



Introduction

- Educational background



- Career

< 2011



2011 to 2013



2014 >



- Hobbies



Topic: Mining tweets with R (and Gephi)

What makes twitter so powerful and what are some data mining techniques that can be applied to manage risk



A quick look at Twitter: Twitter is an effective communication channel

People use Twitter on desktops, tablets and mobile devices, to share events, news and **real-time conversations**. Deemed the 'sms' of the internet

- There are 100 million active Twitter users (3rd biggest social network)
- On a busy day, Twitter sees about 175 million tweets.
- If Twitter were a country, it would be the 12th largest in the world.
- There are more searches on Twitter each month than Yahoo and Bing—combined (24 billion vs. 4.1 billion and 9.4 billion, respectively).
- 92 percent of Twitter users say they would retweet “interesting” content.



Source:

http://www.prdaily.com/Main/Articles/20_illuminating_facts_about_Twitter_2012_edition_12138.aspx

<https://business.twitter.com/twitter-101>

Mistakes from the past: Tweets travel quickly and globally



- The headlines on ESPN was **quickly** retweeted (shared) **globally**
- Various trends ('#') was created such as **#racism** and was a trending topic for a few hours on twitter
- You may not be on Twitter, but a hot topic may still end up in Twitter (because it's a social network)

Source:
<http://www.sbnation.com/nba/2012/2/18/2807696/espn-chink-in-the-armor-headline-jeremy-lin/in/2573882>



Success from the past: Engaging the community at real time



- President Obama used Promoted Trends ([#Forward2012](#)) to communicate his messages to millions, and influencing what voters were talking about
- The most retweeted Tweet in history (of Twitter) until Ellen DeGeneres came along

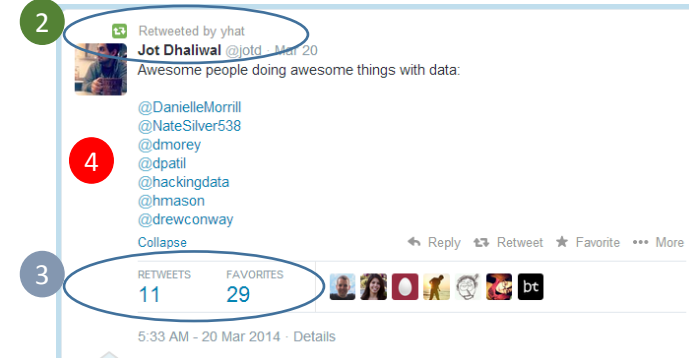
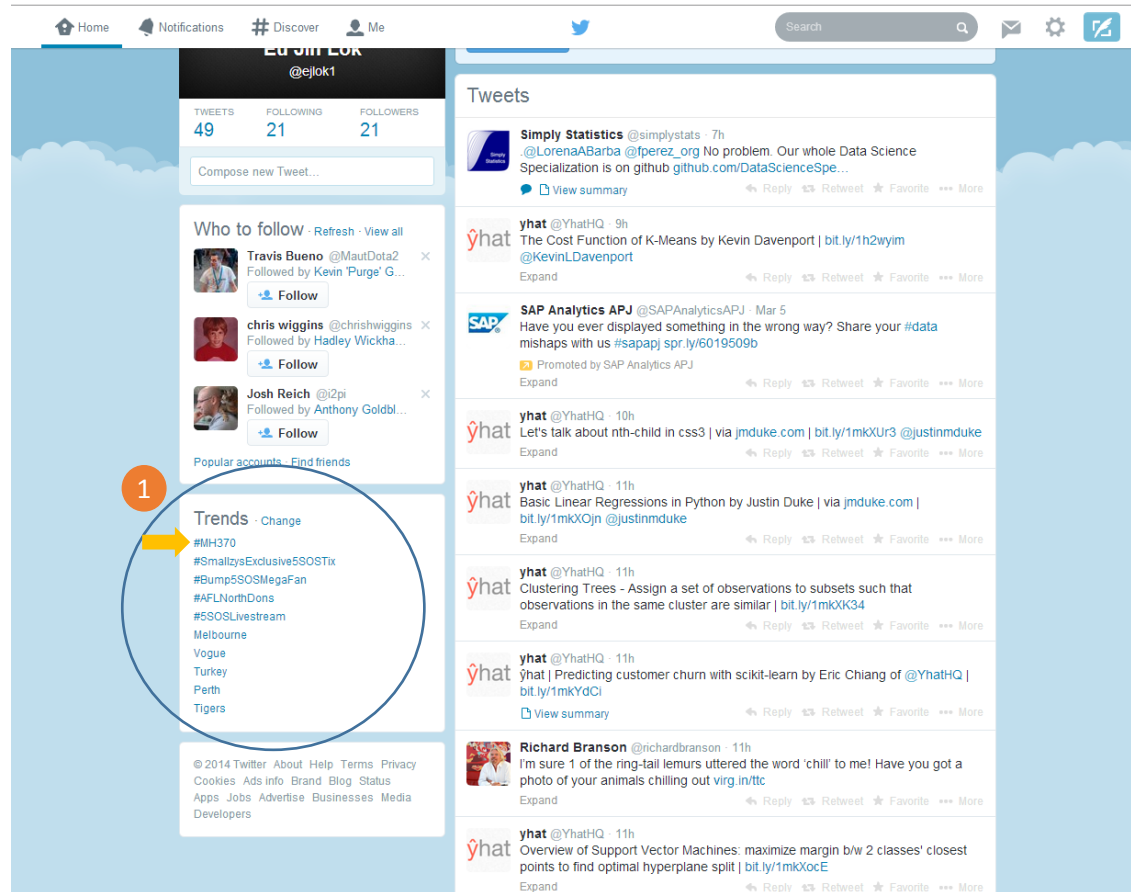
Source:

<https://business.twitter.com/success-stories/barack-obama>

It took less than an hour for Ellen's to beat Obama's retweet record and hit over 1.2 million retweets



A quick look at Twitter



- 1 Trending topics (Popular '#')
- 2 User retweets original tweet
- 3 Popularity of tweet
- 4 Mentions



Real life example: @MAS



Key questions to answer

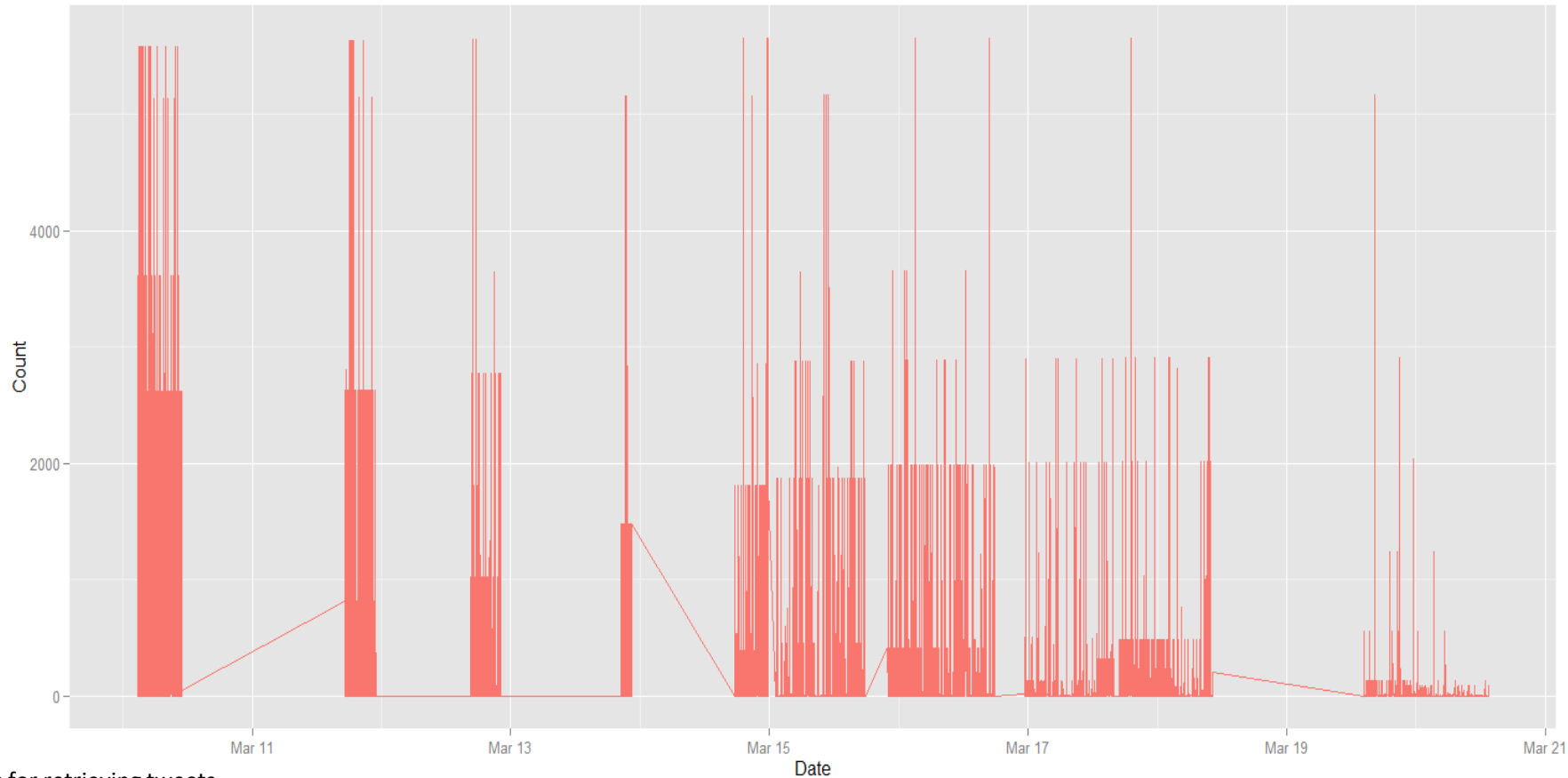
1. What are people tweeting about the MAS brand?
2. What are the popular trending topics?
3. How much brand damage has this tragedy done to the @MAS brand?
4. Who are the influential users in this network of tweets? Who is the Ellen DeGeneres of this network?

- NOTE: This presentation will focus on technical applications, not conclusions about the MH370 tragedy. This merely serve as an interesting example and is by no means a reliable interpretation of the missing MH370 tragedy, and no malice to any party or individual is intended



