**BigML 10-605**

**Homework 6: Approximate PageRank**

**Daniel Ribeiro Silva**

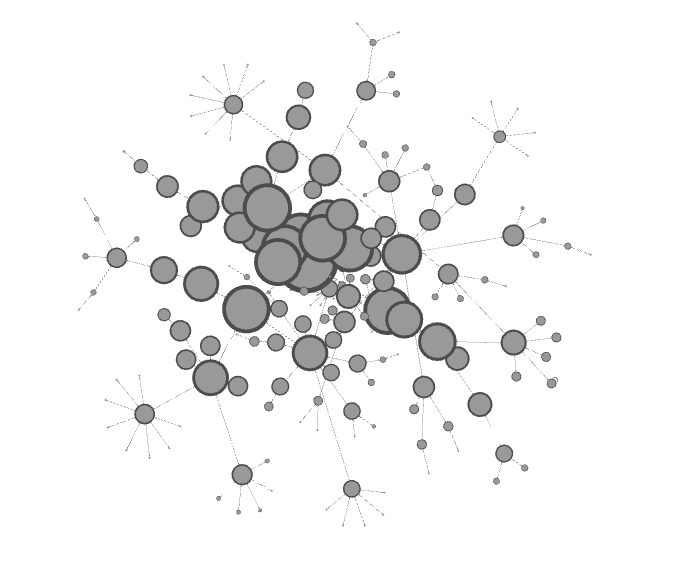
**drsilva**

**Question 1**

Seed: “Machine\_learning”

Nodes: 177

Edges: 357

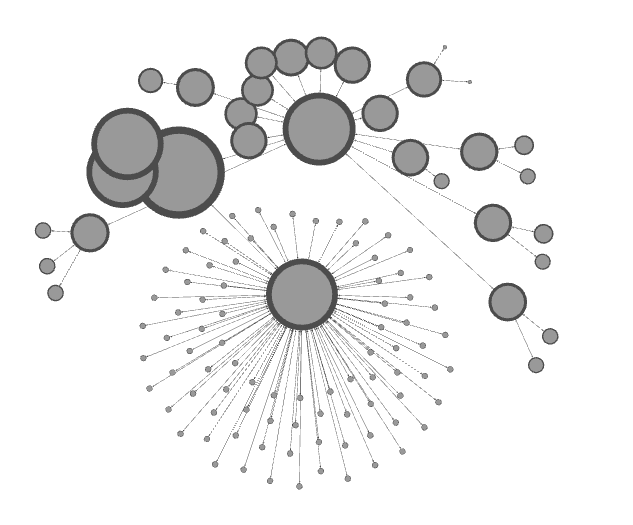


**Question 2**

Seed: “Hipster\_%28contemporary\_subculture%29”

Nodes: 118

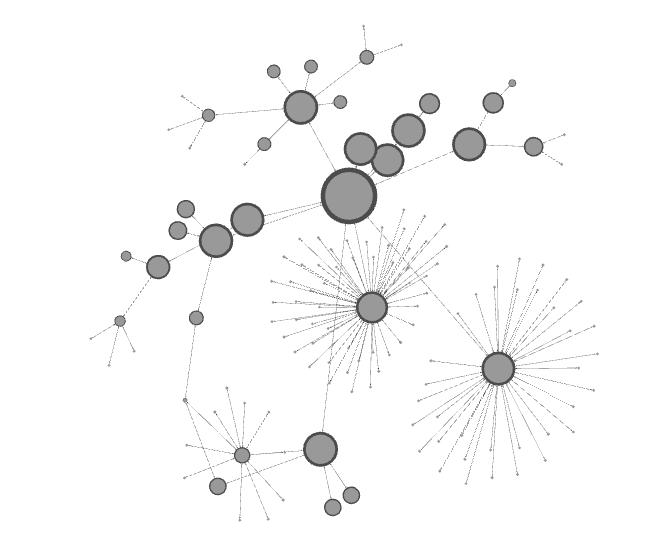
Edges: 234



Seed: “Topographic\_isolation”

