SOCIAL MEDIA ANALYTICS

A PROJECT REPORT

For the course

CSP 301: DESIGN PRACTICES IN COMPUTER SCIENCE

INDIAN INSTITUTE OF TECHNOLOGY

SUBMITTED BY:

PALAS (2011CS10239)

ABHISHEK NIMESH (2011CS10203)

VISWA TEJA (2011CS10260)

LOKESH YADAV (2011CS10226)

BONAFIDE CERTIFICATE

Certified that this Project Report

"SOCIAL MEDIA ANALYTICS"

is the bonafide work of
"PALAS, ABHISHEK NIMESH, VISWA TEJA, LOKESH YADAV",

who carried out the project work under my supervision.

Signature: .....
Name: .....
Date: .....

TABLE OF CONTENTS

INTRODUCTION

Social network analysis (SNA) is the methodical analysis of social networks. Social network analysis views social relationships in terms of network theory, consisting of *nodes* (representing individual actors within the network) and *ties* (which represent relationships between the individuals, such as friendship, kinship, organizational position etc.) These networks are often depicted in a social network diagram, where nodes are represented as points and ties are represented as lines.

Social media analysis takes this a step further, analyzing how certain topics grow in popularity and spread over social networks. Analysing things such as popularity of topics, trending topics, per topic communication over time, relation between network clusters and topic spread, offer a good method of succinctly analyzing a social media platform.

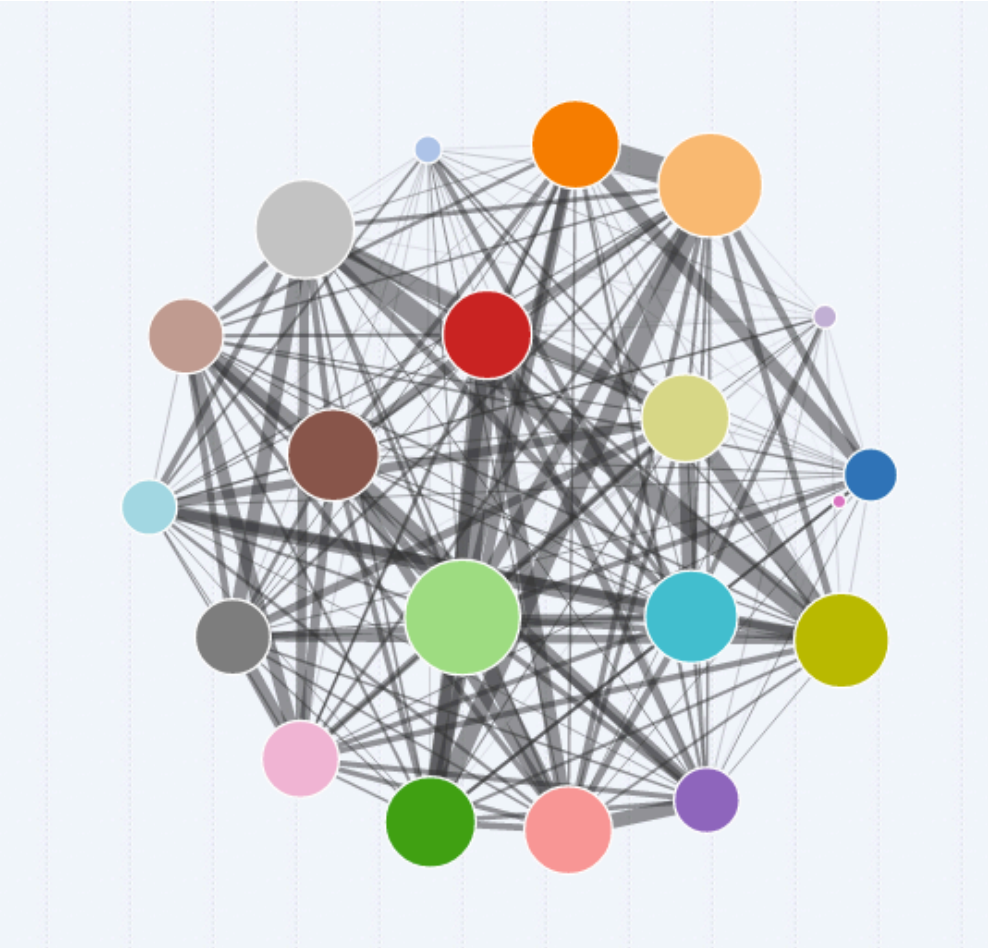
Method of Analysis

The methods adopted are of four types –

a) We look at the data of each week and spot the trends- which topics are popular, which topics are trending and so on. Then we compare the data collected and see if we can draw reasonable conclusions from them.
b) To see how communication activity in the social network changes over time, we constructed an interactive timeline graph that allows us to view the trends in communication growth , either for the whole lifetime of the social media website or on a monthly/weekly basis.
c) We implemented an interactive motion graph to track the trends in topic-wise communication over time. It has three visualization modes – bubble visualization, bar graph visualization, and line visualization from which we can derive many conclusions. The motion graph has a play-pause feature that enables analysis over time.
d) We constructed a location X location matrix, which can be used to compare the ratio of intra to inter communication. It can also be rearranged to draw other conclusions.