Overview of the tweets downloaded overtime on any tweets on the topic @MAS

Retweet counts overtime

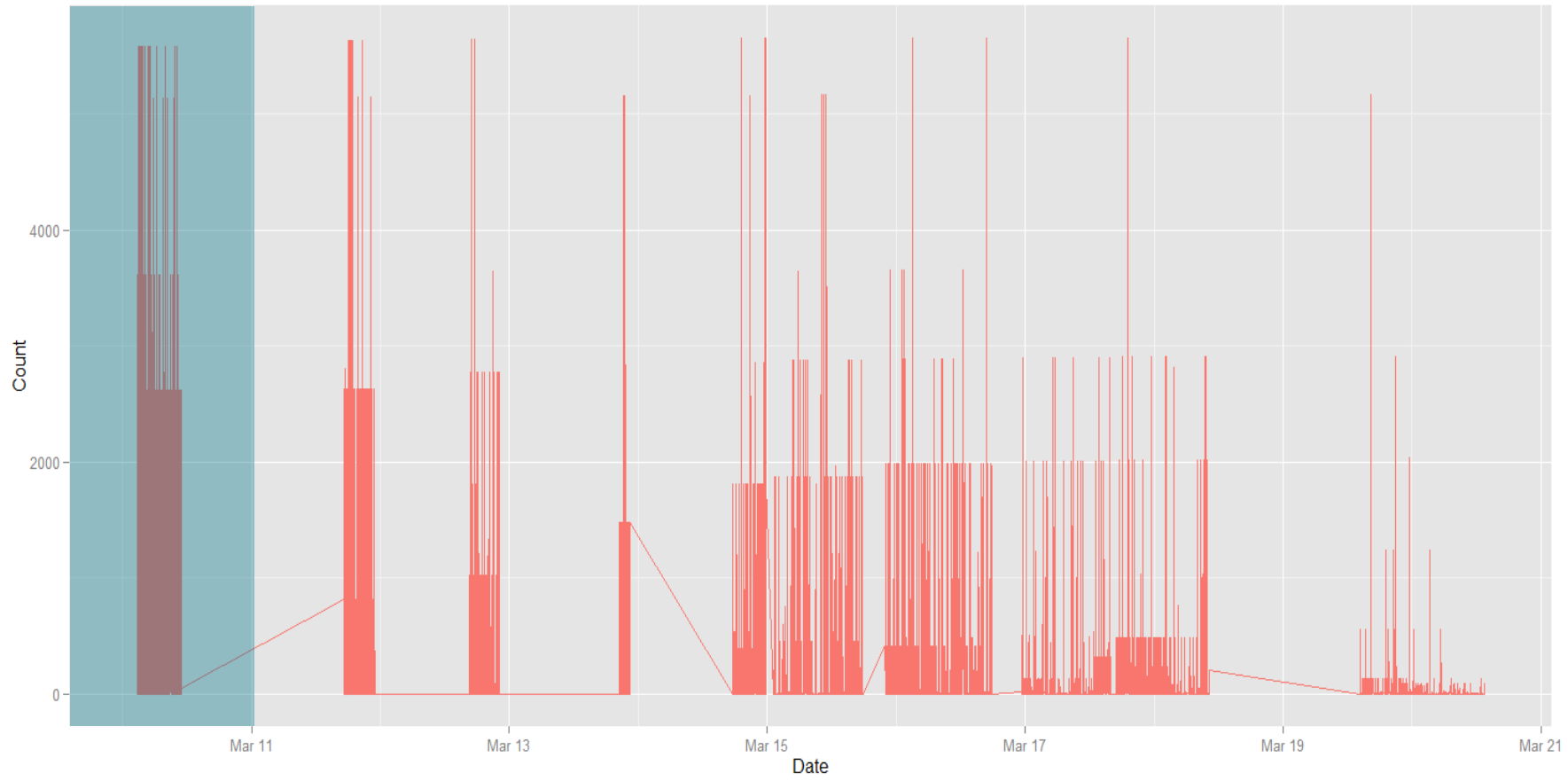


twitterR: for retrieving tweets
ggplot2: for timeseries

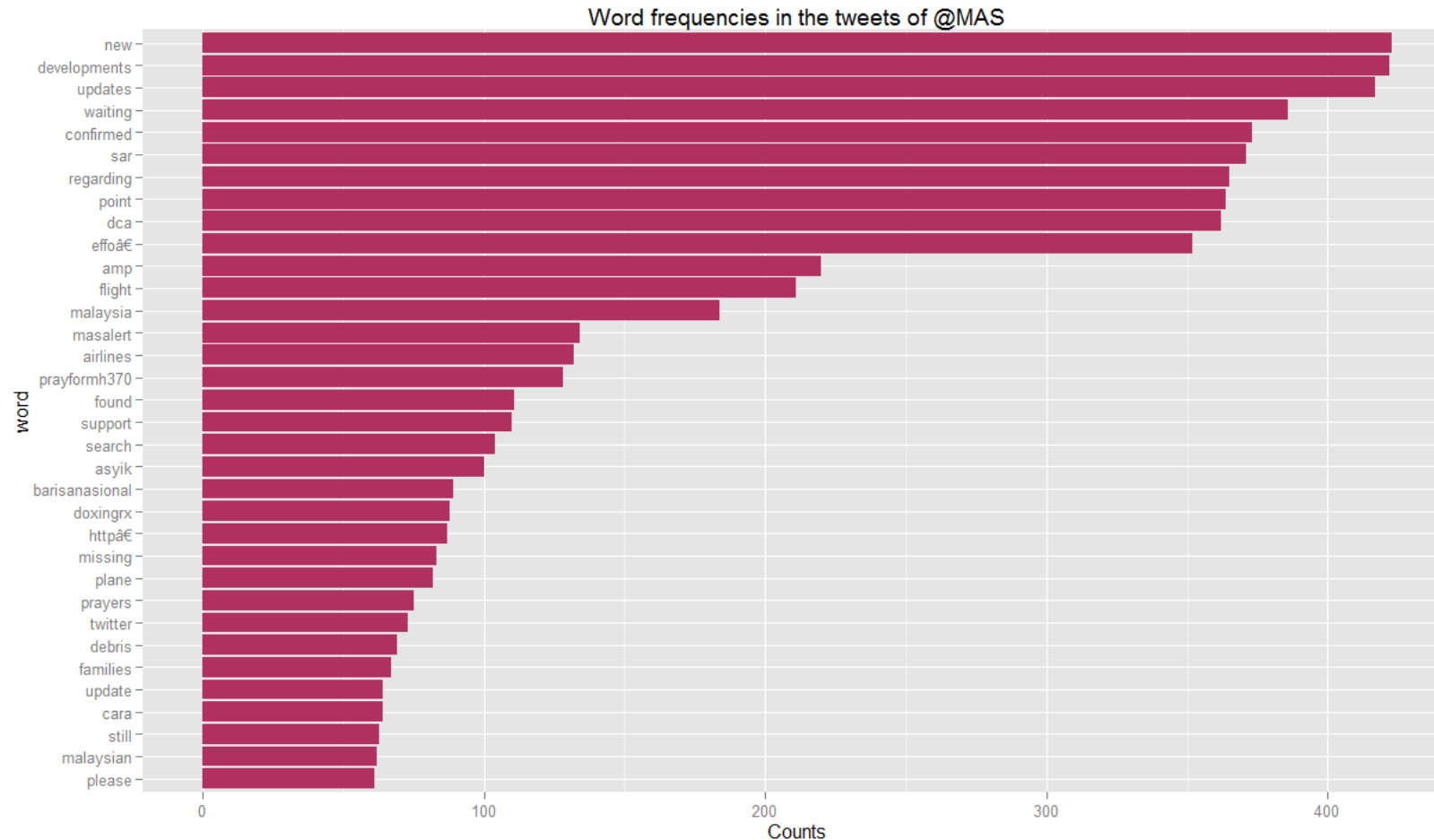


10th March 2014

Retweet counts overtime



10th March 2014: Highest tweet (retweet) comes from the official @MAS account



10th March 2014: Highest tweet (retweet) comes from the official @MAS account



A word cloud (pic above) is a visual representation of text which gives greater prominence to words that appear more frequently in the tweets



wordcloud



Malaysia Airlines
@MAS · Mar 10

At this point, no further developments regarding MH370 has been confirmed. We are waiting for new updates from **DCA on the SAR** efforts.

Collapse
↩ Reply
↻ Retweet
★ Favorite
⋮ More

RETWEETS
2,654

FAVORITES
601

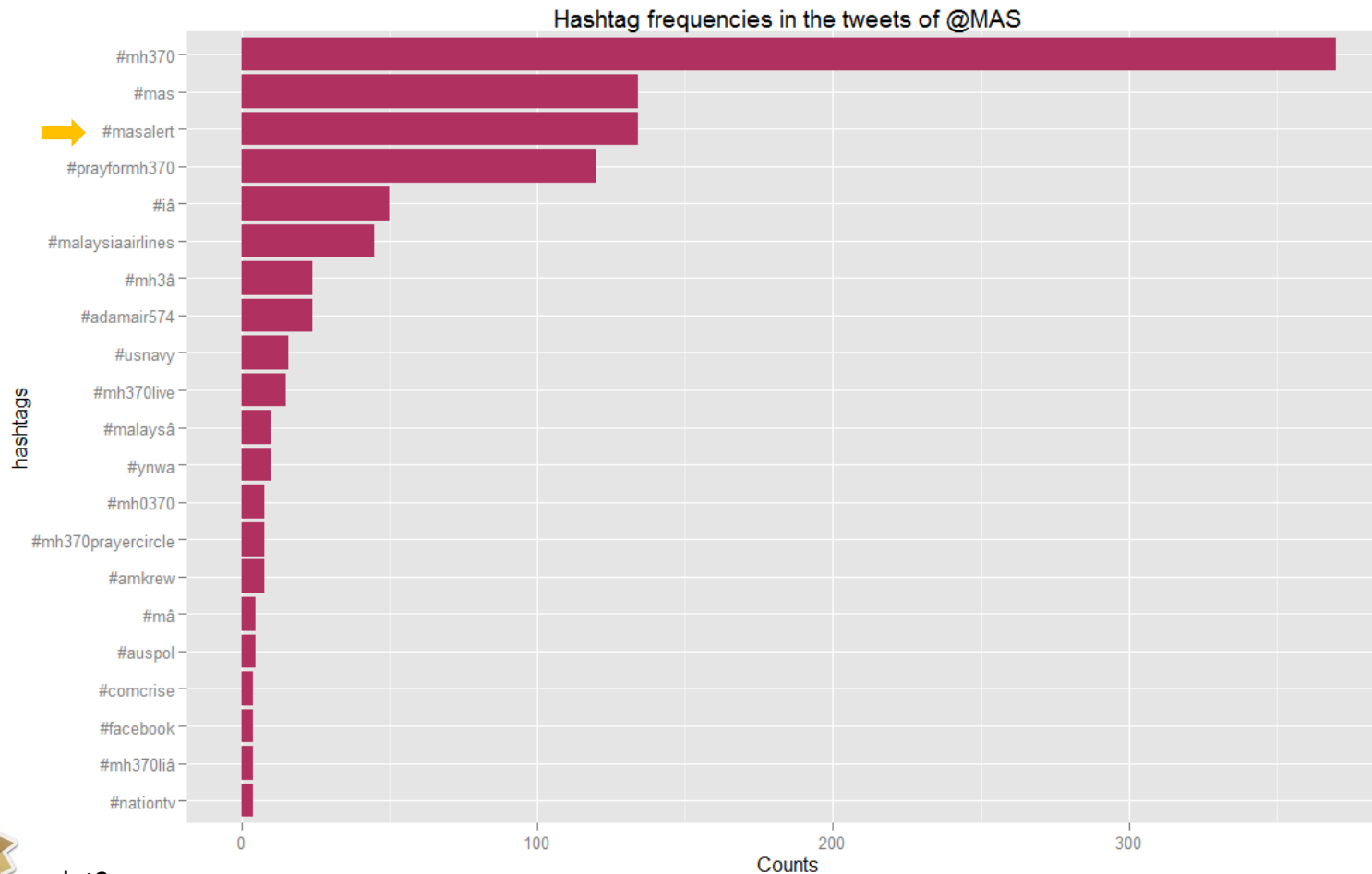


12:57 AM · 10 Mar 2014 · Details

Reply to **@MAS**

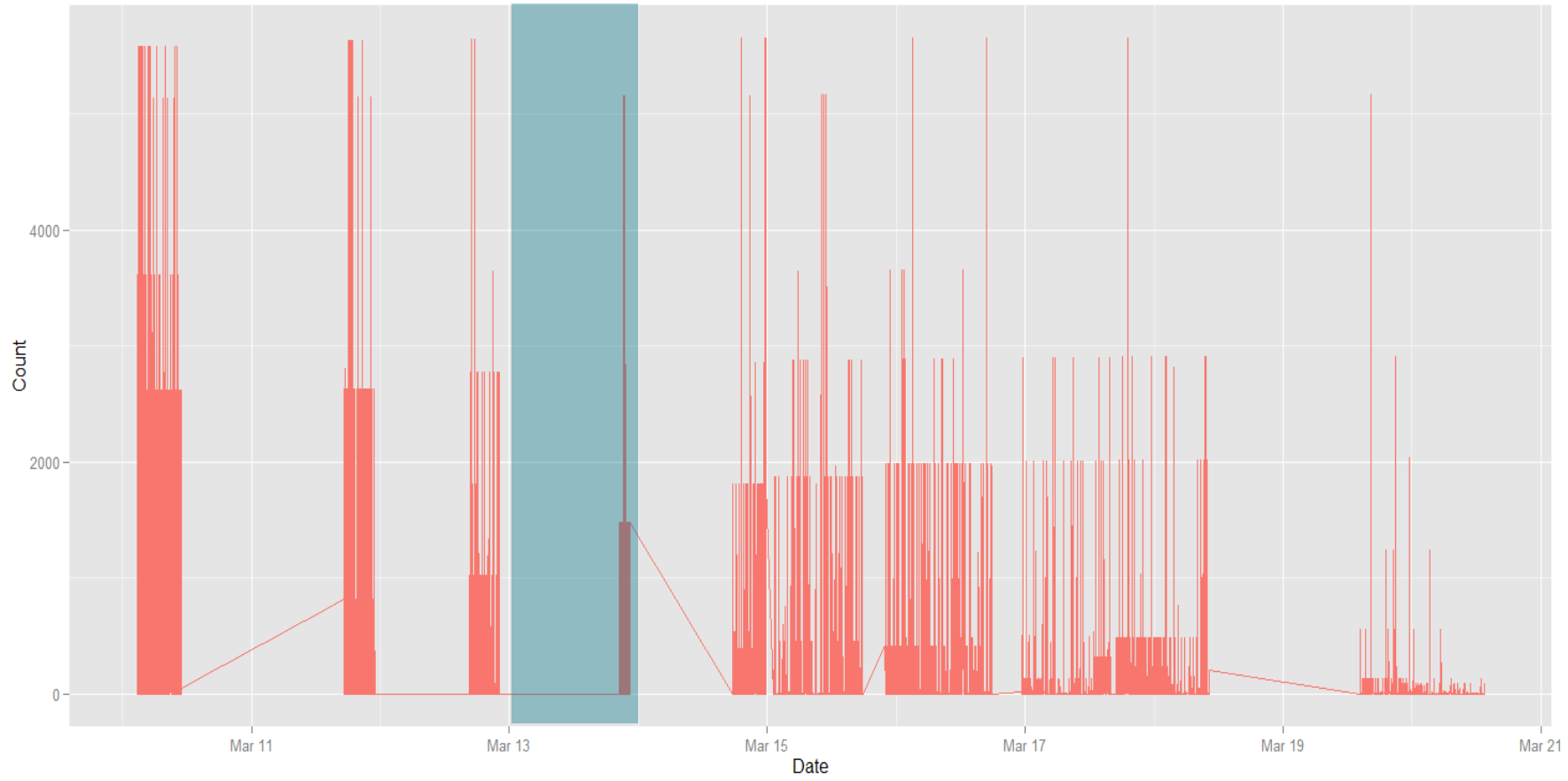


10th March 2014: Most popular trending topic is the #MH370 and by a mile, followed closely by #MASalert



