

**Read Me**  
**Project: Anomaly Detection**  
**Unity Id: ndgandh2**  
**Paper 3 - Event Detection in Time Series of Mobile Communication Graphs**

**Description:**

The main goals of the project is to implement anomaly/event detection for time varying real world graphs.

**Files Includes With This Project:**

- |                  |                             |
|------------------|-----------------------------|
| 1. P4_ndgandh2.R | 4. clusterCoeff_anomaly.txt |
| 2. Read Me       | 5. egonet_anomaly.txt       |
| 3. Report        | 6. degreeofnode_anomaly.txt |

**Environment variable settings (if any) and OS it should/could run on :**

Operating System : Windows 8.1

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/P4_ndgandh2.R')`

This command compiles the R file

2. `findAnomaly(graphDir = "path-to-graph-dir",graphName = "Enron Graph")`

This command runs the ESCG-R function which implements the algorithm provided in Paper 3 for anomaly detection for given time series graph and generates anomaly detection graphs and output files in the current directory for the features namely,

- 1) Degree of the node
- 2) Clustering Coefficient of the node
- 3) Edges in the egonet of the node.

**Arguments:**

graphDir	path to the directory containing the time varying graphs
graphName	name of graph whose anomalies are to be detected

**Instruction on how to interpret results:**

**Name of the output files :** “degreeofnode\_anomaly.txt” , “clusterCoeff\_anomaly.txt”, “egonet\_anomaly.txt”

**Output File Format:**

The first line indicates the total number of anomalies that the algorithm detected, and the following lines are a list of anomalous time points. List contains all of the anomalous time points if there are fewer than 10, the top 10 if there are fewer than 100, or the top 10% if there are more than 100.

**Sample Input:**

Input are a series of directed graph files per day which indicates the communication between two nodes on a particular day whose format is as follows:

```
107  13
22   3
103  3
106  101
```

Rows contains the vertex id pair which has an edge between them indicating the text communication between two nodes.

**Sample Output:**

Output is a text file containing anomalous nodes which is as follows:

```
4
3
91
466
427
```

Here, firstline “4” indicates the number of anomalies and the lines following it indicates the time window starting with the given day in the times series which are anomalous.

**References:**

1. Akoglu, Leman, and Christos Faloutsos. "Event Detection in Time Series of Mobile Communication Graphs." (n.d.): n. pag. Web.
2. "EgoNet." *Wikipedia*. Wikimedia Foundation, 11 Jan. 2014. Web. 01 Nov. 2014.
3. "Clustering Coefficient." *Wikipedia*. Wikimedia Foundation, 24 Oct. 2014. Web. 01 Nov. 2014.
4. "Package 'egonet'." (n.d.): n. pag. Web.
5. "Home - RStudio." *RStudio*. N.p., n.d. Web. 01 Nov. 2014.