

DS256:Jan17 (3:1)

Assignment Discussions

Yogesh Simmhan

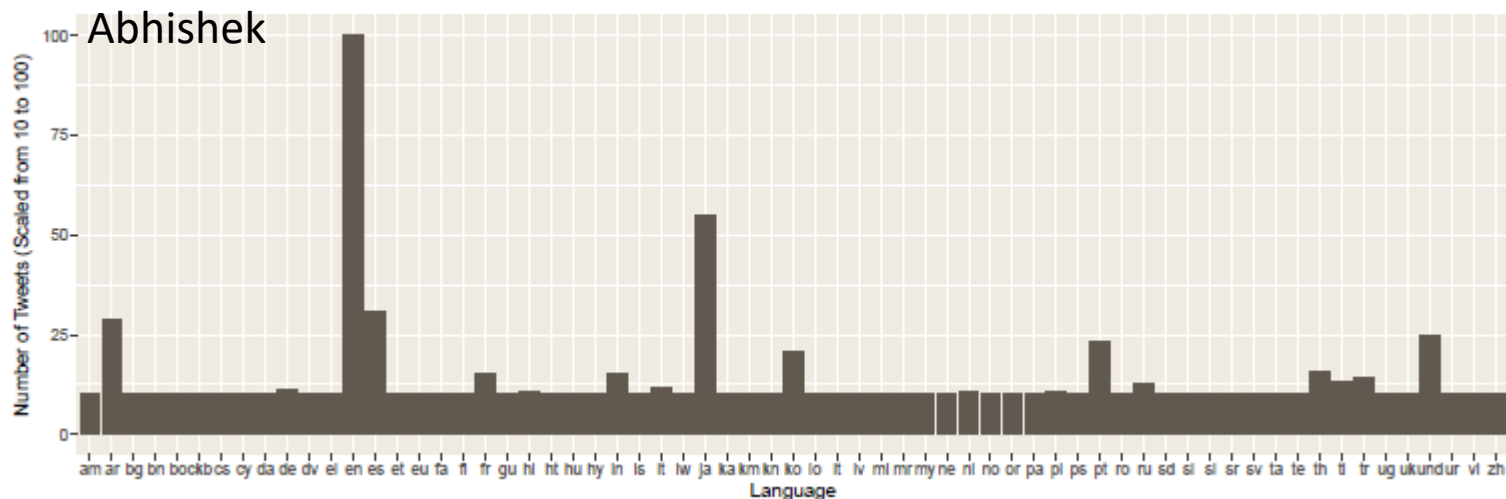
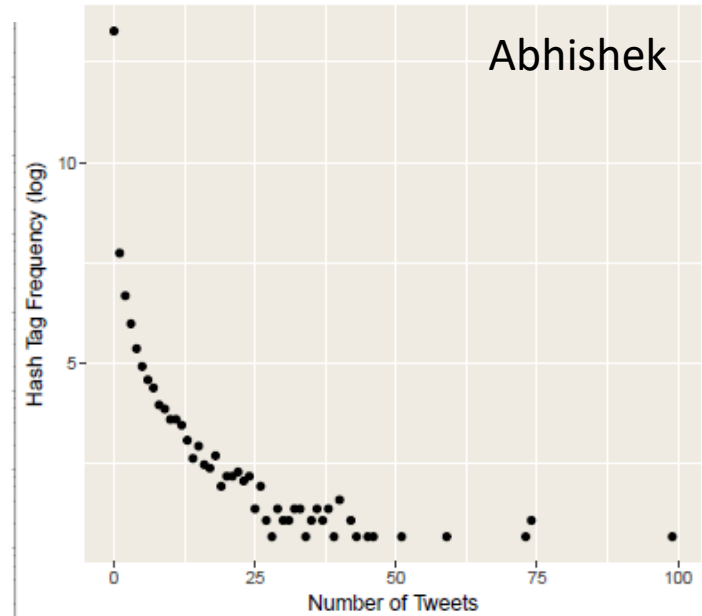
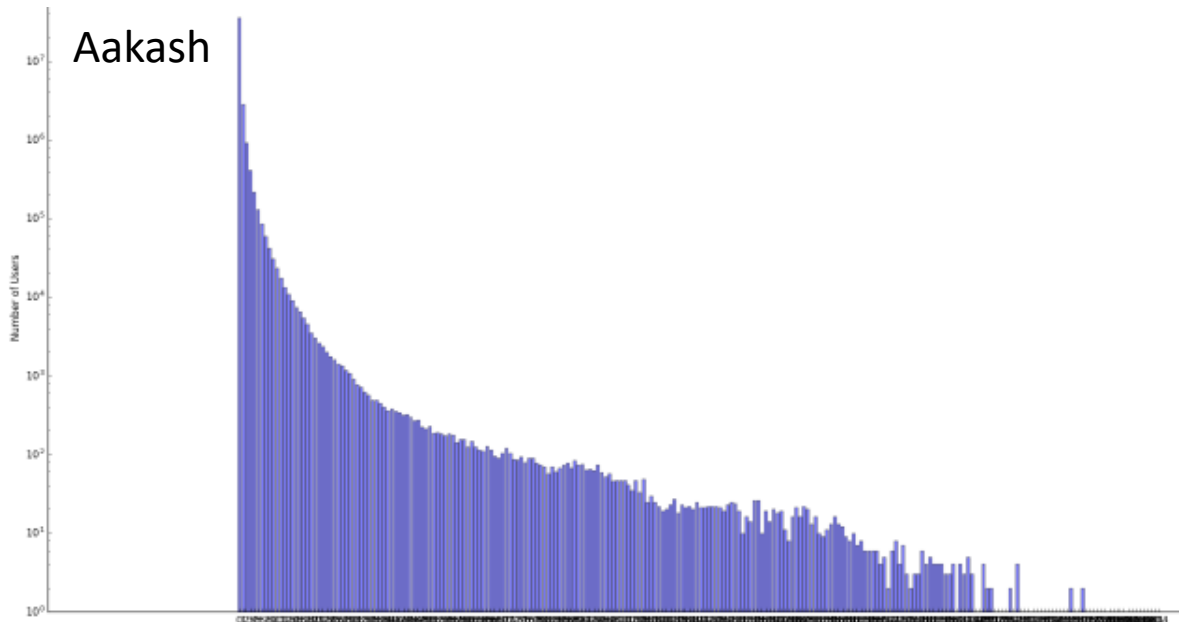
11 Apr, 2017



