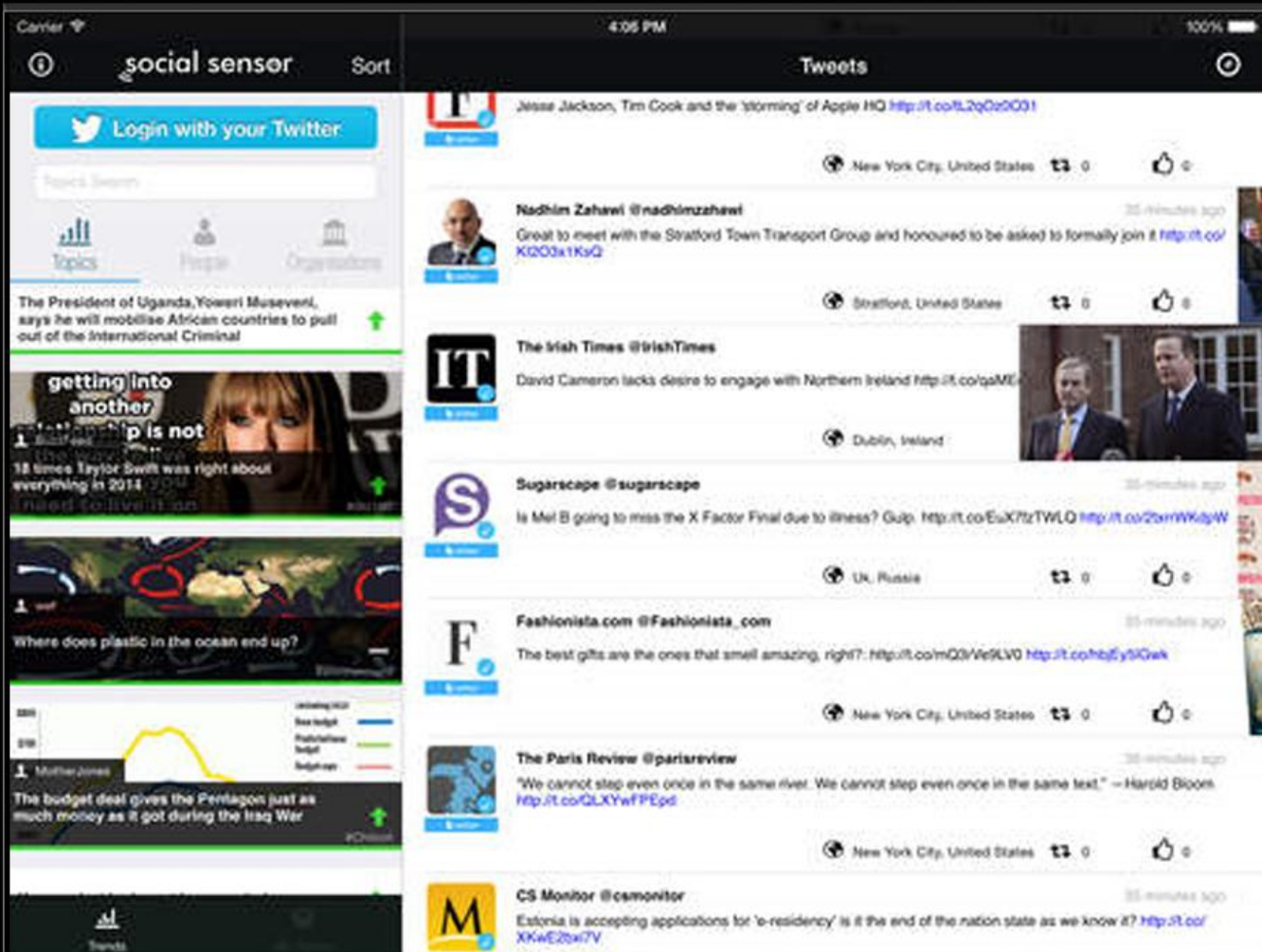
[API](#)

[Insights](#)

[News App](#)

# Here's What Matters

NewsWhip tracks the sharing of all content on social networks. Our platform instantly identifies trending stories, engaging writers and key influencers in thousands of niches and locations.





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)



# 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



CITY UNIVERSITY  
LONDON

- Providing editors with information on trends in **popularity** and **sentiment**
- Alerting newsrooms to **ongoing developments** in running stories and providing contacts and content
- Giving journalists information on the **reliability** of contributors and the **veracity** of content







CITY UNIVERSITY  
LONDON

## Tools for computational news detection & verification

- **Social and professional contexts**
- **How they work.**
- **Biases?**
- **Changes in journalistic work and outputs**





