



Mapping the UK Twittersphere

Steven McDermott – s.e.mcdermott@leeds.ac.uk – @soci
<http://snacda.wordpress.com>
University of Leeds

Overcoming the Lack of Demographic Detail by Collecting
Active Twitter Use by targeting Postcodes

Top UK Twitter Users

- Twitter now has 10m user accounts in UK - 15 May 2012.
- The UK is the fourth-largest country for Twitter users in the world, with 80% accessing it with mobile phones.
- The Top 10 UK tweeters and their followers (drawn from worldwide audience of 140 million) are...
 1. Stephen Fry (actor, writer) – 4.3 million
 2. Jessie J (singer, judge on The Voice) – 4.2m
 3. Lily Rose Cooper [formerly Allen] (singer) – 3.5m
 4. Niall Horan (One Direction singer) – 3.3m
 5. Liam Payne (One Direction singer) – 3.1m
 6. Emma Watson (actor) – 2.8m
 7. Rio Ferdinand (footballer) – 2.7m
 8. Eddie Izzard (comedian, actor) – 2.4m
 9. Ricky Gervais (comedian, actor) – 2.3m
 10. Holly Willoughby (TV host) – 2.3m

Is Analysis Beyond the Celebrity Sphere Possible?

- Getting beyond the celebrity obsession that has transferred across from mainstream media?
- 10 million user accounts are not necessarily active.
- Active users rather than user accounts.
- Some form of demographic data.
- Geographically located in the UK.
- And focused on Gender, Age and Health questions.

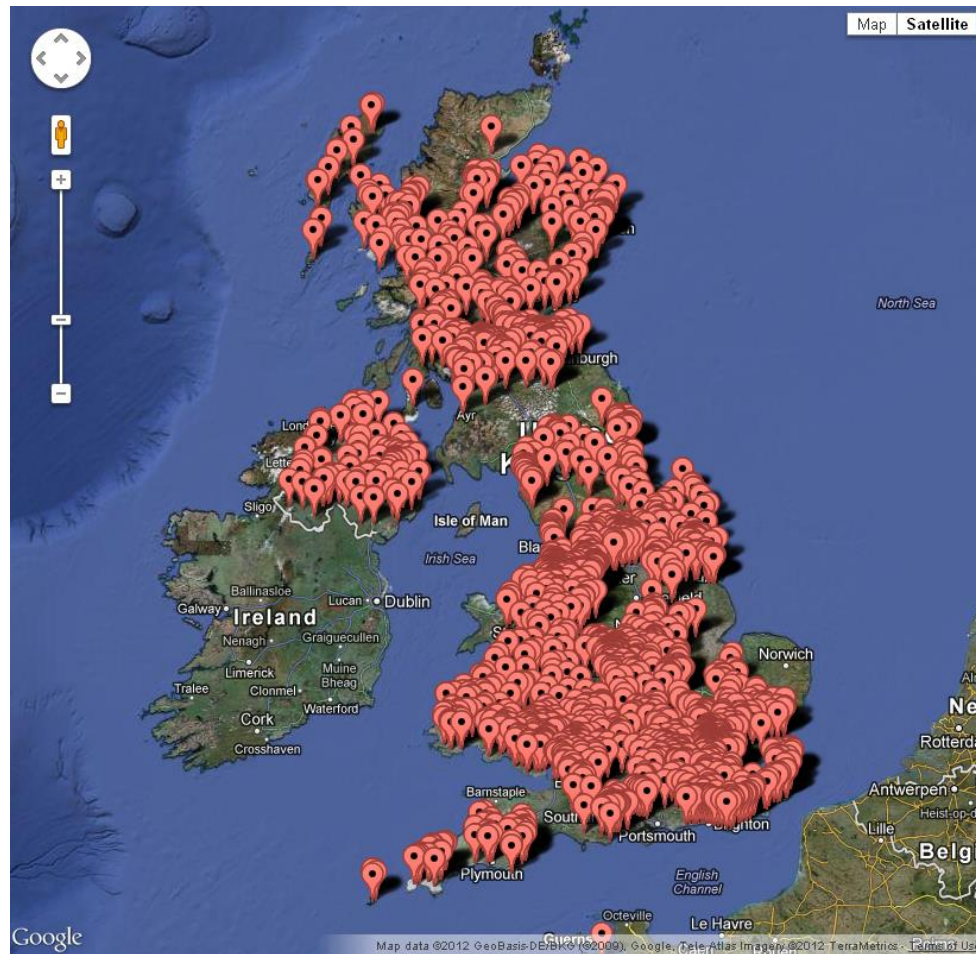
Gender and Age in Health Care Provision

- How do people talk about older people in texts distributed by the various types of media in the United Kingdom?
 - Such talk, if it exists, may be a contributing factor to discrimination in practice.
 - The archived debates and discussion on age and gender discrimination will be analysed to see...
1. How are debates on age and gender discrimination represented in the various types of media?
 2. How do organisations that claim to represent the elderly frame their engagement with them online?
 3. Do health care professionals talk differently than **the public** about age and gender discrimination?