13th March 2014

Retweet counts overtime

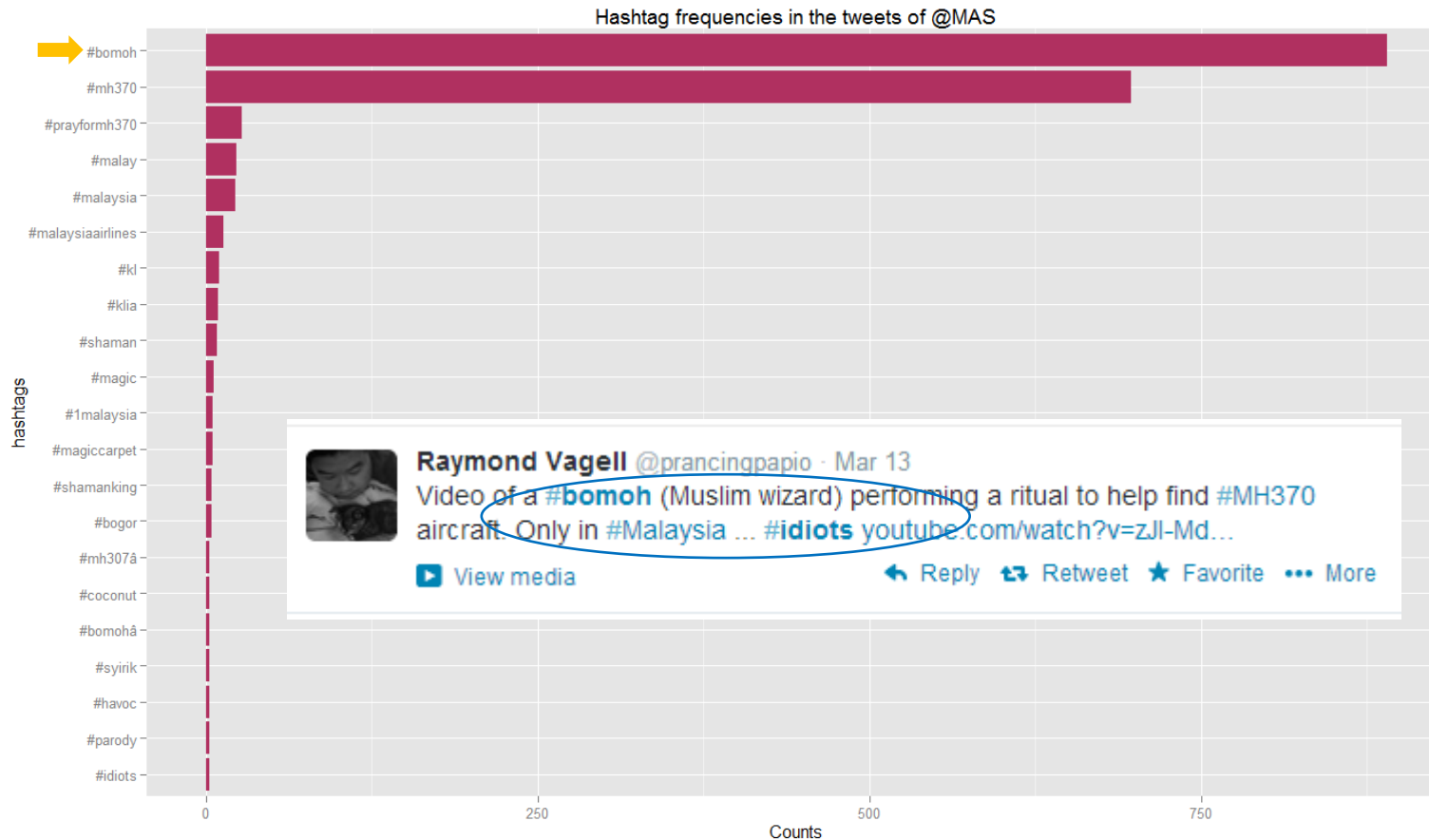


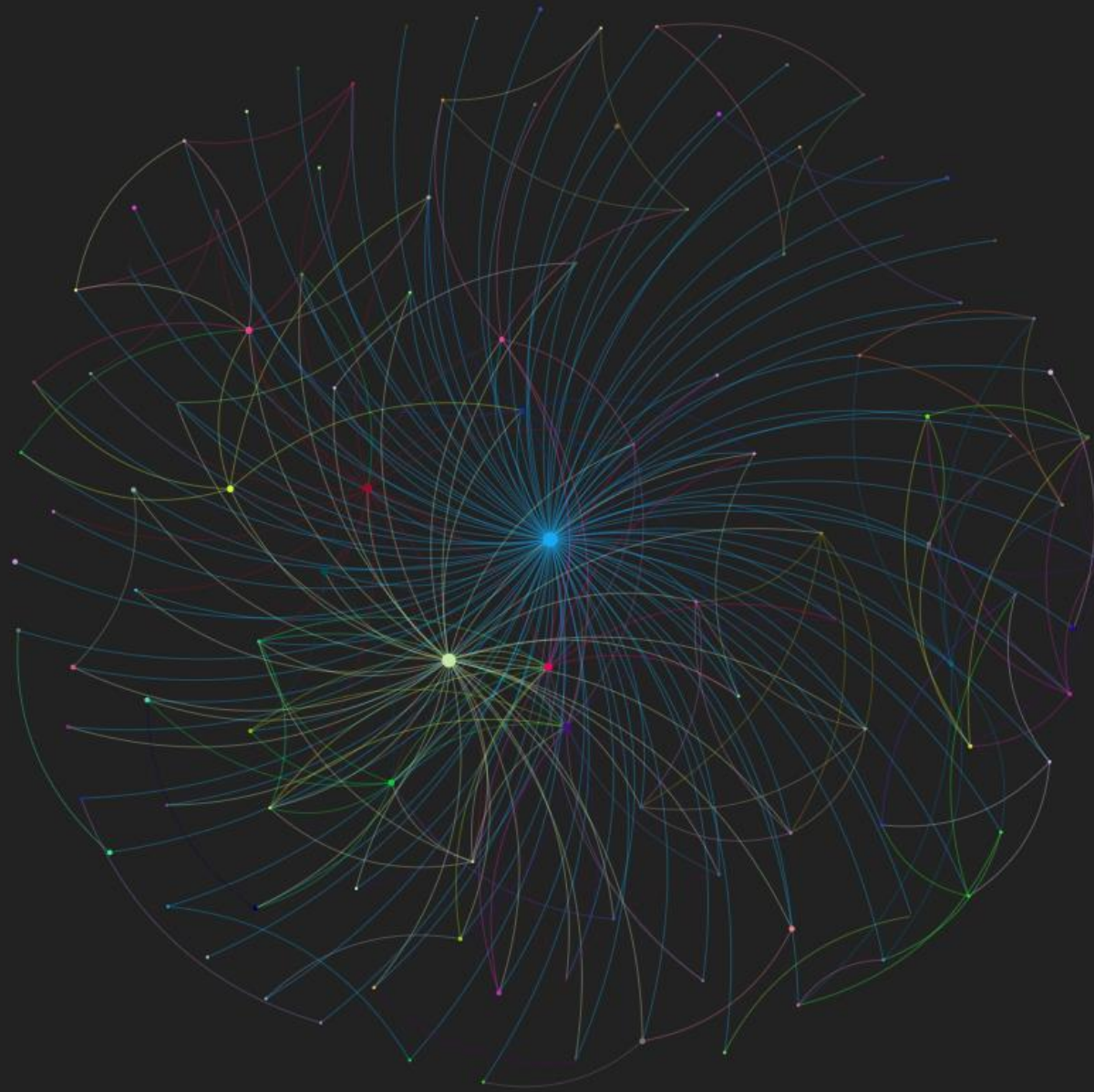
13th March 2014: The #Bomoh incident became the trending topic for the day. Not tweeted by @MAS

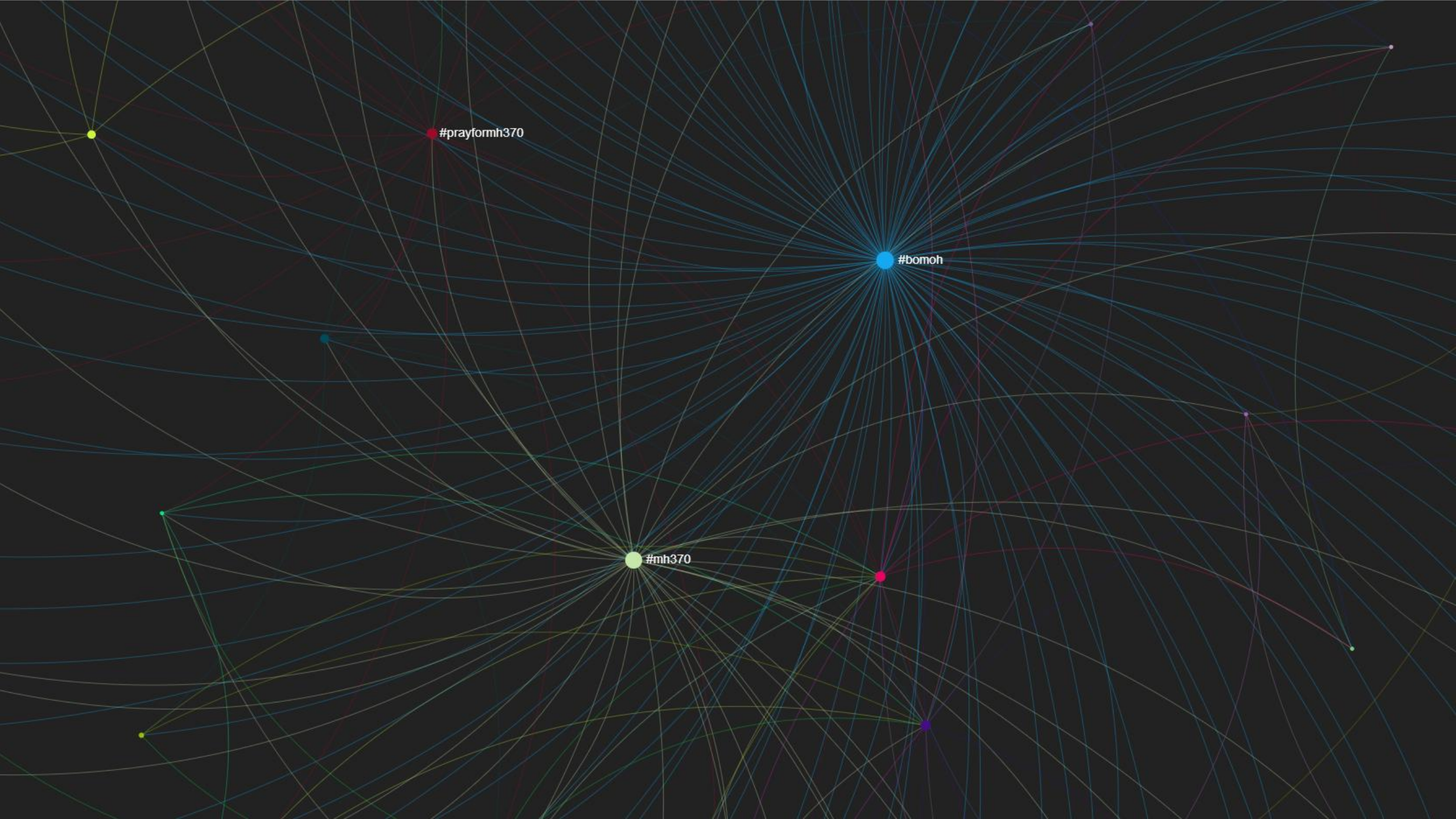


13th March 2014: The #Bomoh incident became the trending topic for the day. Not tweeted by @MAS

Most tweets have multiple trending topics included. Is there a way to visualise how different trends are linked?

