Spread of Scientific Rumor

Referred Paper :
The Anatomy of a Scientific Rumor

Problem Statement

- We have to model the spread of rumour about the discovery of Higgs Boson (God Particle) during the period 1st July 2012 to 7th July 2012 over the twitter network .
- Basically ,We have to do the temporal statistical analysis i.e plot the graph between number of user believing that this rumour is true versus time.

Dataset :

- Dataset Used : Higgs Twitter dataset
- The Higgs dataset has been built after monitoring the spreading processes on Twitter before, during and after the announcement of the discovery of a new particle with the features of the elusive Higgs boson on 4th July 2012.The messages posted in Twitter about this discovery between 1st and 7th July 2012 are considered.

Methodology :

We are using SI model of disease spreading i. e susceptible - infected model with slight refinement.

SI model states that if a user is active i.e tweeting or retweeting about discovery of higgs boson at least once in a period , is infected through out then. So, number of infected user will be monotonically increasing .

So,In our final model we are considering at any given instant of time a non active node will transform to an active node with some phase varying constant (Lambda).

Work Achieved :

- Obtained higgs-activity_time dataset from Stanford website and queried it to obtain only retweets with a python script along with timestamp.
- We have also gone through the Paper “The Anatomy of Scientific rumour ”, as a reference for our project.

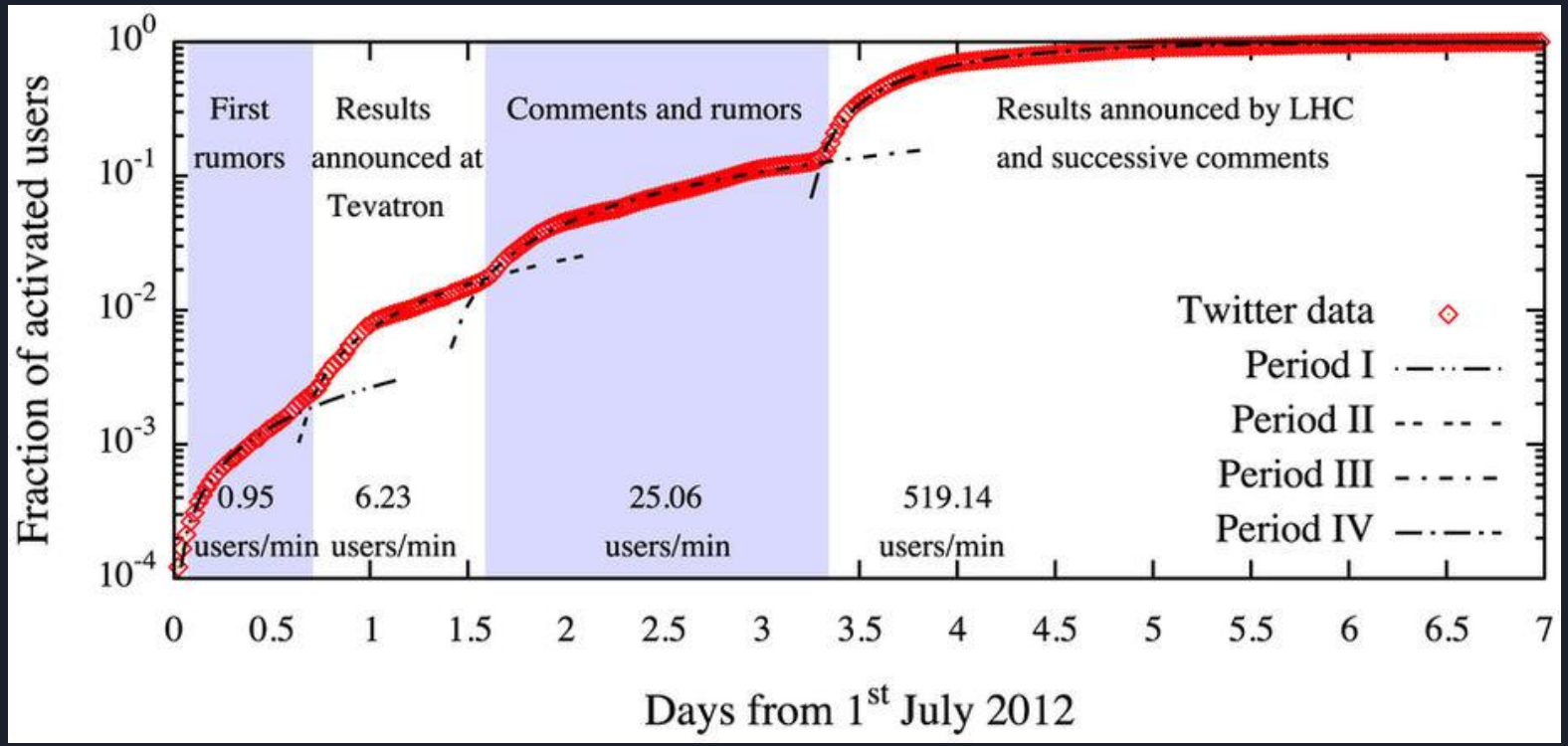
Work Achieved : Cont

- We have implemented those probabilistic formula for determination of active and non active user in the given network at any instant of time.
- Finally with the help of plotting library like matplotlib , We have plotted the result obtained i.e graph between number of active user versus time.