Assignment Alpha



Simple Statistics on Tweets





Temporal Statistics

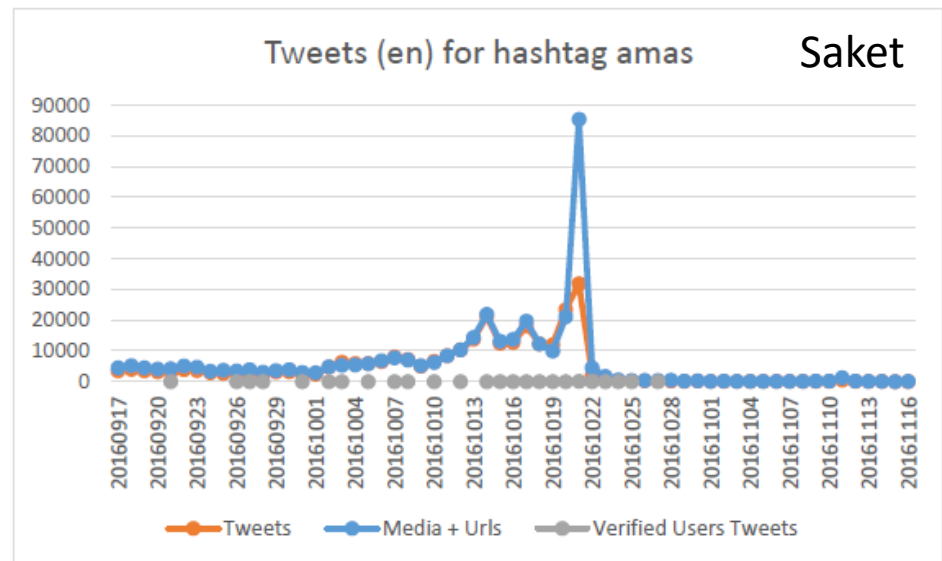
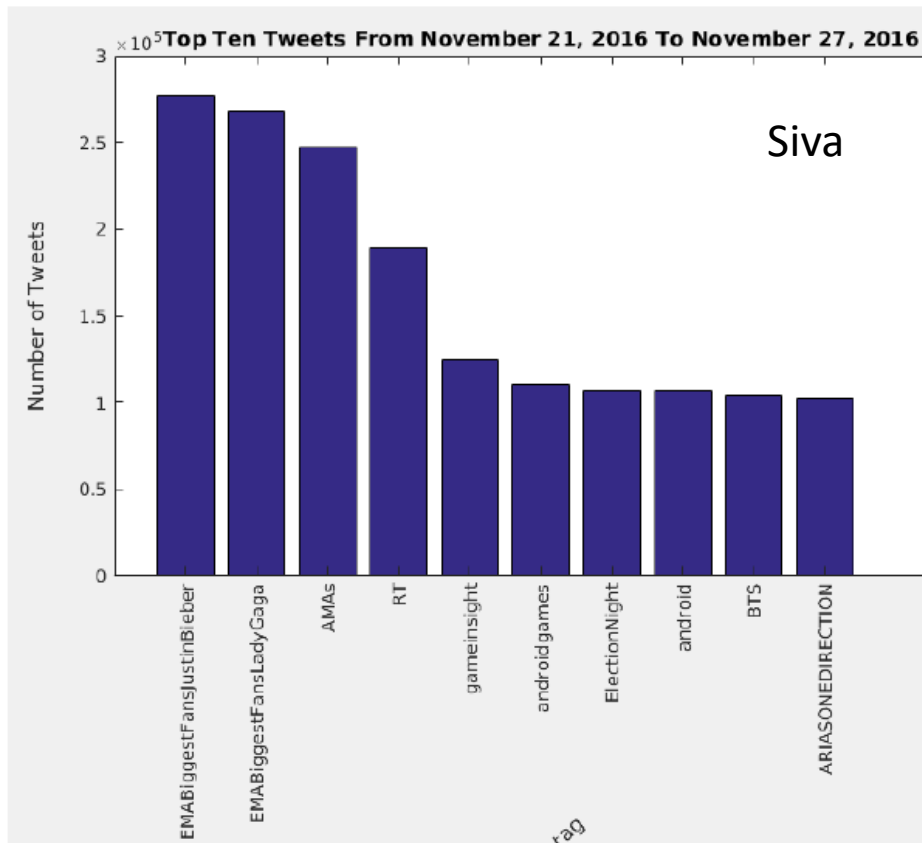


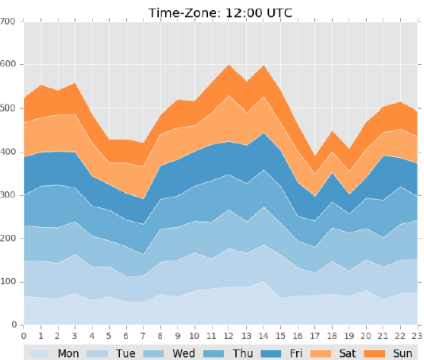
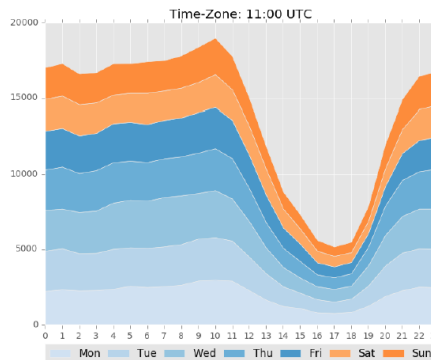
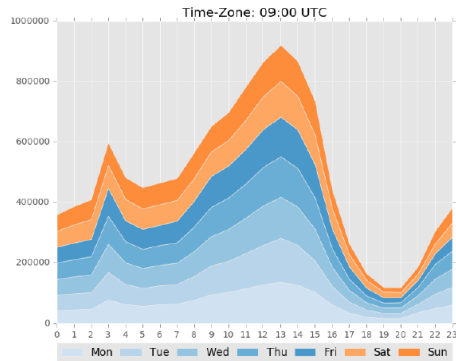
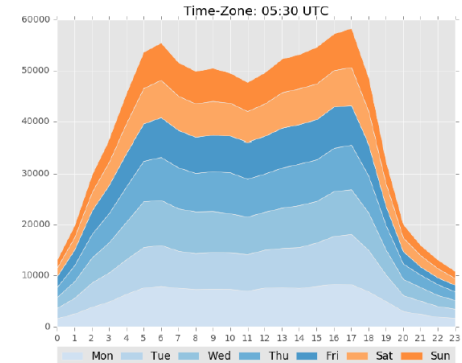
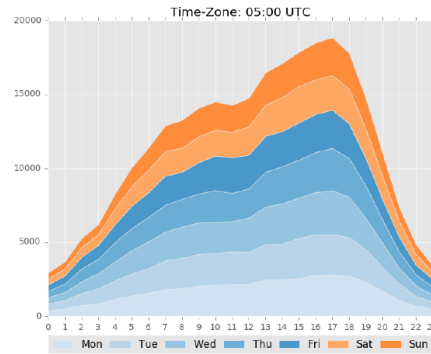
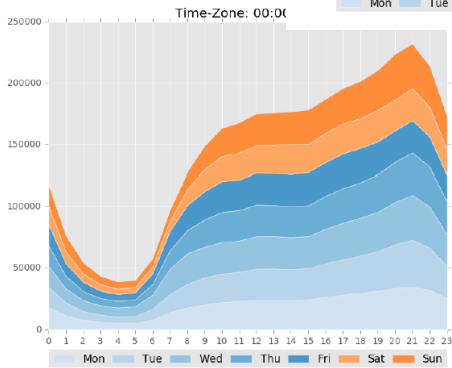
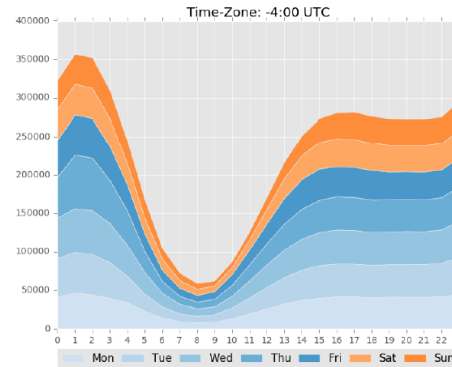
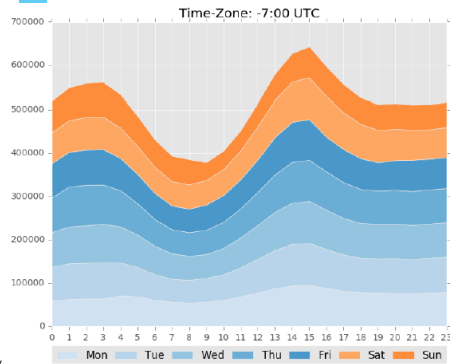
Figure 5.a.1: Tweets per day in lang en for hashtag amas. We notice that all the plots above have a spike around 19th, 20th & 21st Oct 2016.



Temporal Statistics

When do people work/sleep across the world?

