# Read Me
# Project: COMMUNITY DETECTION
# Unity Id: ndgandh2
# Paper 3 - Large-Scale Spectral Clustering on Graphs

**Description:**
The main goals of the project  is to implement Community Detection using Efficient Spectral Clustering on Large Scale Graphs Algorithm.

**Files Includes With This Project:**

1. P2_ndgandh2.R
2. Read Me
3. amazon.comm.small
4. installDependencies.R
5. amazon.graph.small
6. communities.txt

**Environment variable settings (if any) and OS it should/could run on :**
Operating System : Linux Ubuntu 14.04
OS Type: 64-bit
Processor: Intel Core-i5

**Softwares to be installed:**
I have implemented Community Detection using serial program in R.
To install R in Windows:

1. Download R from the following link:
   http://cran.r-project.org/bin/windows/base/
2. Double click on the downloaded R-3.1.1-win.exe file and follow the instructions.

Install RStudio in Windows:

1. Download RStudio from the following link:
   http://www.rstudio.com/products/rstudio/download/
2. Double click on the downloaded .exe file and follow the instructions.

**Instruction to run the program:**
Open any R development environment (RStudio) and type the following commands

1. source('path-to-file/installDependencies.R')
   This command installs all the packages required for the RScript to run

2. source('path-to-file/P2_ndgandh2.R')
   This command compiles the R file

3. ESCGR(graphFile = "path-to-graph-file/amazon.graph.small",supernodes = 600, iterations = 3, communities = 216)
   This command runs the ESCG-R function which implements the ESCG-R algorithm for given parameters and generates an output file “communities.txt” in the current directory.

**Read Me**
**Project: COMMUNITY DETECTION**
**Unity Id: ndgandh2**
**Paper 3 - Large-Scale Spectral Clustering on Graphs**

**Arguments:**

| | |
|---|---|
| graph-file | name of file whose communities are to be detected |
| supernodes | numerical value which defines the number of supernodes to be sampled for ESCG-R algorithm |
| iterations | numerical value which defines the number of times the supernodes are to be regenerated for efficiency |
| communities | numerical value which defines the number of communities to be detected |

**Instruction on how to interpret results:**
**Name of the output file :** “communities .txt”.
Output file is a text file containing the communities and the nodes in each particular community. Sample output file is show in the Sample output topic. Each line in the file indicates a community and each number in a line indicates the vertex in that community.

**Sample Input:**
Input is a graph file whose format is as follows:

```
 V1   V2
2501 2022
  0    1
  2    3
  2    4
  2    5
  2    6
```

First row gives the number of vertices and the number of edges in the graph. Rows after that contains the vertex id pair which has an edge between them.

**Sample Output:**
Output is a text file containing the community nodes which is as follows:

```
1079 1080 1081 1082 1083 1084 1085 1630
159 1312 1313
804 1921 1922 1923 1924 1925 2208
932 947
```

Each line indicates a community and each number in a line indicates the vertex in that community.

**Read Me**
**Project: COMMUNITY DETECTION**
**Unity Id: ndgandh2**
**Paper 3 - Large-Scale Spectral Clustering on Graphs**

**For Efficient Performance:**
I have implemented Efficient Spectral Clustering on Graphs (ESCG-R) algorithm for community detection where in Super nodes are selected randomly and regenerated for efficient supernode selection. By varying the parameters of the algorithms like supernodes and iterations, efficient community detection is possible and precision and recall can be effectively achieved w.r.t. ground truth communities.

**References:**


1. Large-Scale Spectral Clustering on Graphs. http://jialu.cs.illinois.edu/paper/ijcai2013-liu.pdf
2. http://cran.r-project.org/bin/linux/ubuntu/README
3. http://en.wikipedia.org/wiki/Singular_value_decomposition