Broke  
1<sup>st</sup> or  
1<sup>st</sup>=

Event	Newsire	Twitter	Facebook	Google Plus
Uruguay legalizes cannabis	23:49 Dec 10	11:55 Dec 11	01:21 Dec 11	02:35 Dec 11
Pope person of year	11:50 Dec 11	12:48 Dec 11	12:55 Dec 11	16:03 Dec 11
Bubonic plague outbreak	17:24 Dec 10	19:30 Dec 11	18:00 Dec 12	00:22 Dec 11
Golden Globe nominations named	13:18 Dec 12	13:22 Dec 12	13:23 Dec 12	14:32 Dec 12
Jang Sung-taek executed	18:20 Dec 12	21:35 Dec 12	23:00 Dec 12	22:19 Dec 12
Google removes privacy feature	09:25 Dec 12	09:48 Dec 12	07:25 Dec 14	09:50 Dec 12
Car bomb in Mali	09:48 Dec 14	09:48 Dec 14	19:18 Dec 14	14:04 Dec 14
Peter O'Toole dies	17:17 Dec 15	18:21 Dec 15	21:20 Dec 15	01:18 Dec 16
Chinese on moon	13:38 Dec 14	13:38 Dec 14	16:39 Dec 14	17:42 Dec 14
Jameis Winston wins Heisman	02:01 Dec 15	03:19 Dec 15	03:59 Dec 15	02:37 Dec 15
Jane Fontaine dies	00:42 Dec 16	01:05 Dec 16	02:29 Dec 16	06:01 Dec 16
Google buys Boston Dynamics	06:27 Dec 14	08:40 Dec 14	05:04 Dec 15	08:10 Dec 14
Michelle Bachelet elected	22:09 Dec 15	03:56 Dec 16	05:38 Dec 16	22:28 Dec 15
RnR Hall of Fame inductions	04:25 Dec 17	04:37 Dec 17	07:43 Dec 17	13:07 Dec 17
Santiago wins Miss World	12:55 Dec 17	12:55 Dec 17	12:45 Dec 17	19:41 Dec 17
Two sentenced for drug smuggling	16:00 Dec 17	17:46 Dec 17	16:00 Dec 17	19:18 Dec 17
Angela Merkel reelected	14:03 Dec 17	09:24 Dec 17	13:44 Dec 17	07:21 Dec 18
Australia wins Ashes	05:45 Dec 17	06:14 Dec 17	06:24 Dec 17	06:05 Dec 17
Gonzalo Inzunza Inzunza killed	19:26 Dec 18	20:26 Dec 18	00:54 Dec 19	00:36 Dec 20
Pussy Riot given amnesty	22:03 Dec 9	13:36 Dec 9	18:12 Dec 11	12:43 Dec 18
Ronnie Biggs dies	07:43 Dec 18	07:04 Dec 18	12:45 Dec 18	10:33 Dec 18
Target credit cards compromised	22:07 Dec 18	22:07 Dec 18	13:01 Dec 19	22:39 Dec 18
Apollo Theatre ceiling collapses	20:30 Dec 19	20:31 Dec 19	21:08 Dec 19	09:40 Dec 20
Canada overturns prostitution laws	11:00 Dec 19	17:20 Dec 19	09:15 Dec 20	07:38 Dec 20
Alan Turing pardoned	21:19 Dec 23	21:19 Dec 23	22:36 Dec 23	23:24 Dec 23
Shinzo Abe visits Yasukuni shrine	02:11 Dec 26	01:52 Dec 26	05:15 Dec 26	03:22 Dec 26
Joaquin Guzman heart attack	08:15 Dec 26	11:22 Dec 26	00:10 Dec 27	17:41 Dec 28
MS King Seaways catches fire	23:23 Dec 28	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.

## Representative non-events:

- J'aime pas Bieber, 1D le rap et plein d'autres conneries. Vous pouvez m'amener 500 haters je changerai pas d'avis.
- This wine is going down a lil to smoothly. Here comes trouble.
- LIMA HARI BULAN LIMA ! KEK SEBESAR GUNUNG ! kena belajar buat
- kek ni, tinggal 2 bulan jea lagi -.-'
- RT ZorianRamone: Happy Bday

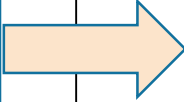
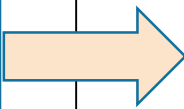
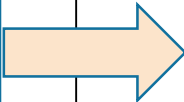
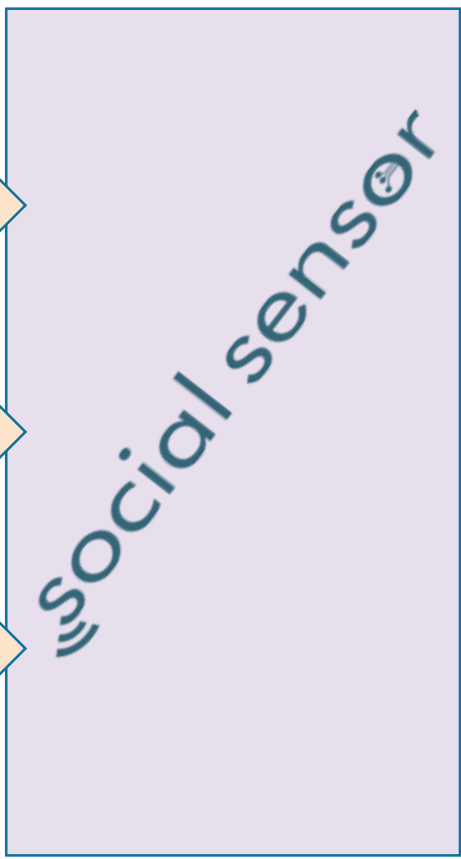
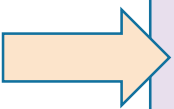
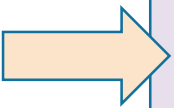
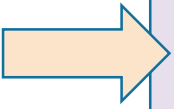
**Less than 5% of Tweets carry newsrelated content**

## Event detection in Twitter

- Running traditional First Story Detection systems on Twitter produces a mass of false positives
- Less than 1% of events detected in Twitter are news related

## Examples of false positives

- Juicy Couture, Ed Hardy, Coach, Kate Spade and many more! Stay tuned for more brands coming in <http://. . .>
- i lovee my nephew hair :D
- Going to look at houses tomorrow. One of them is & right behind Sonic Taco Casa. If I live there, I might weigh 400 lbs within a year.
- Hope a bad morning doesnt turn into a bad day...