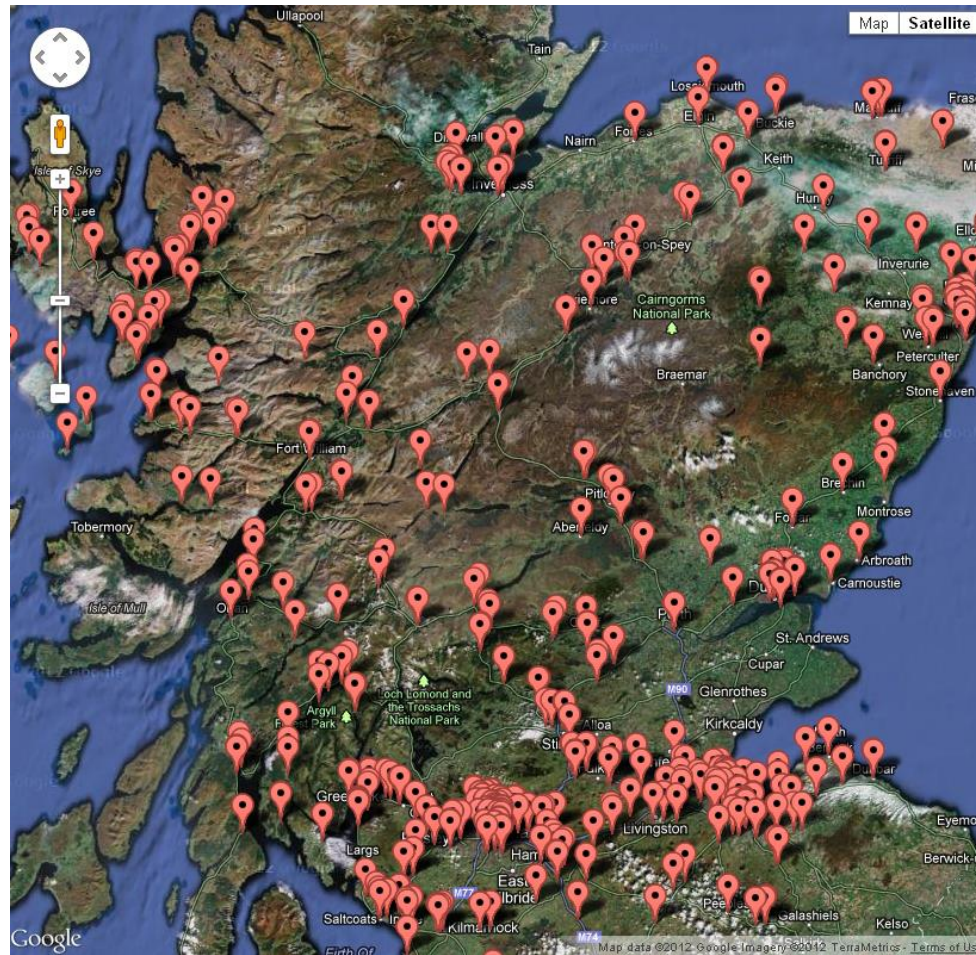
Public Talk on SNS

- How to collect **public talk** about elderly women and health care?
- Only focusing on the UK
- Not only celebrity chit chat or noise but targeting specific key words and phrases.

