

## SOCIAL MEDIA ANALYSIS

Social media data scraping refers to the process of extracting the content and metadata of user submitted activities. This includes information about the author, time, date, location, and content of tweets, Facebook posts, and other social media. When these data are geographically referenced (or geotagged), analyses can then have a geographic or spatial dimension to detect patterns by generating maps. This assignment is worth up to 10 points.

To complete this assignment, follow the steps below:

1. Manual content analysis (e.g. Sentiment analysis of social media data)

Who are the people giving CFA exam and are they discussing course material in the group?

There is a group of student appearing for CFA exam 2013 which well represented and diverse across the country. Now to analyze the group the data was taken from facebook using application NetVizz.

Analysis of Facebook data using Gephi has following report.

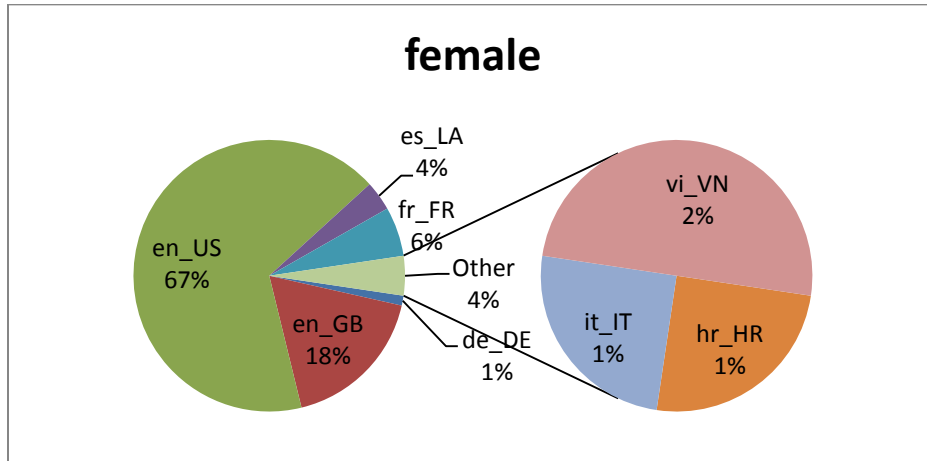
Facebook group used for analysis is

<https://www.facebook.com/groups/CFA.Level1.June2013/>

Analysis of Group:

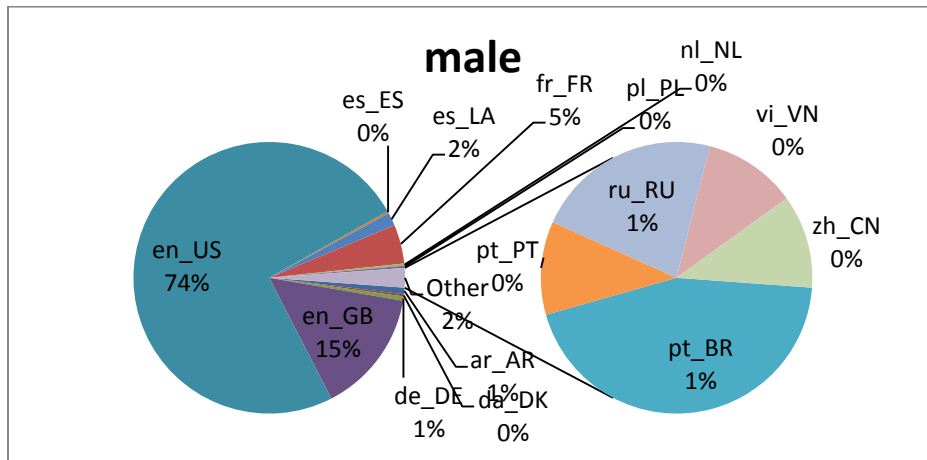
Female representation in group stacked by country in data and pie chart shown below:

Count of nodedef>name VARCHAR	Column Labels	
Row Labels	female	Grand Total
de_DE	1	1
en_GB	15	15
en_US	57	57
es_LA	3	3
fr_FR	5	5
hr_HR	1	1
it_IT	1	1
vi_VN	2	2
<b>Grand Total</b>	<b>85</b>	<b>85</b>



Which show Charted Financial Analysis level 1 Exam group 67% female are from US and 18% from great Britain, France 6%, Latin America 4%, other 4%.

While if we segregate data by male then 74% are from US and 15% from great Britain and France 5%, latin America 2% and others 2%.