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

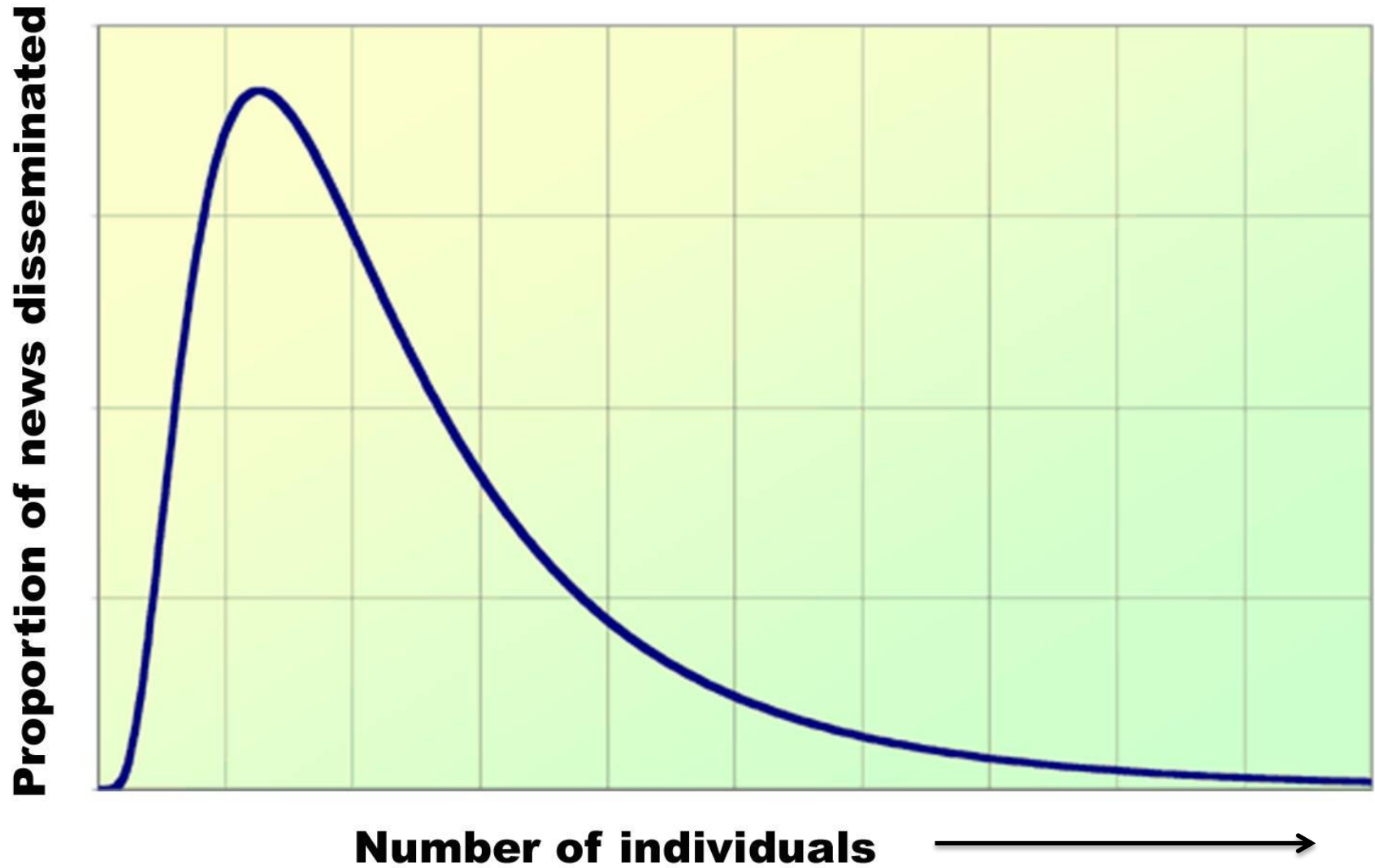
TABLE IV  
EXAMPLE RESULTS AUTOMATICALLY DETECTED BY THE BNGRAM METHOD

#	Detected topic	Corresponding story	Sample tweet
<i>FA Cup</i>			
1	liverpool gets ambushed kalou defence box mazy run before @chelseafc great shoot #cfcwembley #facup	Salomon Kalou has an effort at goal from outside the area which goes wide right of the goal. Shot by Frank Lampard missed to the left of goal.	@chelseafc: Great mazy run by Kalou into the box but he gets ambushed by the Liverpool defence before he can shoot #CFCWembley #FACupFinal (SL)
2	mikel yellow card	Booking The referee shows Mikel a yellow card. Direct free kick taken by Daniel Agger..	@chelseaindo: 37mins Chelsea still lead 1-0. And yellow Card for Mikel
3	half time wembley chelsea lead liverpool final cup 1-0 through goal ramires early #cfc	Half time.	@premierleague: It's half-time in the FA Cup final at Wembley and Chelsea lead Liverpool 1-0 through an early goal from Ramires. #cfc #lfc #facupfinal
4	over line saved super cech claiming went @chelseafc carroll header liverpool #cfcwembley #facupfinal sl	Liverpool nearly score Andy Carroll takes a shot. Petr Cech makes a fantastic save.	@chelseafc: Carroll header is saved on the line by super Cech but Liverpool claiming it went over. #CFCWembley #FACupFinal (SL)
5	goal #cfcwembley #facupfinal sl @chelseafc chelsea	Didier Drogba scores.	@chelseafc: Chelsea goal #CFCWembley #FACupFinal (SL).