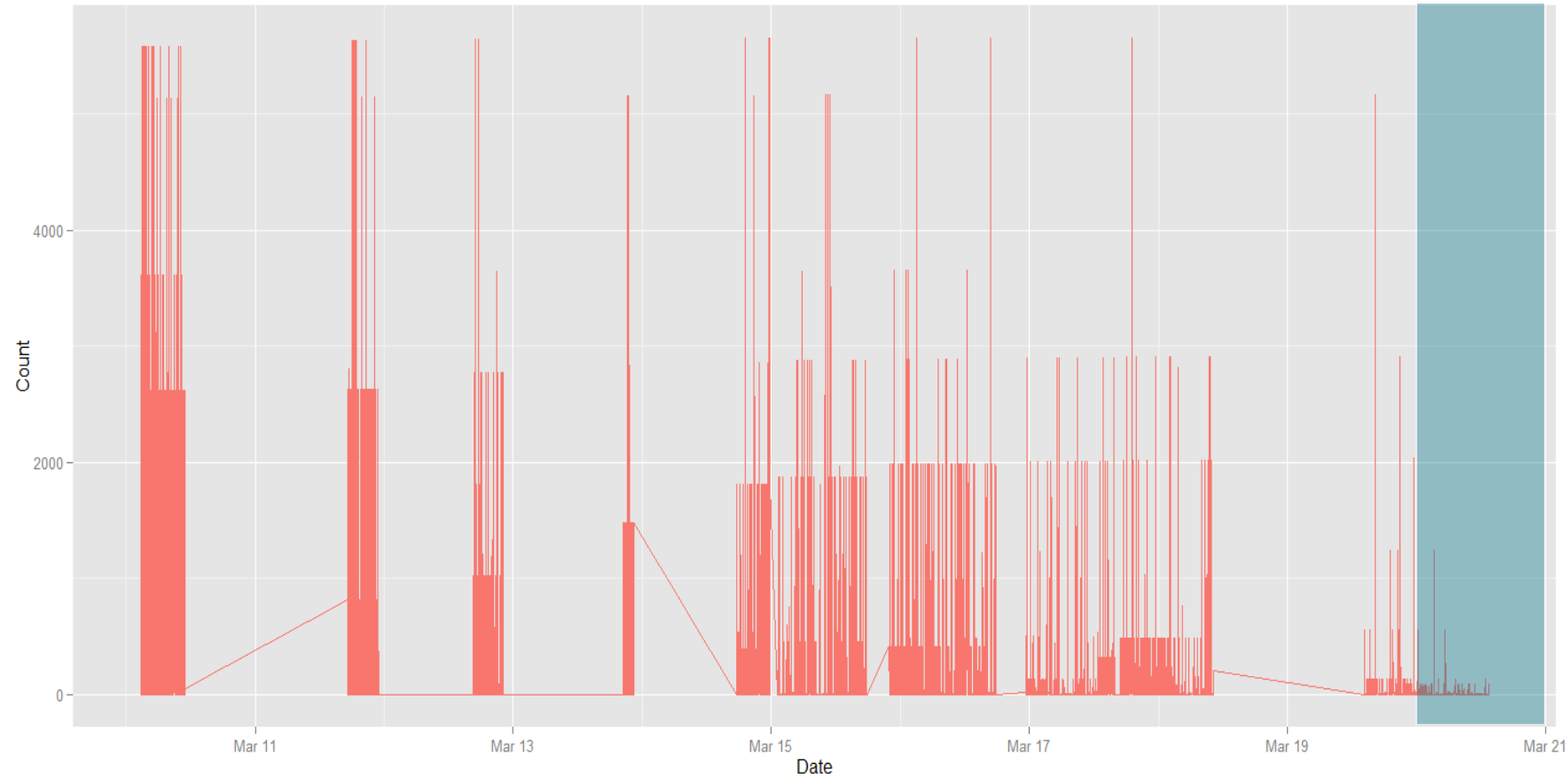




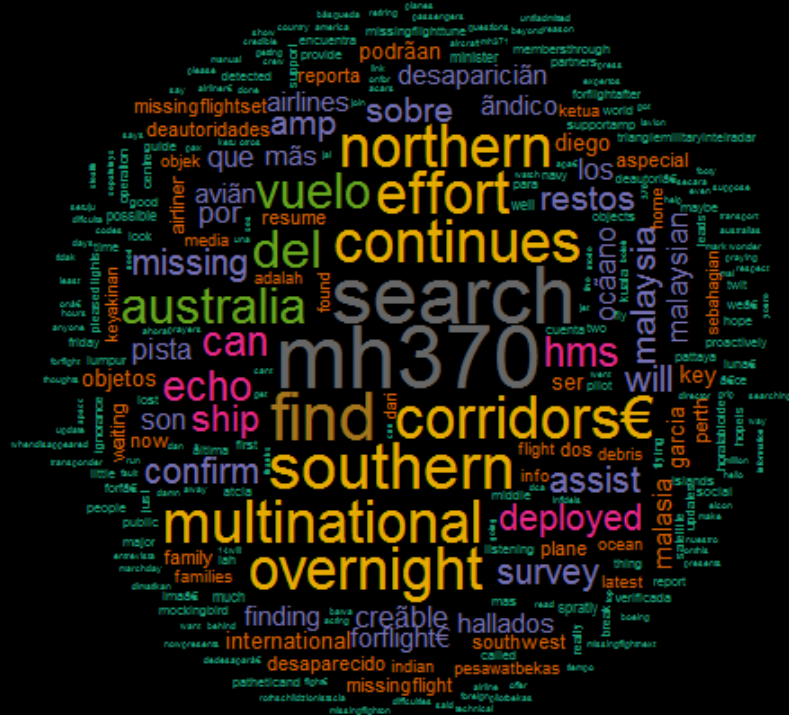


Fast forward to the 20th March 2014


Retweet counts overtime



20th March 2014: Flurry of tweets from different sources around the same topic



Related Searches: #mh370

501Awani @501Awani · Mar 20
 **#MH370** DS Hishammuddin: Confirmed that Msia sending 2 aircraft to Kazakhstan; the UK is planning to send 1 ship to the **southern corridor**
 Expand ↩ Reply ↻ Retweet ★ Favorite ⋮ More

 **H2O Comms** @H2OComms · Mar 20
18 ships are in **Southern Corridor**: 6 frm Msia, 1 from Australia, 5 frm China & 6 frm Indonesia #MH370 @HishammuddinH2O

Expand ↩ Reply ↻ Retweet ★ Favorite ⋮ More

sky
NEWS
BREAKING NEWS

Sky News Newsdesk @SkyNewsBreak · Mar 20

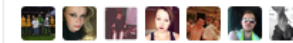
Malaysian Minister: 18 ships, 29 aircraft and six helicopters on ships in search for #MH370, UK planning to send ship to **southern corridor**

Collapse

Reply Retweet Favorite ... More


62 RETWEETS


FAVORITES
15







8:41 PM - 20 Mar 2014 - Details






Reply to @SkyNewsBreak

 **Georges J Danton** @GeorgesJDanton · Mar 20
@SkyNewsBreak Pilot and Co-Pilot of MH370 going through security prior to flight [youtube.com/watch?v=YVqvex...](https://www.youtube.com/watch?v=YVqvex...) ...

 View media

 Reply  Retweet  Favorite ... More

 **MEDIAtweet** @newstweete · Mar 20
@SkyNewsBreak Video Satellite imagery of the missing Malaysia Airlines MH370: youtu.be/UZdpk3DjER0?a #MH370

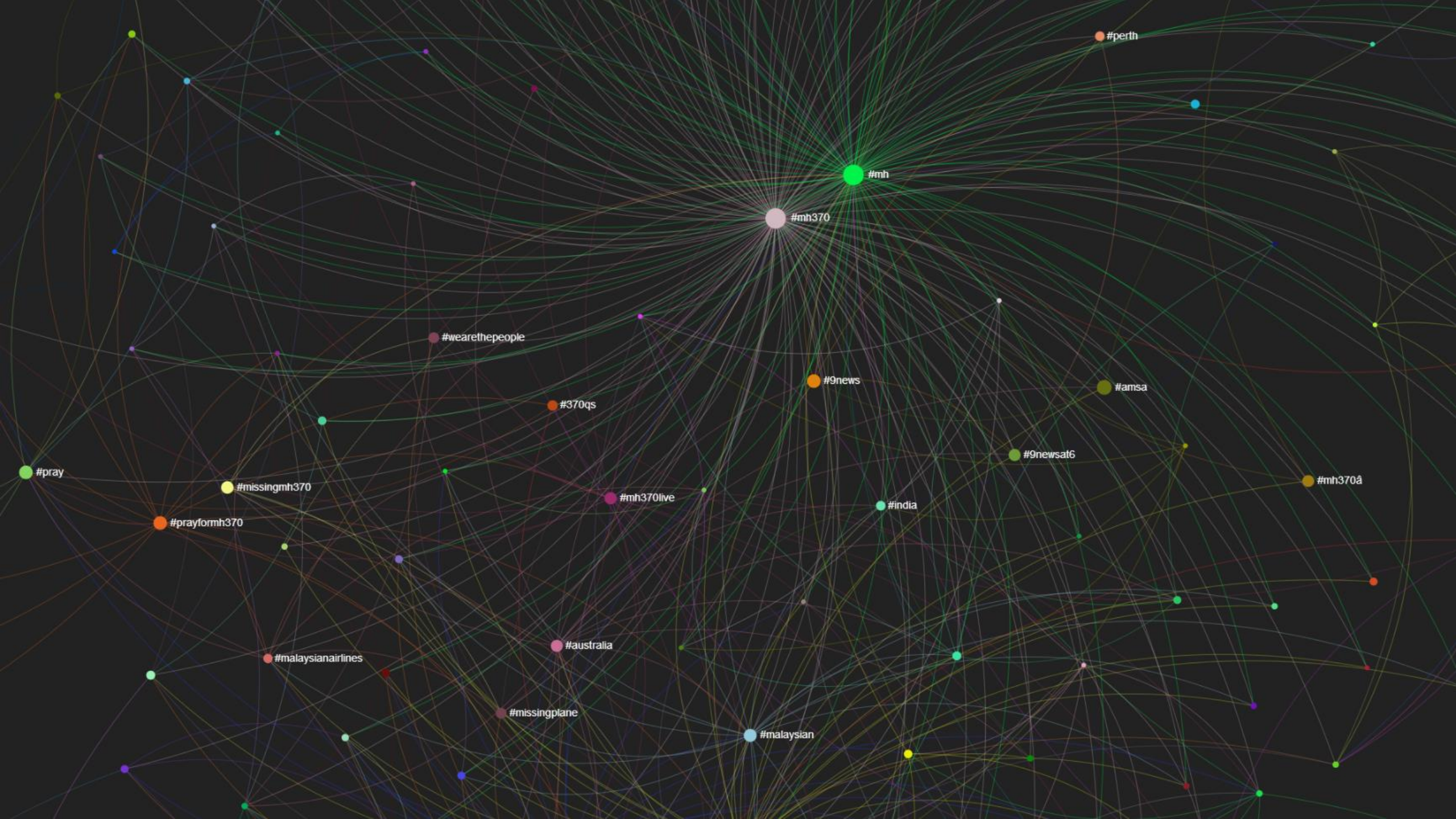
 View media  Reply  Retweet  Favorite  More

Malaysia Airlines @MAS · Mar 20
 [#MASalert] The multinational effort to find MH370 continues overnight in both the Northern and Southern corridors - bit.ly/MH370updates
 Collapse Reply Retweet Favorite More

238

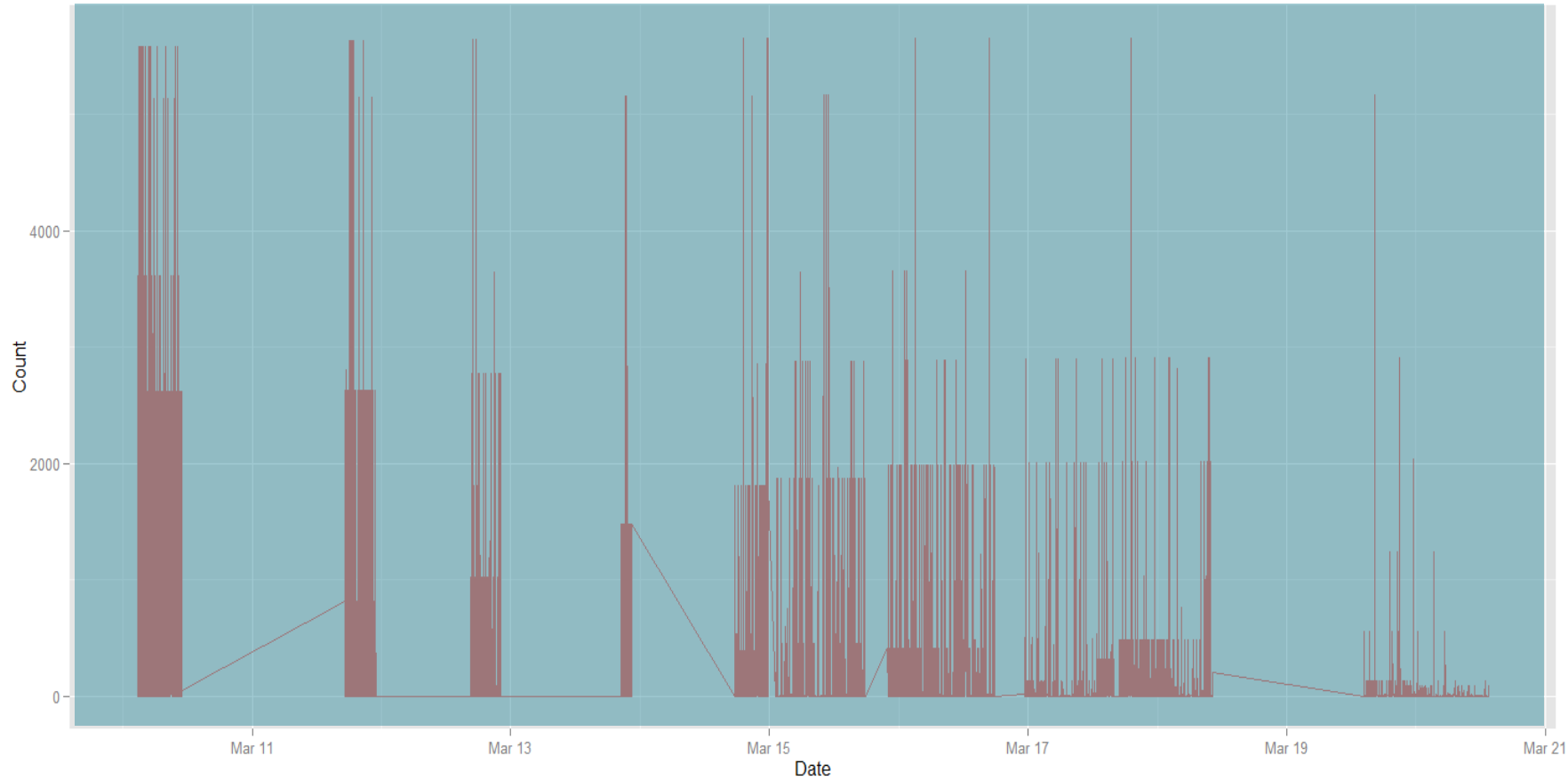
FAVORITES
101





Consolidated View: Time to answer the questions

Retweet counts overtime



The 4 key questions

1. What are people tweeting about the MAS brand?
2. What are the popular trending topics?
3. How much brand damage has this tragedy done to the @MAS brand?
4. Who are the influential users in this network of tweets? Who is the Ellen DeGeneres of this network?