Social Media Visualisation

I. Force-Directed Cluster Layout

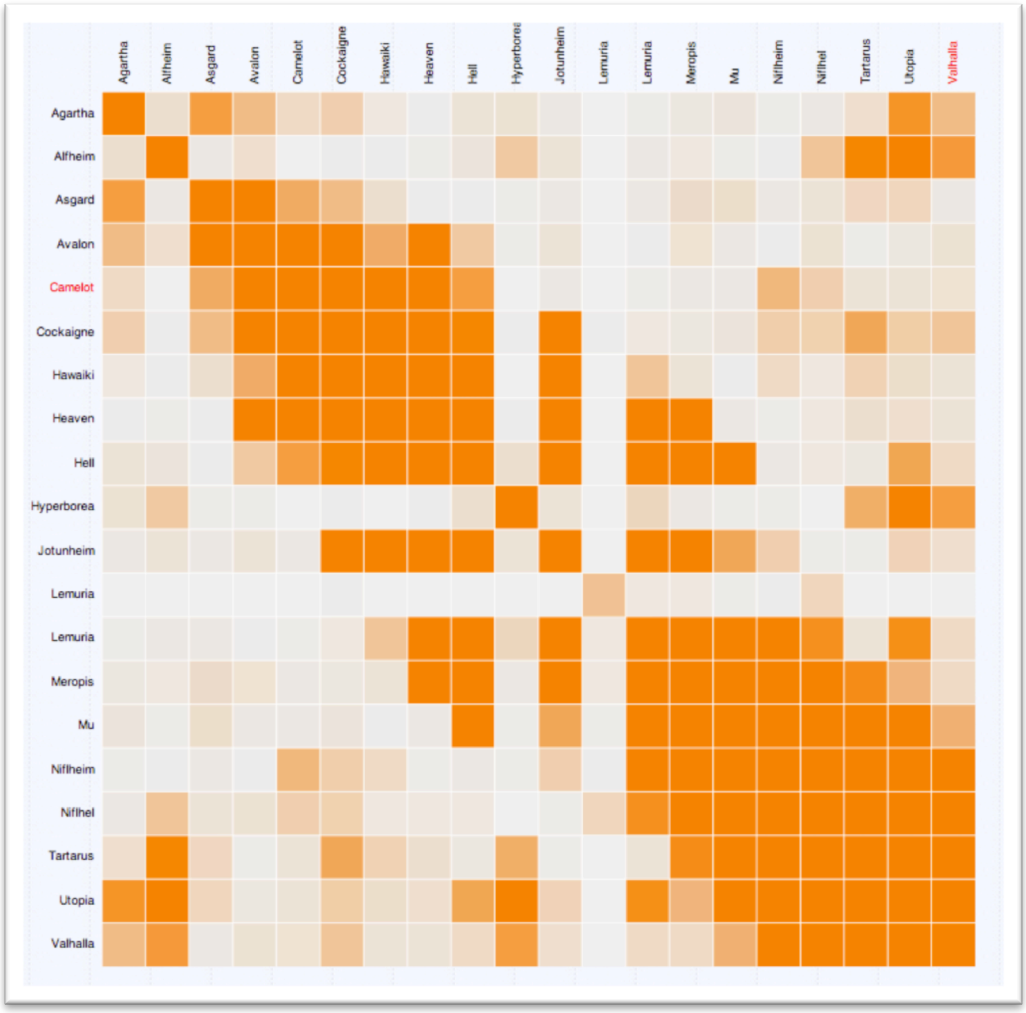


The above visualization clusters the nodes belonging to the same location into a bigger node and demonstrates inter-edges between the location. The size of the node is dependent on the number of nodes it contains and the width of edges depends on the number of edges between the corresponding locations. Upon double-clicking a cluster node, the nodes contained in that cluster become visible.