 social sensor



## Dispersion of news on social media





Refresh

## Trending in social media

Filtered by the last **1 day** based on **US Trends**

## Trending topics

Cameron: Fox News commentator was an idiot



politicshome
 20
 1
 #foxnewsfacts
 #foxnewsfacts
 #birmingham
 94% positive

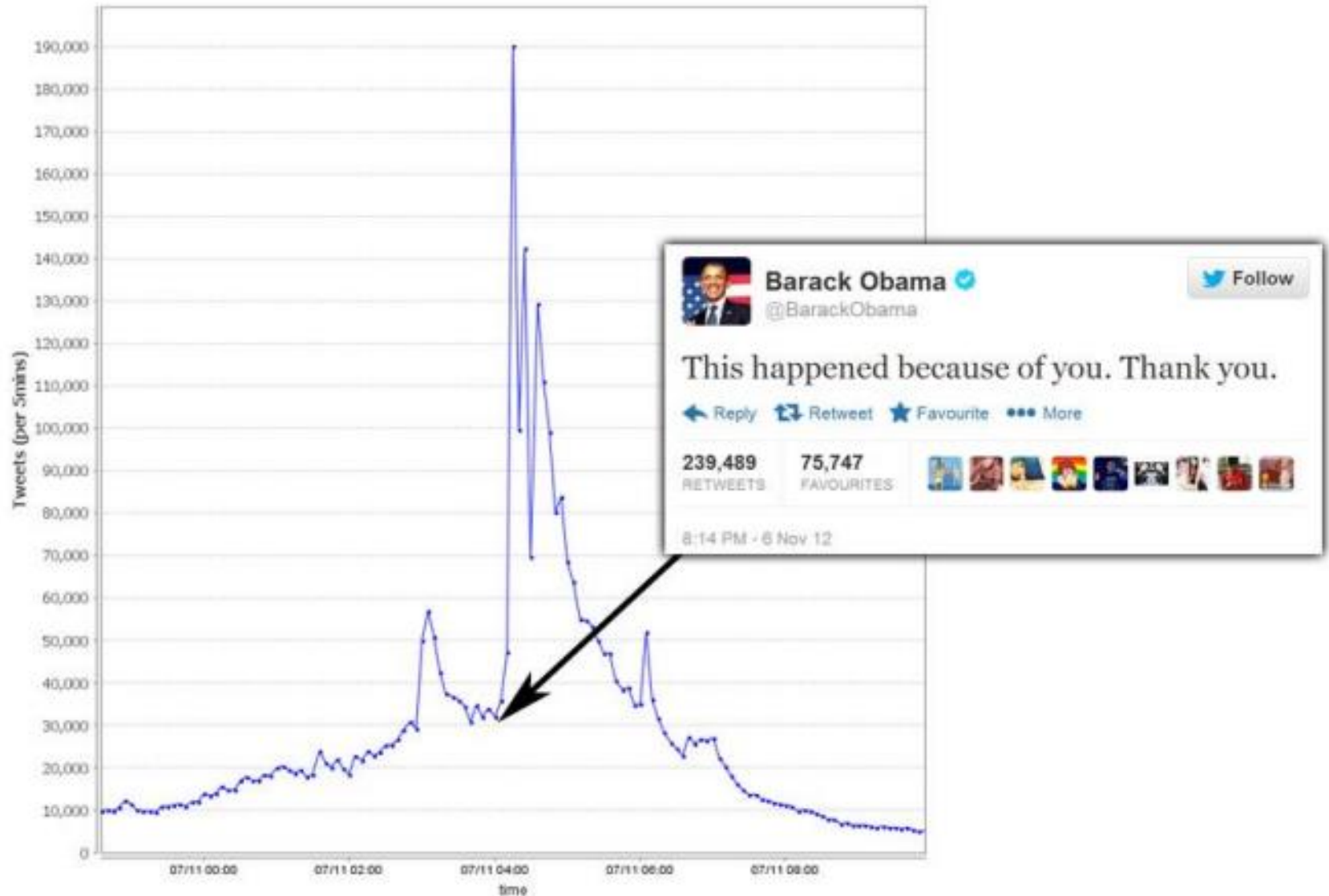
Ukraine bonds sink again as reserves evaporate



Human rights: weapons exports and extremism looks at a Saudi blogger given 50 lashes for 'insulting Islam'



# Burstiness



# THUNDERCLAP AMPLIFIES YOUR MESSAGE WITH THE POWER OF THE CROWD



watch how



Watch Video



Clapping for Photographers



Local Internet Choice



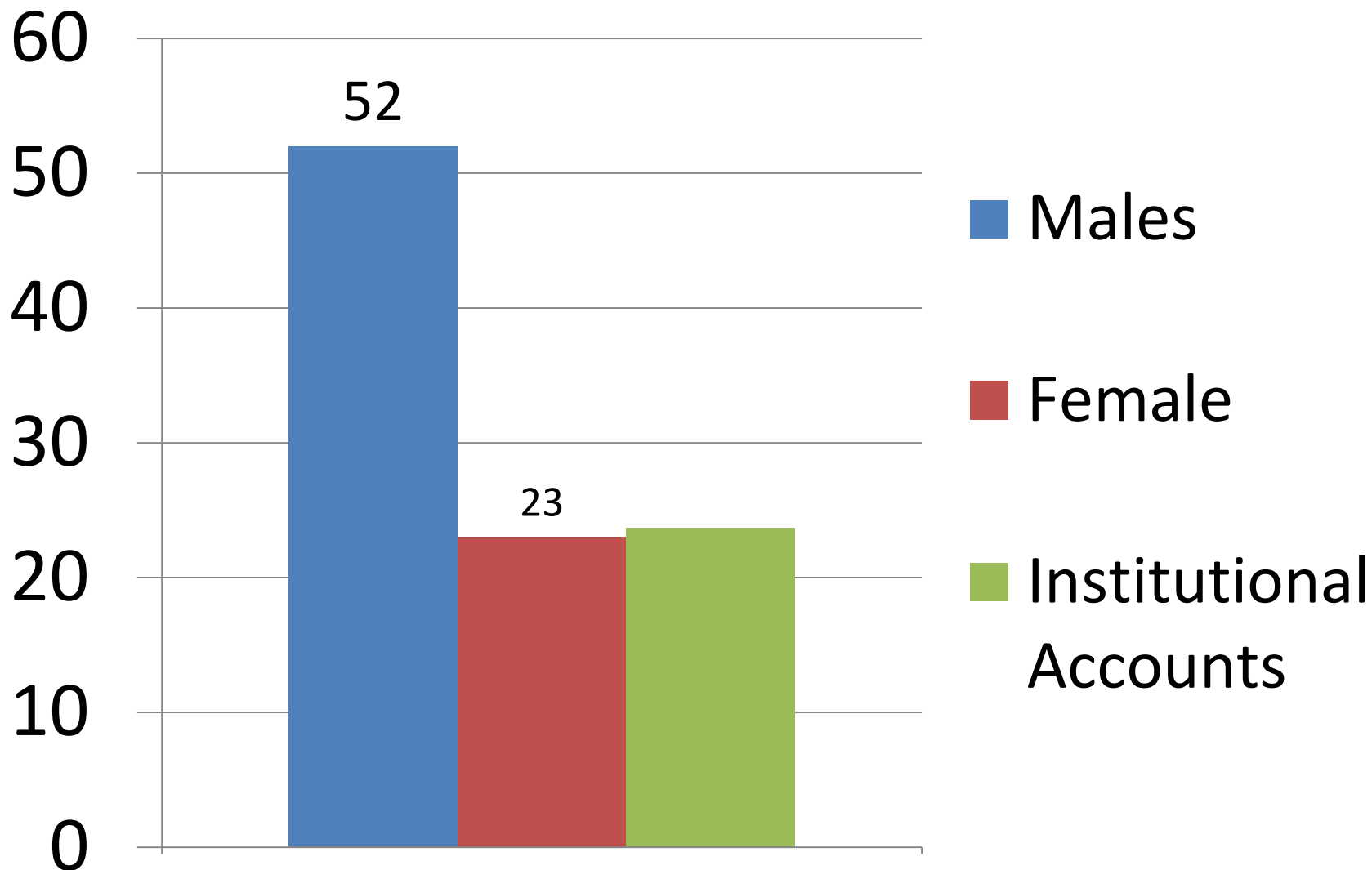
Let's make some noise!