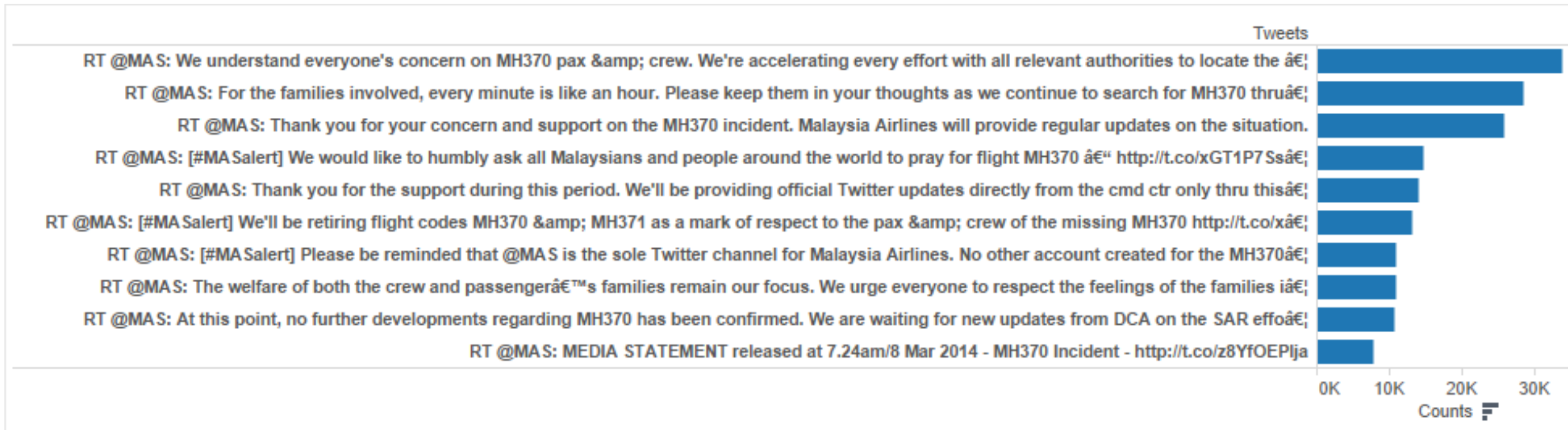
The 4 key questions

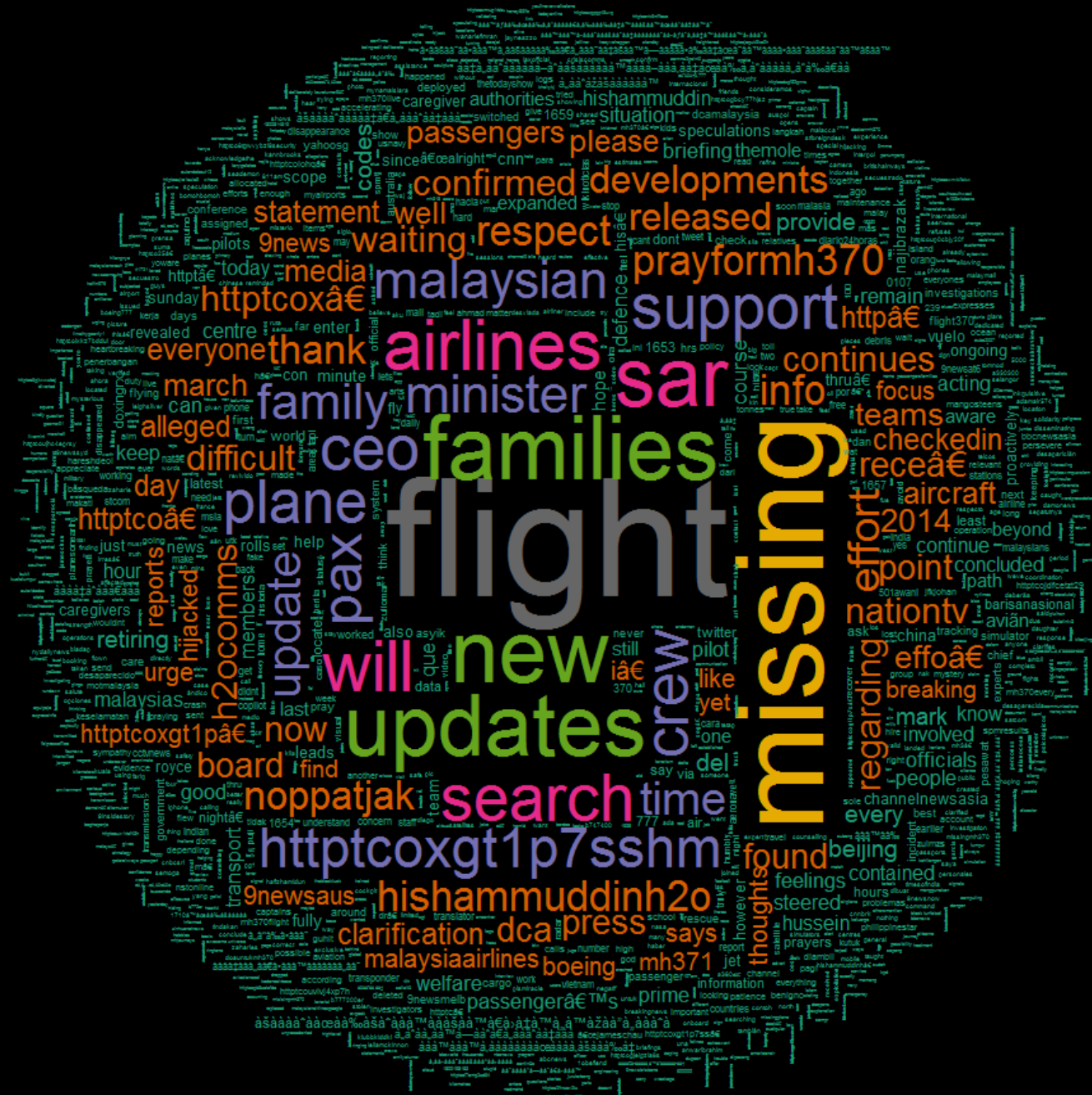
1. What are people tweeting about the MAS brand?
2. What are the popular trending topics?
3. How much brand damage has this tragedy done to the @MAS brand?
4. Who are the influential users in this network of tweets? Who is the Ellen DeGeneres of this network?



What are people tweeting?: Popular tweets (retweets) are MAS official tweets

- No damaging tweets within the top 20



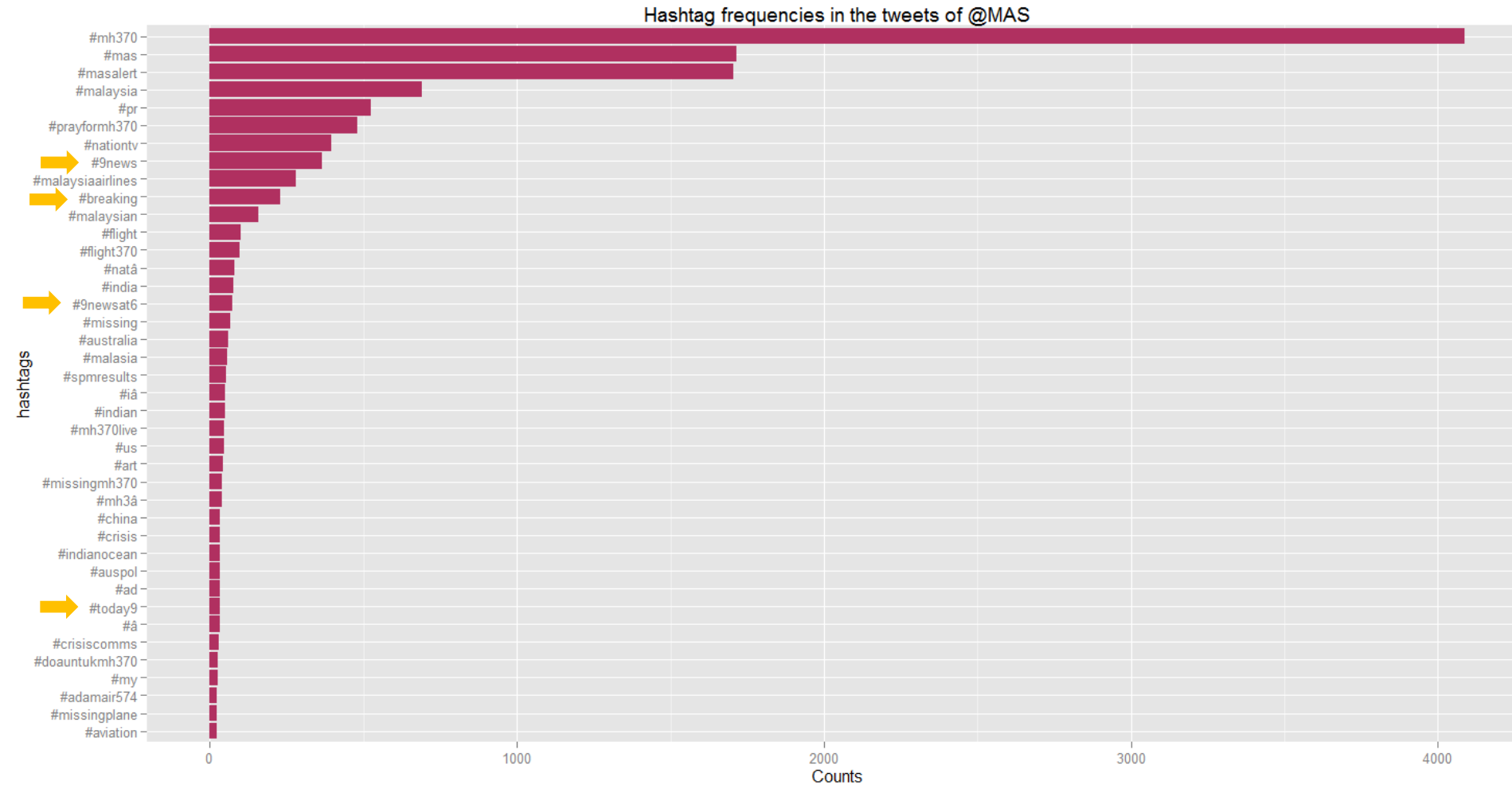


The 4 key questions

1. What are people tweeting about the MAS brand?
2. **What are the popular trending topics?**
3. How much brand damage has this tragedy done to the @MAS brand?
4. Who are the influential users in this network of tweets? Who is the Ellen DeGeneres of this network?

