Bachelor of Information Technology

**DS6501: Social Data Analytics**

**Assignment 2: Social Network Analysis**

**BD21338_**

**Due date and time**

5 pm Friday 5th June 2020 (Week 13)

**Submission details**

Use the Drop Box feature I have set up called ‘Assignment 2 - Drop Box’ in the Assignment section of my Moodle site. You should submit your group’s analysis report as described in task 9 and in the Assignment Deliverables section of this document. Also include the R script in your submission that contains the commands you used to visualise and analyse the Facebook network. This script should be zipped with your report and uploaded into Moodle as a single file. **Only 1 submission per group is required.** Please include the names of all students within your group on the cover sheet of your report.

**Extensions**

Extensions of time will only be granted for students who have an acceptable documented reason for not completing the assessment by the specified due date.

**Grading**

This assignment is worth 30% of the total module.

The assignment will be marked out of 100.

To pass a grade of 50% must be achieved.

**Terms**

See details of terms in the Bachelor of Information Technology handbook.

**Group Work**

1. You will work in a self-selected group to complete this assignment. Group size should not exceed 3 students. Alternatively, you may choose to work alone on this assignment.
2. ***A group contribution form*** must be completed by your group and submitted with the assignment. You can find the form on my Moodle site in the Assignments section.

**Purpose**

This assignment requires you to work collaboratively in a group to visualise and analyse a network of Facebook users. The analysis will include describing links and nodes, assigning attributes to the network, simplifying the network by filtering into groups, detecting communities and calculating measures of centrality and density. The results of your analysis will be documented within a report, as described in the Tasks section below.

**Brief**

Your task is to analyse and visualise a network of Facebook users using social network analysis techniques.

**Tasks**

Create an R script that will perform the following:

1. Import two csv files that contains the links and nodes of a network of Facebook users. The files are called **FB\_Links** and **FB\_Nodes**, which you will find in the Assignment 2 section on Moodle. Create an igraph object based on these files using the setting of **T** for the **directed** parameter.
2. Inspect the attributes of the network using the **E** and **V** function and use the **as\_data\_frame** function to describe the nodes and links. Plot the network using the default settings and describe the main problem with this plot in terms of its readability.
3. Plot the network again using the following attributes settings; the colour of the nodes should be blue for male Facebook users and red for female (hint: use the **ifelse** function – see p23 of the **R Studio Student Guide** on Moodle), the node size should be set to value of the friend\_count attribute divided by 70 and the arrow.size attribute should be reset to an appropriate value that improves the clarity of the plot.  
     
   Describe how the network diagram has been improved and identify the node with the largest friend count by visually inspecting the network.
4. Using the **group** attribute, plot a series of networks where each plot contains only Facebook users that belong within a particular group. For example, one plot will display the network containing only Facebook users within group B. You should plot a total of 8 networks for this task - one network plot for each group. You should adjust the node size to a more appropriate value by dividing the friend\_count value with a smaller number than one used in task 3. Each plot should be displayed using the layout style of **layout\_with\_fr.**  
     
   By visually inspecting the network plot for group M, identify the largest node (the Facebook user with the highest friend count). By examining the number of links each node has in this plot, would you agree that this user is the most important node in this network?
5. Re-plot the network containing group F Facebook users so that only males within that group are displayed. Perform the same task again, but this time only plot the females in this group. Both plots should be displayed using the layout style of **layout\_with\_fr.**
6. Using the **cluster\_optimal** function, detect the communities in each plot you created in task 5. Both plots should be displayed using the layout style of **layout\_with\_fr.**
7. Simplify the original network containing all Facebook users so that only those nodes with a degree of greater than 10 are plotted. In this simplified network use the **hub\_score** and **authority\_score** functions to calculate these scores. Create two network plots – one showing the hubs and another showing the authorities. The node size in each of these plots should be 10 times the value of the hub and authority scores. Your plots should be appropriately labelled and displayed using the layout style of **layout\_with\_fr.**  
     
   By visually inspecting the network displaying hubs, identify the 5 most important female Facebook users.
8. For the simplified network created in task 7, calculate the betweeness (ignoring loops) of each node and edge density value for the network. Identify the node with the highest betweeness value in this network.
9. Prepare an analysis report that documents the findings of your analysis and observations, as performed in the above tasks. Your report should present all the plots and results of the analysis generated by the functions you have used.

**Assignment Deliverables**

Each group should submit the following items zipped as a single file (only **one** submission per group is required):

1. An R script that contains all the commands used to analyse the network of Facebook users. The script should include comments that help to explain the tasks performed.
2. A word-processed Analysis Report that contains:

* An overview (executive summary)
* The results of each stage of the analysis performed in task 2 to task 8, including all plots and observations
* An interpretation of the results of the analysis, as described in the Tasks section above
* A conclusion section that summarises what you have learnt from analysing this network

**Marking Schedule**

**Student Id’s:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item** | **Mark Allocation** | **Max Mark** | **Student Mark** | **Comments** |
| 1. | Overview/ Executive Summary | **5** |  |  |
| 2. | Importation of the two csv files and the creation of a directed igraph network from these files | **5** |  |  |
| 3. | Network attribute inspection as described in task 2 | **5** |  |  |
| 4. | Re-plotting of the network with adjusted attribute settings as described in task 3 | **10** |  |  |
| 5. | Plotting a series of networks where each plot contains only Facebook users that belong within a particular group, as described in task 4 | **10** |  |  |
| 6. | Re-plotting the network containing group F Facebook users by gender, as described in task 5 | **10** |  |  |
| 7. | Detection of communities in each plot created in task 5, as described in task 6 | **5** |  |  |
| 8. | Simplifying the network and calculating hub and authority scores and displaying these details in two networks plots, as described in task 7 | **10** |  |  |
| 9. | Calculating measures of centrality and density, as described in task 8 | **5** |  |  |
| 8. | An analysis report that documents the findings of your analysis and observations, as described in Task 9 and Deliverables section including:   * Visualisations * Observations and Interpretations * Conclusion | **10**  **10**  **5** |  |  |
| 9. | Comments in R Script that documents the functions used to perform network analysis | **5** |  |  |
| 10. | Overall presentation of report (contents page, numbering, layout, grammar and punctuation etc.) | **5** |  |  |
|  | **Total** | **100** |  |  |

**Appendix**

The following details may help to provide a ‘back story’ for the Facebook network you are analysing.

**Groups in the Facebook Network (FB\_Nodes.csv)**

**B – Book Club**

**C – College**

**F – Family**

**G - Graduate School**

**H - High School**

**M – Music**

**S – Spiel**

**W - Work**