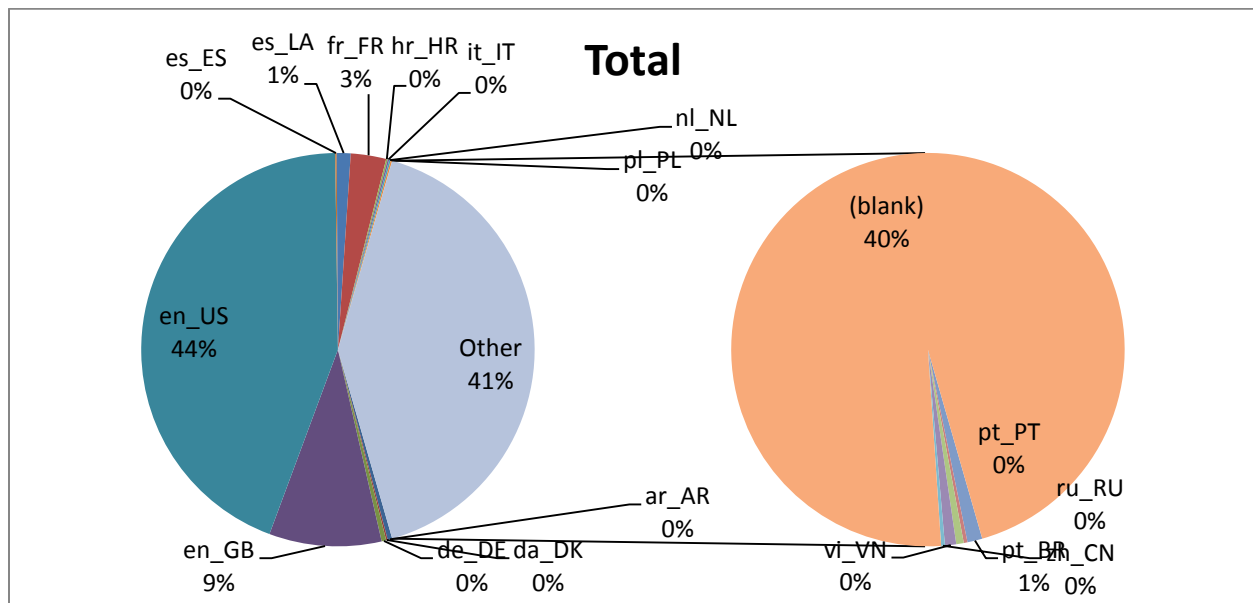


While there is almost 50% data where there is no information about user locale

Row Labels	Count of nodedef>name VARCHAR
ar_AR	3
da_DK	1
de_DE	3
en_GB	73
en_US	349
es_ES	1
es_LA	9
fr_FR	23
hr_HR	1
it_IT	1
nl_NL	1

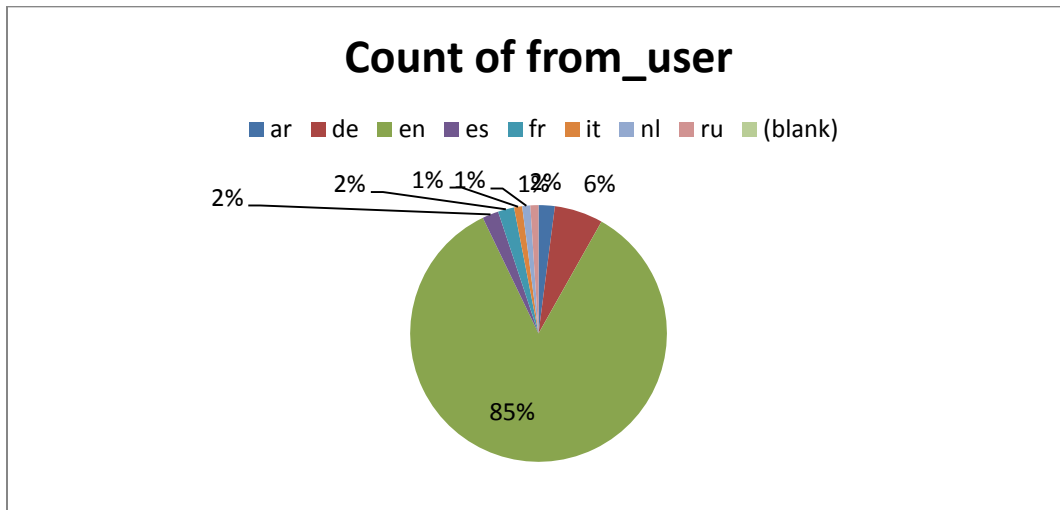
pl_PL	1
pt_BR	4
pt_PT	1
ru_RU	2
vi_VN	3
zh_CN	1
(blank)	314
Grand Total	791

Because of entry of these blank fields other the which takes up almost 50% data point the average whole data changes like avg US representation in groups drop from 65-70% range to 44% as 40% data is blank. As shown in graph.



