**Question 18**: How many personal networks are there?

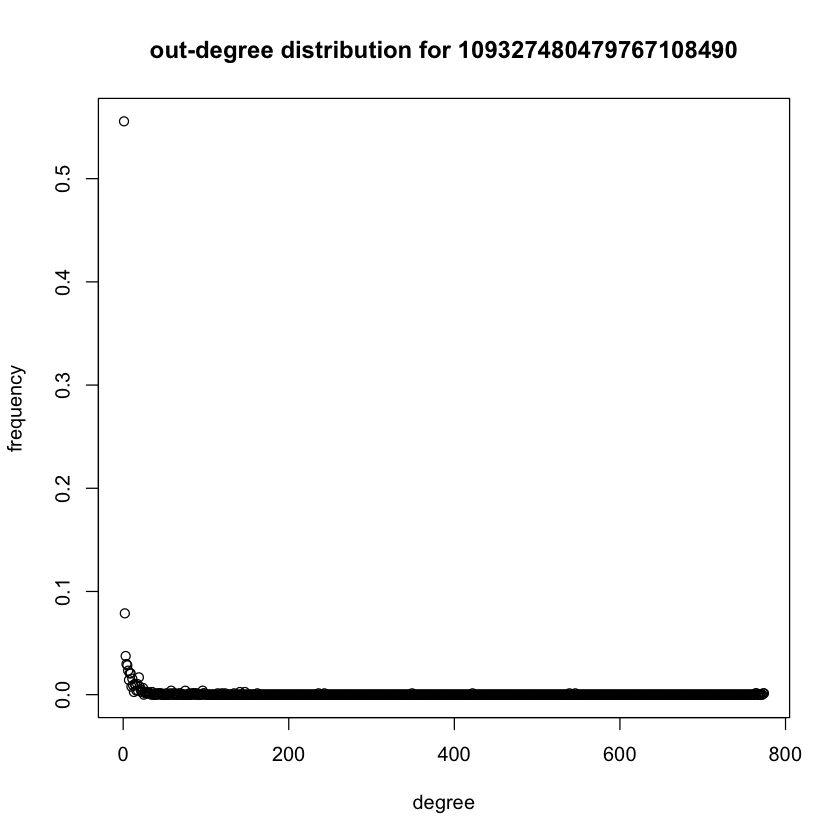
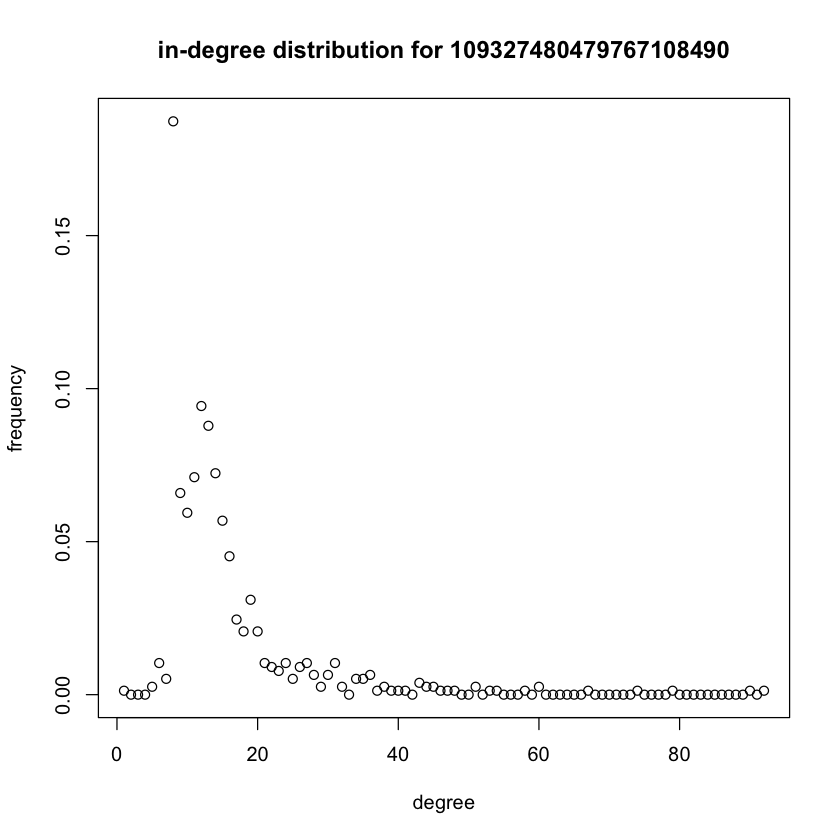
**Result:**