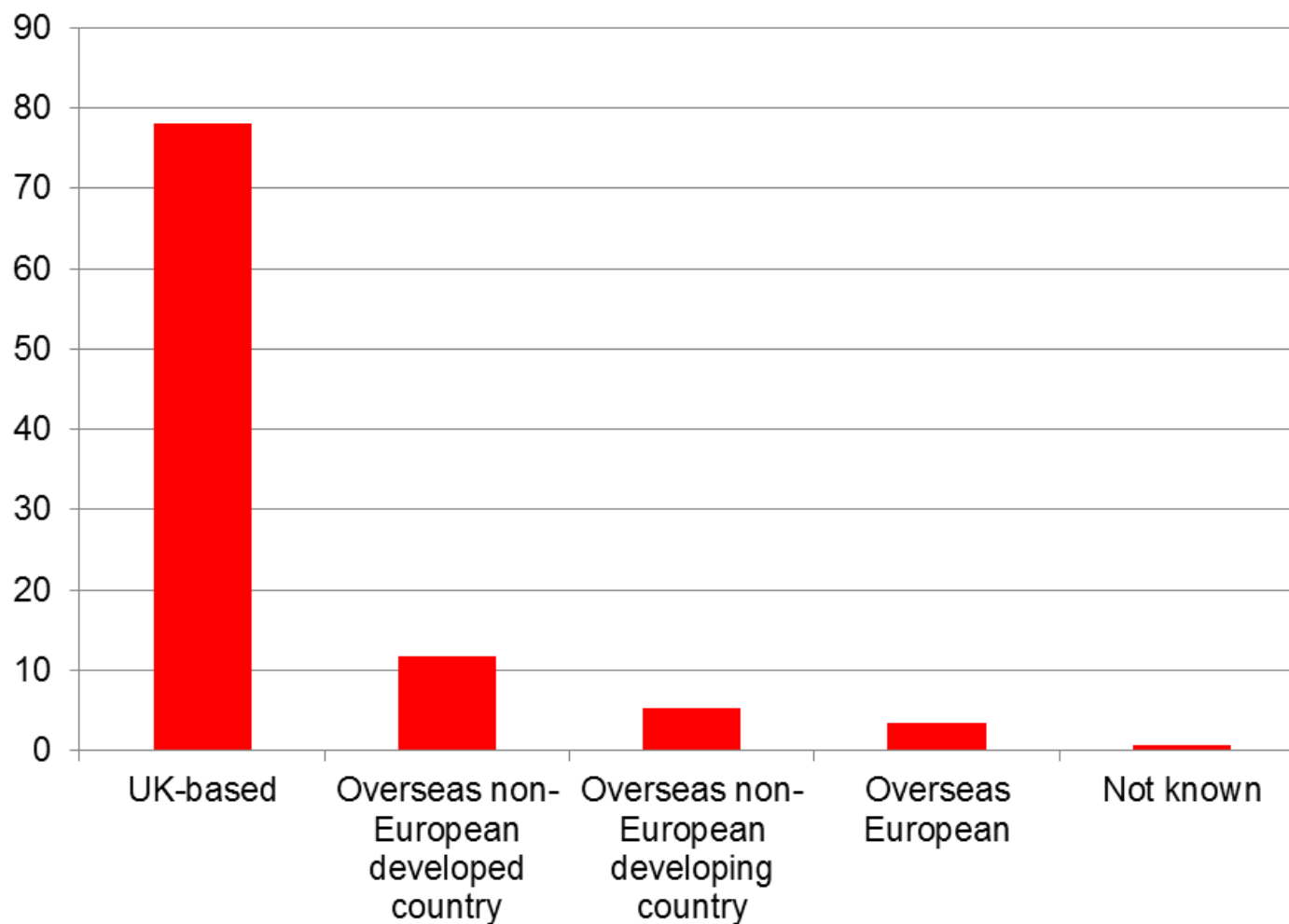
#ConflictOfInterest

► LEARN MORE ABOUT THUNDERCLAP

## Who are the 'news hounds'?

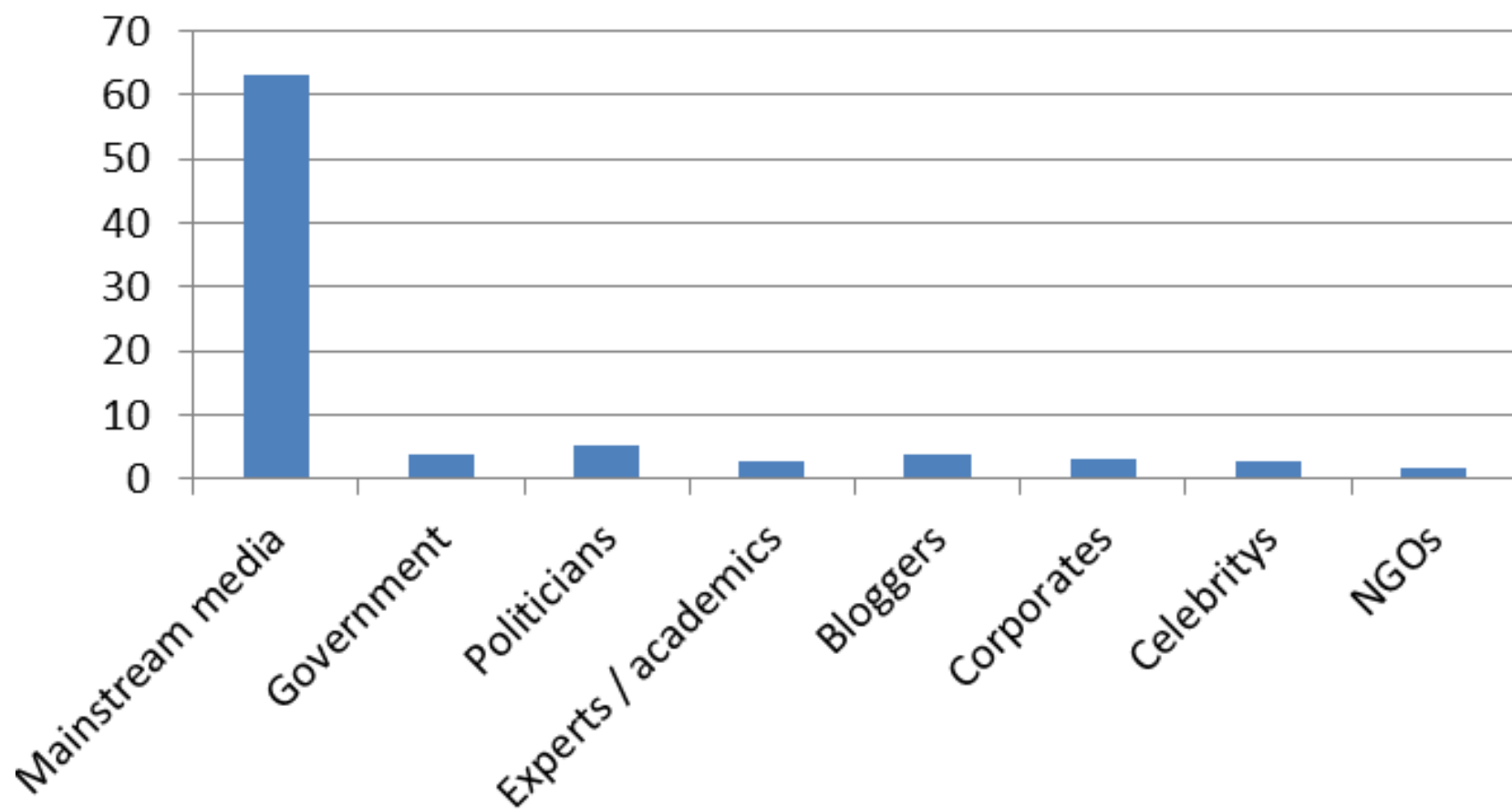


## Who are the 'news hounds'?





## Who are the 'news hounds'?