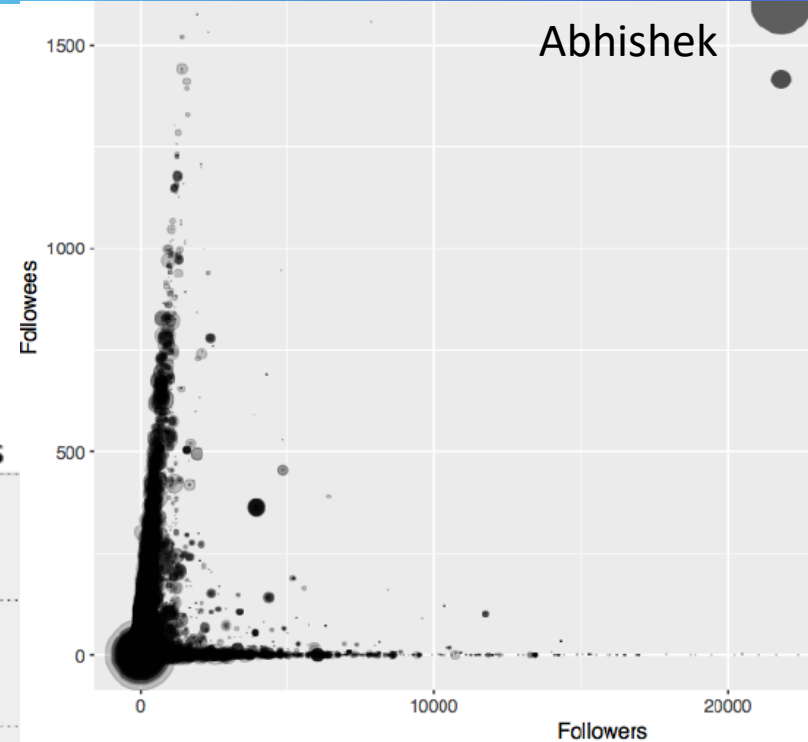
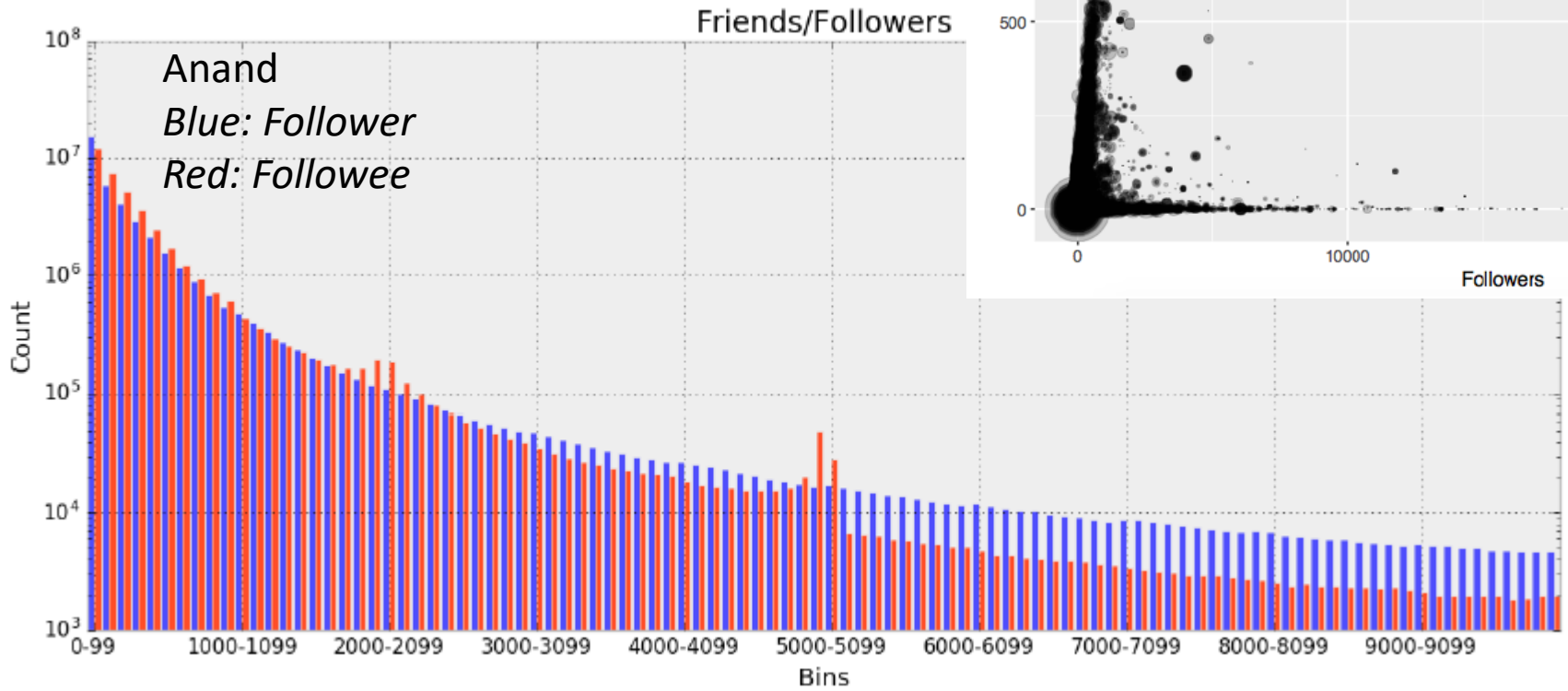
Anand





