WORKPLAN FOR THE STREAM PROJECT (HPC & DE)

ISHA AGARWAL

14075063

B.Tech Part – III

Problem Definition –

To identify the underlying community structure in the dataset and form a unified graph representation of the relations.

Input and Expected Output -

Input : 5 Twitter Datasets as input will be used. Each Twitter dataset has 9 different views in sparse matrix form.

Expected Output : k nearest neighbour graph (for different values of k) which is directed and unweighted which will show the community structure of the dataset.

Tools and Methods to be used -

- Unsupervised Learning will be used to find the communities.
- Cosine similarity
- K nearest neighbor sets
- SVD – Singular Value Decomposition (Factorization of the Matrix)

Week wise progress –

Week 1 : Study and analysis of the existing literature and doing research and collecting information on the required topic.

Week 2 : Exploring the dataset and community finding algorithms.

Week 3-4 : Implementation of the algorithm to produce the rank vector for all users.

Week 5-6 : Implementation of the SVD(Singular Value Decomposition) to get the k highest rank users.

Week 7-8 : Construction of the graph

Deliverables -

Graph which shows the communities structure of the dataset for different values of k.