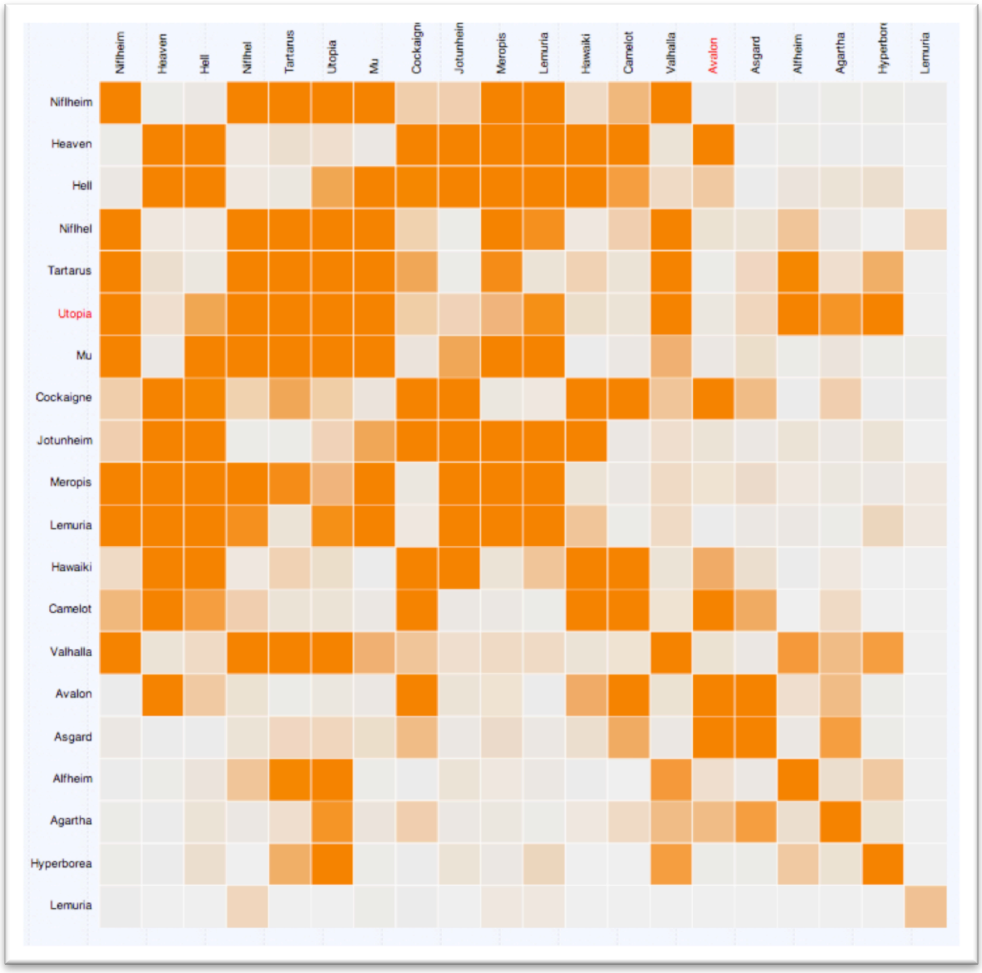
II. Cluster Matrix Visualisation

A network can be represented by an adjacency matrix, where each cell ij represents an edge from vertex i to vertex j. Here, vertices represent locations in a social network, while edges represent communication activity.
Given this two-dimensional representation of a graph, a natural visualization is to show the matrix! However, the effectiveness of a matrix diagram is heavily dependent on the order of rows and columns: if related nodes are placed closed to each other, it is easier to identify clusters and bridges.

In this visualization, the nodes are clustered according to their locations and an intensity map in the cluster x cluster matrix is drawn, darker shades representing more interaction between corresponding nodes in the matrix. The clusters can be rearranged either alphabetically or depending upon the frequency of their communication.



The above matrix has the locations arranged alphabetically. A dark band is observed along the diagonal. This indicates that intra-location communication activity is much higher than inter-location communication activity.

