Project Title: Analysis of Citation Network
Team Members: Ansa Ephraim, Aditya Garg, Aashutosh Trivedi
Summary by Aashutosh Trivedi

Paper 1:
The structure of scientific collaboration networks: M. E. J. Newman

- The author considers 2 scientists to be connected to each other if they have co-authored a paper together. The paper argues that most people who have written a paper together will know one another quite well. Thus it is a moderately stringent definition

- The author has constructed collaboration graphs for scientists in a variety of fields. The data come from four databases: MEDLINE (which covers published papers on biomedical research), the Los Alamos e-Print Archive (preprints primarily in theoretical physics), SPIRES (published papers and preprints in high-energy physics), and NCSTRL (preprints in computer science). In each case, he has examined papers that appeared in a 5-year window, from 1995 to 1999 inclusive. The sizes of the databases range from 2 million papers for MEDLINE to 13,000 for NCSTRL.

- The distance is calculated based on the co-authoring of the paper. 2 scientists are said to be at distance 1 if they have co-authored a paper together. Any authors who have worked whit these co-authors come at a distance 2 from the original author and so on.

- The paper takes a number of different scenarios such as the number of authors (and the errors such as same author giving different initials or different authors having same initials), Mean Papers per Author and Authors per Paper and number of collaborators.

- The paper also suggests that with a very large database and huge number of connections, the first degree connection takes up around 3/4 part of the database. Also the second connection group is much smaller than the first degree connection group.

- The author concludes that the maximum distance between 2 scientists is an average around 6 irrespective of the size of the database.

Paper 2:
Understanding Importance of Collaborations in Co-authorship Networks: A Supportiveness Analysis Approach

- This paper supports the theory that 2 authors co-authoring a paper means that one author supports the research of the other author. Thus it provides various such supportiveness measures. The definition of supportiveness here is given as "For an author a, the supportiveness from author b to a is used to measure how close the collaborations from b to a."

- The authors develop efficient methods to extract top n most supportive authors in co-authorship networks. They model the supportiveness ranking problem as a reverse k nearest neighbor (k-RNN for short) searching problem on graphs. To better model the co-authorship relation, we use hyper graphs in this paper.

- The authors use harmonic distance measure to calculate the support. It shows that support can differ in different directions based on the number of papers written in total and the number of papers co-authored. Thus the closeness and distance is calculated to be the harmonic mean of the contribution from one author to the other.

- The k nearest neighbor algorithm is used to find the author with maximum support. The algorithm finds the neighbor having the lease harmonic distance and thus concludes that it can be a neighbor. Here one author can have multiple or one nearest neighbor, depending on different factors.

- The authors also have tried to expand the algorithm to work for not just a single vertex but also a group of vertices. They have discussed the efficiency of both problems on large co-authoring networks and come up with very interesting solutions.