Network Statistics

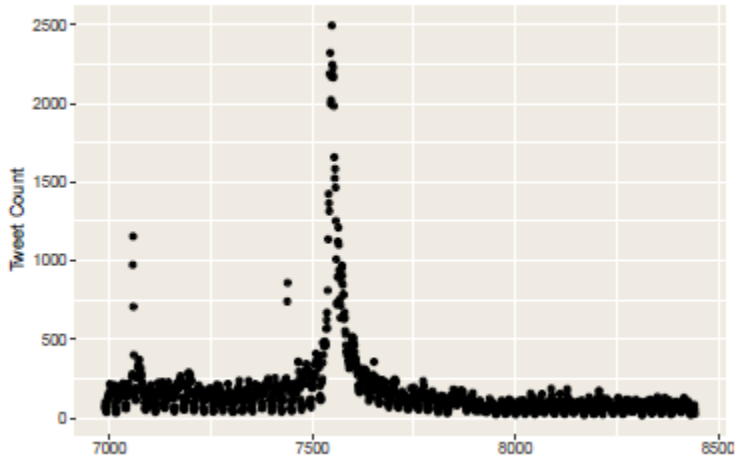
Abhishek





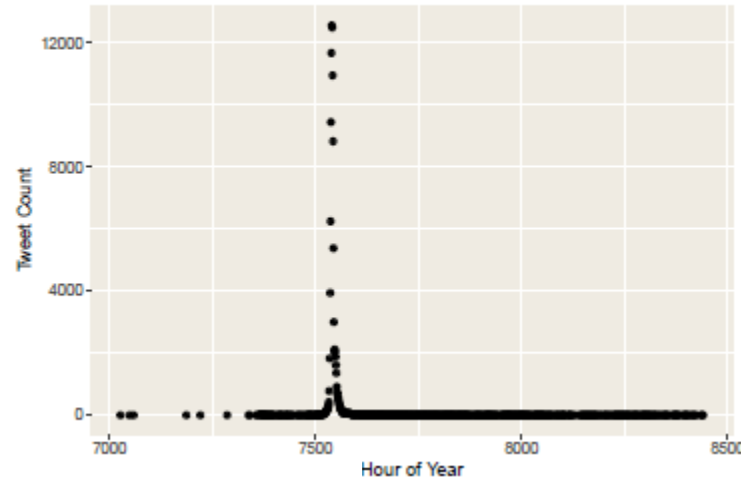
Influence

(a) TrumpInfluence

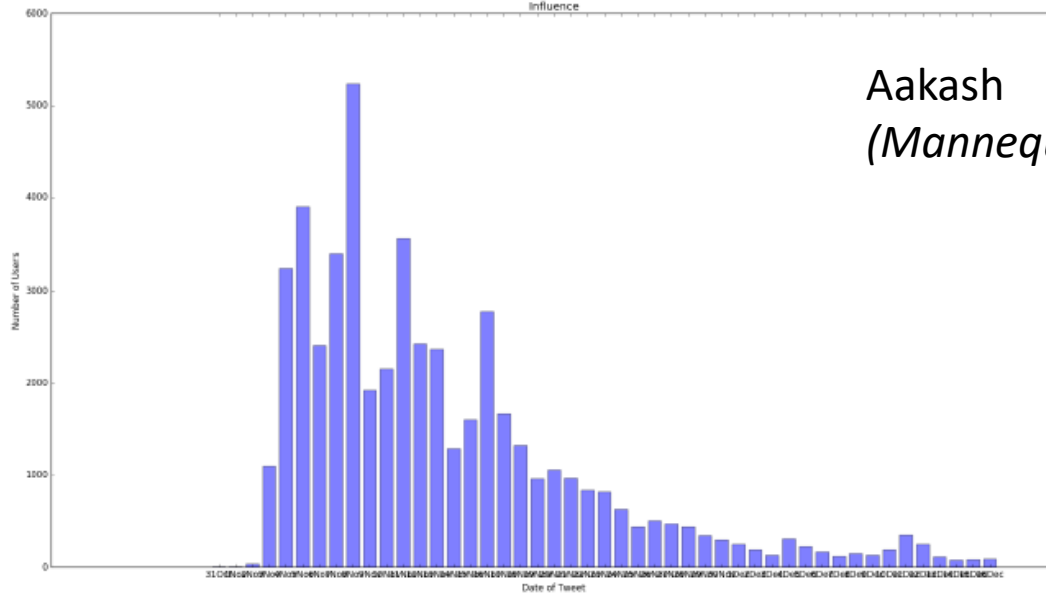


(b) ElectionNightInfluence

Abhishek



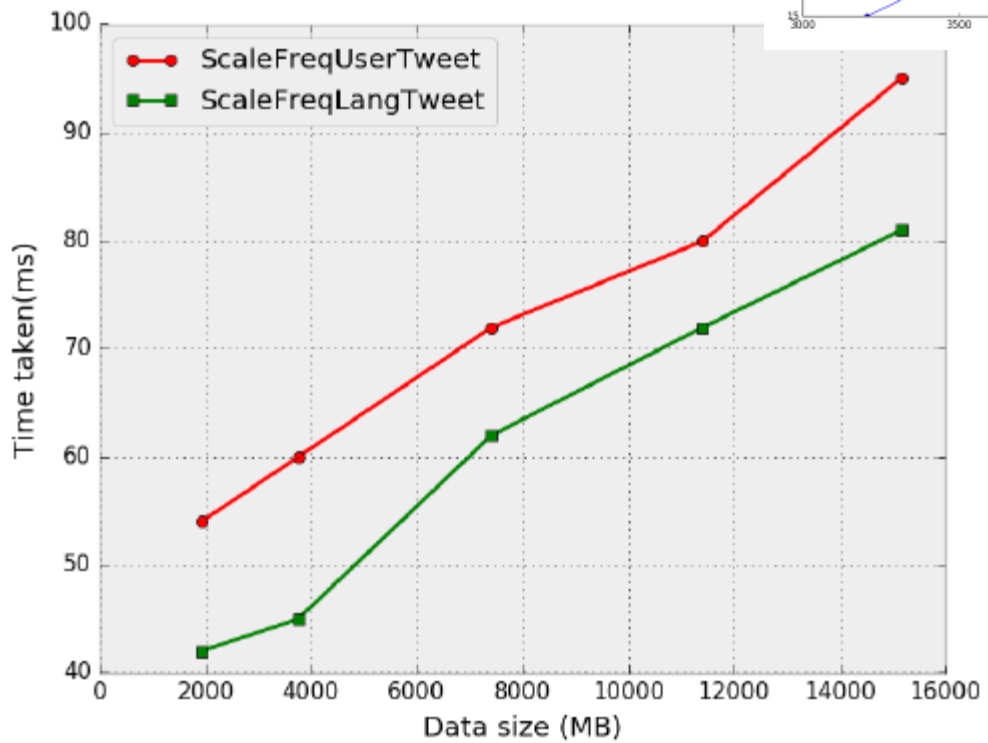
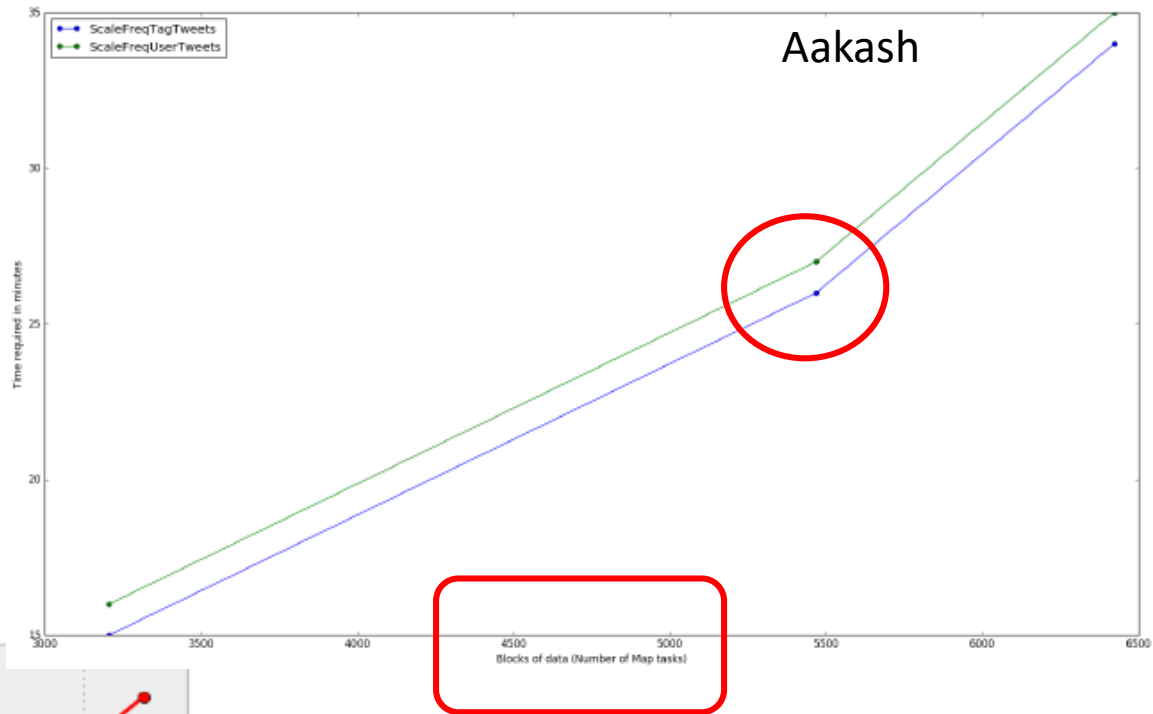
Influence



Aakash
(*Mannequin Challenge*)



Scalability

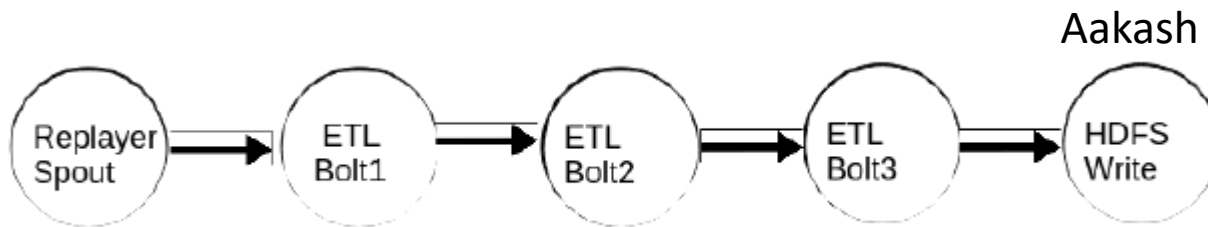
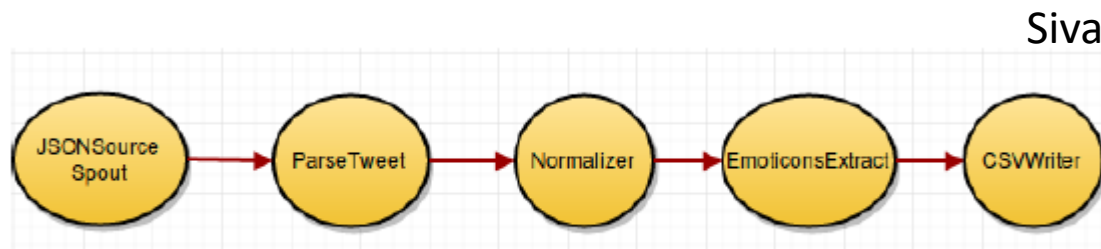
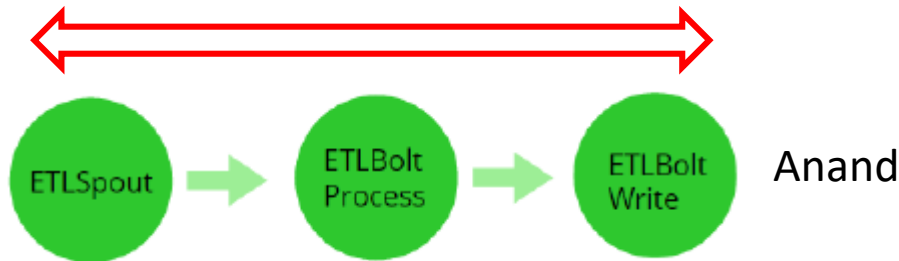




