**6. Use personalized PageRank to study the overlapped communities structures**

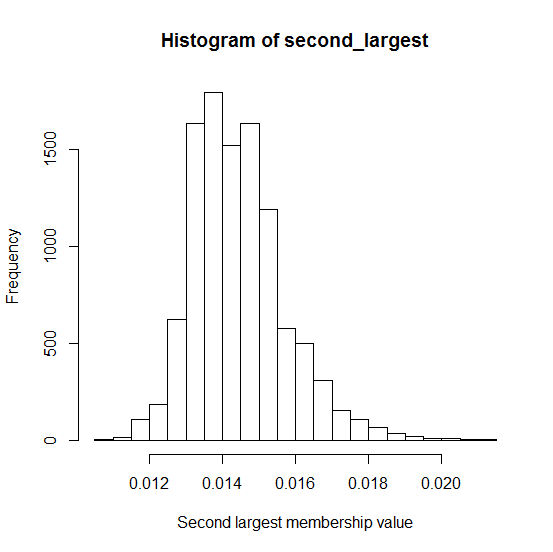
Firstly we used the community information we computed in problem 3, as well as functions in the package of netrw to solve this problem.

By generating the personalized page rank, we then find the visiting probability of each node in the giant connected component. Then we picked those top 30 nodes with largest visiting probability and calculated M associated with them, in which M and m are both matrix of n\_nodes\*n\_communities. Here n\_nodes is the number of nodes and n\_communities is the number of communities generated. Then the threshold was set to determine whether a node belonged to multiple communities. Here we just implemented it by finding the second largest community characteristic in each node vector and set the threshold so that when the value is larger than the threshold it means the node belongs to at least 2 communities.

For fast greedy algorithm, after testing we set the threshold to be 0.02, which yielded 12 nodes belonging to at least 2 communities.

Then we lower the threshold to 0.01 and it turns out to be 10487 nodes belonging to at least 2 communities.

The following figure shows the distribution of number of nodes that belong to at least 2 communities versus the threshold.



Figrue6.1: the distribution of number of nodes that belong to at least 2 communities versus the threshold

We can see from the above figure that when the threshold becomes larger, the number of nodes belonging to 2 communities becomes smaller.