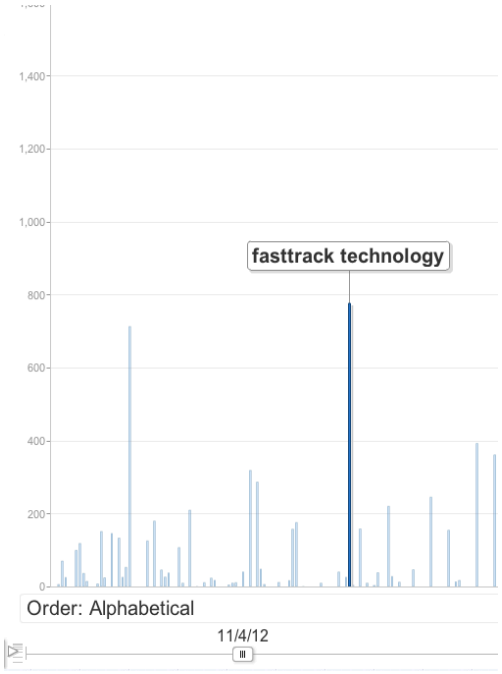


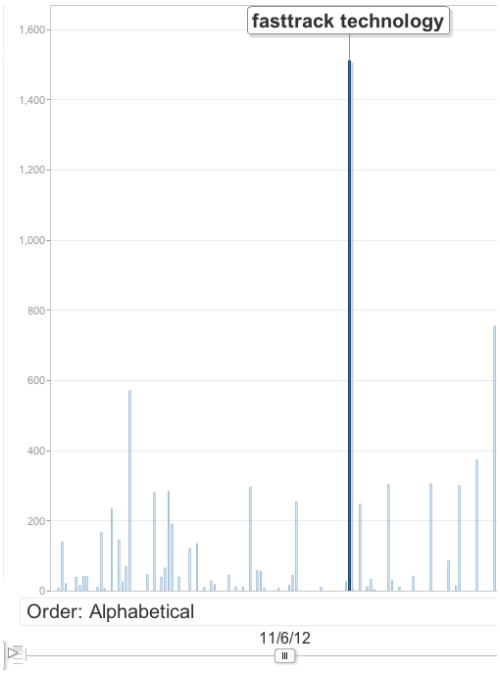
The above matrix has the locations arranged by frequency of communication in that location, with the more interactive locations placed first. A dark band is observed in the upper left quadrant., but the dark band of the diagonal is maintained. This indicates that locations which have higher frequency of communication have both higher intra and inter communication.

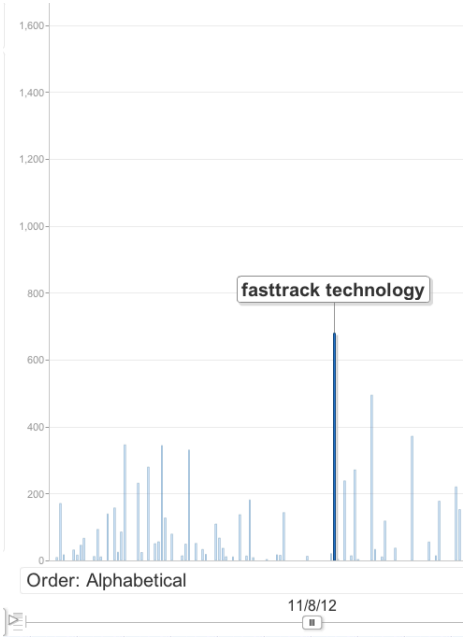
Motion graph Visualisation

Conclusions drawn:

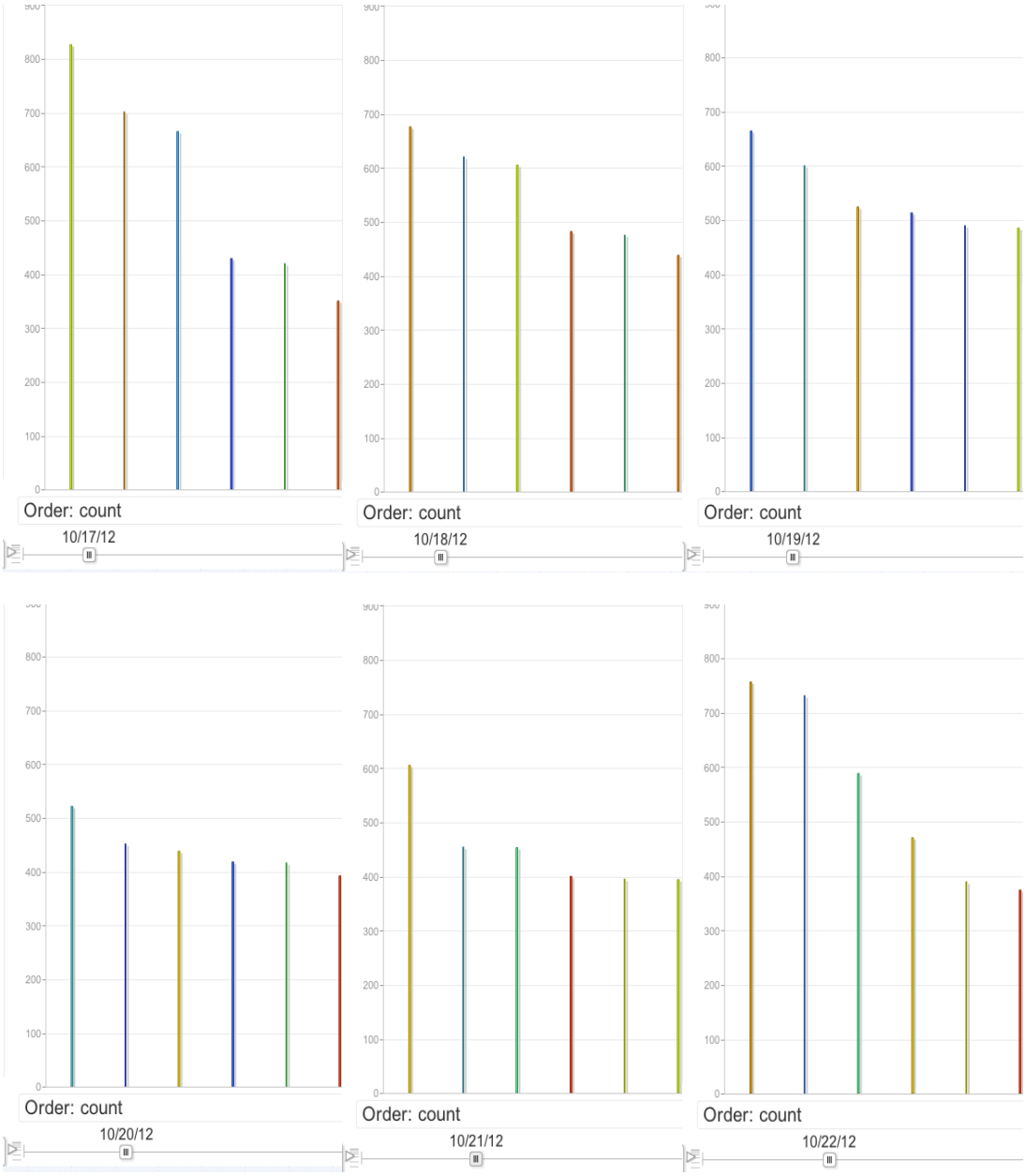
1) The communication activity for a topic, if it becomes popular, peaks very quickly and then decreases equally quickly.