Nodes: 144

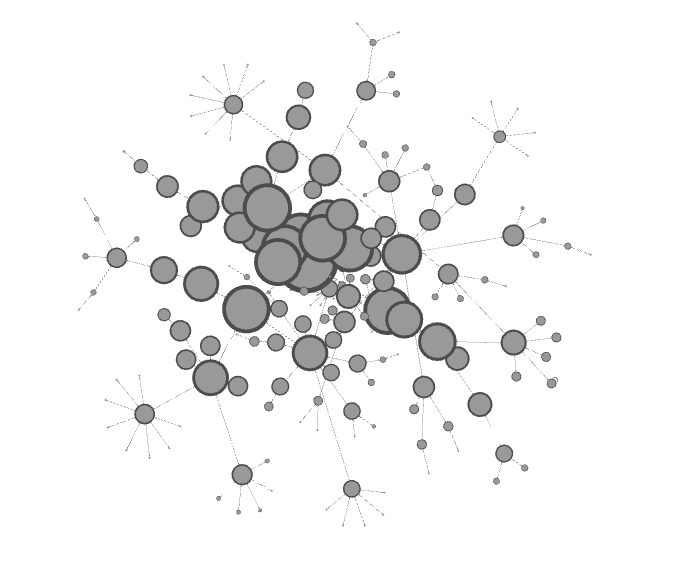
Edges: 288



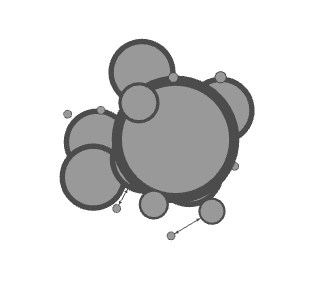
**Question 3**

In order to illustrate how sensitive the resulting graph is for each parameter, I run the script for a variation of each parameter, and it is compared to the “standard” graph (seed: “Machine\_learning”, , ).

Standard graph:



Variation 1: increase by a factor of 100: new (alpha remains the same)



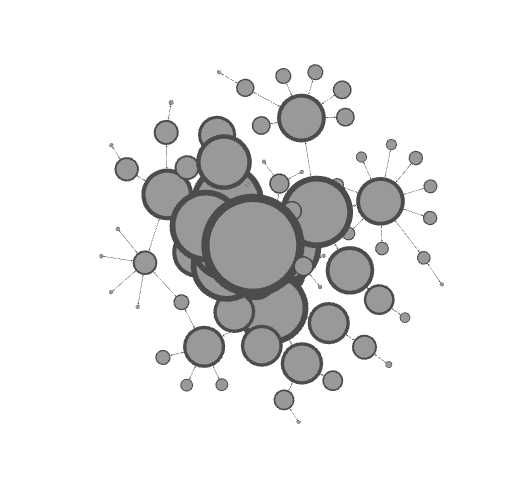
Variation 2: decrease by a factor of 100: new (alpha remains the same)



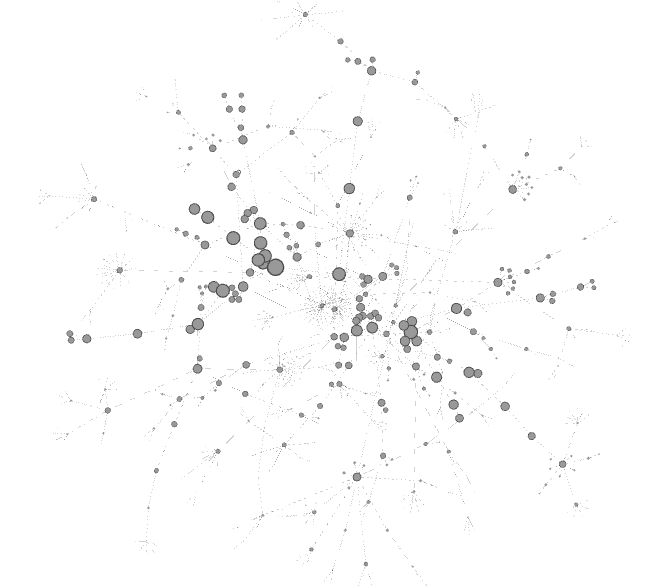
Here we observe that by increasing the value of , we get a smaller graph and by decreasing the value of we get a larger graph. The reason is that represents the threshold for which you start pushing residue from a node to its neighbors. If the threshold is small, then more nodes will push to its neighbors and the scores will be more distributed. On the opposite case, if the threshold is too large, then almost no nodes will have the chance to push into their neighbors, making the graph mostly concentrated around the seed and its direct neighbors.

Now we illustrate the variation of the parameter

Variation 3: increase in by 0.2: new ( remains the same)



Variation 4: decrease in by 0.2: new ( remains the same)



Here we observe that by increasing the value of , we get a smaller graph and by decreasing the value of we get a larger graph. The reason is that is associated with the probability of jumping to a neighbor node or staying at the same node. Larger alpha represents a higher chance of staying at the node. If we observe the Approximate PageRank algorithm, by increasing we decrease the amount of residue that is transferred to the neighbors. As a consequence, the graph will tend to be not very sparse and concentrated around the seed. On the other side, if alpha is small, then more residues are transferred to the neighbors, and so one. That means that the graph will spread more (less sparse) and as a consequence we’ll have a larger graph.

**Question 4**

I discussed with Jeff Gee solutions for optimizing the code