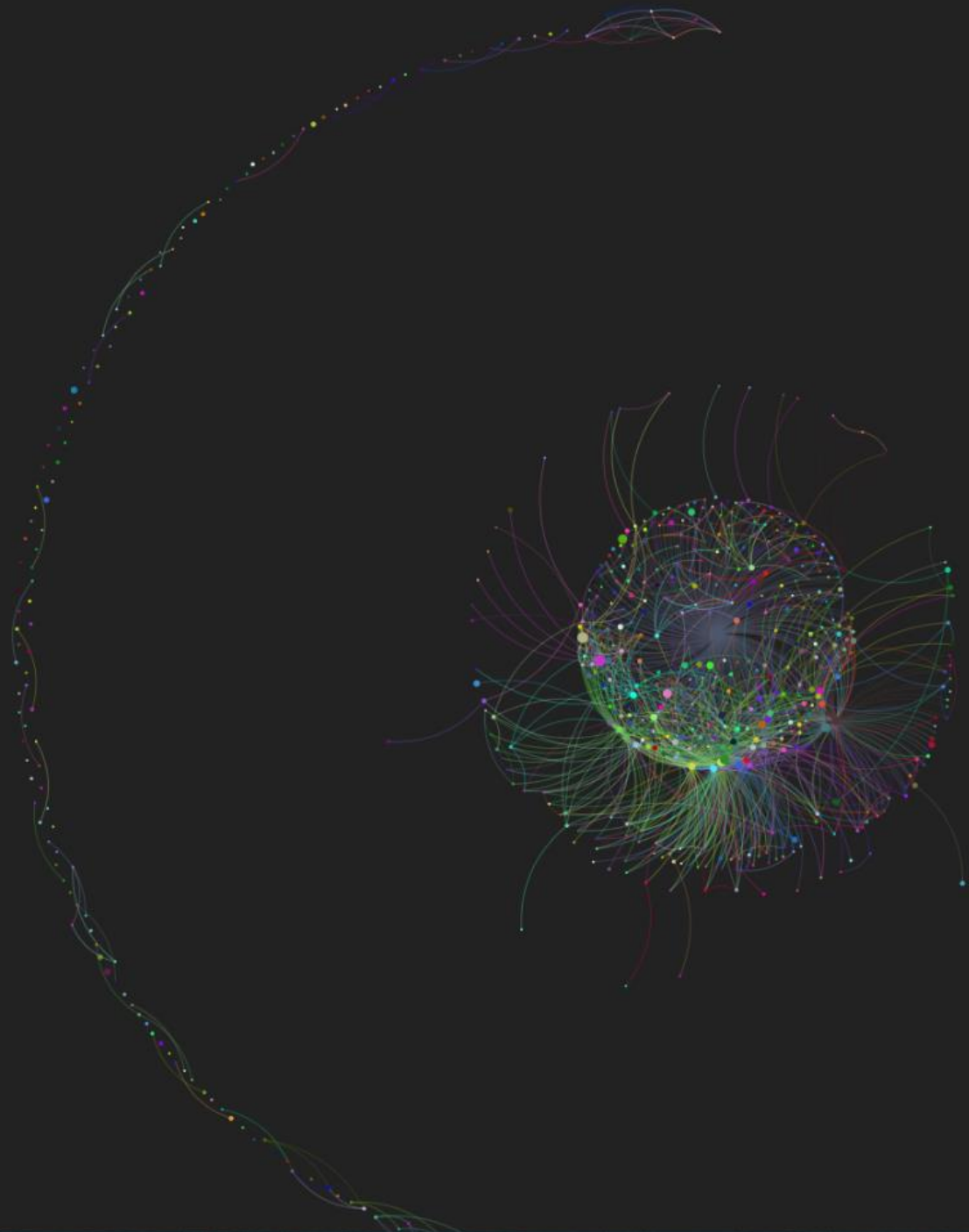


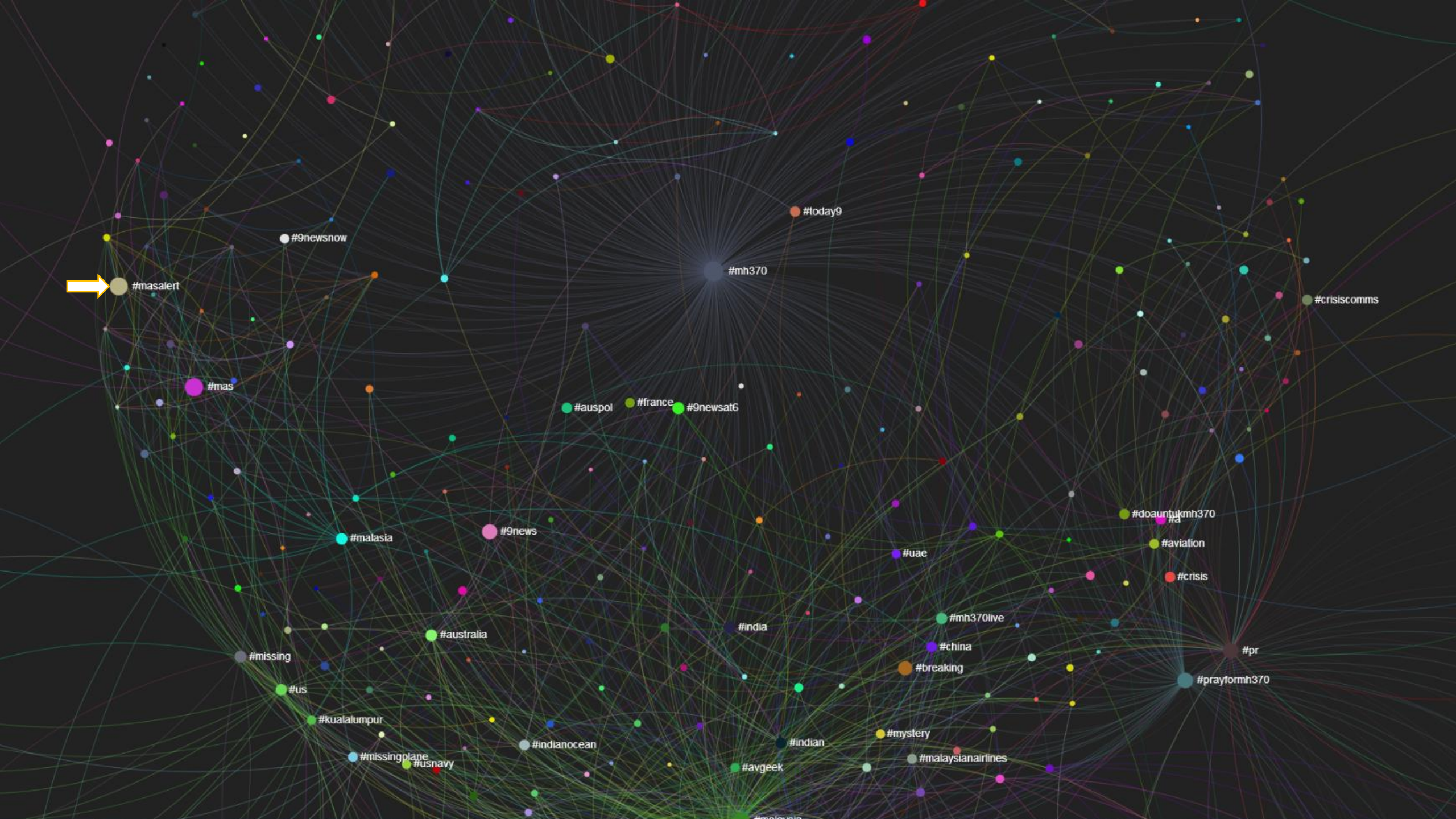
What are the popular trending topics?: #masalert is the 3rd most popular trending topic being used. No damaging trending topics



Also notice that there are lots of news channels from Australia







The 4 key questions

1. What are people tweeting about the MAS brand?
2. What are the popular trending topics?
3. How much brand damage has this tragedy done to the [@MAS](#) brand?
4. Who are the influential users in this network of tweets? Who is the Ellen DeGeneres of this network?



How much brand damage done to the @MAS brand?: Sentiment Analysis

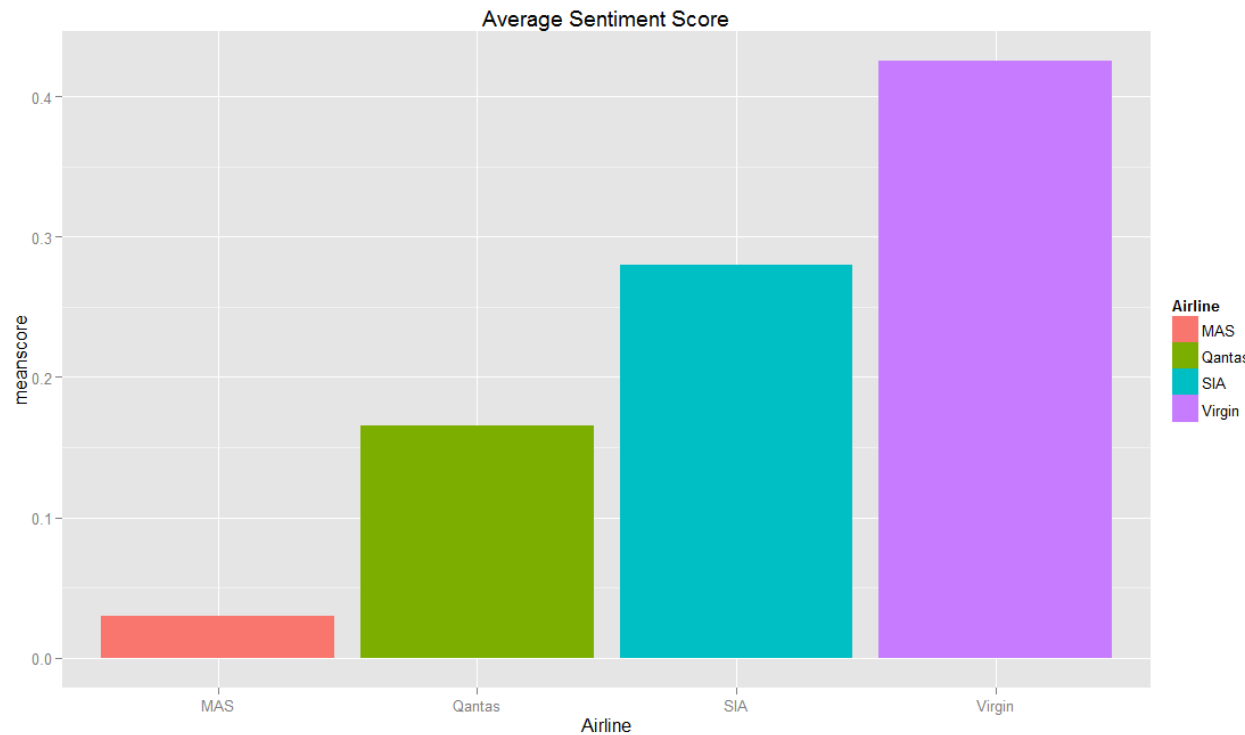
- Jeffrey Breen's method (<http://jeffreybreen.wordpress.com/2011/07/04/twitter-text-mining-r-slides/>)
- Simple algorithm which assigns a score by counting the number of occurrences of 'positive' and 'negative' words in a tweet. These 'positive' and 'negative' words is compiled by academic researchers
- Example of how it works:

Example of tweets about @Qantas	Score
Malaysian amp China students showing support for @MAS in a solidarity football match SussexUni	2
I cant believe @MAS are still using the #MH370 callsign absolutely disgraceful are they idiots r wat aircraft aint even found yet #avgeek	-2
RT @MAS No new update as yet but the search amp rescue teams are hard at work Next tweet in an hour or two depending on developments	0



How much brand damage done to the @MAS brand?: @MAS has the lowest sentiment score, followed by @Qantas

- Using other airlines as benchmark



ggplot2



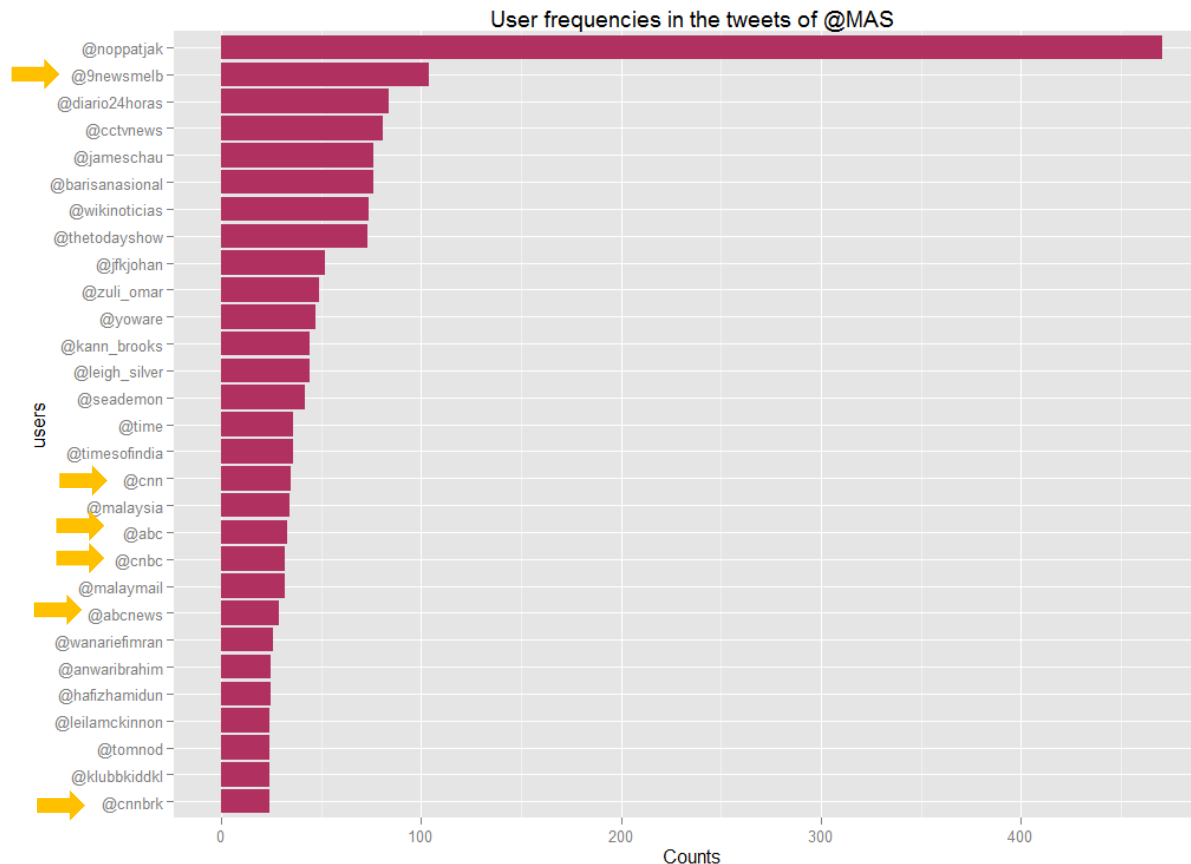
The 4 key questions

1. What are people tweeting about the MAS brand?
2. What are the popular trending topics?
3. How much brand damage has this tragedy done to the @MAS brand?
4. Who are the influential users in this network of tweets? Who is the Ellen DeGeneres of this network?