Time Period

- **Period I:** Before the announcement on 2nd July, there were some rumors about the discovery of a Higgs-like boson at Tevatron;
- **Period II:** On 2nd July at 1 PM GMT, scientists from CDF and D0 experiments, based at Tevatron, presented results indicating that the Higgs particle should have a mass between 115 and 135 GeV/c^2 (corresponding to about 123–144 times the mass of the proton)
- **Period III:** After 2nd July and before 4th of July there were many rumors about the Higgs boson discovery at LHC
- **Period IV:** The main event was the announcement on 4th July at 8 AM GMT by the scientists from the ATLAS and CMS experiments, based at CERN, presenting results indicating the existence of a new particle, compatible with the Higgs boson, with mass around 125 GeV/c. After 4th July, popular media covered the event.

Graph in Research Paper



Formula Used

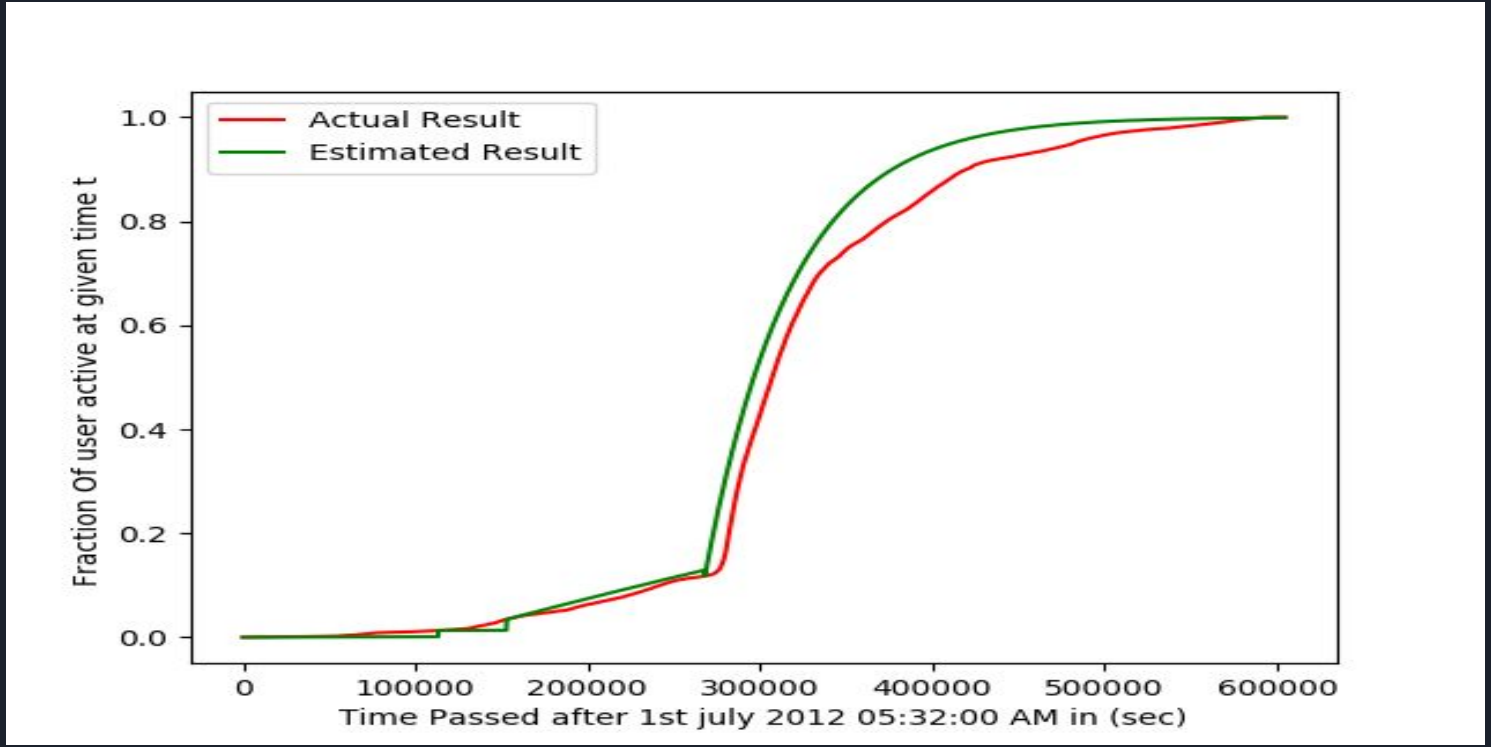
$$A^*(t+\Delta t) = A^*(t) + \lambda^*[N - A^*(t)]\Delta t, \quad (1)$$

$$\frac{da^*(t)}{dt} = \lambda^*[1 - a^*(t)], \quad (2)$$

$$a^*(t) = 1 - [1 - a^*(t_k)]e^{-\lambda^*(t-t_k)}, \quad (3)$$

WHERE
a*(t) = Fraction of Active Users
a*(tk) = Fraction of initial active users at the start of time period
T = Time
Tk = Initial time of give time period
Lambda = Activation Rate

RESULT


Actual Result
Estimated Result
Fraction Of user active at given time t
1.0
0.8
0.6
0.4
0.2
0.0
0
100000
200000
300000
400000
500000
600000
Time Passed after 1st july 2012 05:32:00 AM in (sec)

Code Files

Network.py :- It is used to create our estimated graph using the given formula with the help of matplotlib

Dataset_beautifier.py :- Extracted the relevant information from the given dataset for various phases

Actual_result.py :- It extracts information from the above obtained dataset which comes after refinement

Test.py :- This merges all the above files and plots the relevant graph

Team Members :
Group - 12

1. 1501CS14 : Ashutosh Parashar
2. 1501CS37 : Rijul Dhir
3. 1501CS16 : Ayush Mishra

Thank You