University of British Columbia Okanagan
COSC 421 / 521 Network Science (W2023)
*Assignment 1 – **45 points***

Due by Friday, September 22, 2023, 11:59 pm Kelowna time. Late submissions will not be accepted. You must upload a single pdf file. Feel free to ask the TA (during the lab) or the instructor (during office hours) for help with the assignment.

Your assignment should look like a professional report with your name and ID on the first page, pages be numbered, questions answered in the correct order. Graphs must be of at least one-quarter page size. Provide the code, numerical and graphical results within each question and NOT IN AN APPENDIX. Use MSWord or Latex to write your answers using complete sentences in at least 11 point font size.

Install and upload igraph in the current session if needed:

```
> install.packages("igraph")  # Put the package name in quotes!
> library(igraph)  # Do not put the package name in quotes!
```

Data: You will collect your own data from this assignment. Students may work in groups of **two** students maximum. If you are working with another student, both students should submit the same document. Both students must submit the assignment. Both students will get identical marks. This is to avoid any confusion with canvas declaring "no submission". Write the name of your partner on the report if working with one.

Data Collection: This exercise is about the flights network of the following Canadian airports with the international airport codes as shown in the following table.

airport	**code**
Victoria	YYJ
Vancouver	YVR
Abbotsford	YXX
Kamloops	YKA
Kelowna	YLW
PrinceGeorge	YXS
Cranbrook	YXC
Edmonton	YEG
Calgary	YYC
Fort McMurray	YQU
Regina	YQR
Saskatoon	YXE
Brandon	YBR
Winnipeg	YWG
Hamilton	YHM
Toronto	YYZ
Kitchener	YKF
Ottawa	YOW
ThunderBay	YQT

```
Montreal          YUL
QuébecCity        YQB
Fredericton       YFC
Greater Moncton   YQM
Halifax           YHZ
Charlottetown     YYG
StJohns           YYT
Stephenville      YJT
DawsonCity        YDA
Whitehorse        YXY
Yellowknife       YZF
```

Task 1: Enter the airports' name and code data in a text or a csv file and load the data in R as a data frame using one of the following.

```
nodes <- read.csv("filename.csv" , header=TRUE)
nodes <- read.table("filename.txt",header=TRUE)
```

Task 2. Our definition of a flight link between two airports is "two airports are connected if there is a direct (non-stop) flight between the two (either way) today". You will collect data using information from the internet. An airport's home page contains a list of incoming (arrivals) and outgoing (departures) flights for the day (typically for the next 24 hours). You can collect data using either arrivals or departures. The flights are directed but we count a connection either way.

Go to each of these airports' home pages one-by-one and find if there is a connection (edge) between the given airport and other 29 airports in the list. Store the edge information in two columns (with headers `airport1 airport2`) where each data row contains the names of two connected airports.

Use the airport names exactly as they appear in your nodes list in Task 1. Don't record a connection between airports a and b twice if there are more than one flights. Store the data in a text or a csv file with a name of your choice and load the edge list in R by
`edges <- read.csv("filename.csv" , header=TRUE)`

Task 3. Use graph_from_data_frame to create a graph object

```
airports <- graph_from_data_frame(edges, nodes, directed=FALSE)
vertex_attr(airports) <- list(name = nodes$code)
```

Answer the following questions.
**Clearly state in your report the date of data collection

Question 1. How many nodes and edges are in this graph? (4 points)
Question 2. Plot the network graph using the airport names or codes as node labels. Try using some layout options to get a better plot than using the default layout. Your plot should cover at least half the page. (10 points)

Question 3. Compute and report the degree of the airports in your network and the mean degree. Which two airports are the most connected and which two are the least connected? (10 points)
Question 4. Save the degree sequence in a vector and make a histogram of the degree distribution. (10 points)
Question 5. Get and report the adjacency matrix of your network and save it as a matrix object. Check that the adjacency matrix is symmetric. (10 points)

What to submit

- The data file (csv file) that contains the list of edges in task 2 above. Your assignment will not be graded without this file
- The pdf file/report that contains answers to the 5 questions listed above. **Clearly state in your report the date of data collection**

**For following instructions and completing all the tasks (1 points)

Some R commands for igraph usage

```
vcount(graph.name) # count of the vertices
ecount(graph.name) # count the number of edges
plot(graph.name) # to plot the network graph
V(graph.name) # to print the vertex list
E(graph.name) # to print the edge list
degree(graph.name) # to print the degree sequence for undirected graphs
deg <- degree(graph.name) # store the degree sequence as a vector
mean(deg) # compute the mean degree value
degree(graph.name, mode = "in") # in degree sequence for directed graphs
degree(graph.name, mode = "out") # out degree sequence for digraphs
get.adjacency(graph.name) # to print the adjacency matrix
A <- as.matrix(get.adjacency(graph.name)) # save the adjacency matrix
isSymmetric(A) # to check symmetry
plot(g, vertex.size=8, edge.arrow.size=0.5, vertex.label.cex=1)
```