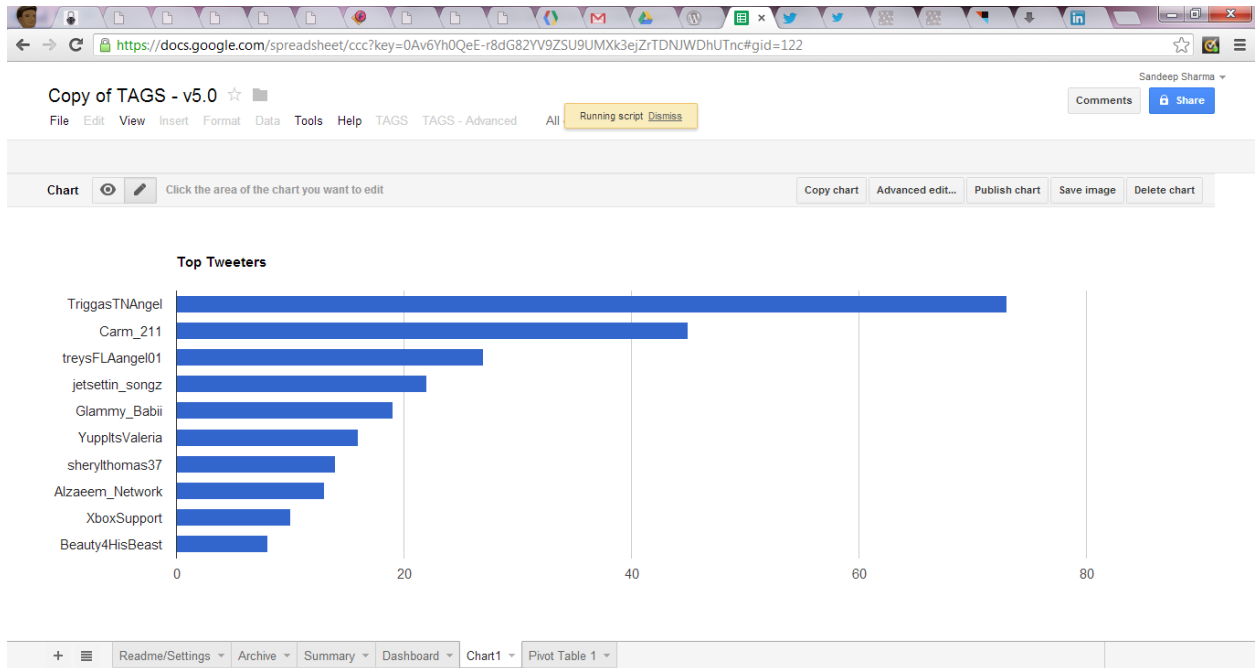
- Sentiment analysis using on-line sentiment analysis tools (e.g. Sentiment analysis of Twitter data).

Analysis of twitter data for search term network reveal 83 % are English speaking who are tweeting about network. Data exported to Excel and then report created out of it. Data extracted using TAGS connecting to twitter to fetch data.