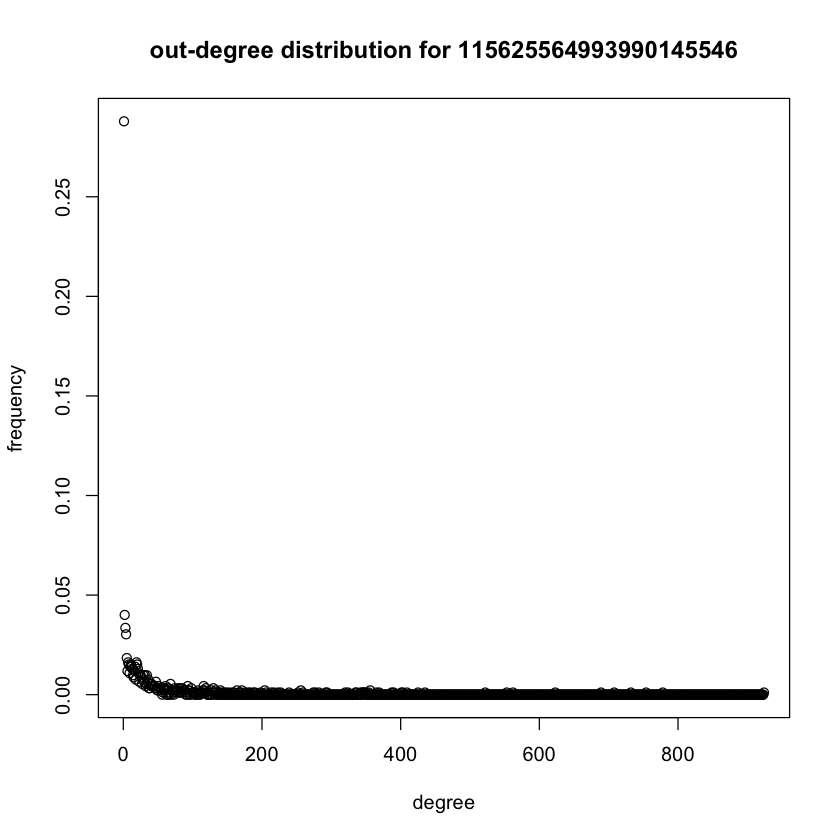
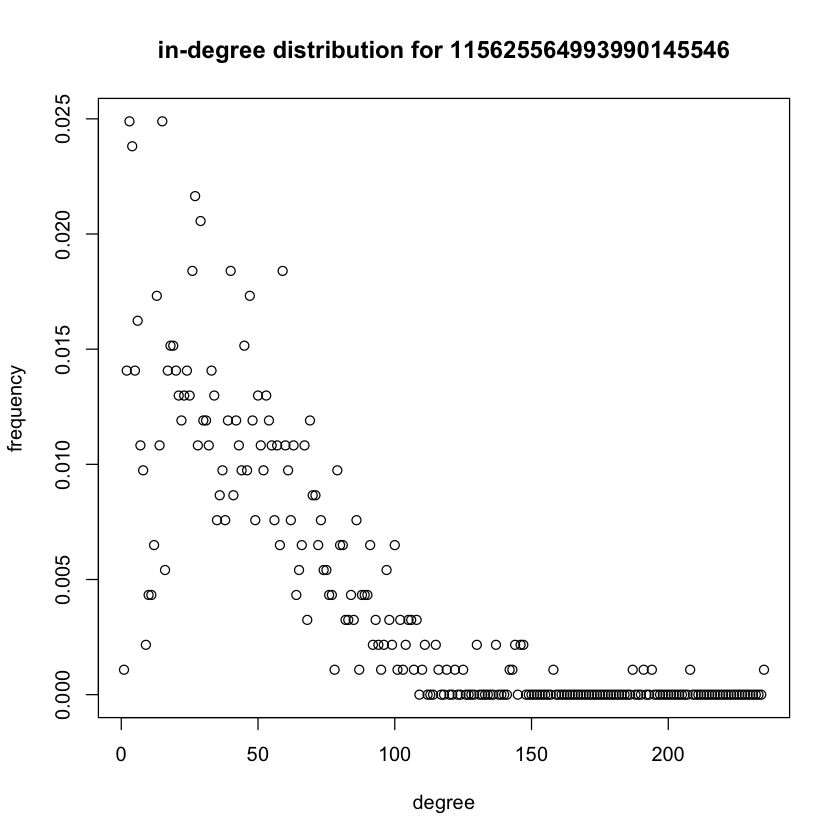


