**ECE 232E - Project 2**

**Social Network Mining**

*Zhechen Xu (805030074), Yufei Hu (404944367), Qi Zeng (204946904), Liangkun Zhao (204947003)*

*May 2, 2018*

# **Part 1: Facebook network**

**1.1 Structural properties of the facebook network**

**Question 1:**

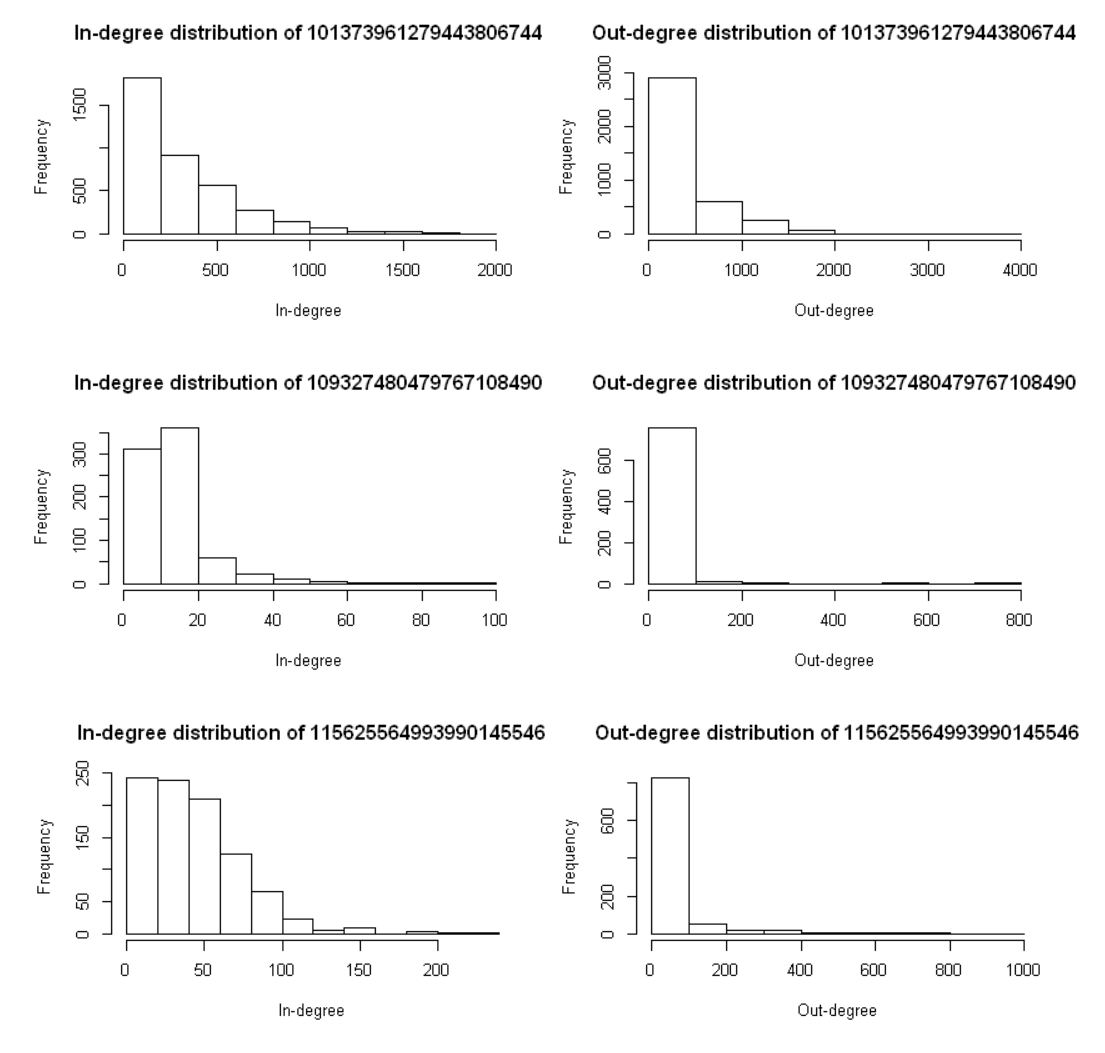
# **Part 2: Google+ network**

**Question 18:**

In this part, the properties of a Google+ directed network is fully explored. After loading and constructing the network, the community structures of some certain personal networks are extracted based on Walktrap community detection algorithm. Finally, the chosen networks are further analyzed in terms of homogeneity and completeness.

According to the simulation results, there are **58** personal networks that have more than 2 circles.

