Homework1

Marco Primo And Pedro Ferreirinha
April 23, 2022

1 QUESTIONS AND ANSWERS

1.1 QUESTIONS

1 - Show the degree of each node and make a plot of its (normalized) degree distribution.
2 -Calculate the diameter and the average path length of the network.
3 - Calculate the local clustering coefficient of each node and the average local clustering coefficient of the entire network
4 - Calculate the (normalized) betweenness centrality and closeness centrality of each node.

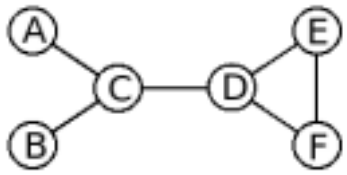

Figure 1.1: network represented by the following graph:

1.2 ANSWERS

1 -The nodes in Figure 1.1 have the following degree A→{1} B→{1} C→{3} D→{3} E→{2} F→{2}

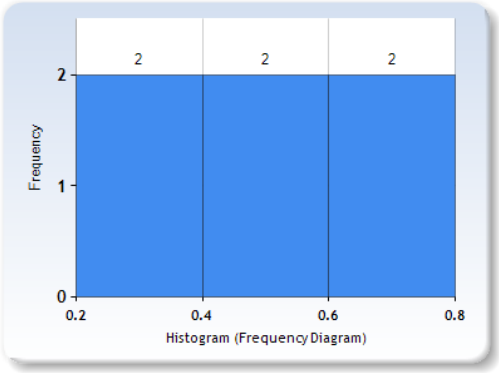

Figure 1.2:

2 - Using matrix Table 1.2 which gives the distance between all nodes is possible to calculate the Diameter and the average path length. The Diameter is then → 3 and the Average path length is 58/30 → 1.93333333333 .

–	2	1	2	3	3
2	–	1	2	3	3
1	1	–	1	2	2
2	2	1	–	1	1
3	3	2	1	–	2
3	3	2	1	2	–

Table 1.1: Distance matrix

3 - average local clustering coefficient of the entire network is 1/6 *(1/3+1+1) →0.39 and the local clustering coefficient of each node is:

C_A 0

C_B 0

C_C 0

C_D 0.333

C_E 1

C_F 1

4 - The closeness centrality of each node is:

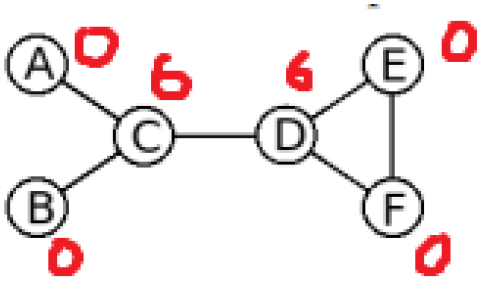

Figure 1.3:

normalized betweenness centrality is = 0.6 0.6

C'_A 5/(2+1+2+3+3) = 0.45

C'_B 5/(2+1+2+3+3) = 0.45

C'_C 5/(1+1+1+2+2) = 0.71

C'_D 5/(2+2+1+1+1) = 0.71

C'_E 5/(3+3+2+1+2) = 0.45

C'_F 5/(3+3+2+1+2) = 0.45

2 GLOBAL CLUSTERING COEFFICIENT

The global clustering coefficient (or transitivity) is based on triplets of nodes. A triplet is three nodes that are connected by either two (open triplet) or three (closed triplet) undirected ties.

The local clustering coefficient of a vertex (node) in a graph quantifies how close its neighbours are to being a clique (complete graph). The network average clustering coefficient is the average of the local clustering coefficients of the all the vertices.

According to our search, it's worth noting that this metric places more weight on the low degree nodes, while the transitivity ratio places more weight on the high degree nodes.

This is due to the fact that the global clustering coefficient gives an indication of the whole network, while the average of local clustering focus on the proportion of links between the vertices within its neighborhood. Meaning the high degree nodes will be much more important (and have more weight) in the global clustering coefficient.

3 QUESTIONS AND ANSWERS USING GEPHI

1. What is the number of airports and flights in the network?

a) 3147 airports and 66679 flights, number of nodes and edges after importing the file and not merging.