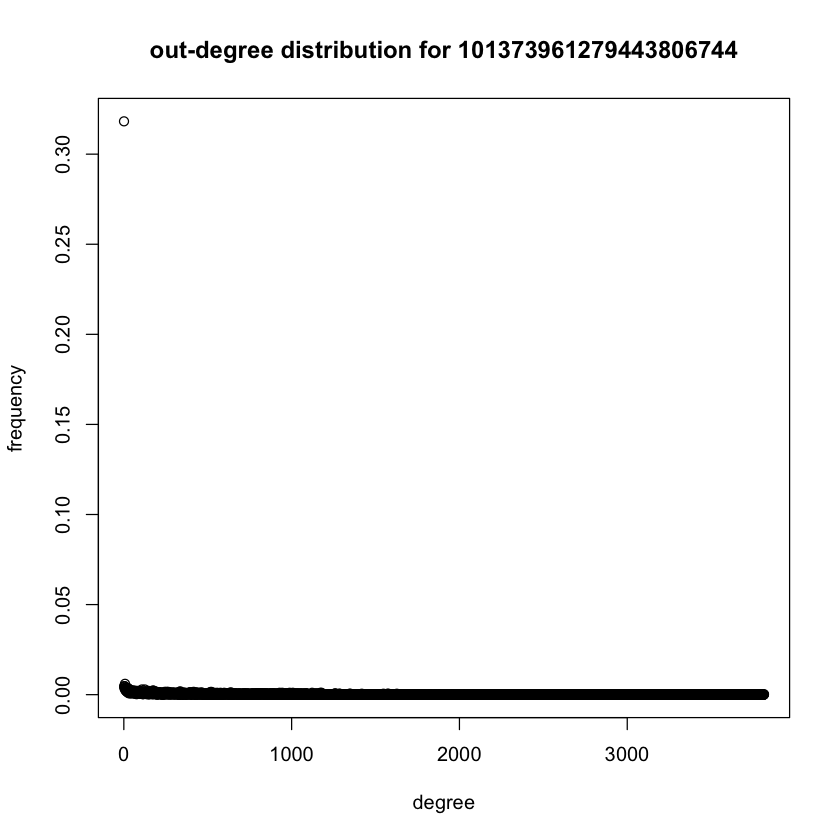
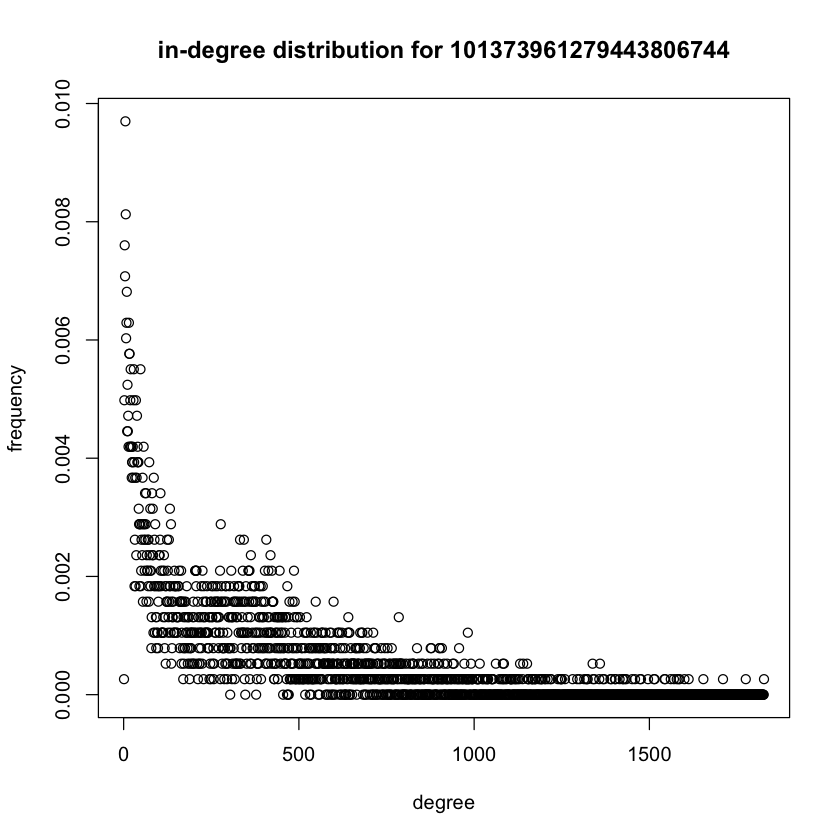
**Analysis:**