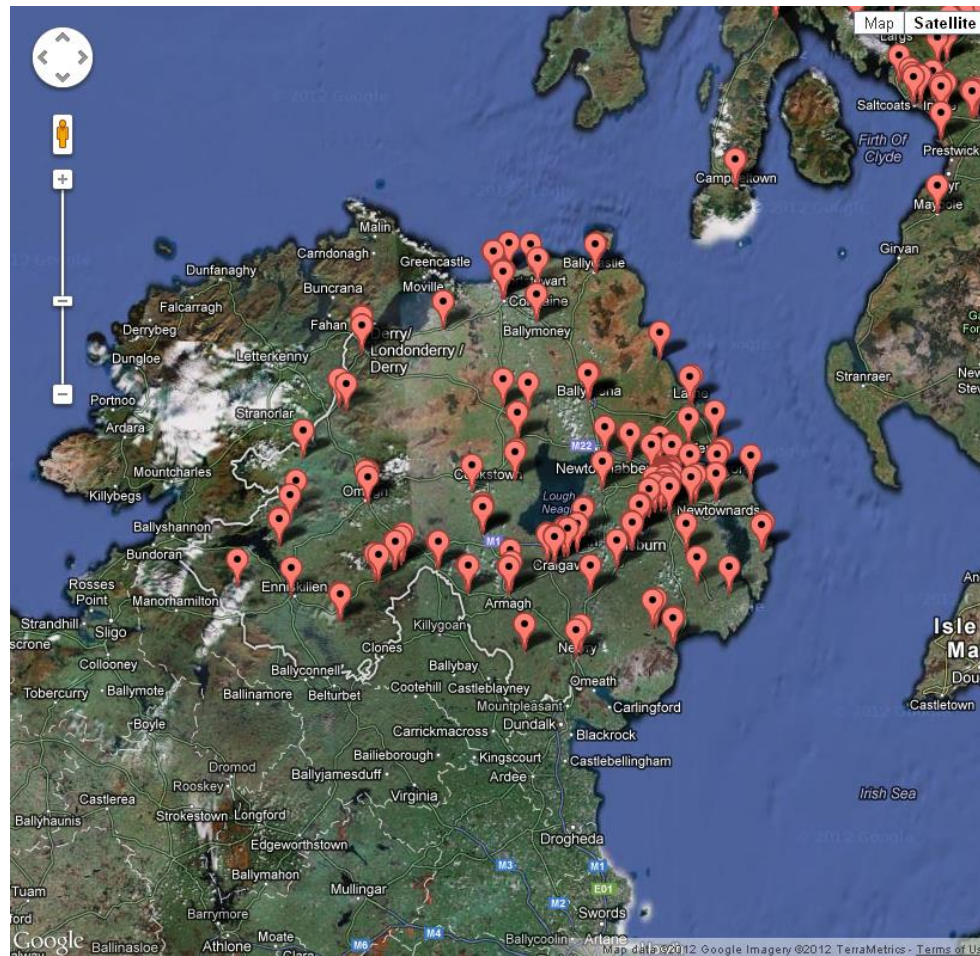
5,900 Random Postcodes



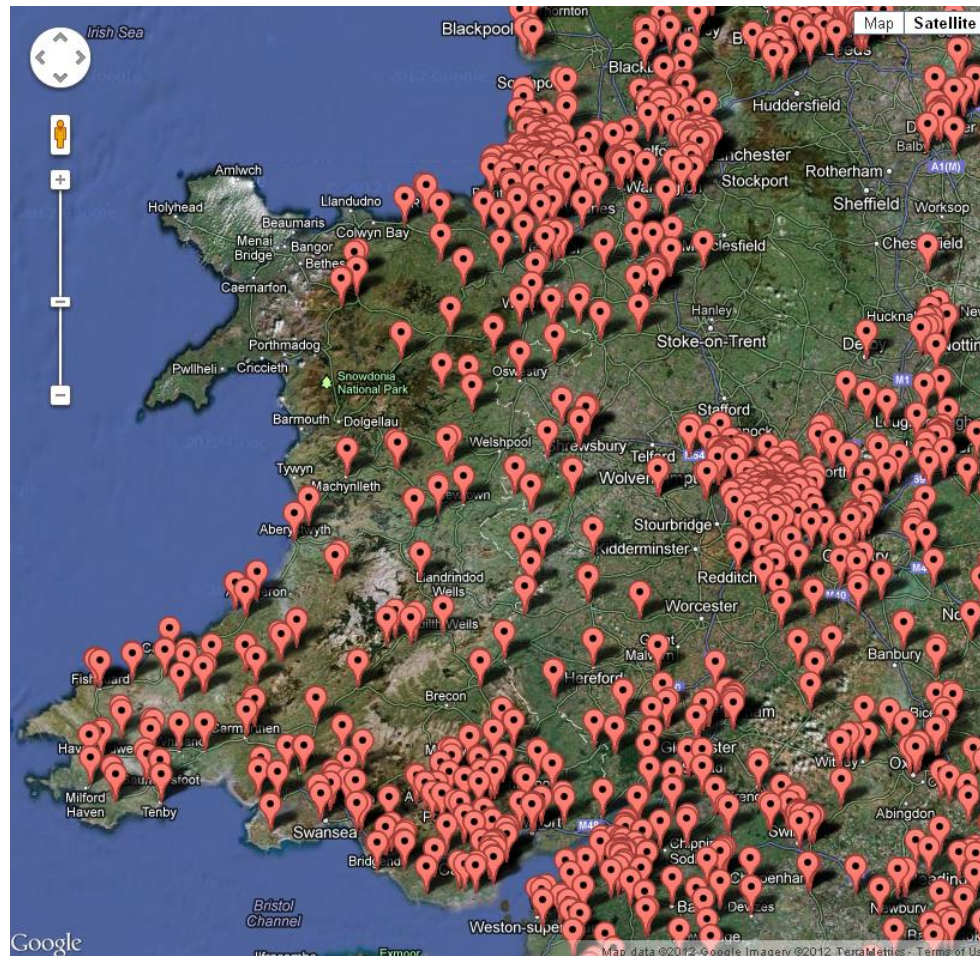
Scotland



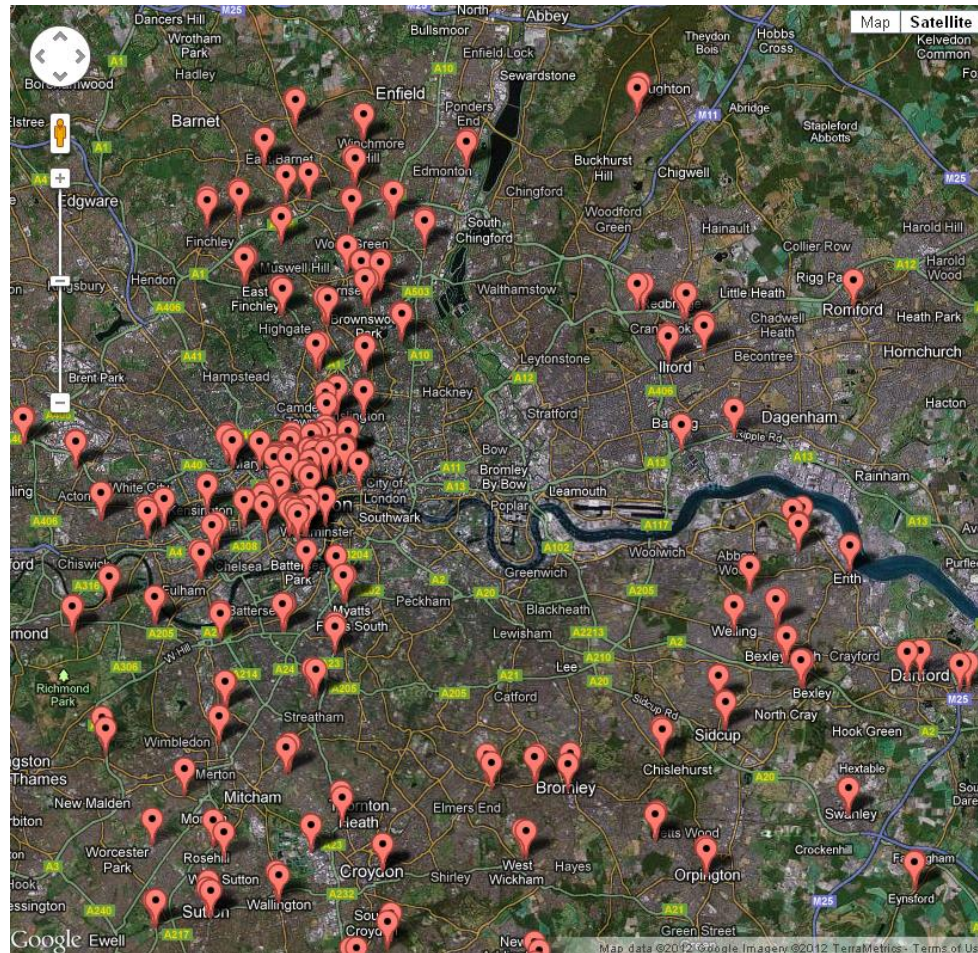
North of Ireland



Wales



Greater London



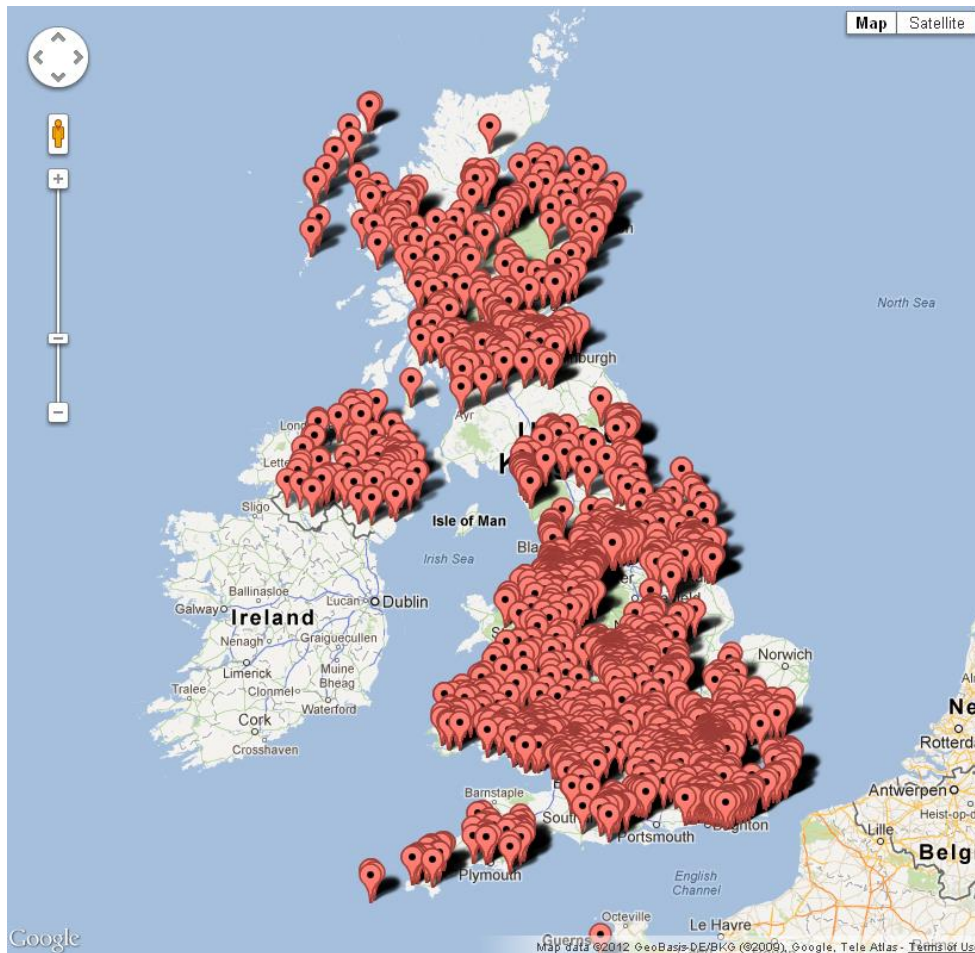
Postcodes converted to Geocodes

- AB10 1AA = 57.1482317,-2.0966478
- Radius = 20 miles – this can be increased or decreased
- Returning 300 per search = 300
- <http://search.twitter.com/search.atom?q=&lang=en&geocode=57.1482317,-2.0966478,20mi&rpp=300>