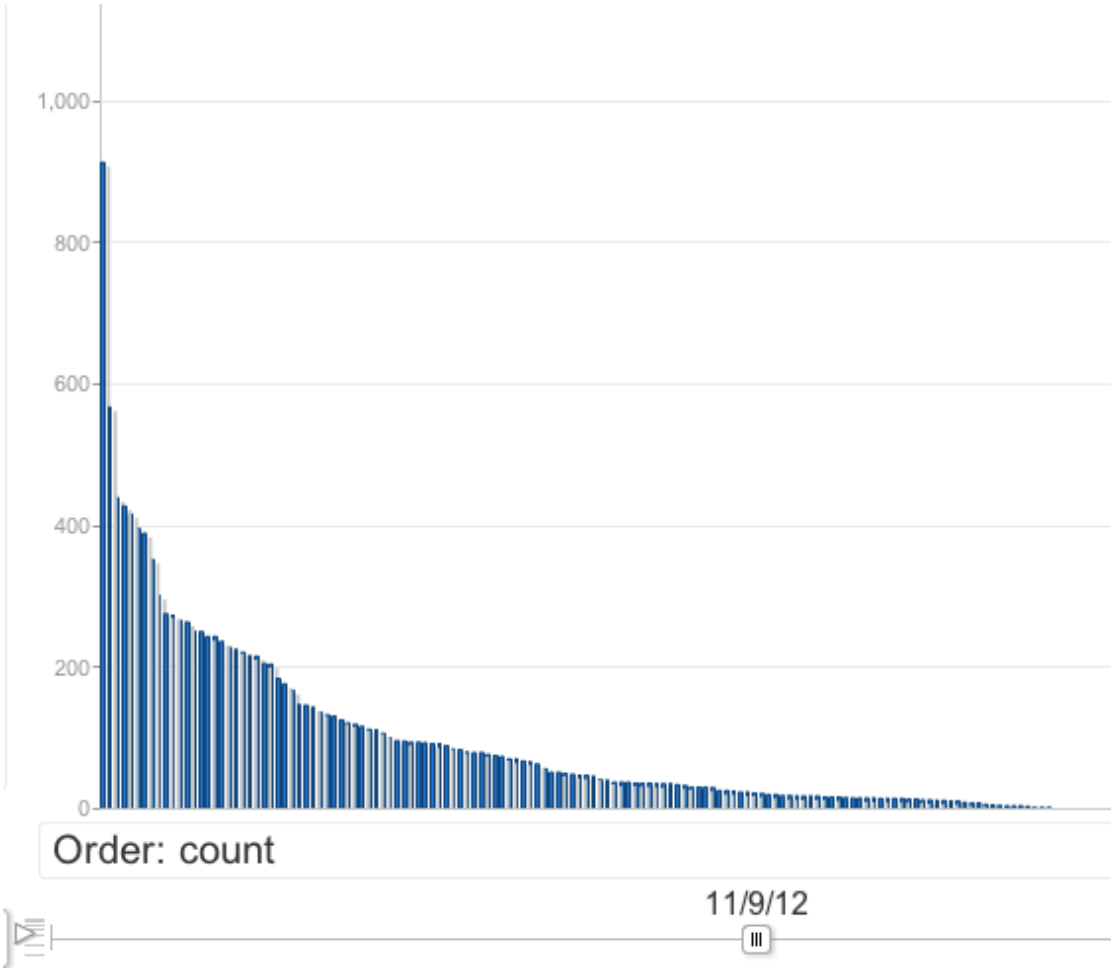


2) The most popular topic of the day changes almost every-day.