As result shown above, there are 57 personal networks who have more than 2 circles.

**Question 19**: For the 3 personal networks (node ID given below), plot the in-degree and out-degree distribution of these personal networks. Do the personal networks have a similar in and out degree distribution?







**Analysis:**

 If a network is [directed](https://en.wikipedia.org/wiki/Directed_graph), meaning that edges point in one direction from one node to another node, then nodes have two different degrees, the in-degree, which is the number of incoming edges, and the out-degree, which is the number of outgoing edges.

We plot the in-degree and out-degree distributions of the 3 required personal networks. They have similar part in both in and out degree distribution, and they vary in in-degree distribution.

For in-degree distribution:

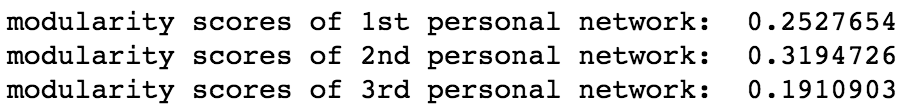
1. They differ in maximum frequency: first one has a maximum frequency around 0.18, second one has a maximum frequency around 0.025, third one has a maximum frequency around 0.01.
2. The number of nodes are different in 3 personal networks, the first one has a smallest minimum and the third one has a maximum number.
3. They are similar in the degree value that have maximum value of frequency, which is around 10.
4. They are similar in their shape. They are all experience an increase in frequency and then decrease in frequency with the increasing of degree, its maximum value are around degree =10.