- If postcodes have demographic information associated with them then can't that help geographically locating the origin of the tweets and have the postcode data associated with them.
- Postcodes have socioeconomic data – associated with the population residing there.
- As 80% use their mobiles to tweet – can't assume they actually reside there.

Key Words Searched For

- [http://search.twitter.com/search.atom?q=elder* OR aged OR older OR oldest OR "nursing home*" OR women OR woman OR female* OR man OR men OR male*&geocode=51.5761974,0.0576684,20mi&rpp=300](http://search.twitter.com/search.atom?q=elder* OR aged OR older OR oldest OR \)
- Similar strings were created for all 5,900 postcodes.

Resulting in...



23/01/2013 – 30/01/2013

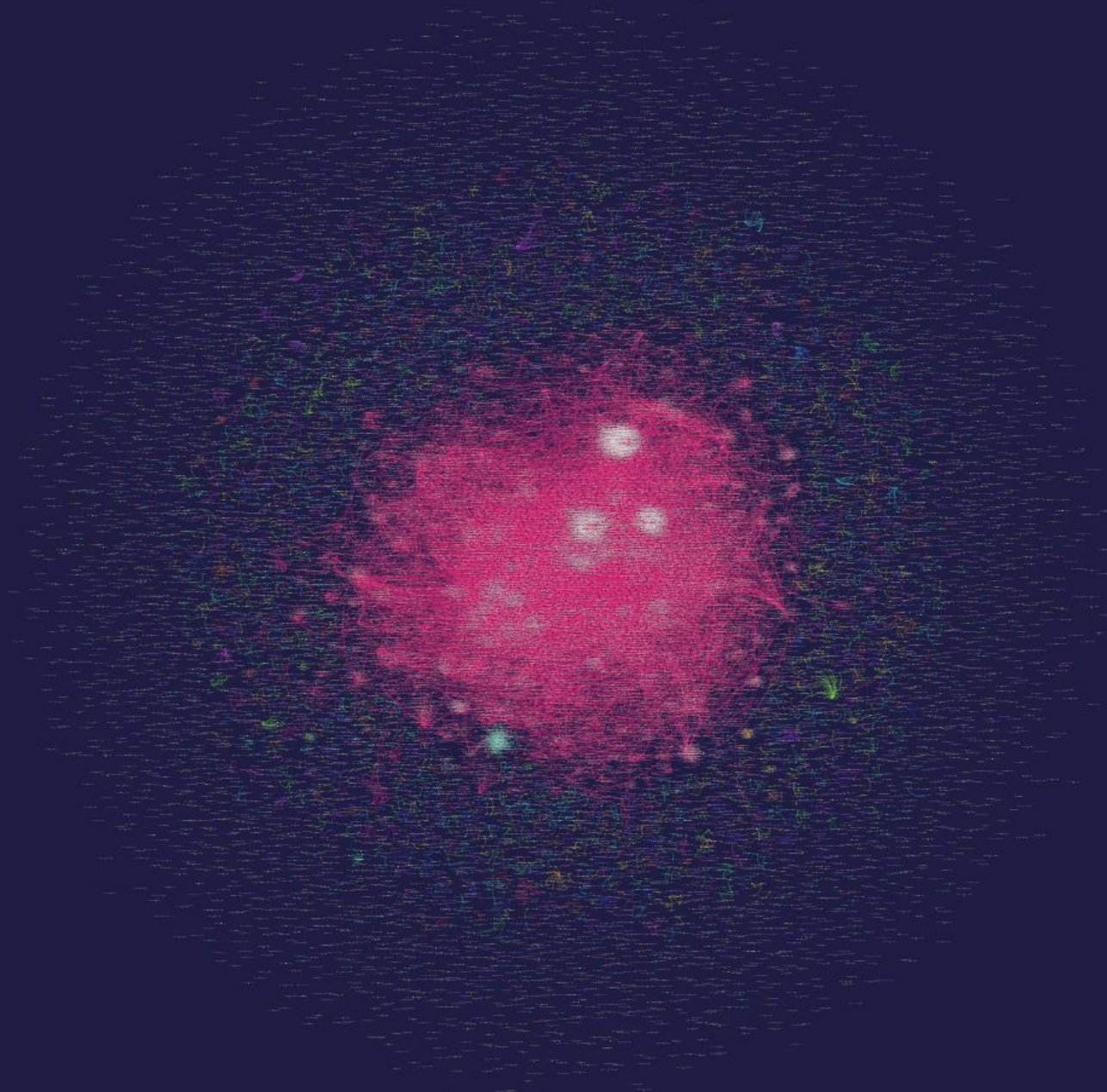
22,833,946 tweets
containing at least
one search term -
that can be traced
back to a
postcode.