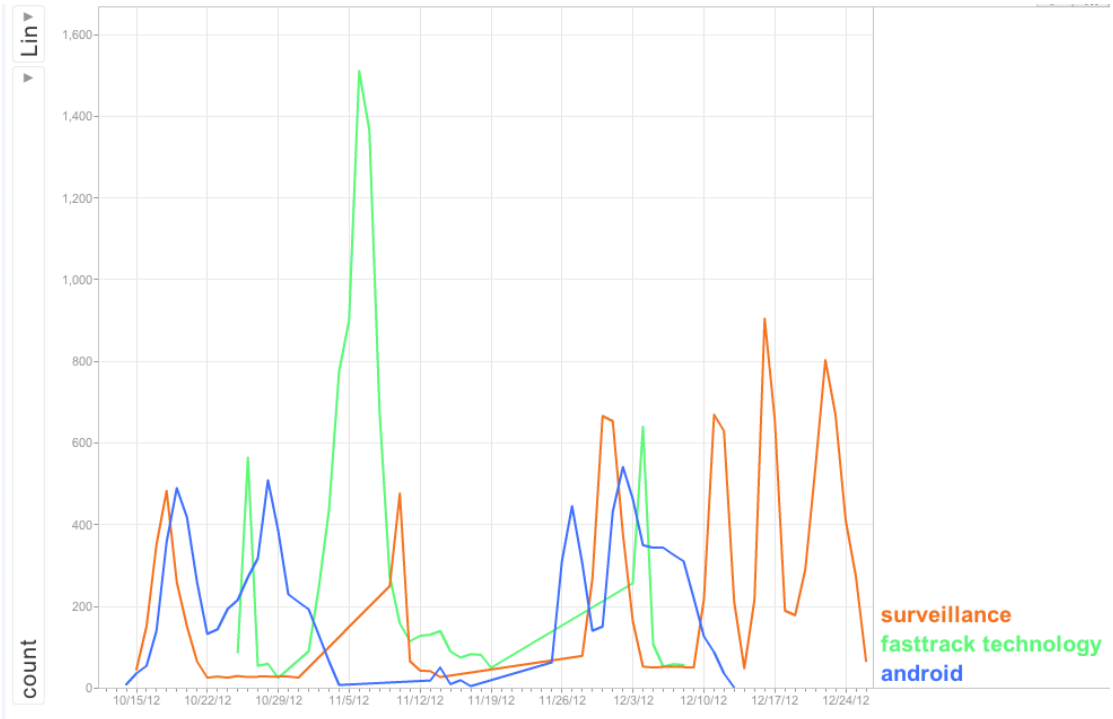


In the visualization, each topic is assigned its own unique color. When we order the topics by count, it is seen that the topic with most communication changes every day as can be seen from different colors of the first bar.

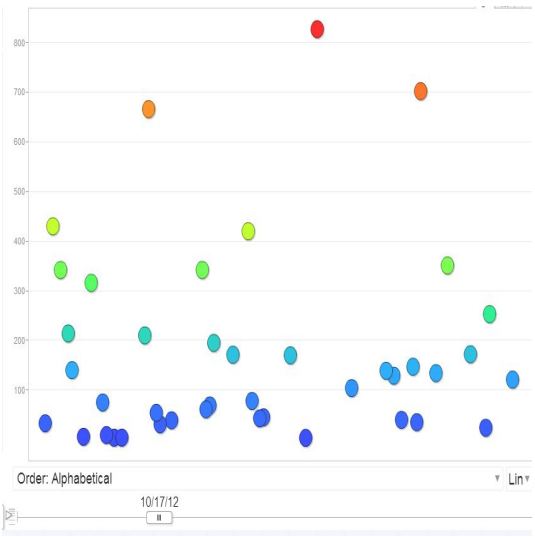
3) Less popular topics account for a large percentage of the total number of topics. Also, the popularity of different topics varies almost logarithmically.