Assignment Beta



Extract, Transform, Load

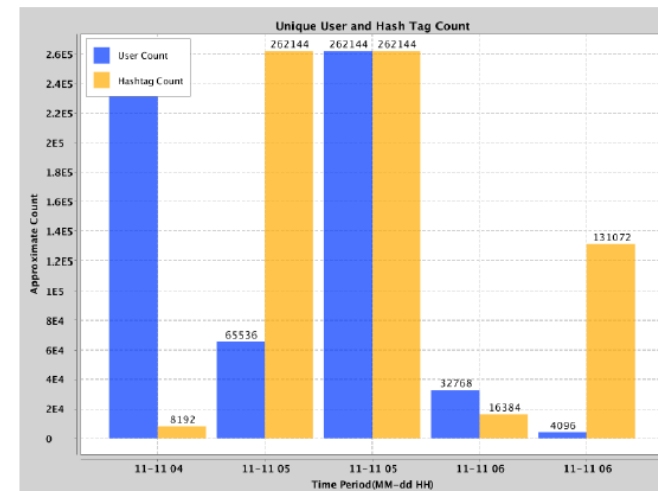
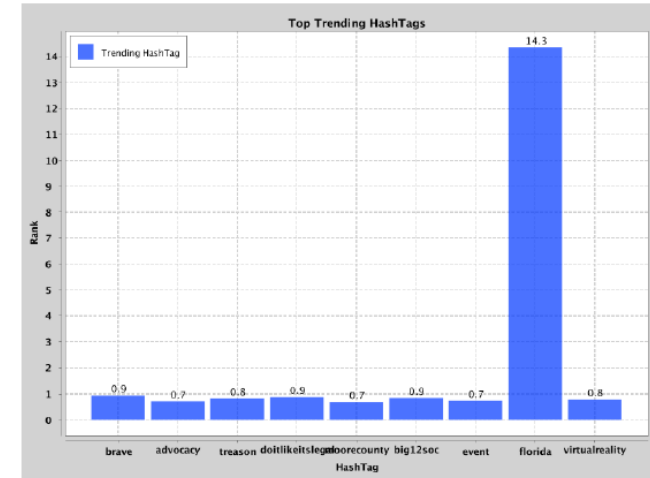
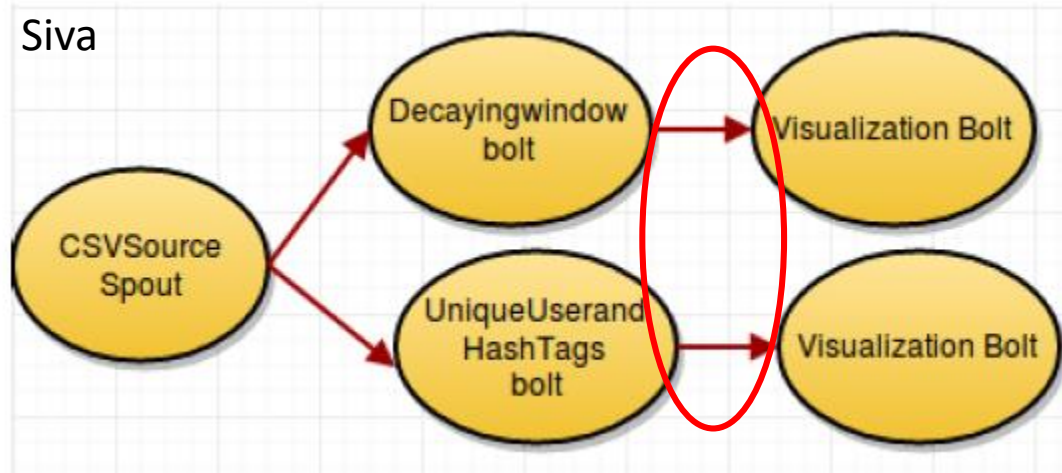


- ETL Bolt1: Extracts characteristics like userID, tweetID, timezone, etc
- ETL Bolt2: Performs NLP
- ETL Bolt3: Constructs the CSV
- HDFS Write: Writes the CSV to HDFS



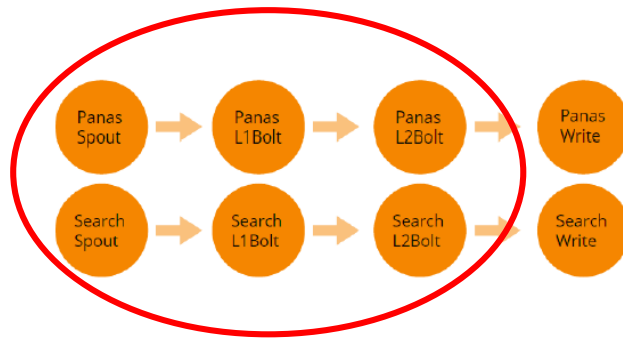
Realtime Statistics over Tweets

Siva

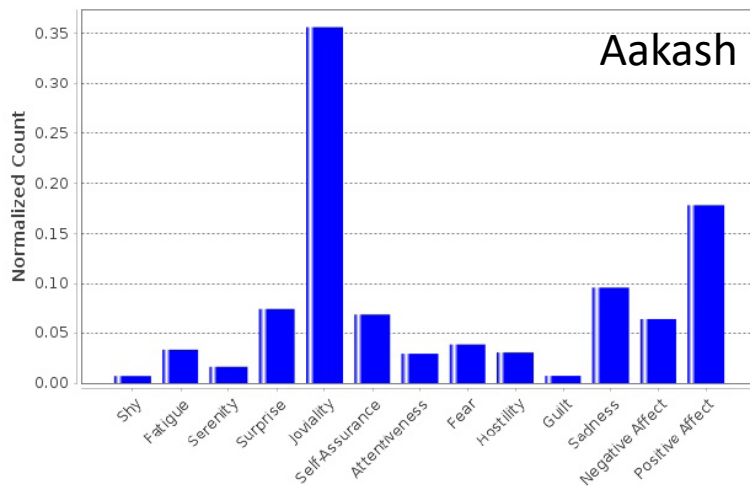
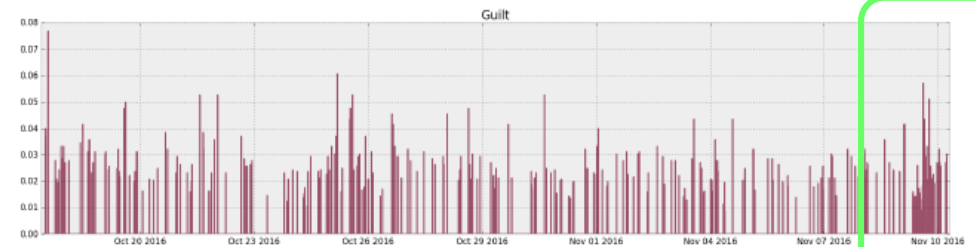
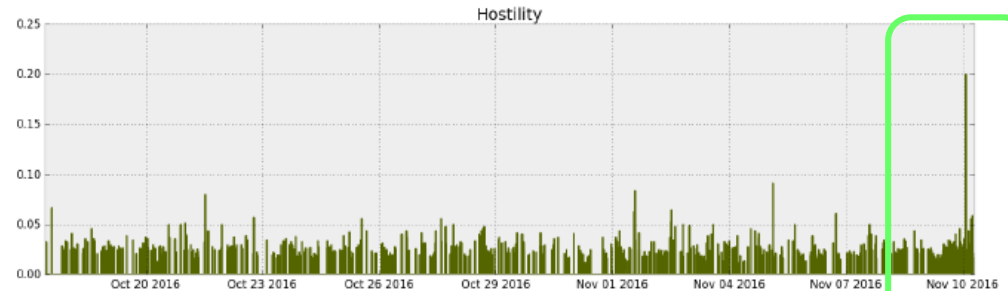
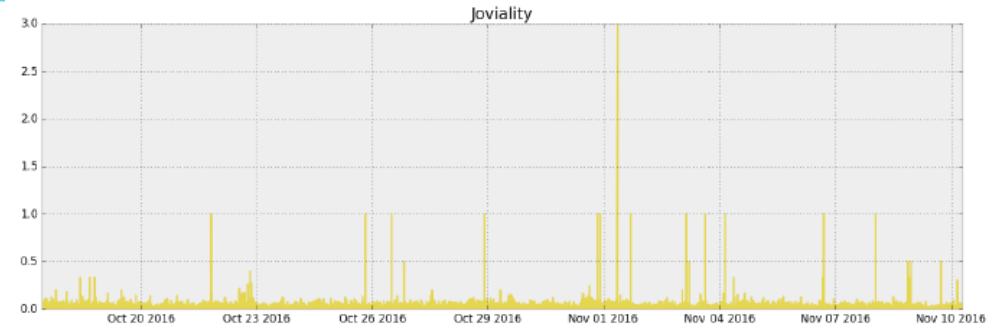