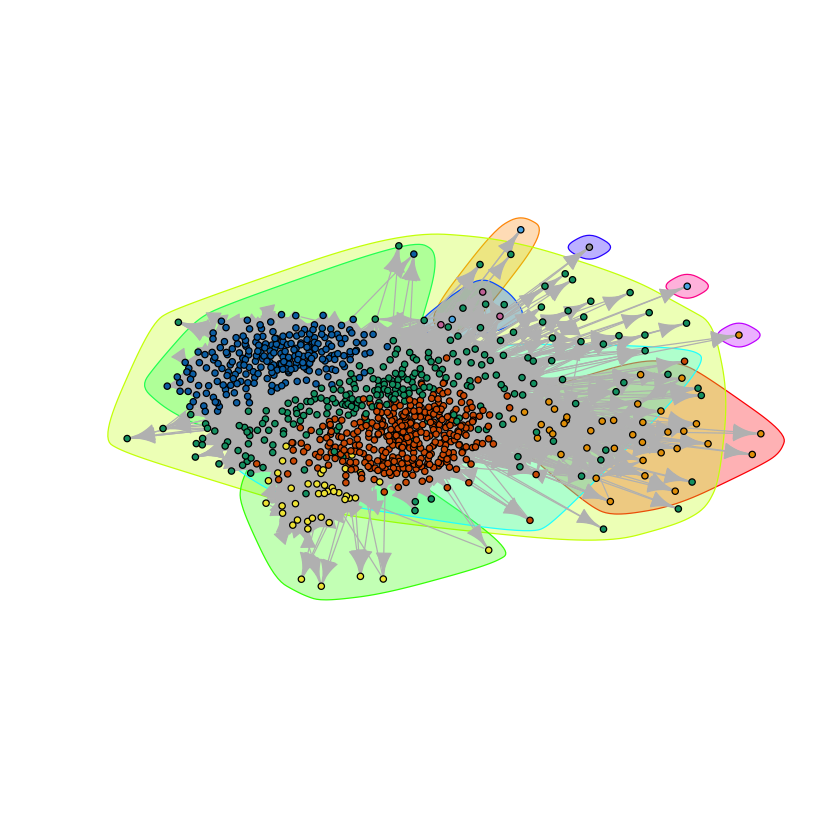
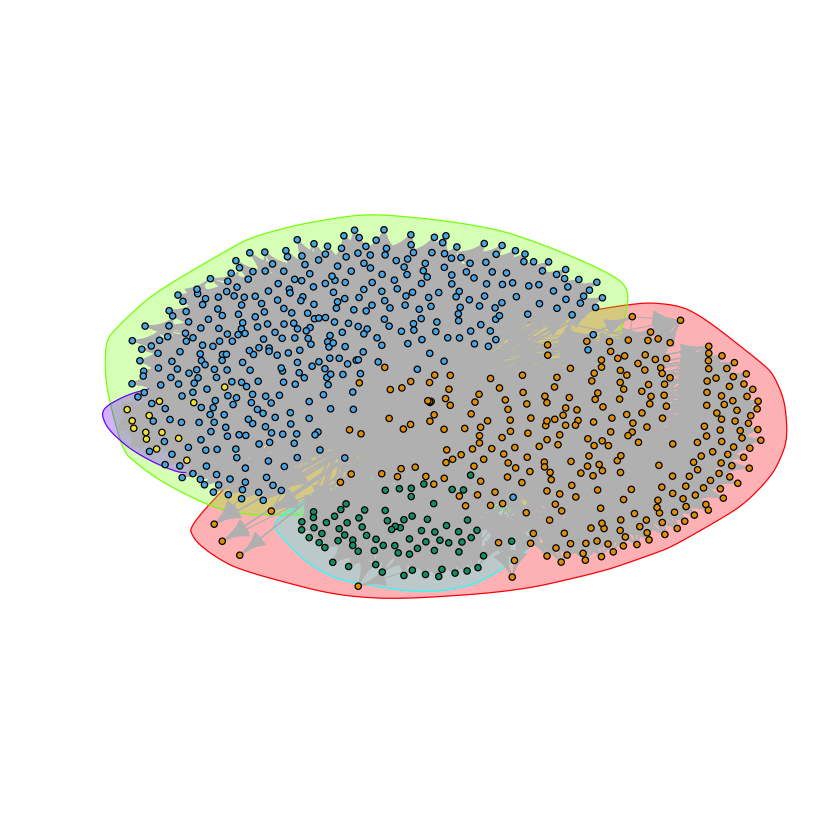
For out-degree distribution:

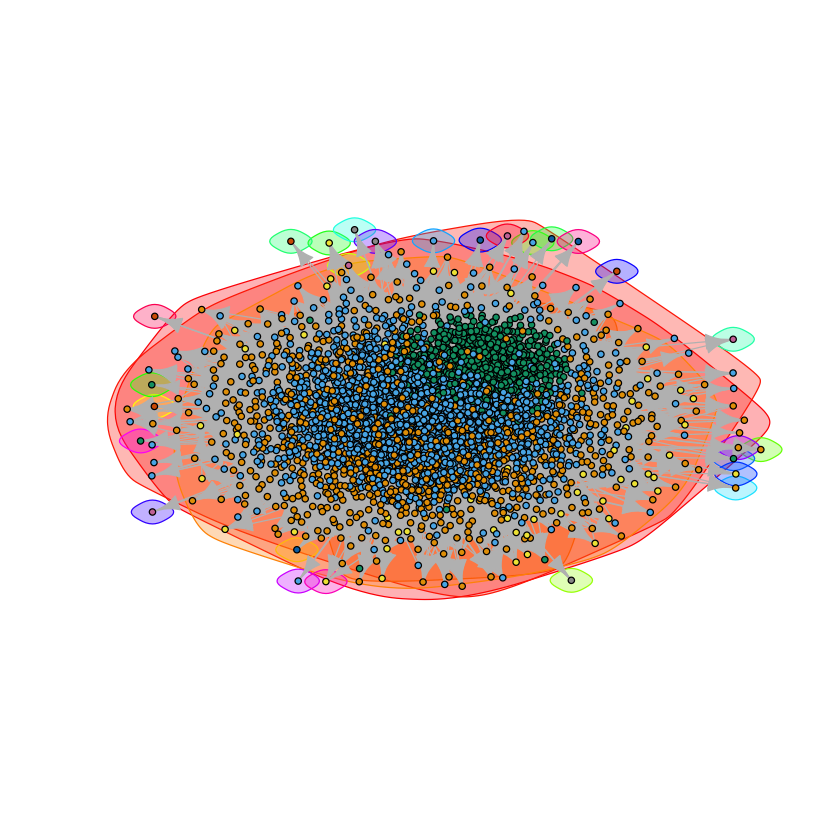
1. They all have a maximum value of frequency around 0.3-0.5 at degree=0.
2. They all decrease with degree increasing and equals to 0 when they decrease to 0.

**Question 20**: For the 3 personal networks picked in question 19, extract the community structure of each personal network using Walktrap community detection algorithm. Report the modularity scores and plot the communities using colors. Are the modularity scores similar? In this question, you should have 3 plots.

**Result:**







**Analysis:**

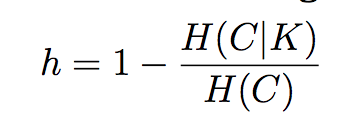
Modularity is one measure of the structure of [networks](https://en.wikipedia.org/wiki/Complex_network) or [graphs](https://en.wikipedia.org/wiki/Graph_(discrete_mathematics)). It was designed to measure the strength of division of a network into modules (also called groups, clusters or communities). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.

As we can see in result, the modularity scores are different in three cases. The 2nd personal network has highest value of modularity score and the 3rd personal network has lowest value of modularity score, which shows the 2nd personal network has strongest connections between nodes while the 3rd personal network has sparsest connections between nodes.

**Question 21**: Based on the expression for h and c, explain the meaning of homogeneity and completeness in words.

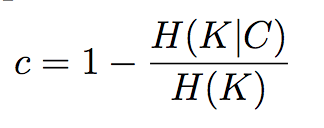
**Analysis:**

Homogeneity indicates every node in community should belong to a circle. In this case, H(c|k) describes the entropy circle given its community. H(c) describes the total entropy of the circle.



Thus h describes the degree of each community contains only nodes of a single circle in a network. With the increasing of h, the network’s community contains only members of a single circle becomes more likely. If the node within the same community have multiple circle information, h will become small given the same total entropy.

Completeness means that all nodes in a circle should be classified to same comminity. In this case H(k|c) describes the entropy of the community given the nodes’ circle information. H(k) describes the total entropy of community.



Thus, c describes the degree of all nodes of a given circle are assigned to the same community in a network. With the increasing of c, the network’s nodes of a given circle are assigned to the same community becomes more likely. If the nodes belonging to the same circle are in different communities, c will become small given the same total entropy.

**Question 22**: Compute the h and c values for the community structures of the 3 personal network (same nodes as question 19). Interpret the values and provide a detailed explanation.

**Result:**

For 1st personal network:

Homogeneity= 0.8518851, completeness=0.3298739

For 2nd personal network:

Homogeneity= 0.4518903, completeness= -3.4239623

For 3rd personal network:

Homogeneity= 0.003866707, completeness= -1.504238388

**Analysis:**

1st personal network has the large homogeneity and small completeness, it indicates that for each community in the 1st personal network, it’s most likely for its nodes in a single circle, while nodes belonging to the same circle are in different communities.

2nd personal network has a negative completeness. It shows that H(K|C)>H(K) in this case, which means there are some nodes belong to different circles.

3rd personal network has the small homogeneity and a negative completeness, it indicates that for each community in the 3rd personal network, it’s most likely for its nodes in different circles, while some nodes belong to different circles.