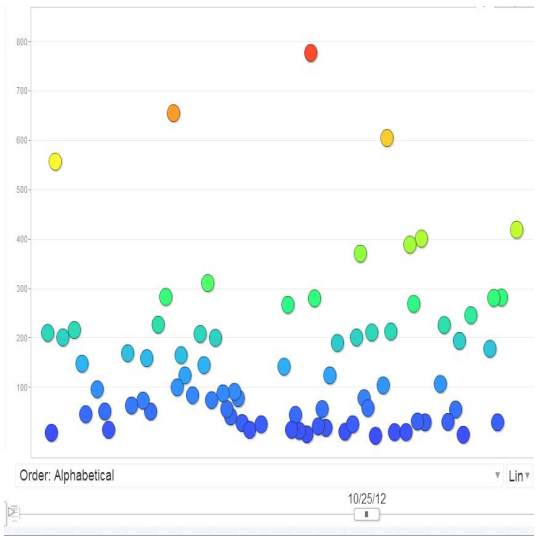


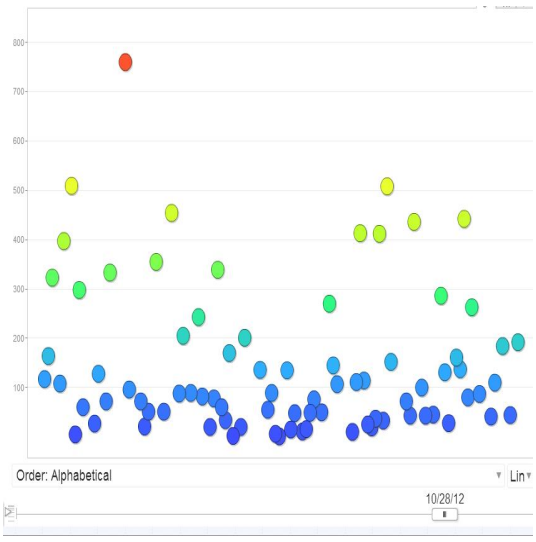
4) The popular topics follow similar curves.

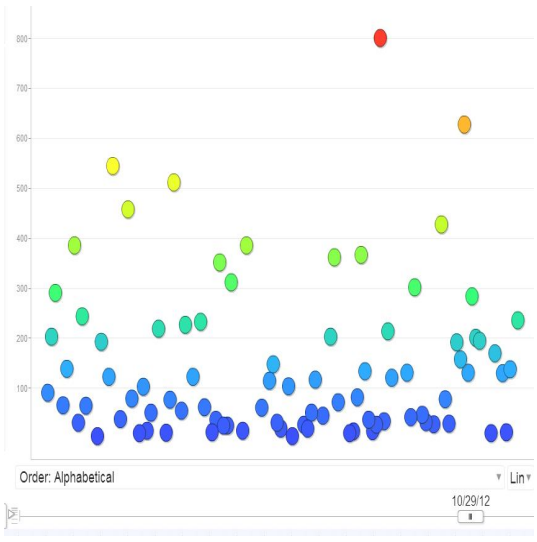


5) At any time, there is usually only one very popular topic.





