In Graphs and Networks, a node can be part of multiple communities. In simple words, let’s take Facebook friend list of a person ‘X’ and divide the list into communities such as school friends, gym friends, college friends, friends from work and so on. A friend can belong to one or more communities. This project aims to find out the overlapping communities in the given network.

**Input Data**

There are two main datasets used in this project.

1. Youtube.edgelist file that serves as network
2. Three community files: ground truth, neighborhood seed, 20 percent seed

The data sample for this project is obtained from <http://snap.stanford.edu/data/com-Youtube.html>

**Implementation**

The algorithm implemented to identify communities is BIGCLAM (Cluster Affiliation Model for Big Networks), which is a better version of AGM (Community Affiliation Graph Model). The steps involved are

1. **Factor Matrix Initialization**: A factor matrix of size N X K is initialized, where N is the number of nodes in the community and K is the number of communities in the groud truth file. The values are populated with three ways, random generation, as per 20 percent seed communities and neighborhood seed communities. The rest of the values in the factor matrix for the latter two methods is populated using conductance concept. Each node’s conductance with respect to the available communities is measured and the community with lower conductance value is assigned to the particular node.
2. **Factor Matrix Factorization**: Factor Matrix values are optimized using BIGCLAM V2.0 Algorithm.

Reference link for BIGCLAM V2.0 : <https://youtu.be/Y78Kugdq24I>

1. **Assignment of Communities**: Based on the factor matrix values, nodes are assigned to the communities if the value is greater than a threshold value.

The threshold value is 

In the end, recall is calculated for the detected and ground truth communities and the three methods of initialization are compared.