# *Realtime and Big Data Analytics Project Proposal*

## *Part 1. General Information*

Team Members:                         Ansa Ephraim, Aditya Garg, Ashutosh  Trivedi

Project Title:                           Analysis of Citation Network

Project Description:

We access citation database of conferences over a period of time. Some analytics we hope to derive are

1) Potential Hot Topic : Paper with the least citation distance from all the published papers.

2) Citation distance between 2 members on the same panel

3) Citation distance between member of panel and published author

## *Part 2. General Data Source Information*

| Data Sources<br><br>-Citation Network | Data Source Description | Data Size<br><br>MB |
|---|---|---|
| http://snap.stanford.edu/data/cit-HepTh.html | Paper citation network of Arxiv High Energy Physics Theory category | 36.3 |
| http://www.cs.cornell.edu/projects/kddcup/datasets.html | 29,000 hep-th papers with 1.7 gigs of data | 36.37 |
| http://people.cs.umass.edu/~mccallum/data.html | Simulated/Real/Aviation/Auto UseNet data | 1.02 |
| http://konect.uni-koblenz.de/networks/subelj_cora | Cora citation network datase | 2.05 |