726,315 active
user accounts.

All **5900** search
strings were called
3.4 times over 7
days in Jan 2013.

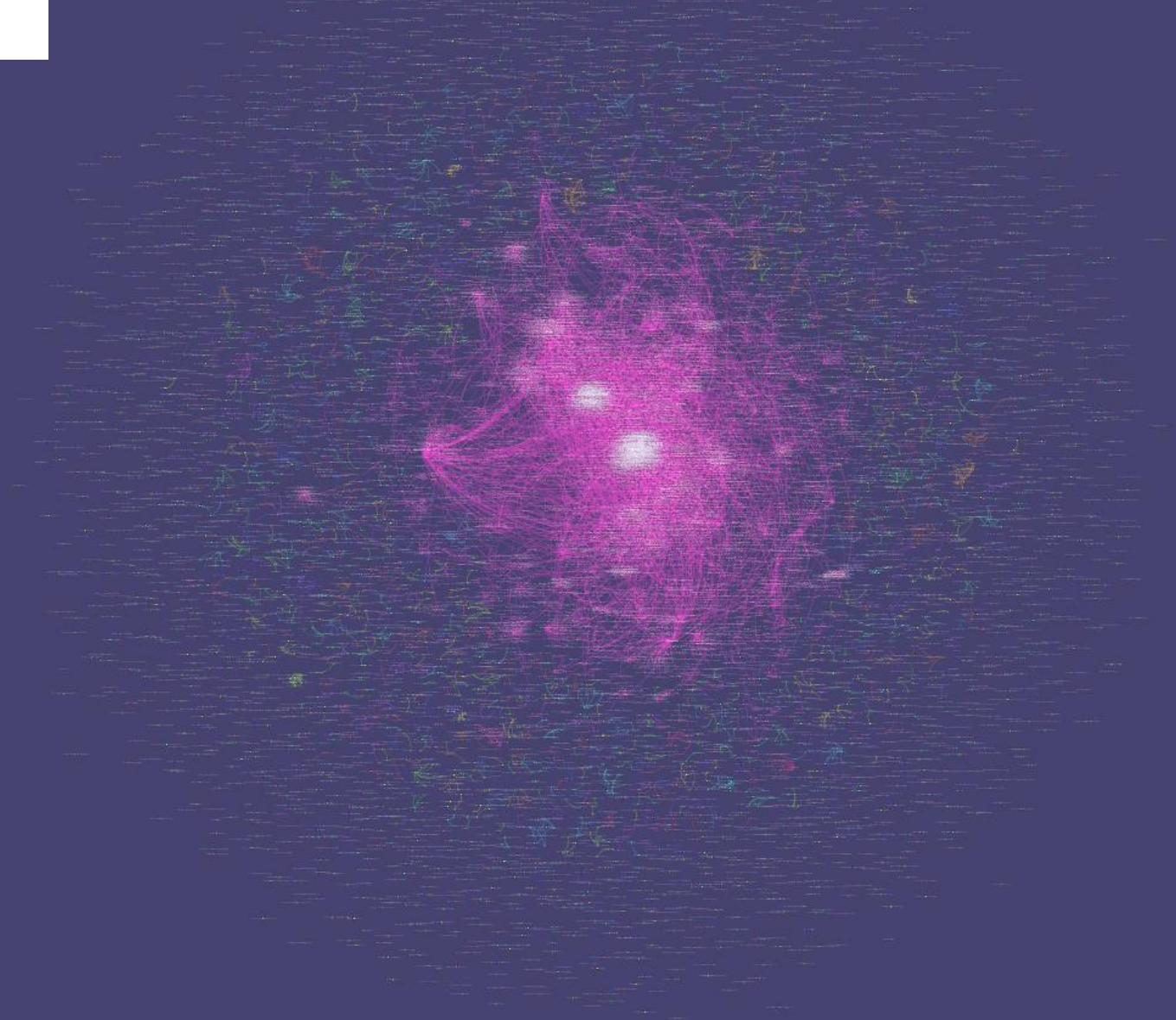
Quantitative Analysis

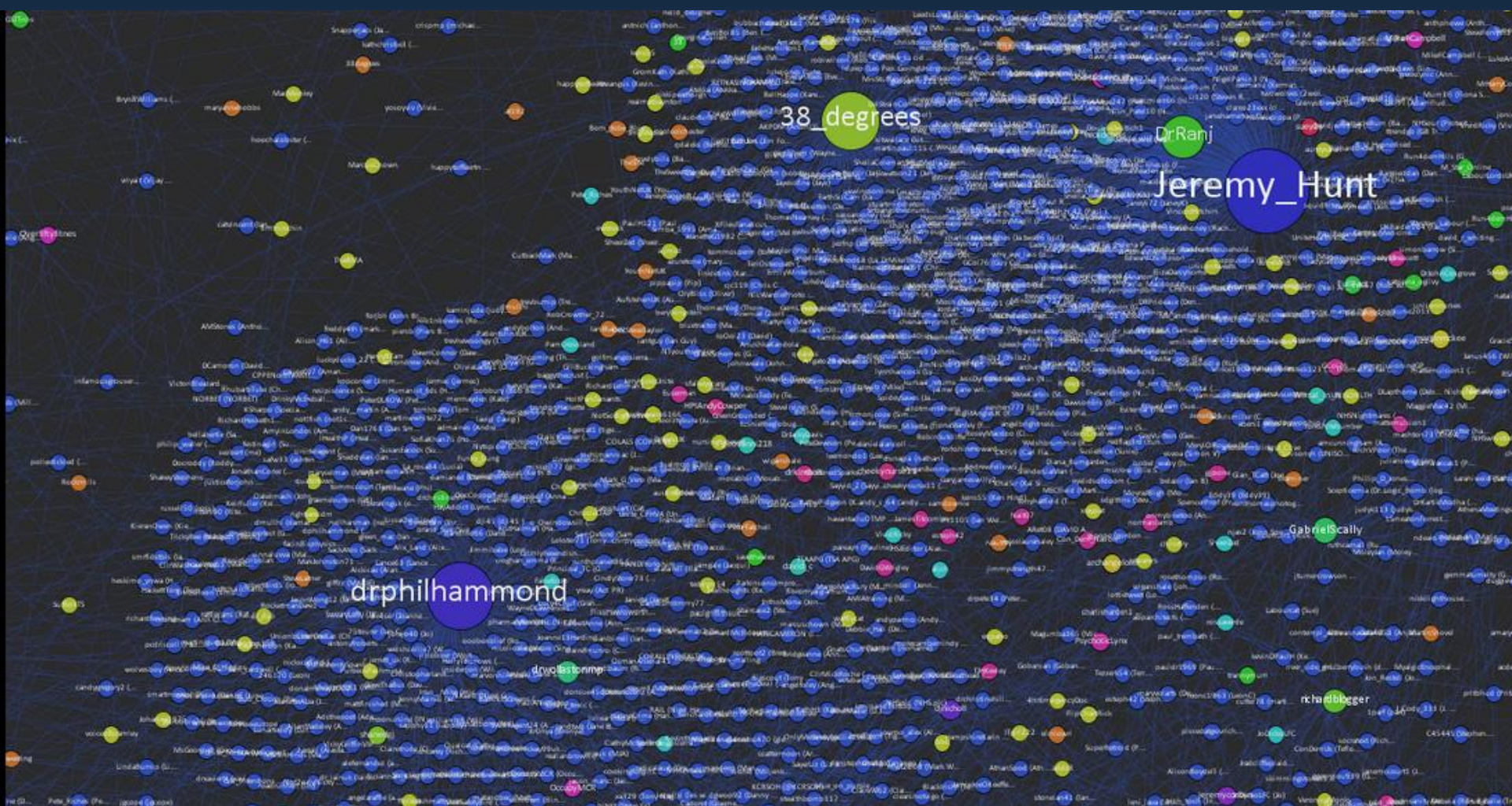
- Webometrics – network @user connections – used to create the following maps...
- Quantitative – Mozdeh can perform Time Series and Sentiment Analysis
- Time series graph of the tweets
- Identifying keyword spikes
- Spam filtering and removal
- SentiStrength Analyze ???



