Pattern Mining for Social Networking Data

Mahesh Hingane
April 22nd, 2013

Introduction
… of the project

- Social impact of social networking
  - Facebook, Twitter, LinkedIn, FourSquare

- User data availability
  - Activities, Preferences, Geographical information

- Opportunities for data mining
  - Advertising, Social studies

- Project – **Pattern Mining for Social Networking Data**

Project Work

- Twitter
  - Large user base

- Hashtags
  - 'Key' to track the tweets
  - Messages related to particular topic

- Geographical information tracking

Project Work *[Contd]*

- Twitter API

  – Tool to collect information

- Repeated search queries to collect data

- Look for *#bigdata* in the dataset

- Geographical information tracking

- Technologies – Python, MySQL

Current Results

- Date range: March 25th, 2013 – April 19th, 2013

- 94365 unique tweets

- 29354 unique users

- On an average ~3629 tweets per day

Current Results *[Contd]*

- Frequency analysis (most frequent words):

#EMCfaces	#cloud	#analytics	#sxsw	#gitpro

- Users retweeted the most number of times:

| @IBMbigdata	300
@WSJ (Wall Street Journal)	241
@BigDataBorat	220
@kdnuggets (Gregory Piatetsky)	190
@gigaom	176

Current Results *[Contd]*

- Date Trends:

| April 16th	#cloud, #cloudexpo, #analytics, #virtualization, …
April 15th	#cloud, #analytics, #virtualization, #nube, …
April 14th	#cloudcomputing, #openstack, #sdn, #ecommerce, …
April 13th	#cloudcomputing, #analytics, #cloudexpo, #hadoop, …
April 12th	#cloudcomputing, #analytics, #cloudexpo, #businessobjects, …

… and so on.

Current Results *[Contd]*

- Retweet map

  - NetworkX DiGraph, Graphviz
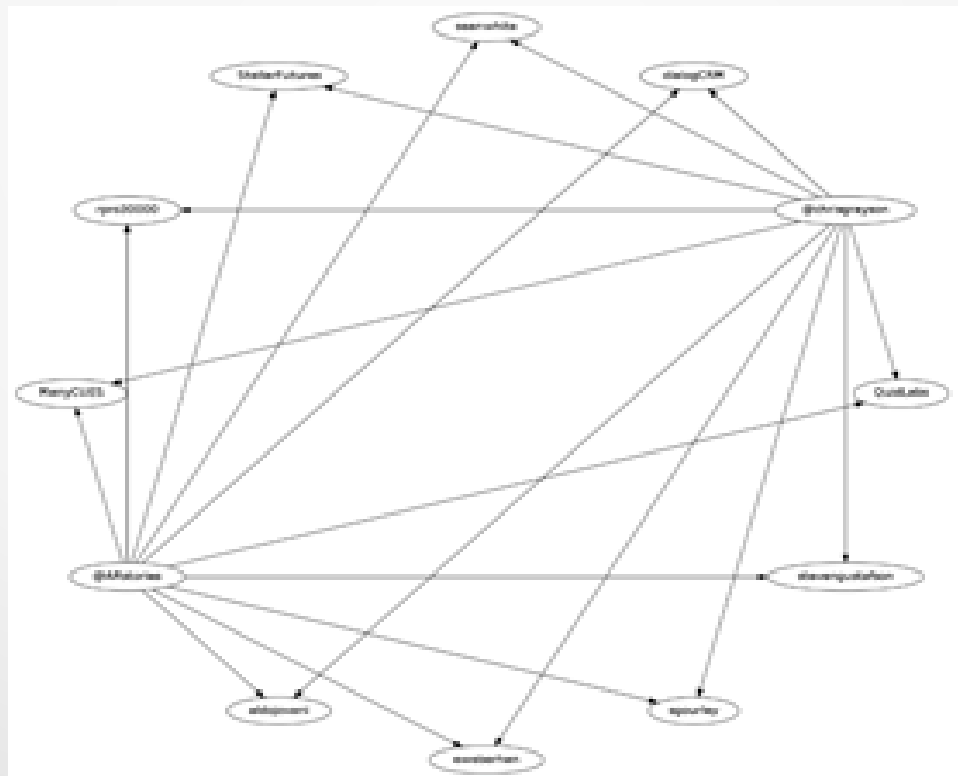
Current Results *[Contd]*