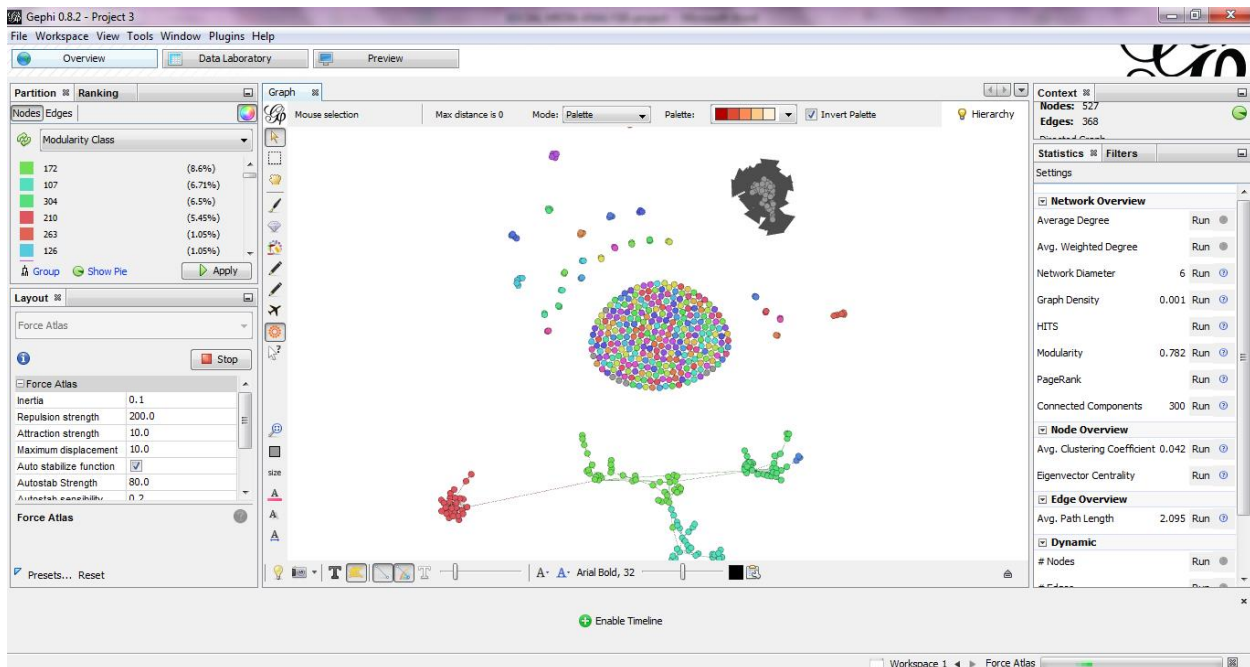
Here is pivot of data grouped by language/locale for search term Network.

Row Labels	Count of from_user	Count of source	Count of user_followers_count
ar	2	2	2
de	6	6	6
en	83	83	83
es	2	2	2
fr	2	2	2
it	1	1	1
nl	1	1	1
ru	1	1	1
(blank)			
<b>Grand Total</b>	<b>98</b>	<b>98</b>	<b>98</b>