2. On average, an airport has how many outgoing flights? And to how many different airports?
   a) each airport has in average 21.054 outgoing flights - importing the file and not merging, filters intra-edges select out-degree and calculate the average degree.

      each airport has in average 11.529 outgoing flights to different airports - merge the network through sum and filter intra-edges select out-degree and calculate the average degree.

3. What is the diameter and average path length of the network?
   a) the network diameter is 13 and the average path length is 3969 - calculate with statistics

4. What is the pair of airports with more flights between each other (and how many flights)?
   a) Chicago O'Hare International Airport - Hartsfield Jackson Atlanta International Airport -> 39 paths, import file merging through sum, going into data laboratory and searching the ones with more weight.

5. List the top-3 of the airports that have flights to the highest number of other airports.
   a) Frankfurt am Main Airport (239), Charles de Gaulle International Airport (237), Amsterdam Airport Schiphol (232). This was done by importing and merging, calculating out-degree and seeing the ones that had the highest values out of all the node.

6. List the top-3 of the airports with highest normalized betweenness centrality.
   a) Los Angeles International Airport (0.0851819763708826), Charles de Gaulle International Airport (0.07194737632863432), London Heathrow Airport (0.061904393935910365). Import without merging, calculating avg path length with centralities normalized and check the top 3 with highest betweenness centrality.

7. Consider Ted Stevens Anchorage International Airport. What is its global ranking in terms of betweenness centrality and out-degree? Can you explain the discrepancy? Indicate another airport with the same kind of behavior (high betweenness centrality but relatively low out-degree).
   a) 8th place in betweenness centrality and 272nd in out-degree. the discrepancy is based on this airport being used in quite some stopouts. Faa'a internation airport also suffers from the same problem.

8. List the top-3 of countries with the highest number of airports.
   a) United states (525), canada (204), china (173). Partition by country and checking which ones have the highest number.

9. List the top-3 of airlines with the highest number of flights.
    a) Ryanair (2484), American Airlines (2351), United Airlines (2177). Partition by airline and checking which ones have the highest number.

10. What is the number of domestic flights inside USA ?
    a) 10487. Import file without merging. Partition by country (America) and check how many edges there are.

11. How many airports in China fly to at least 50 other airports?
    a) 21. Import file merging with sum, partition by country (China), add subfilter with the range of out-degree at least 50 and check how many nodes remain.

12. How many flights are there between Brazil and Portugal?
    a) 24 flights. Import file without merging, filter by intra edges between Portugal and Brazil and see the number of edges.

13. Consider a network formed only by Ryanair flights. What is the number of nodes and edges of its giant component? Considering only this giant component, what is the most important airport in terms of closeness centrality?
    a) Import the file without merging, partition by airline (FR), filter the giant component of the partition and see the number of edges (2484) and nodes (176). Calculate closeness centrality and check which one is higher, in this case it's London Stansted Airport with 0.7675438596491229.

14. How many airport are reachable from Francisco de S´a Carneiro Porto Airport in 1 flight? And in at most 2 flights? And in at most 3 flights?
    a) Import the file with merging, ego network with start in node id 1636, in 1 flight (with depth 1) we can reach 61 airports. In 2 flights we can reach 75 airports and in 3 flights we can reach 2376 airports.

15. Create an image showing the flight network between american and canadian aiports with more than 100 destinations in the global network. The size of the nodes should reflect the global betweenness centrality, and their colors should be different for each time zone. Nodes should be labeled with the city name. Try to make your image as comprehensible and aesthetically pleasing as possible.
    a) Import the file without merging, filter with intersection between the intra edges of United Stated and Canada and the range of out-degree at least 100. The size of the nodes reflect upon the betweenness centrality and their colors are different for each time zone (tz). Nodes are labeled with the city name. The obtained result is shown in Figures 4.2 - 4.3 - 4.4.

4 Erdos-Renyi Model

4.1 Was the plot what you were expecting? What is the shape of it? At what average degree values do you notice something happening?

Yes it was what we were expecting. The curve shape generated by the graph is very similar to the curvature shown in the teacher's slides for an different network. Figure 4.1 shows this similarity. The shape of the curve resembles the shape of the logarithmic function of the genus log(x) [1-20].