# News hounds scoring system

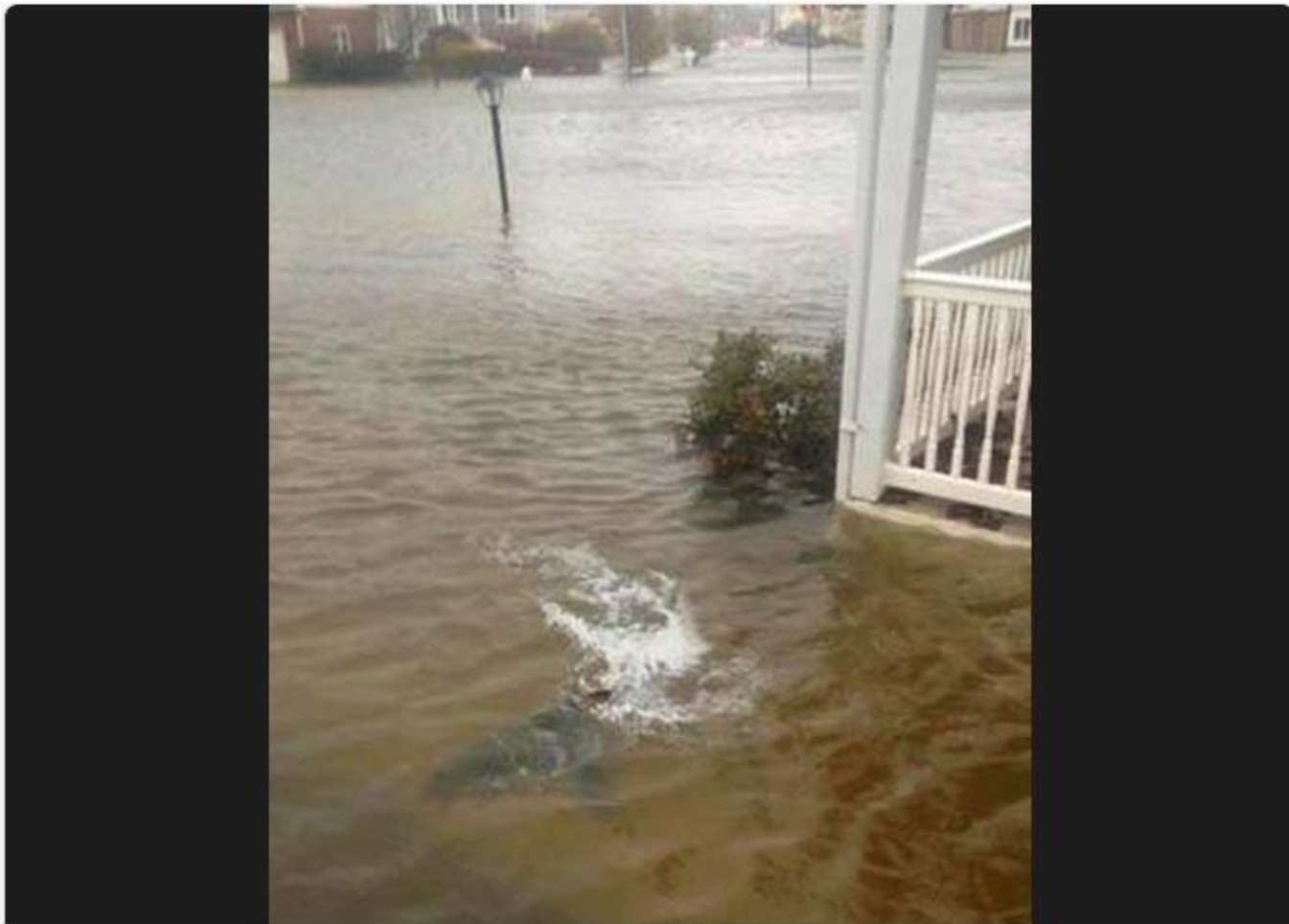
Criteria	Score
On 'seed' list?	150
Each seed that follows them	5
Each seed they follow	2
Send at least 10 tweets per day	50
Verified with Twitter's blue tick	25
Presence on at least 50 Twitter lists	25