Filtered to show only
top 10000 connected
component. March
2013

Gephi

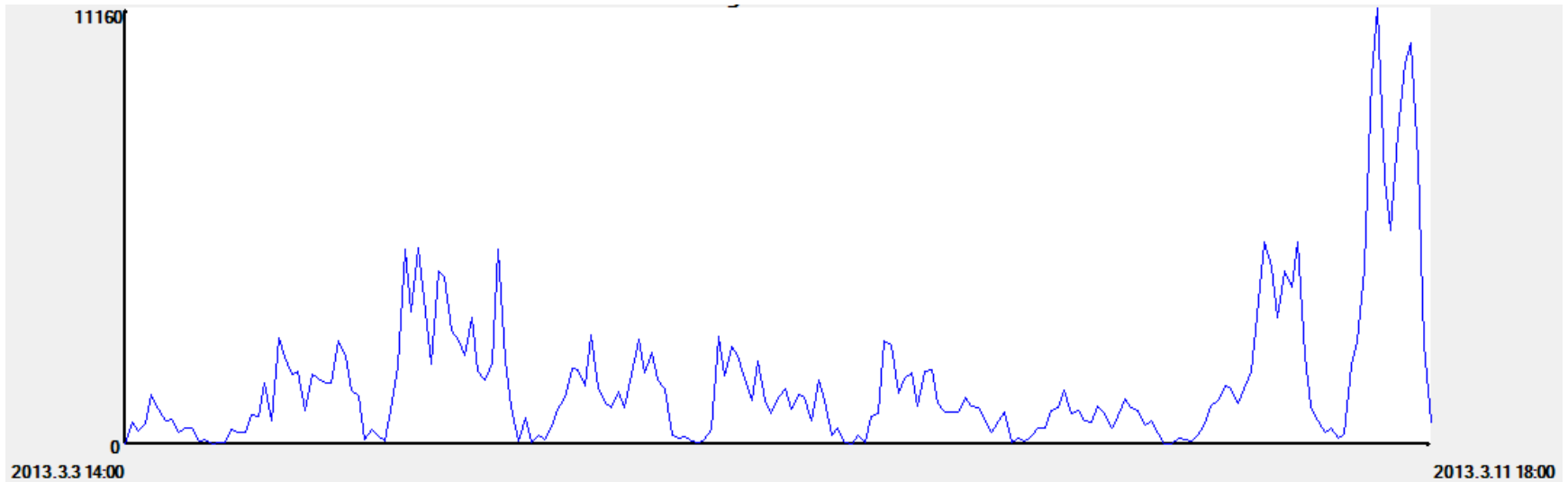




Sentiment Analysis

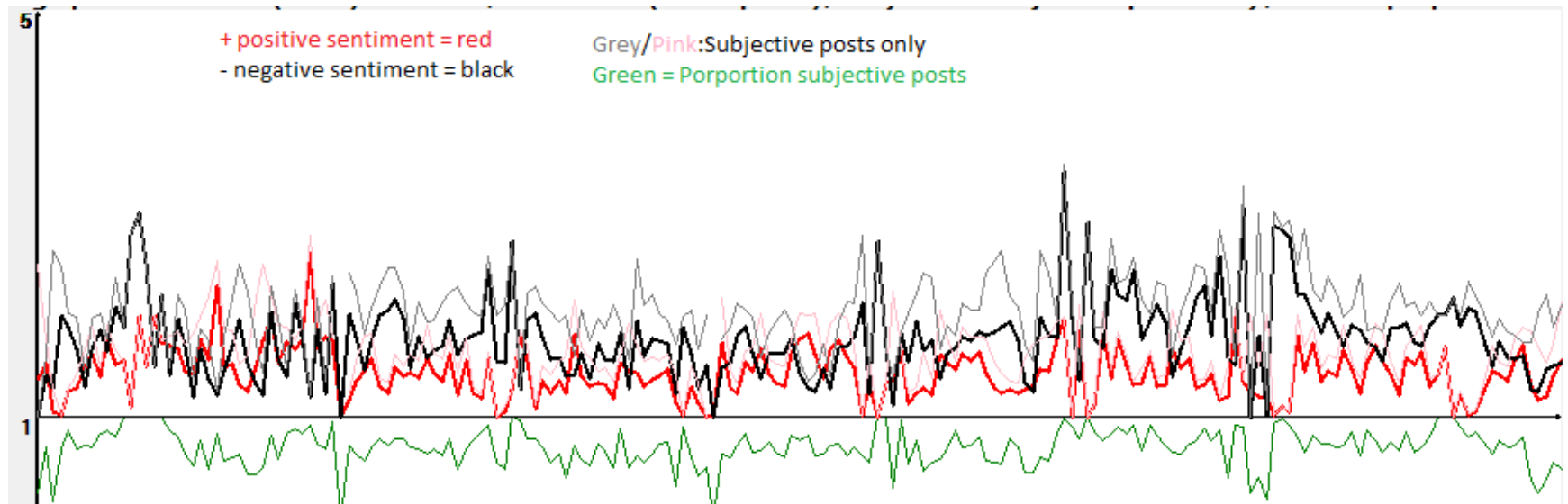
- Leximancer - £500 to purchase and txt to be converted to .doc
- QDA miner – Admin access required
- Automap – installed – However 12gbs is too large a file
- Nvivo is also unable to handle 12gbs.
- SentiStrength

SentiStrength



Time series graph of the tweets
Identifying keyword spikes

Sentiment Analysis of 410,000 Tweets containing the term NHS in March 2013



SentiStrength

Geocoded Data and Postcode Data

- If postcodes have demographic information associated with them then can't that help geographically locating the origin of the tweets and have the postcode data associated with them.
- Postcodes have socioeconomic data – associated with the population residing there.
- Can I say that the Tweets are **REPRESENTATIVE?**