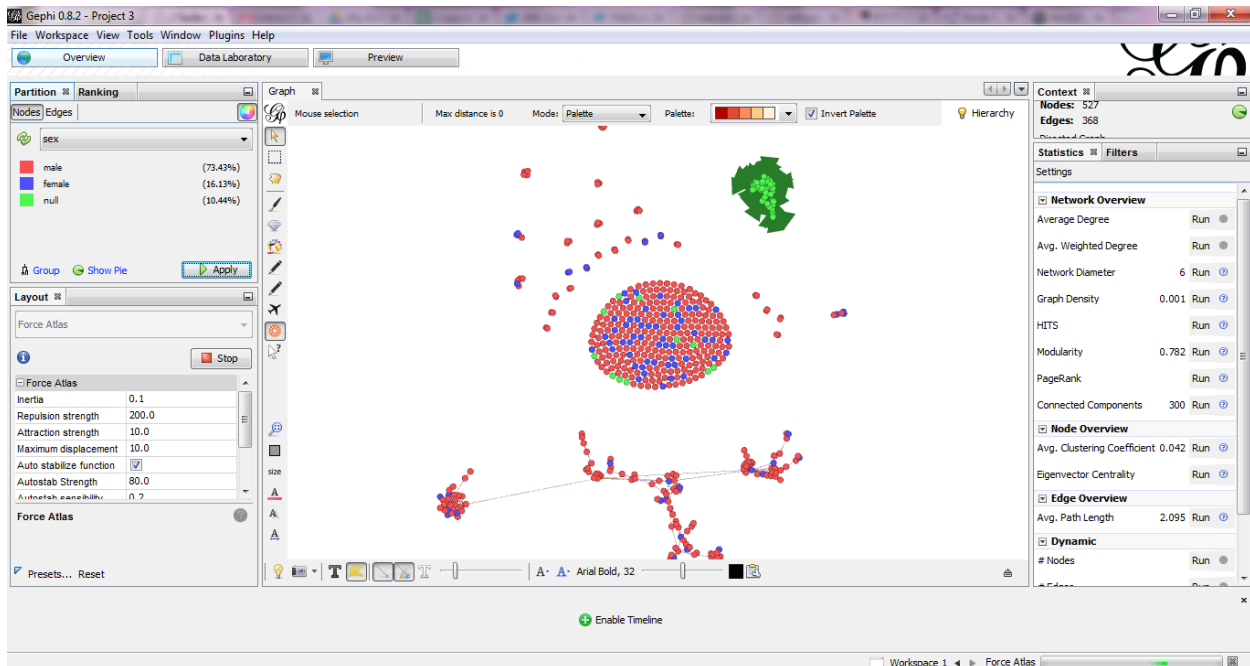


Network analysis of social network data (e.g. using Gephi for analyzing Facebook data) user data was taken from TAGS into .gdf format which was further taken as source in Gephi for analysis. And graph was drawn showing network diagram where linkage of various user to each other is clearly shown in the map.

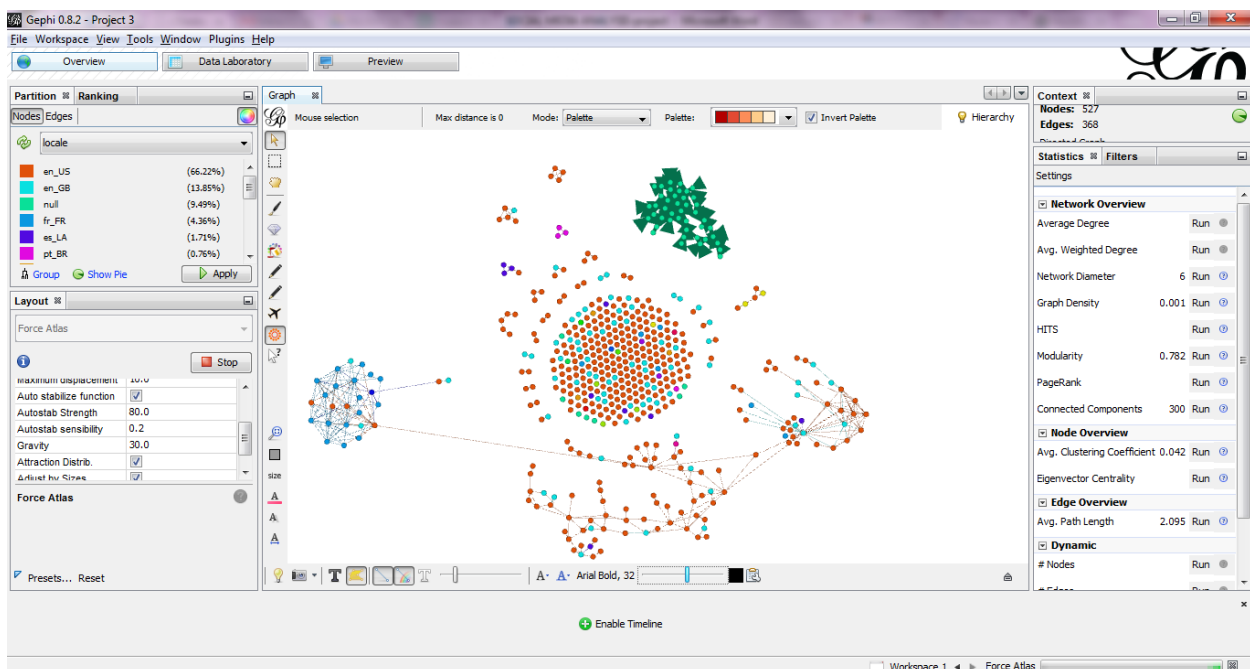
Data segregated by Modularity class.



Segregation by Male/female:



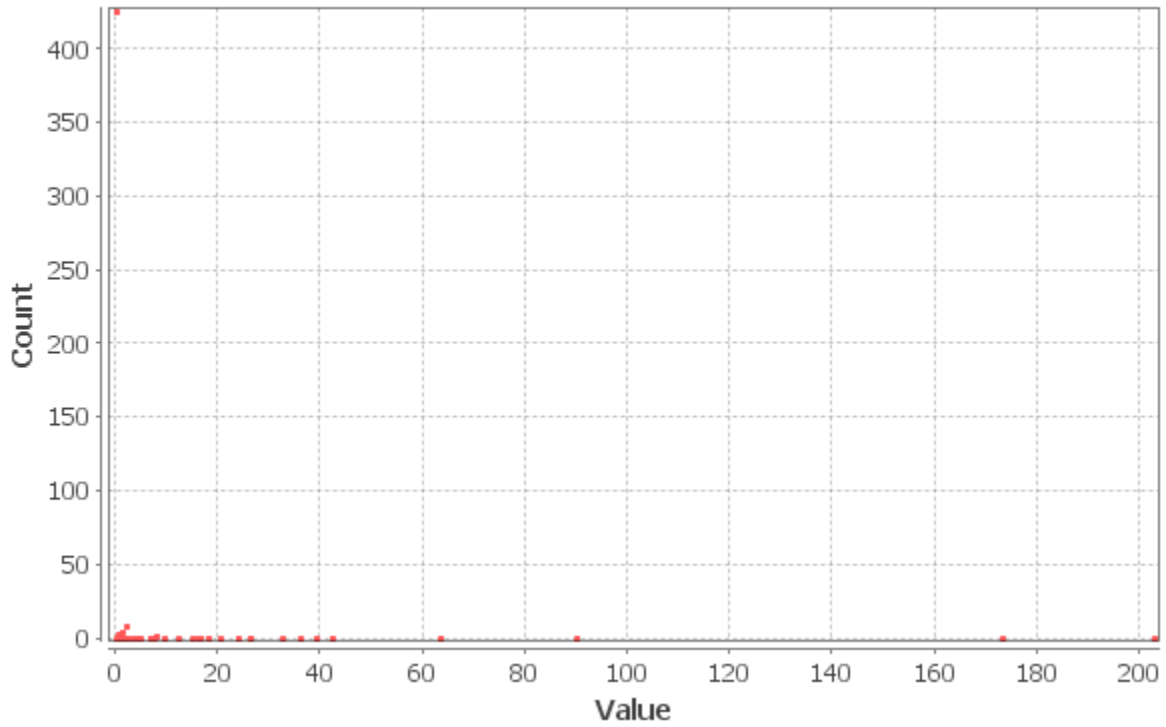
The network search data segmented by geographic region: color representing region like orange for US. Here attractive forces are distributed along outbound links to push periphery and represent authorities at centre.



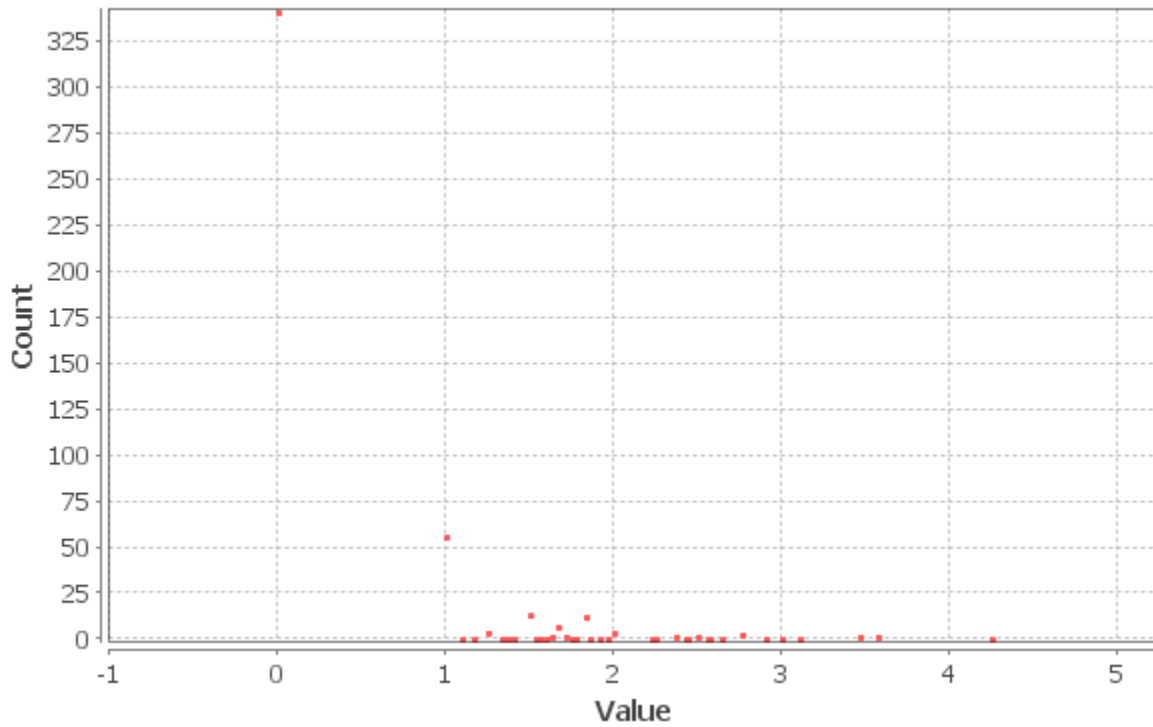
Analysis shows amount of users the analysis shows the tweet how closely they are associated with each other like where u can see 1 user\_476 tweet dominating with more than between centrality count of more than 202 counts.