**Question 19:**

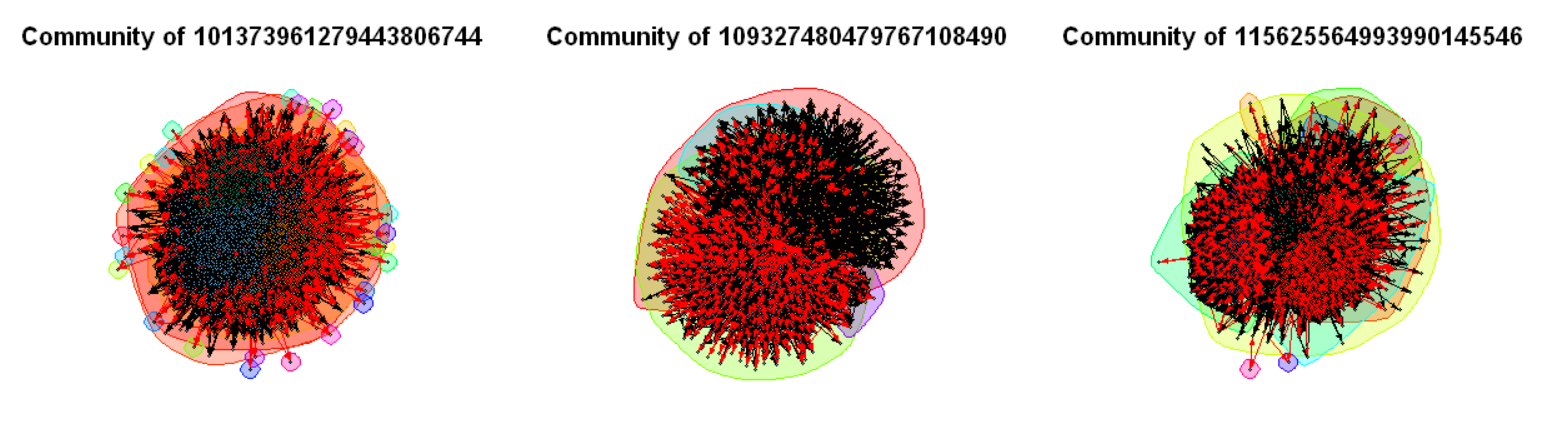


**Figure 1:** In and out degree distribution of the chosen personal networks

The degree distributions of the three chosen personal networks are quite different comparing to each other. The personal network for node *101373961279443806744* has much more and much bigger in and out degrees based on the histogram. The personal network for node *109327480479767108490*, on contrary, has the least number of in and out degrees and with the smallest scales. However, the majority of the nodes in the three personal networks all have relatively small in and out degrees.

**2.1 Community structure of personal networks**

**Question 20:**



**Figure 2:** Communities of the chosen personal networks

**Table 1:** Modularity scores of the chosen personal networks

|  |  |
| --- | --- |
| **Network** | **Modularity scores** |
| 101373961279443806744 | 0.191090 |
| 109327480479767108490 | 0.252765 |
| 115625564993990145546 | 0.319473 |

The modularity scores for the three chosen personal networks are quite different. It can be seen both from the Figure and Table above that the personal network of node *115625564993990145546* has the densest connections while *101373961279443806744* has the sparsest connections.

**Question 21:**

**About homogeneity:**

Based on the definition, homogeneity reaches its maximum when *H(C|K)* equals to 0 and it reaches its minimum when *H(C|K)* equals to *H(C)* whencircles do not have overlaps*.* The former situation implies the condition when each community only contains nodes belonging to one circle while the latter situation indicates that the community contains no extra information since the distribution of circles within each community is the same to the overall distribution of all circles. This means that homogeneity measures the degree to which the communities contain only nodes of a single circle.

**About completeness:**

Based on the definition, completeness reaches its maximum when *H(K|C)* equals to 0 and it reaches its minimum when *H(K|C)* equals to *H(K)* whencircles do not have overlaps. The former situation implies the condition when all the nodes of a circle are also in a same community. The latter situation indicates that each circle is represented by every community with a distribution same to the circle size. This means that completeness measures the degree to which each community contain all nodes of a given circle.

**Question 22:**

**Table 2:** Homogeneity and completeness of the chosen personal networks

|  |  |  |
| --- | --- | --- |
| **Network** | **Homogeneity** | **Completeness** |
| 101373961279443806744 | 0.003867 | -1.504238 |
| 109327480479767108490 | 0.851885 | 0.329874 |
| 115625564993990145546 | 0.451890 | -3.423962 |

It can be told from the table above that the communities of the personal network of node *109327480479767108490* appears to be less complete and more homogeneous. The scales of homogeneity and completeness seem to go in opposite direction, meaning increasing the homogeneity often leads to a decrease in completeness. Intuitively, when all nodes belonging to a single circle are included in a community, it is highly likely that the community also contains nodes from other circles, thus the community appears to be less homogeneous while more complete.

For the personal network of node *101373961279443806744*, the homogeneity is rather small. The reason is there are 3815 nodes in the original network while only 521 of them are tagged into circles. This causes the communities to contain many untagged nodes, leading to a decrease in homogeneity.

The completeness for *101373961279443806744* and *115625564993990145546* are both negative. This happens when circles have overlapped nodes, meaning some users gave their friends multiple tags. Then according to the definitions of completeness, *H(K|C)* will be bigger than *H(K)*, thus causing the completeness to be negative.