Sentiment Analysis

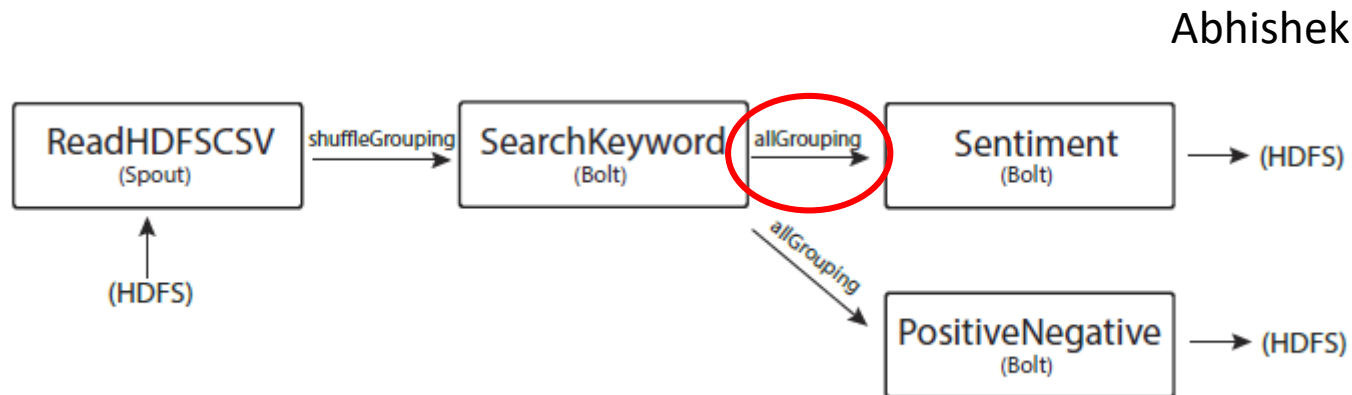


Anand





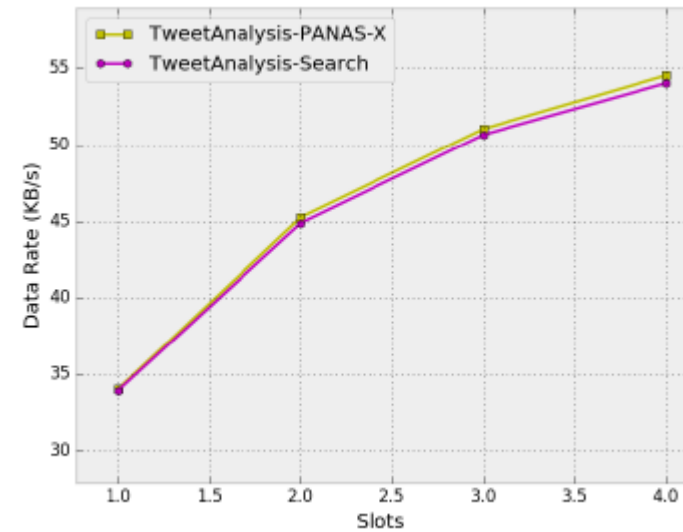
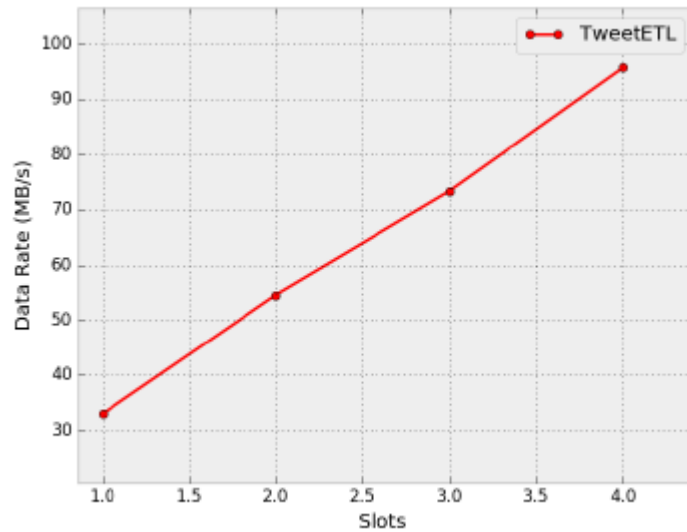
Sentiment Analysis





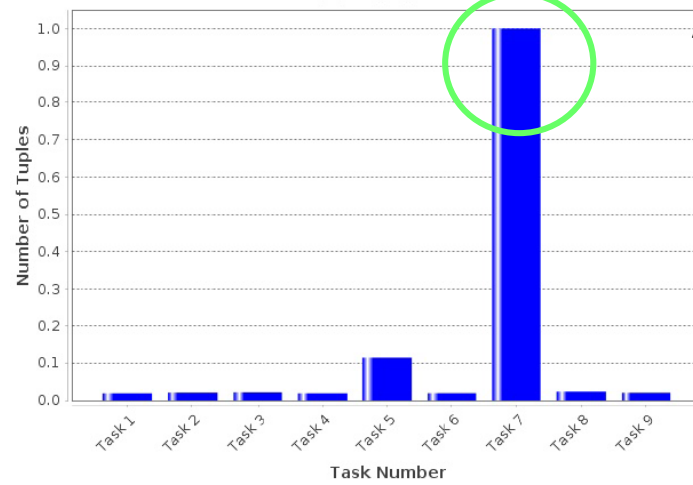
Scalability

Anand



HashTag_FM_9_Nodes

Aakash





Assignment Gamma



PageRank

Anand

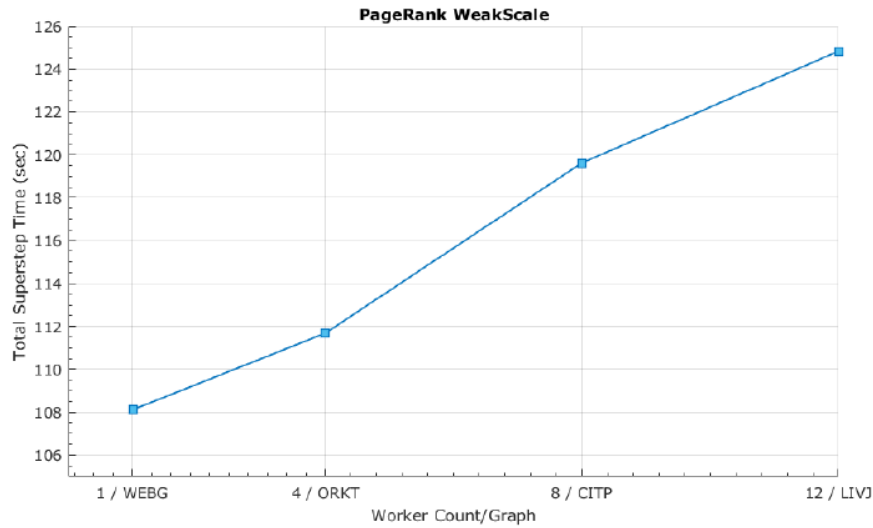


Figure 1: PageRank Weak Scaling

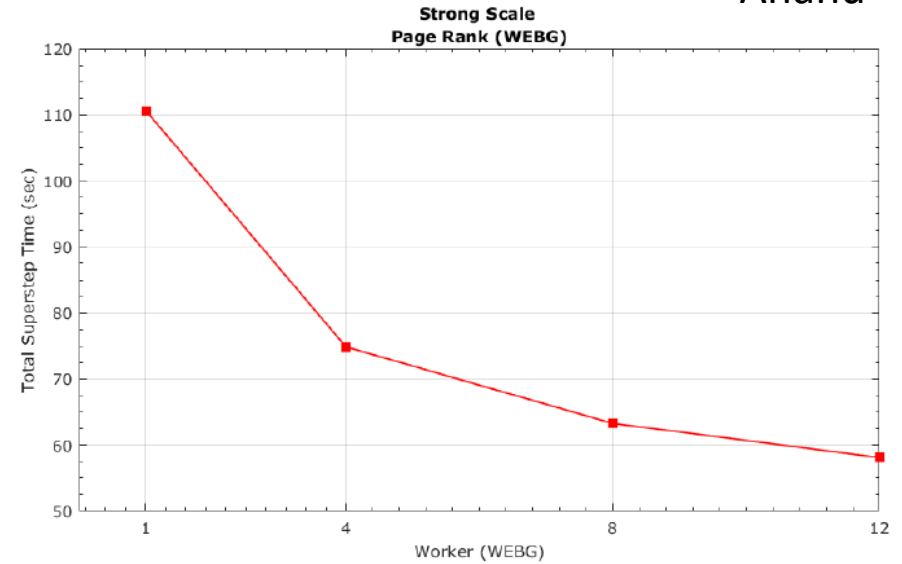
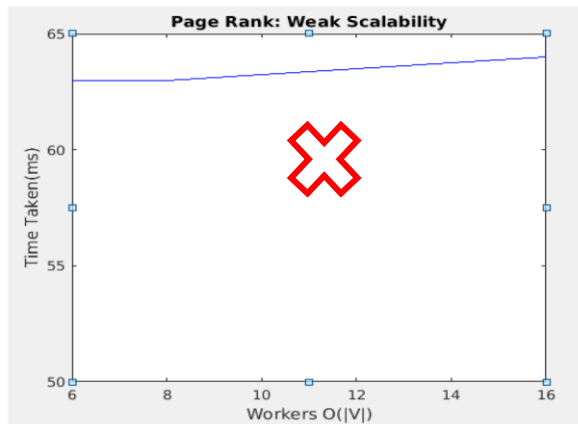