- User locations

  - Matplotlib Basemap
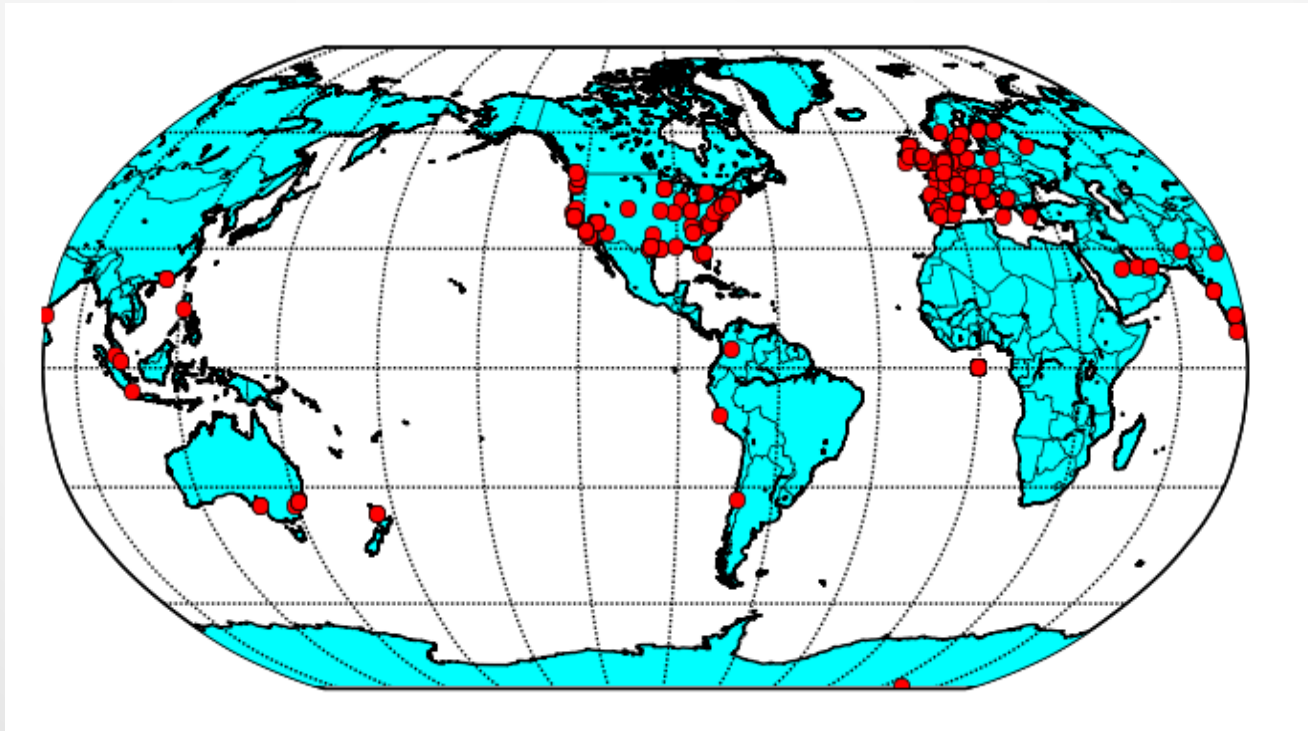
Current work

- Plotting users by their profile locations – JSON queries

- Tracking other tweets by the frequently retweeted users

- Continuing data collection and findings trends in larger data sets

- Identifying date trends in the tweets

Future work

- Data collection for other hashtags – looking for data rich in location-information

- FourSquare API

- Establishing trends in the observed data

Thank you!