# Social Network Analysis
# Assignment 1

Aarthi Reddy

January 22, 2018

DEADLINE: 8am on Feb $14^{th}$, 2018

## 1 Introduction

In this assignment you will be asked to perform some social network analysis on a dataset that is provided to you. You will load the dataset provided, manipulate the csv file, perform computations and display/visualize results.

## 2 Description of Dataset

The network was generated using email data from a large European research institution. We have anonymized information about all incoming and outgoing email between members of the research institution. There is an edge (u, v) in the network if person u sent person v at least one email. The e-mails only represent communication between institution members (the core), and the dataset does not contain incoming messages from or outgoing messages to the rest of the world.

The dataset also contains "ground-truth" community memberships of the nodes. Each individual belongs to exactly one of 42 departments at the research institute.

This network represents the "core" of the email-EuAll network, which also contains links between members of the institution and people outside of the institution (although the node IDs are not the same).

## 3 Objectives

You will perform the following tasks for this assignments:

1. (5 points) Title of app should be "FullName_Social_Network_Analysis"

2. (5 points) User should be able to select and load the two relevant files for this assignment using the app: email-Eu-core-department-labels.txt, email-Eu-core.txt

3. (5 points) Using any package or function, display any n connections from the file, "email-Eu-core.txt"; where n is some number input by the user.

4. (7.5 points) Programmatically compute the number of emails sent by each person; display this information in a tabular format.

5. (7.5 points) Programmatically compute the number of emails received by each person; display this information in a tabular format.

6. (10 points) Display up to 2-hop neighbors of the top 10 from (4) and (5).

7. (10 points) Assume that each email sent or received is a connection. Compute the degree centrality of each person. Display/visualize up to 2-hop neighbors of 10 people with the highest centrality. The degree centrality of a node(person) $i$, can be defined as the total number of nodes connected to node $n_i$. Also, color code nodes according to the department to which they belong.

8. (10 points) Assume that each email sent or received is a connection. Compute the betweenness centrality of each person. Display/visualize up to 2-hop neighbors of 10 people with the highest betweenness. Betweenness centrality, $C_B$ for a node $i$, can be defined as:
$$C_B(i) = \sum_{j \neq k} \frac{g_{jk}(i)}{g_{jk}}, \tag{1}$$
where $j$ and $k$ are other nodes. Also, color code nodes according to the department to which they belong.

9. (10 points) Display/visualize 2-hop neighbors of nodes with the top 10 indegree centrality. Color code nodes according to the department.

10. (20 points) Aggregate the emails sent per person, to the department level. After aggregation, you should have a new table that indicates the number of emails sent and received between each and every department. The table should have three columns. Column A, indicates the department from which emails are originating, Column B, indicates the department to which the emails are being sent, and Column C indicates the total number of emails sent from A to B. Display the table, and visualize the directed connections.

11. (5 points) In a few sentences describe your observations when comparing the visualizations from 7, 8 and 9

12. (5 points) 5 points are allocated to creativity

# 4   Grading

Requirements of your shiny app:

- All computation should be performed as part of the R script and displayed in your shiny app.

- You can assume that I have a copy of the two files . Do not send me a copy when you submit your assignment.

- I will be modifying the input files to reflect a different emailing scenario. Your shiny app should still be able to provide results to the questions/tasks.

- Do not load your jpg/png/pdf/csv or any other type of files. I just want the R scripts to create your shiny app. I will not download anything other than R scripts to run on my computer.

- If submitting multiple R scripts, please zip it up and upload zipped file.

- Be creative.

# 5 Expectations of student

The student is:

- Expected to complete all tasks listed above.

- Expected to submit this mini-project by 8am on day of deadline mentioned at the start.

- Expected to submit original work

- Be aware of the packages you are using to compute centrality and betweenness. Each package computes these metrics differently and hence the values might differ, but the top 10 or 20 should not change.