Figure 2: PageRank Strong Scaling



Siva

Spanning Tree

Aakash

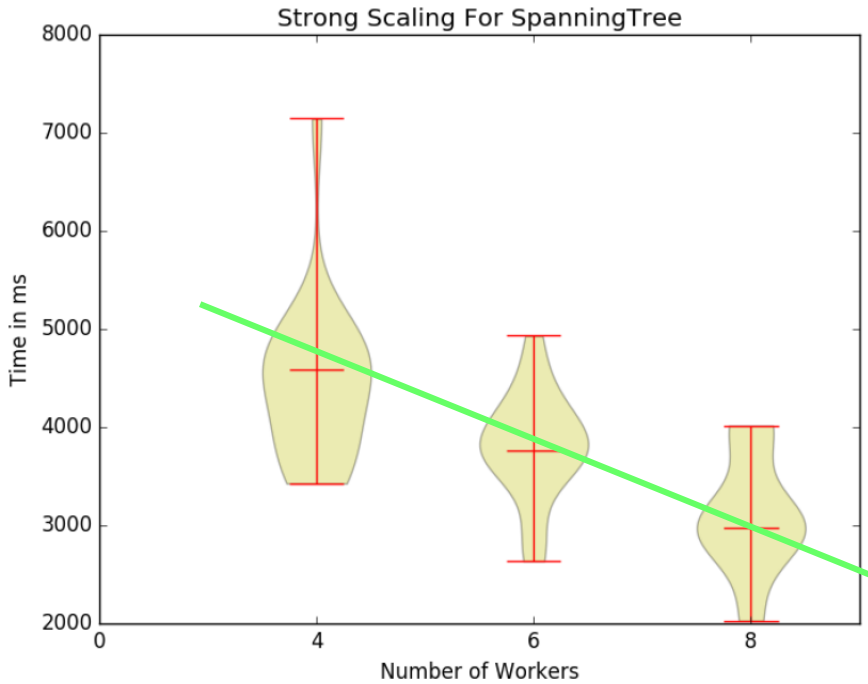


Figure 13: Strong Scaling for SpanningTree Algorithm

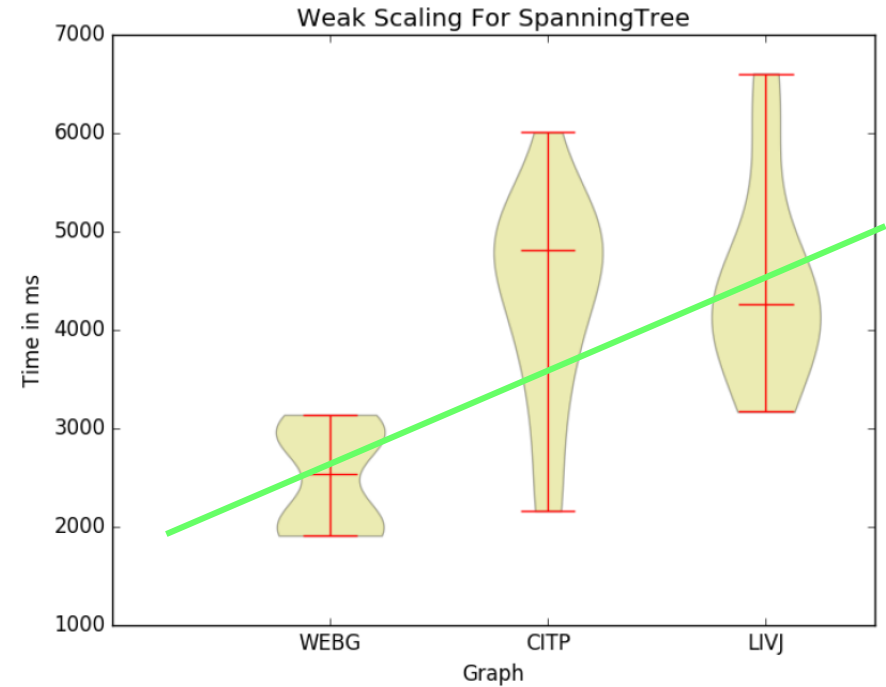


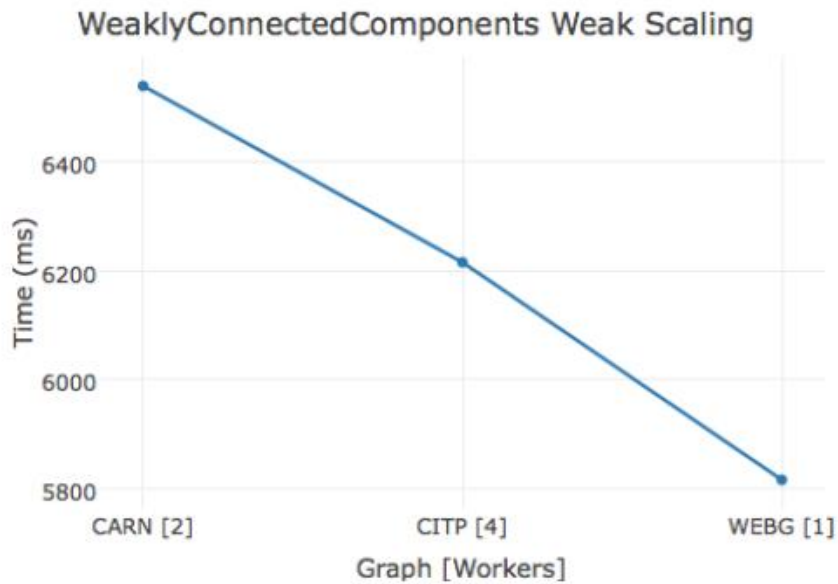
Figure 14: Weak Scaling for SpanningTree Algorithm

Based on V+E graph size?