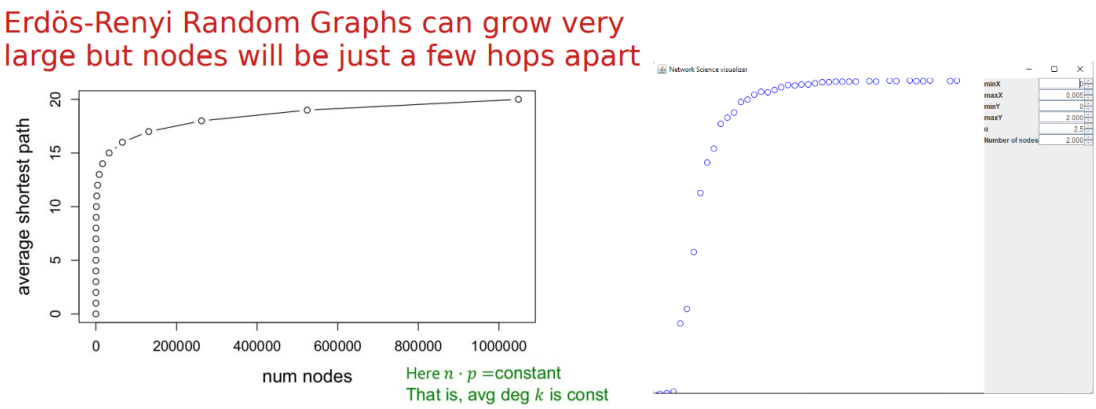

Figure 4.1: Erdos-Renyi Random

In fact it was noticed that when the average degree is between 0.9 and 1.2; around p=0.0005; An almost perpendicular growth of the numbers happens from us in the giant component

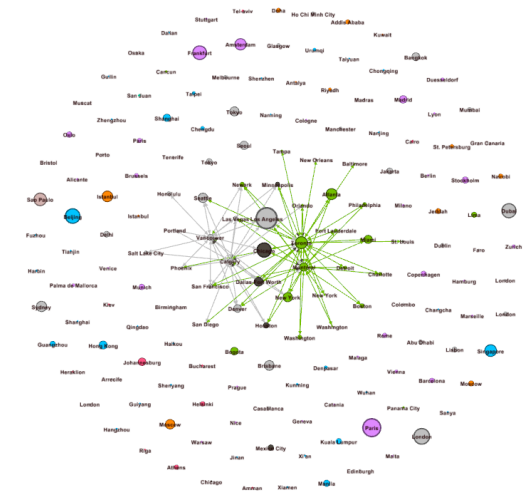

Figure 4.2: flight network between american and canadian aiports zoom=1

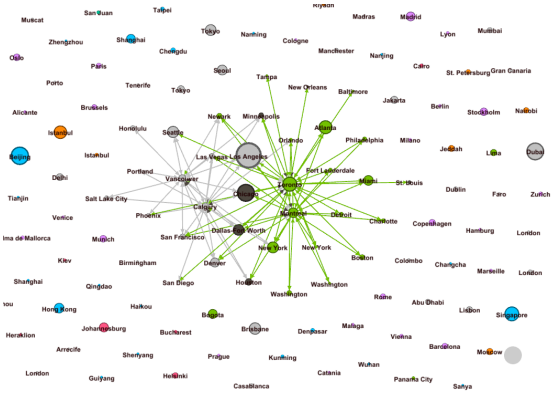

Figure 4.3: flight network between american and canadian aiports zoom=2

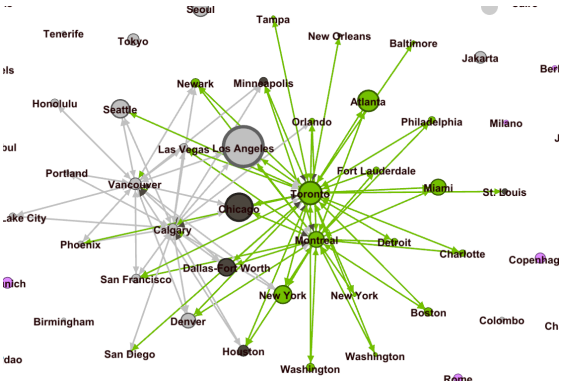

Figure 4.4: flight network between american and canadian aiports zoom=3