Who are the influential users in this network?: Most popular User (other than @MAS) is @noppatjak?

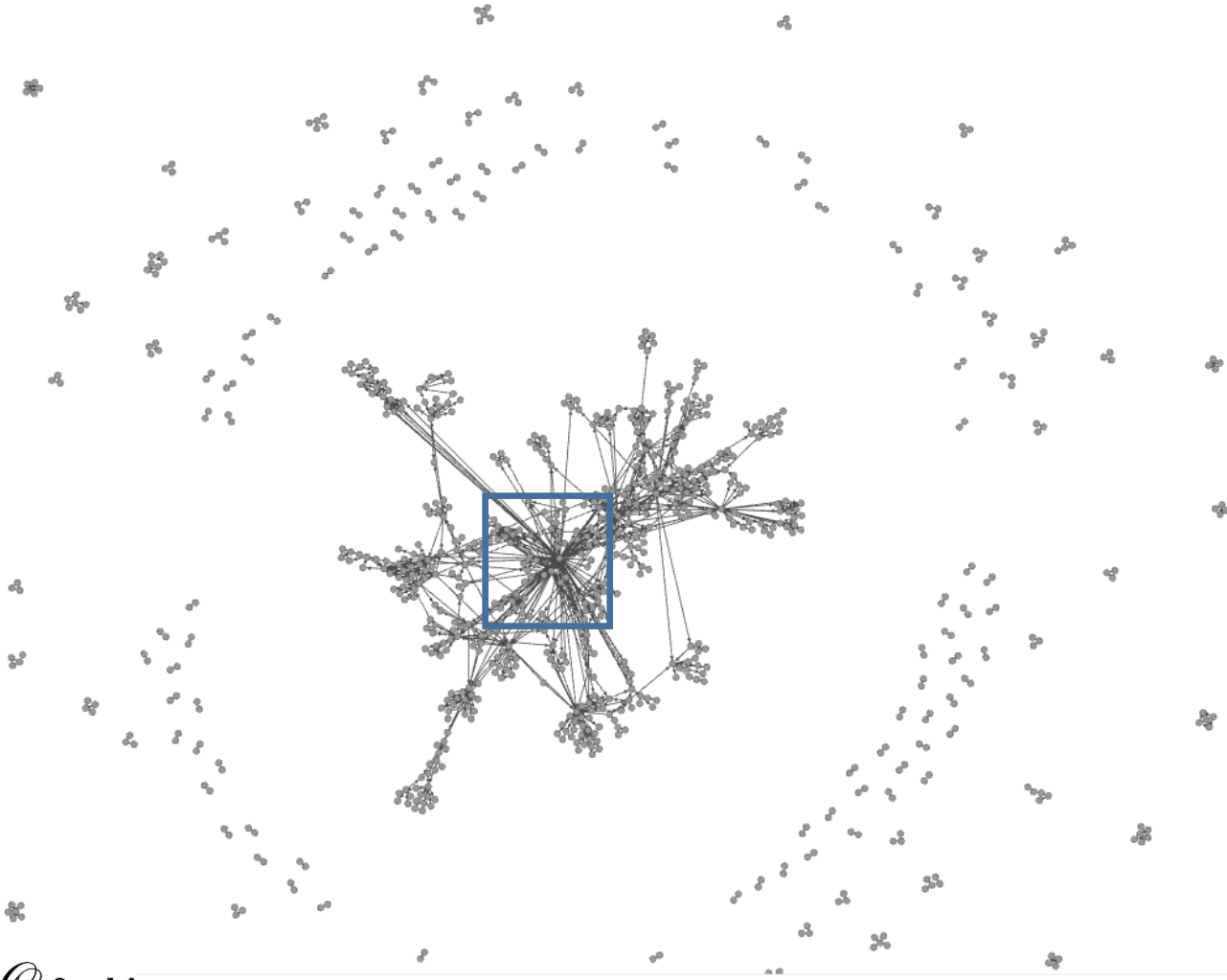
@noppatjak has lots of tweets. So who is @noppatjak and does this mean that he is most influential user after @MAS?



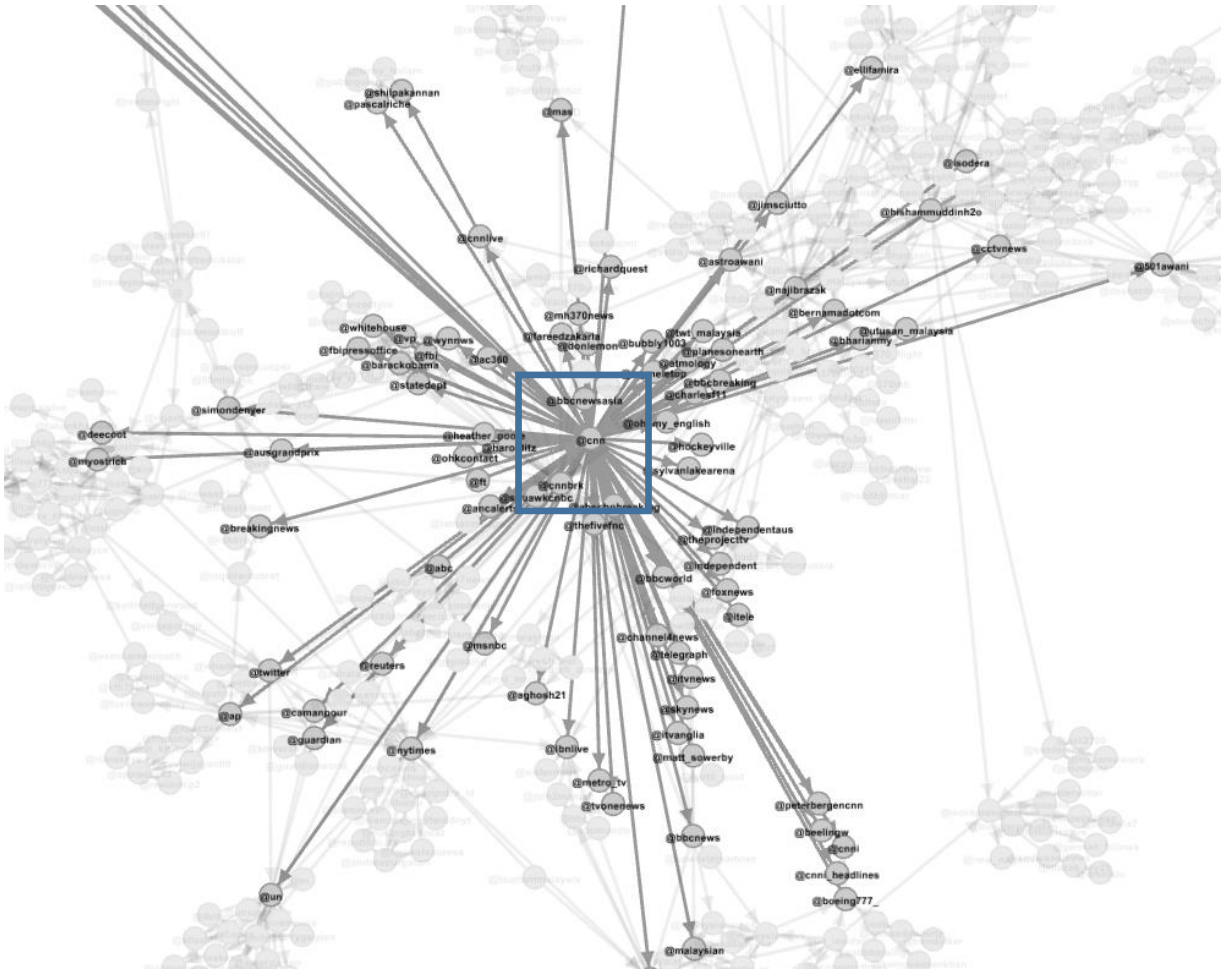
A few news channels made it to the list as well such as @9newsmelb and @cnn



Who are the influential users in this network?: A group of users are key influencers in the network



Who are the influential users in this network?: A group of users are key influencers in the network



Label	▼ Degree	Closeness Centrality	Betweenness Centrality
@cnn	92	2.475	77,386.611
@cnnbrk	46	2.836	11,850.071
@hishammuddin2o	38	2.97	34,961.448
@malaysia	36	3.161	17,966.574
@najibrazak	35	2.893	19,276.265
@501awani	27	3.219	16,666.165
@nytimes	24	2.918	18,403.44

Measure	Explanation	Even shorter explanation
Degree	most direct connections in the network	Popularity
Closeness Centrality	Shortest path to all users (takes few steps to reach any user in the network)	Key connections
Betweenness Centrality	Borker roles. Losing this node will disconnect a large group of users	Mediators



Who are the influential users in this network?: Removing these key users, is like removing vowels from the vocabulary

@CNN is an important user in the network. If we remove @CNN and the other key important users, the network will shrink



Who are the influential users in this network?: Use PageRank to rank the most important user

“PageRank is an algorithm used by Google Search to rank websites in their search engine results. PageRank was named after Larry Page,[1] one of the founders of Google. PageRank is a way of measuring the importance of website pages”

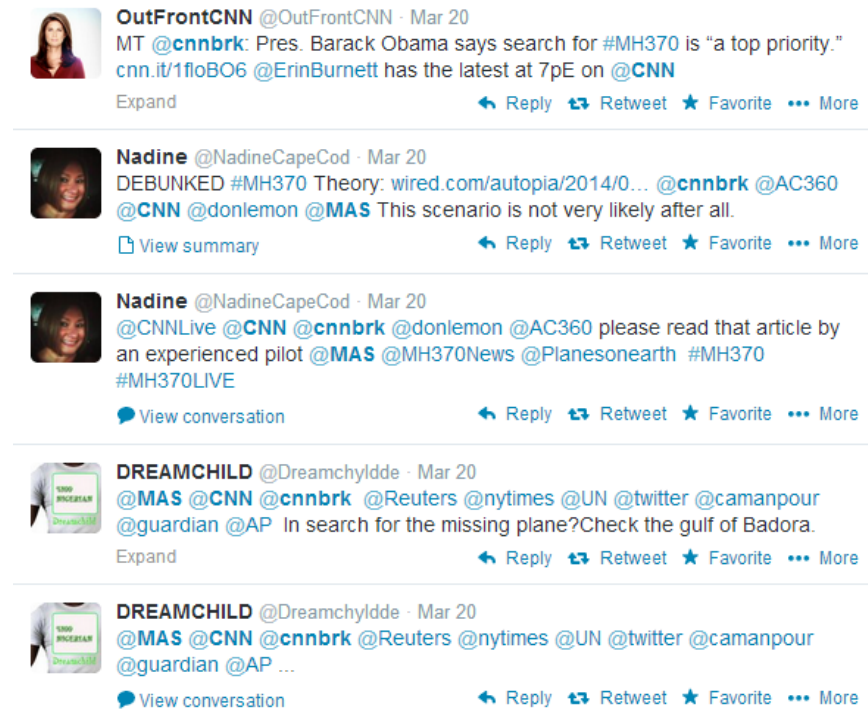
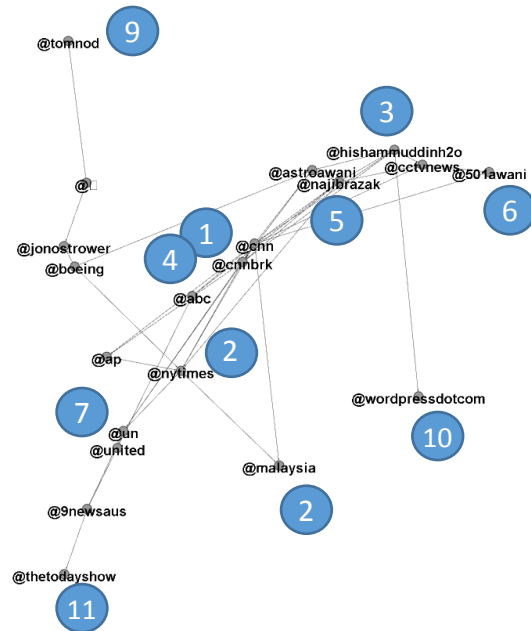


Source: <http://en.wikipedia.org/wiki/PageRank>

Who are the influential users in this network?: @CNN is the 'highest value' user in the network, therefore most influential user

These key users generate many tweets on the same topic, and have a large followers base. This prompts a lot of response all over the world