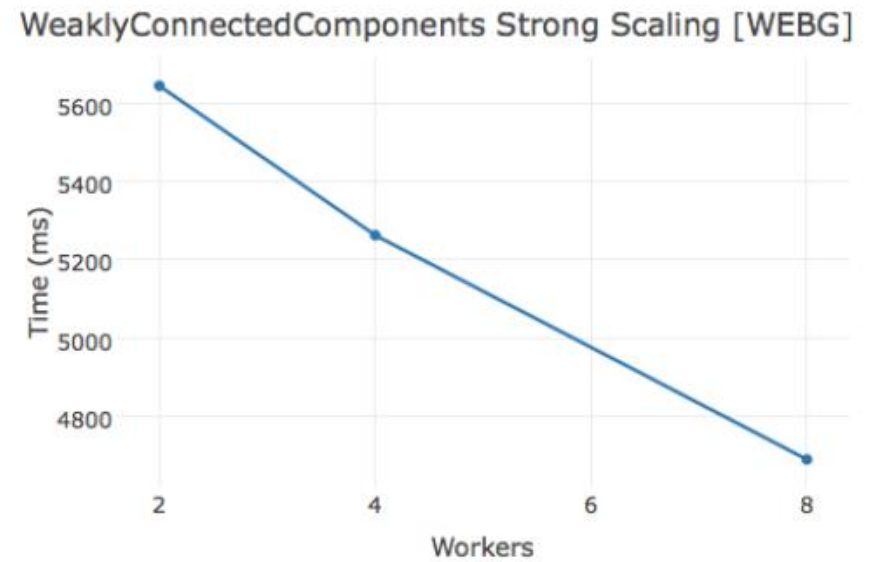


WCC

Abhishek



(a) WeaklyConnectedComponents Weak Scaling



(b) WeaklyConnectedComponents Strong Scaling

Euler Tour

Aakash

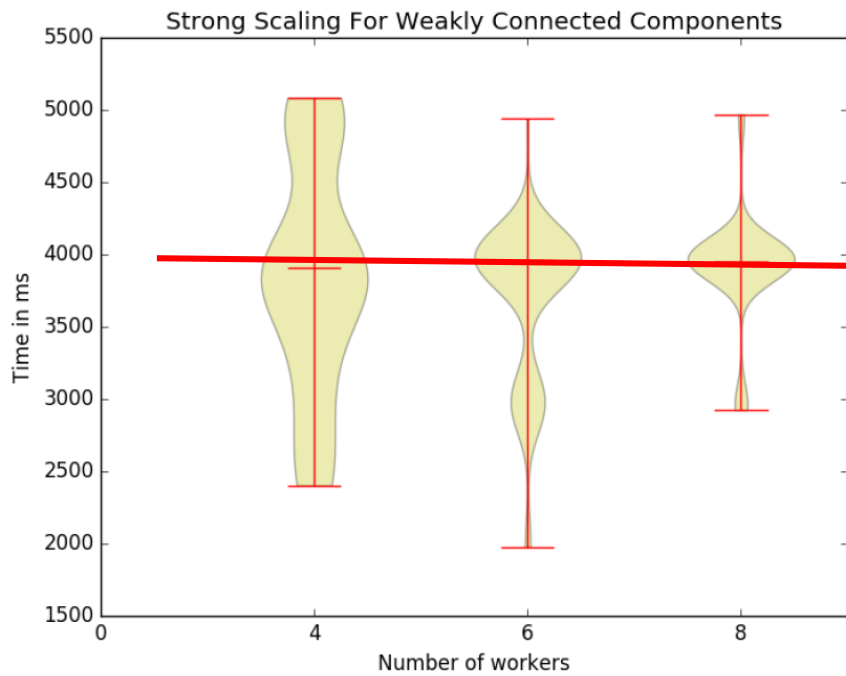


Figure 17: Strong Scaling for Weakly Connected Components Algorithm

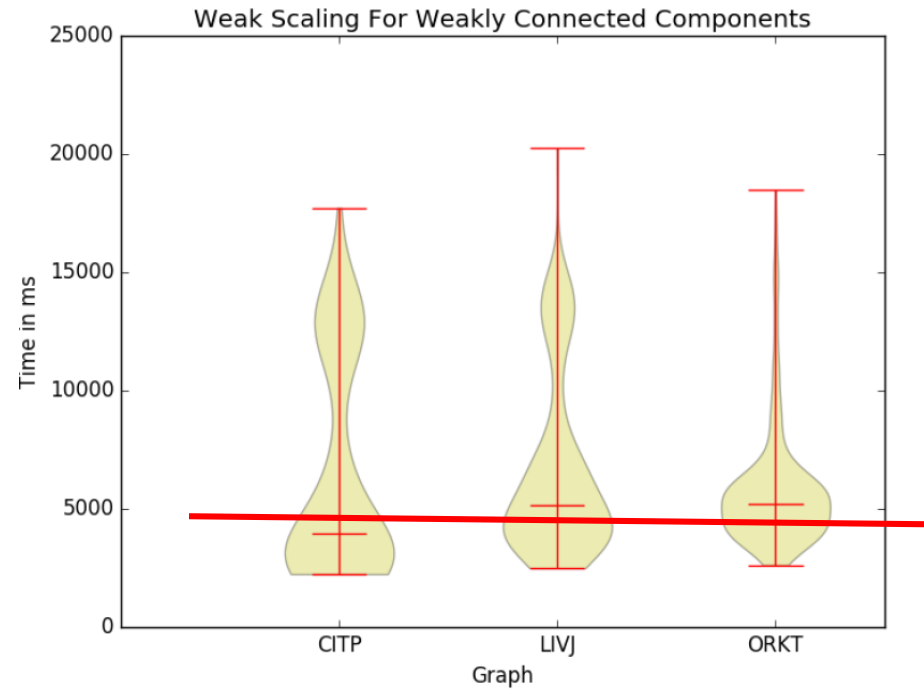


Figure 18: Weak Scaling for Weakly Connected Components Algorithm



Subgraph: PageRank

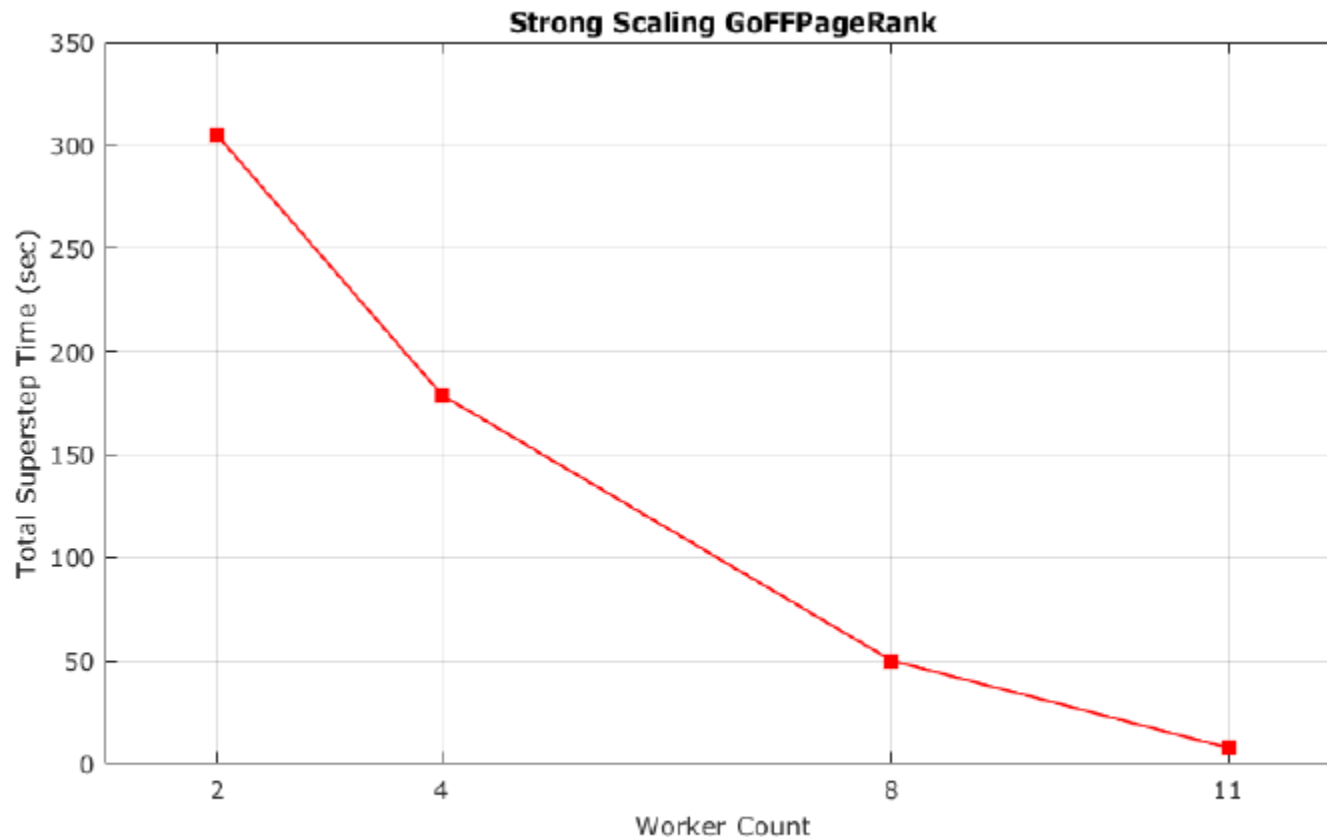


Figure 6: GoFFish PageRank Strong Scaling



Final Exam

- Final exam on **Fri, Apr 28, 2017 2-5PM**
- 200 points
- Syllabus: All topics from slides, lectures

Project

- Source code and report due by **April 25, 11:59PM**
- Demos & Presentation on **Sat, Apr 29, 2017 (10AM-1PM)**, CDS 309
- 300 points