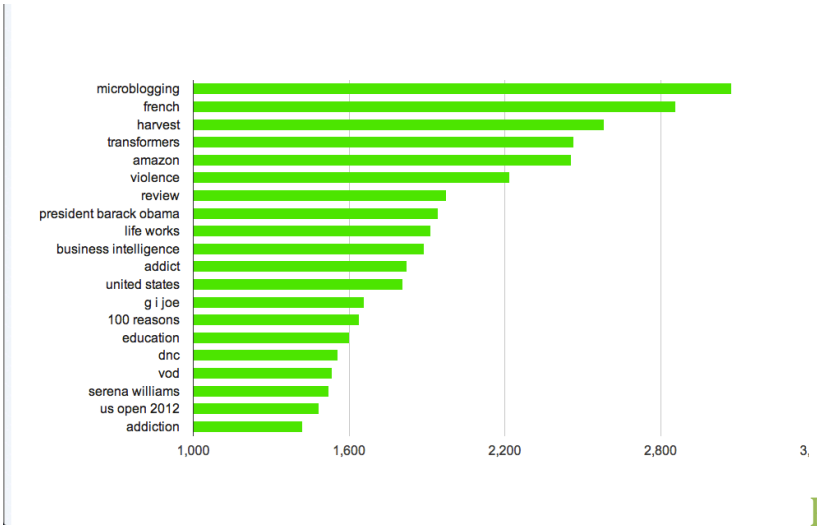




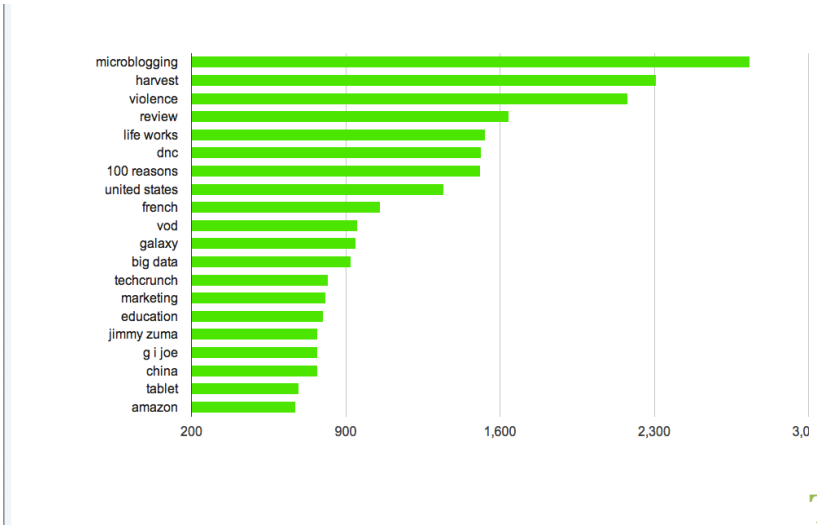
The dark color red color dot denotes a topic of very high popularity. As can be seen, at any time, there is usually one darkly colored red dot.

Trends in Popularity of topics/ Trending topics

The popular/trending topics are shown by a bar-graph. The top 20 in each category are chosen.



Popular Topics



Trending Topics

Q&A:

Q. Do trending topics grow very suddenly?

Ans. Yes. They usually grow in popularity usually very quickly.

Q. What is the lifetime of trending topics?

Ans. Very Short. Usually the topics don't continue to be trending for many weeks.

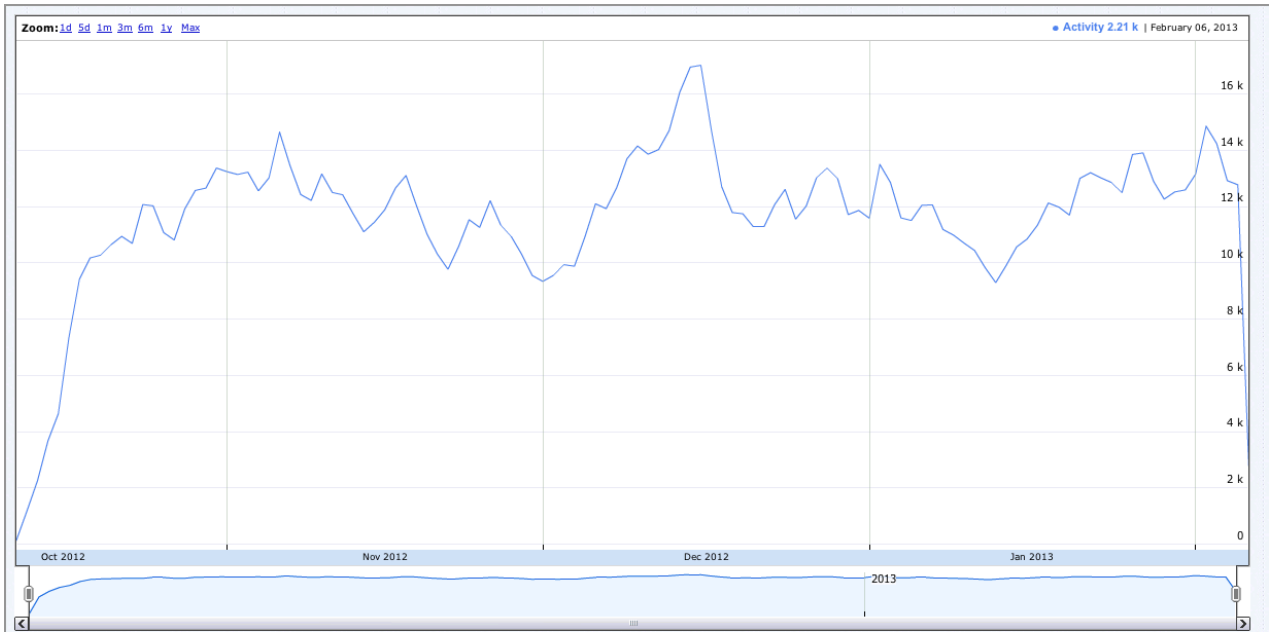
Q. Do trending topics become overall popular topics?

Ans. Some do so, but very occasionally. Usually trending topics do not become overall popular.

Q. How do trending topics affect overall communication in that week?

Ans. Yes. For instance, in a week, the overall communication was 87,279 and the top 20 trending topics between them accounted for 35,361 communications, almost 41% of the total communication.

Overall Communication Visualisation



As can be seen from the activity time line, the communication increased very rapidly in the initial phase of the social media website, and thereafter varied only marginally except during December. (Likely due to it being holiday season)