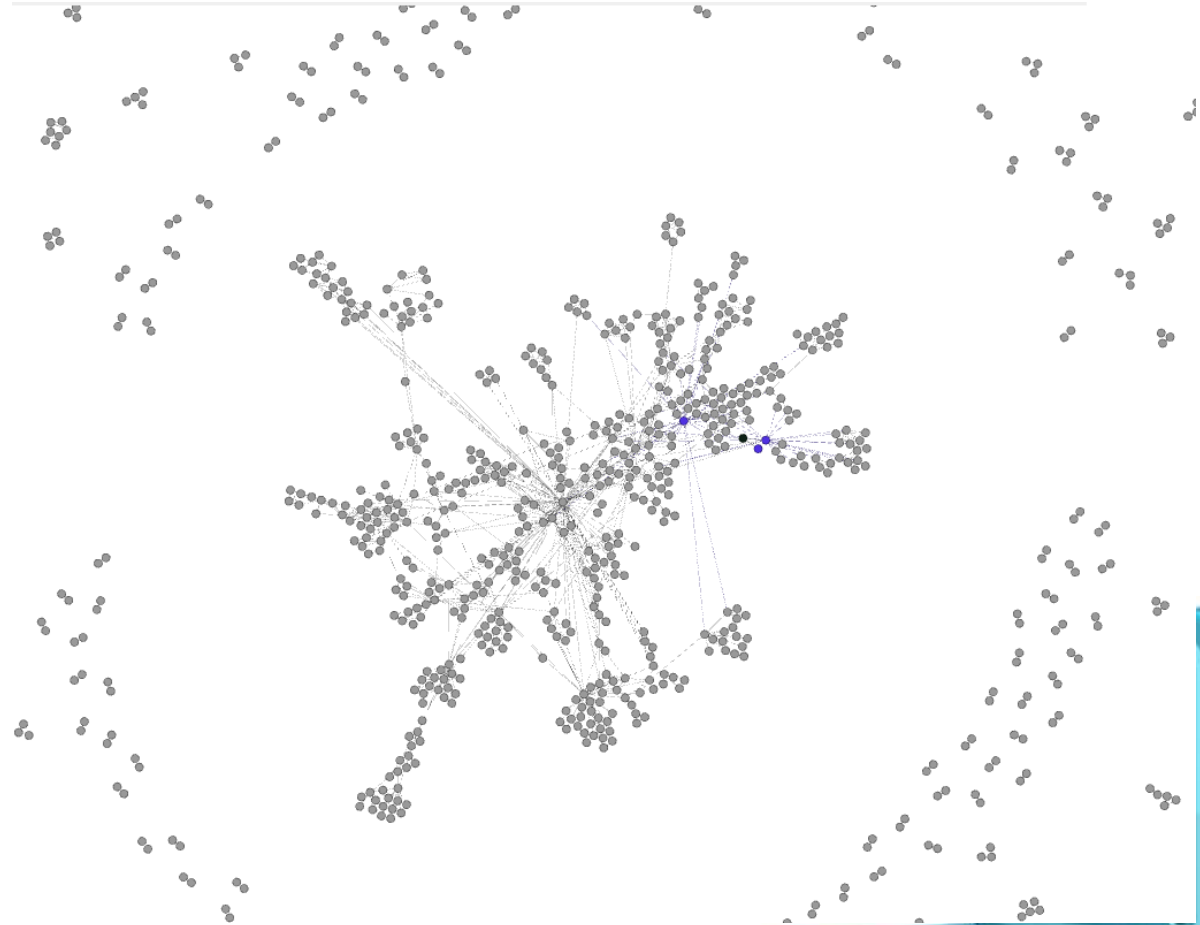
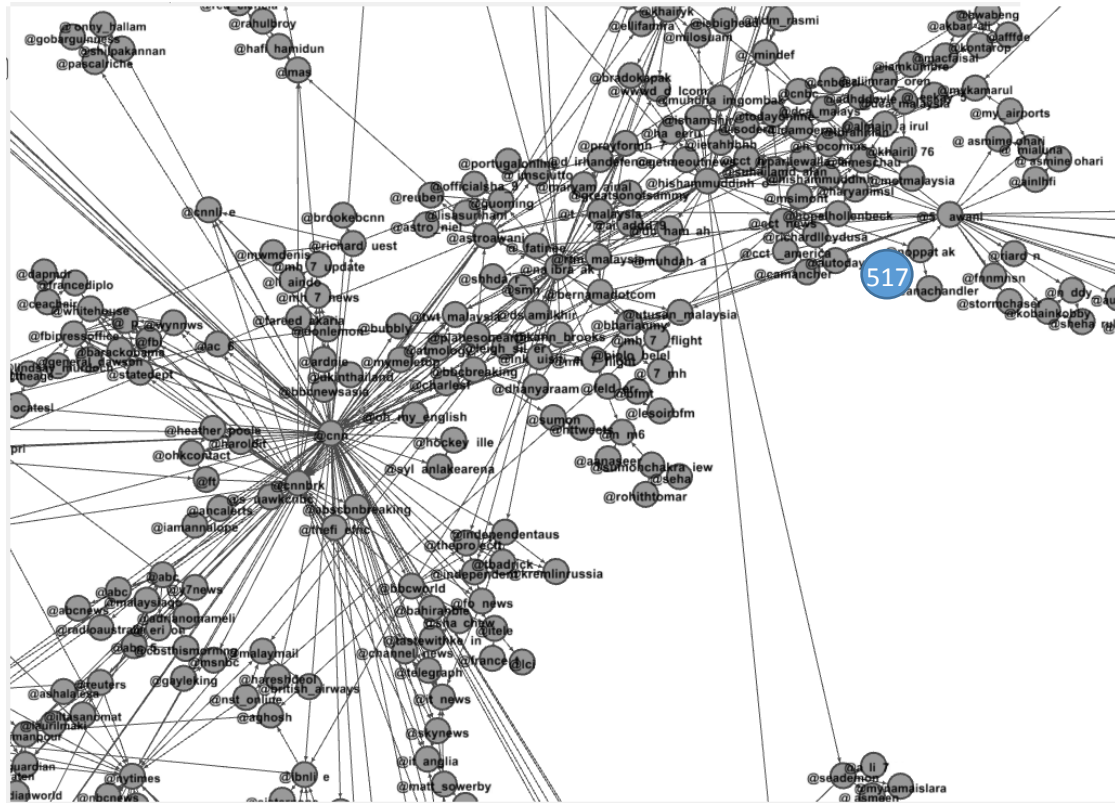
The User rankings



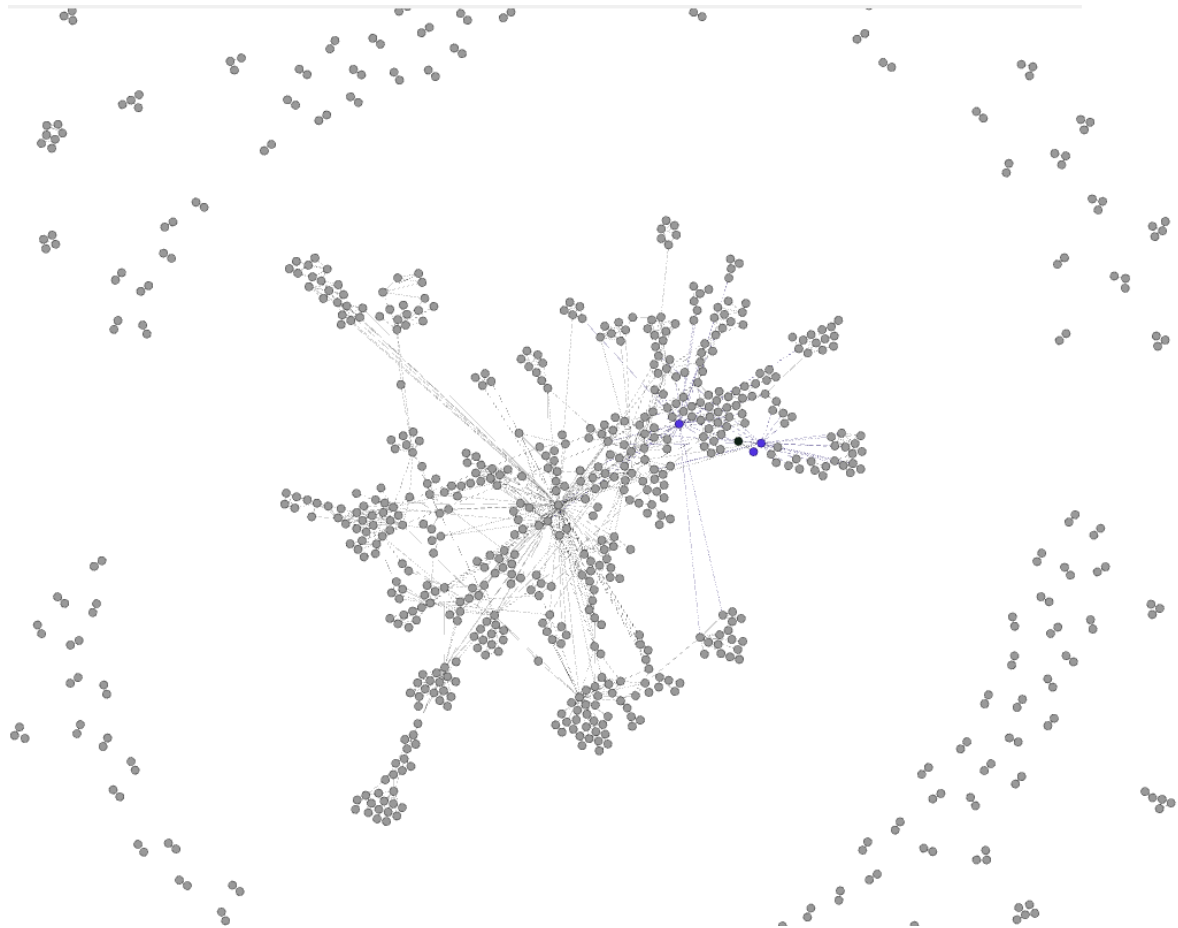
So where's @noppatjak?



where's @noppatjak?: Ranked 517 in the network. Mostly retweets as opposed to being commented on



where's @noppatjak?: Ranked 517 in the network. Mostly retweets as opposed to being commented on



เอม นกพัฒน์จักษ์ @noppatjak · Mar 13
16.54 CEO @mas ยืนยันโรลสอยส์และโบอิง ไม่พบสัญญาณ #MH370 หลังจากเวลา 01.07 น. คำคืนที่เครื่องบินหาย #nationtv
Collapse Reply Retweet Favorite More

RETWEETS 166 FAVORITES 10

8:56 PM - 13 Mar 2014 from Labu, Sepang · Details

Reply to @noppatjak @MAS

cha-aim @callmeaim · Mar 13
@noppatjak
globalnews.ca/news/1202910/o...
จริงรึเปล่าคะ
View summary Reply Retweet Favorite More

เอม นกพัฒน์จักษ์ @noppatjak · Mar 12
16.53 ได้ตรวจ #MH370 ลานี้ หลังมีคำเตือนมาหรือไม่?
CEO @MAS ตอบว่าตามนโยบายแล้วเครื่องบินทุกลำต้องผ่านการตรวจสอบก่อนบิน
#nationtv
Collapse Reply Retweet Favorite More

RETWEETS 121 FAVORITES 6

8:55 PM - 12 Mar 2014 · Details

Reply to @noppatjak @MAS

@noppatjak is a news journalist?



Conclusion



Key questions to answer

- What are people tweeting about the MAS brand?

MAS official tweets. No speculations or damaging tweets

- What are the popular trending topics?

#mh370 and #masalerts. A few news channels as well such as #9news. No damaging trending topics

- How much brand damage has this tragedy done to the @MAS brand?

Some. Currently lowest amongst Qantas, Singapore Airlines and Virgin.

- Who are the influential users in this network of tweets? Who is the Ellen DeGeneres of this network?

CNN and the other news channels. Need to look out for these users



The end! Questions?

- Data available on request!
 - @MAS
 - #Bomoh
 - #mh370
 - #MASalerts
 - #PrayforMH370
- R code will be available on GitHub



Limitations

- Issues with the Twitter API:
 - Maximum of 3200 tweets
 - Expiration dates: Unable to obtain tweets that are too old
 - Inconsistency in the number of tweets downloaded
 - Strange characters will cause an error in download
- Did not look at other trending topics such as [#mh370](#) due to time constraints
- Other languages cannot be analysed (eg- Thai) whilst some other languages such as Malay could be causing a lot of white noise to the analysis
- 'Followers' and 'Following' not covered
- Topic modelling not covered



Thank you!



@ejlok



<http://au.linkedin.com/pub/eu-jin-lok/4/b8b/896>

kaggle

www.kaggle.com/users/3346/eu-jin-lok



<http://www.slideshare.net/EuJinLok>

GitHub



<https://github.com/ejlok1?tab=repositories>



References and other materials

- <http://jeffreybreen.wordpress.com/2011/07/04/twitter-text-mining-r-slides/>
- <http://www.r-bloggers.com/r-text-mining-on-twitter-prayformh370-malaysia-airlines/>
- <ftp://cran.r-project.org/pub/R/web/packages/twitteR/twitteR.pdf>
- <http://www.hashtags.org/business/management/what-does-it-take-for-a-hashtag-to-trend/>
- <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>