# Computational journalism tools

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

Agents for change?

# Verification



**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







LIVE CAM 12:12 PM

58°

# Principles for social media verification

- Content
- Contributor
- Context

# Content



**Brett Lawrie** ✓

@blawrie13

+ Follow

Pretty sure someone just let off a round bullets in eaton center mall .. Wow just sprinted out of the mall ... Through traffic ...

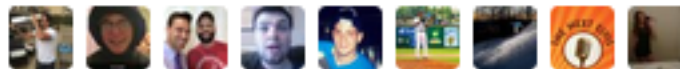


RETWEETS

1,171

FAVORITES

103



3:24 PM - 2 Jun 2012

# Contributor



**Ali Zidan**



**@AliZidanPM**

Prime minister of Libya رئيس وزراء ليبيا ملاحظة : التخريبات التي تحدث في ليبيا #الرئيس او  
#علي\_زيدان هي من كتابة د. علي زيدان شخصيا

Tripoli · pm.gov.ly

**Dr. Ali Zidan – Prime Minister of Libya**



# Context





**Justin Hart**

@justin\_hart



Follow

Wow... the missing kid Sunil Tripathi sure does look like suspect #2...







Justin Hart  
@justin\_hart



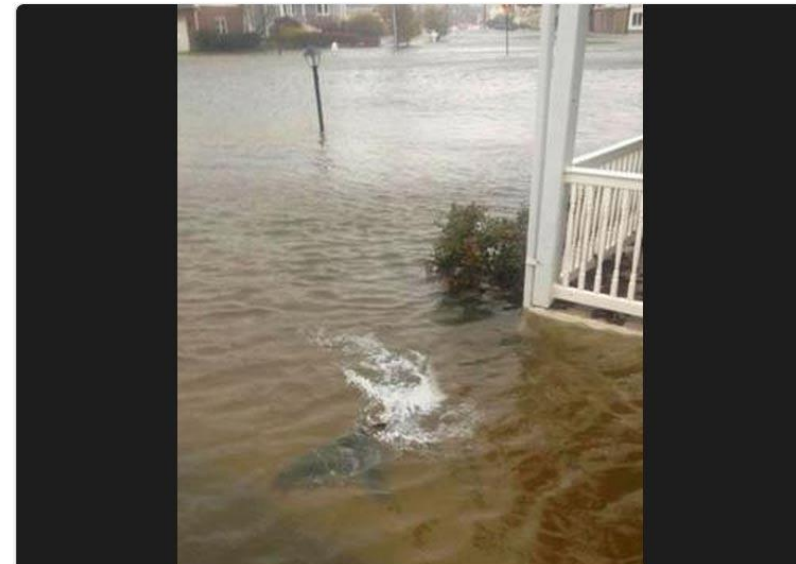
Follow

Wow... the missing kid Sunil Tripathi sure does look like suspect #2...



If training=

And test=



Mind Blowing  
@TheMindBlowing

Follow

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





Harriet Harman @HarrietHarman

*MP Camberwell and Peckham, Labour deputy leader, shadow culture secretary*

Influential on: [Ukraine, ALSIceBucketChallenge, ferguson, 1063mobileradio, BreakingNews]

Location: **Camberwell & Peckham, London**

## Truthmeter details

### Breakdown

#### History

9

Information about how active a given source is on different social media platforms, also combining it with validity information (when did this source appear, how frequently has it been posting)

#### Popularity

5

Information about how many and what kind of people are following the activity of this source, and how many are reading or are recommending this activity to others

#### Influence

5

Information about how much this source triggers activities, such as re-posts, discussions or comments. Are things happening because of what this source is posting?

### Truthmeter rating

#### Overall

5

This is an overall rating taking into account all the information extracted from the history, popularity and influence attributes, giving an overview of the trustworthiness of the contributor

## Contributor's details

Twitter [www.twitter.com/HarrietHarman](http://www.twitter.com/HarrietHarman)

## Contributor's network

Tweets	5369
Followers	59690
Following	1275

## Latest Tweets



Harriet Harman @HarrietHarman



12h



Congrats Duncan Campbell 2014 #TurnerPrize winner @Tate



Harriet Harman @HarrietHarman



17h



..@UKLabour Gordon Brown – A political giant of his generation  
<http://t.co/sKva131kDQ>