### Betweenness Centrality Distribution

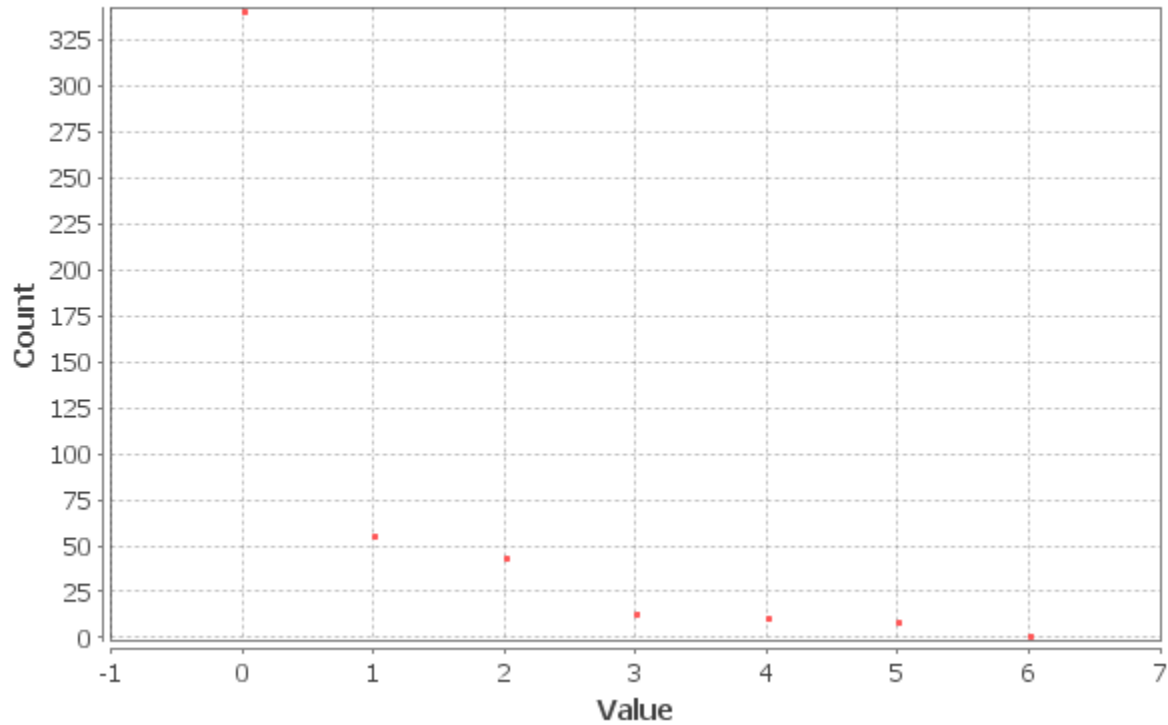


### Closeness Centrality Distribution





### Eccentricity Distribution

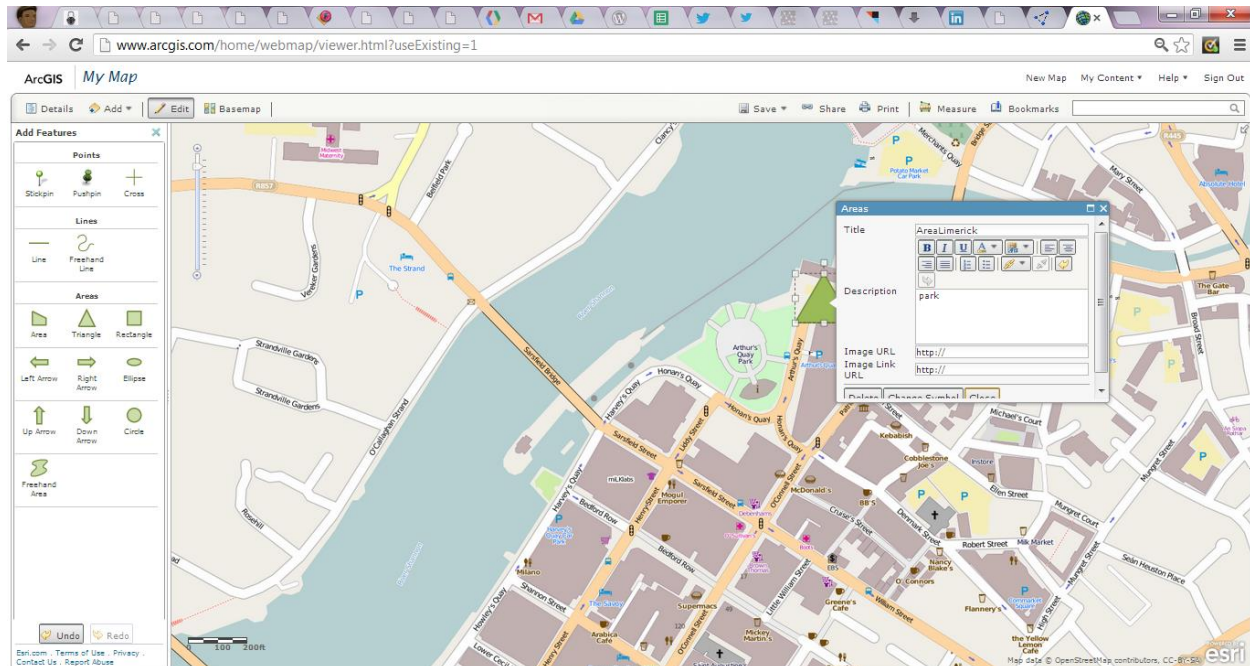


Spatial analysis of social media data (e.g. using ArcGIS or Quantum GIS for spatial analysis of location-based data)

GIS ArcGIS data Dublin tourist map

The screenshot shows the ArcGIS web interface. The search bar contains 'dublin'. A popup window displays the following information for the Guinness Storehouse:

Top 10 Tourist Locations in Ireland: Guinness Storehouse	
Rank	3.00
Tourist Destination	Guinness Storehouse
Townland	St James Gate
City	Dublin 8
Country	Dublin
Country	Ireland
Phone Number	(01)4538364
Opening Hours	<a href="#">More info</a>
Website	<a href="#">More info</a>
Picture	<a href="#">More info</a>
Latitude	53.34
Longitude	-6.29
Zoom to	



1. Based on the data you collected in the social media scraping exercise, decide what method of analysis you wish to undertake. To learn about the techniques that you can use for analyzing your scraped data, please see the case study videos from week 3. Students are encouraged to research other internet sources for additional background on analysis methods. Options for data analysis include

### What you need to turn in:

1. Write a description of your social media analysis answering the following questions:

- The name of the data source (Facebook)
- The name of the source (Facebook – CFA level 1 2013 group).
- The date of the data scraping (May 23, 2013 data was taken)

Summary next page:

Summary:

- In not more than one page, explain the key points or summary of your analysis results. Summary is presented below each analysis of facebook data captured using NetVizz and presented using pivot and using pie chart. Data was of group CFA level 1 2013 where we wanted to analyze the demographic segmentation of user and also segmentation by gender across demographics which was also presented by graph which reveals the highest percentage of users come from US which is around 74% if

we do not take into account blank. There are almost 40% values which does not have locale setting set hence if we take blank values in account the representation of other segment in percentage comes down fro example it was shown in graph that representation of US user come down from 70% to 44%.

- then twitter data captured using TAGS and presented using Gephi to show the closeness of user tweets on subject of networking.

Here tweet data was taken for a particular date for search term Network and data was analyzed for its statistical parameters like closeness, between centrality distribution of tweets these data taken from twitter is fed in Gephi through which graphs were shown to represents the figures.

Pivot of users searching network term show amazing skew of 85% user searching this term in English which was represented in graph. Network is broad term which might mean even network marketing . to computer network so this 85% people searching only in English was amazing fact.

Tourist Map ArcGIS for Dublin city was used. And created for Limerick city park near Shannon River.