# Computing Contributor Credibility

No. of Tweets

HISTORY

---

Frequency

HISTORY

---

No. of followers

POPULARITY

---

No. of follows

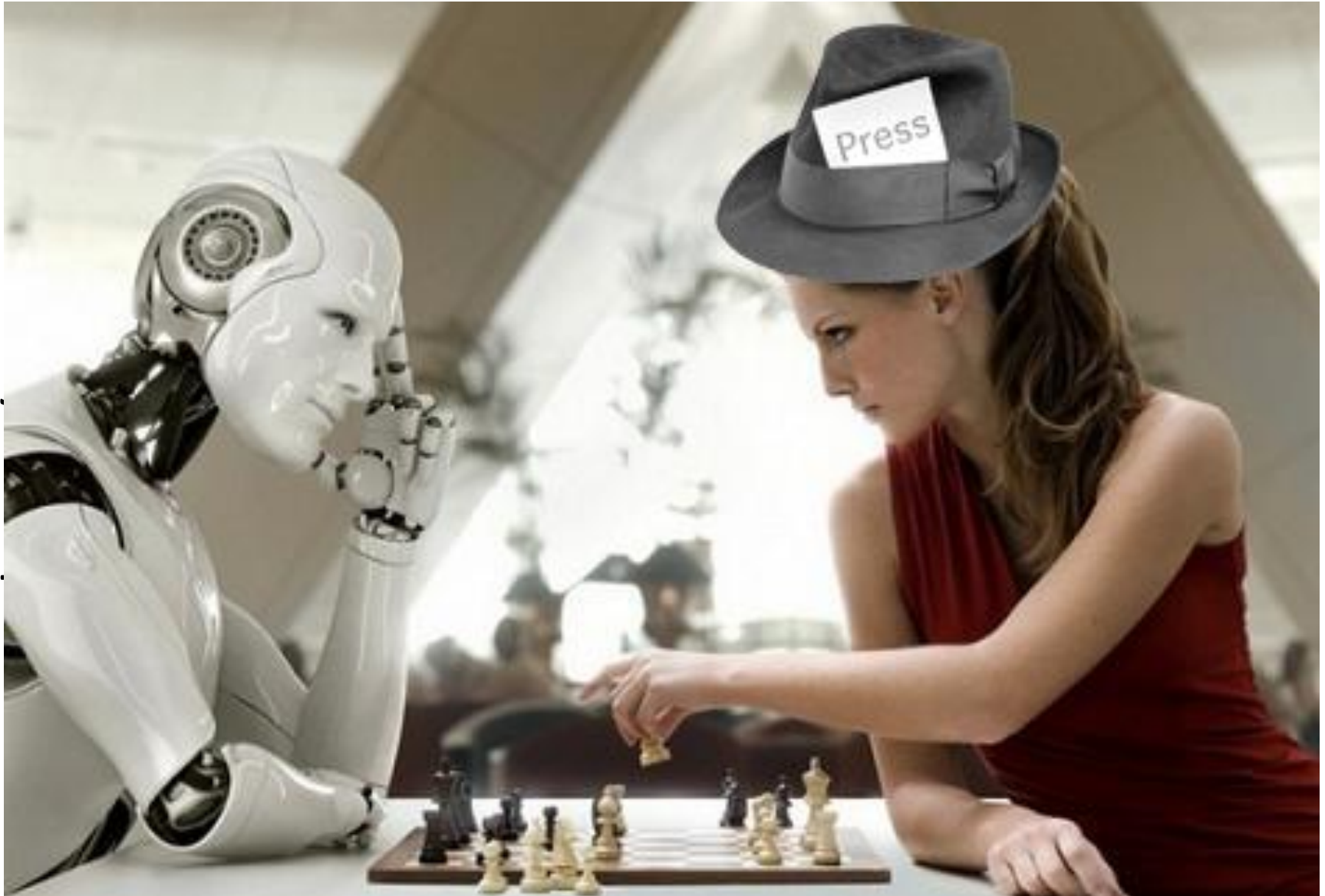
POPULARITY

---

Retweets

INFLUENCE

# Human vs. algorithmic evaluation of social media contributors

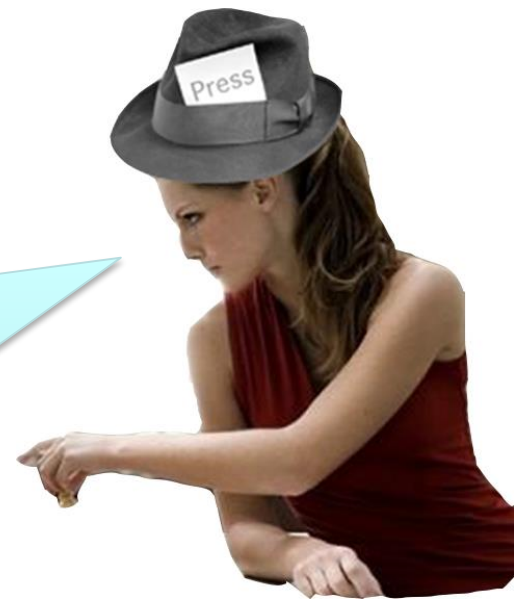






Inactive!

Yeah, but  
check out his  
followers!



**Gordon Corera**

@gordoncorera

BBC Security Correspondent - views expressed are my own not those of the BBC. author of MI6: Life and Death in the British Secret Service

Joined January 2011

TWEETS  
554

FOLLOWING  
408

FOLLOWERS  
3,195

FAVORITES  
1

LISTS  
2

Tweets

Tweets & replies

Photos & videos



**Gordon Corera** @gordoncorera · 10h

No classified material posted and no compromise of #CENTCOM servers in act of 'cybervandalism' US military says



**Gordon Corera** @gordoncorera · 14h



5/10

What? She's  
Deputy Leader of  
the Labour Party!



Social Sensor<sup>BETA</sup>

Search for...



Neil Thurman



Harriet Harman @HarrietHarman

*MP Camberwell and Peckham, Labour deputy leader, shadow culture secretary*

Influential on: [Ukraine, ALSIceBucketChallenge, ferguson, 1063mobileradio, BreakingNews]

Location: **Camberwell & Peckham, London**

## Truthmeter details

### Breakdown

#### History

9

Information about how active a given source is on different social media platforms, also combining it with validity information (when did this source appear, how frequently has it been posting)

#### Popularity

5

Information about how many and what kind of people are following the activity of this source, and how many are reading or are recommending this activity to others

#### Influence

5

Information about how much this source triggers activities, such as re-posts, discussions or comments. Are things happening because of what this source is posting?

### Truthmeter rating

#### Overall

5

This is an overall rating taking into account all the information extracted from the history, popularity and influence attributes, giving an overview of the trustworthiness of the contributor



## **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**  
@neilthurman

**Prof Steve Schifferes**  
[Steve.schifferes.1@city.ac.uk](mailto:Steve.schifferes.1@city.ac.uk)