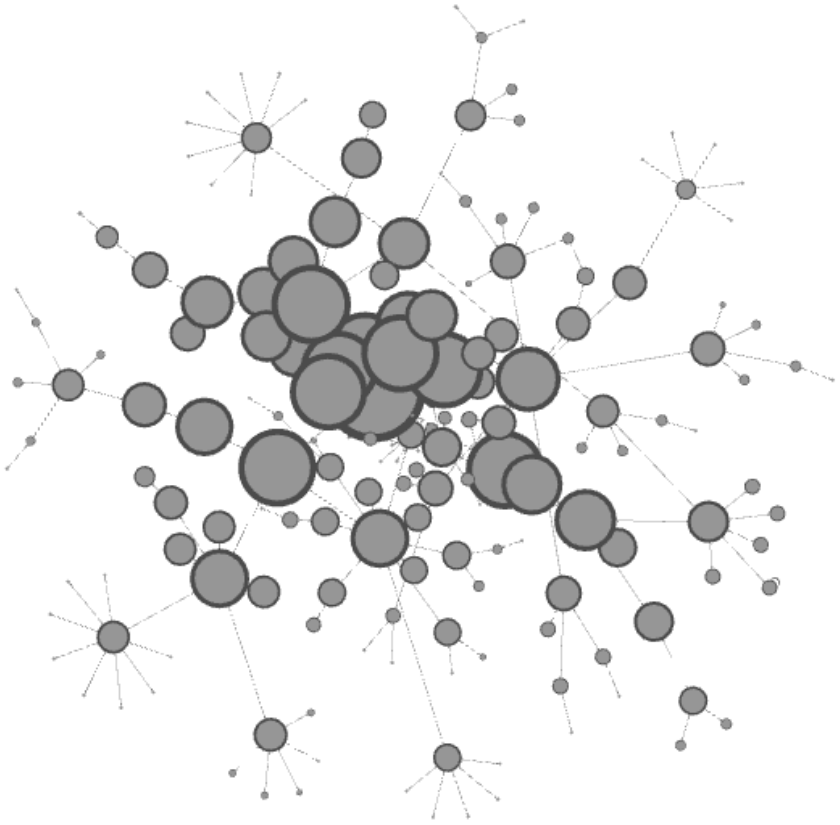
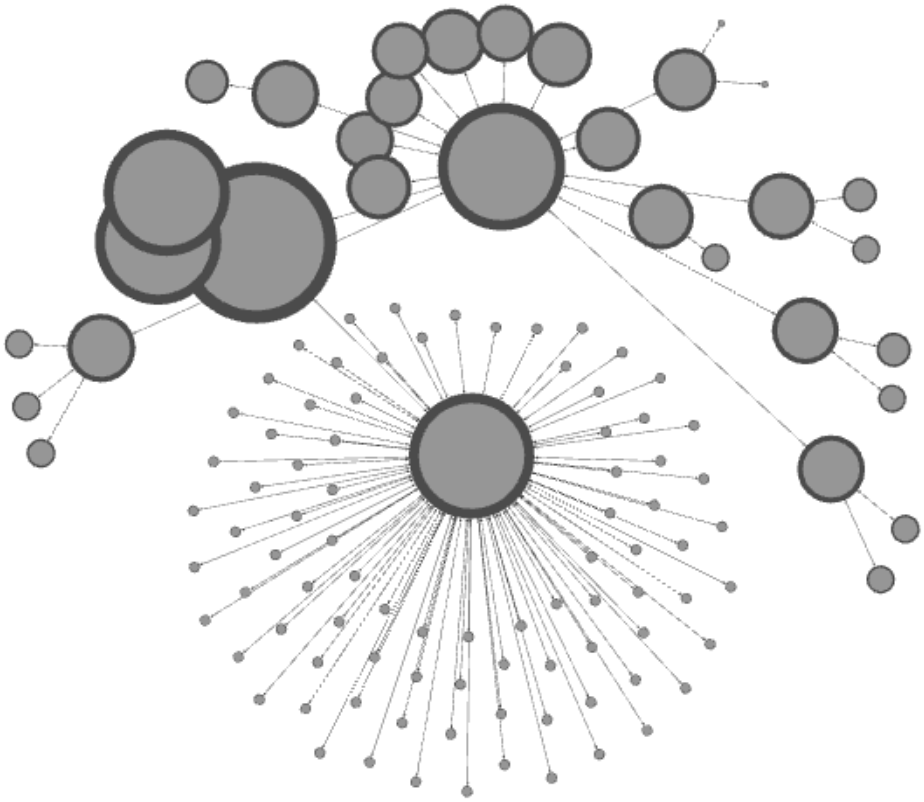
# BigML 10-605
# Homework 6: Approximate PageRank
**Daniel Ribeiro Silva**
**drsilva**

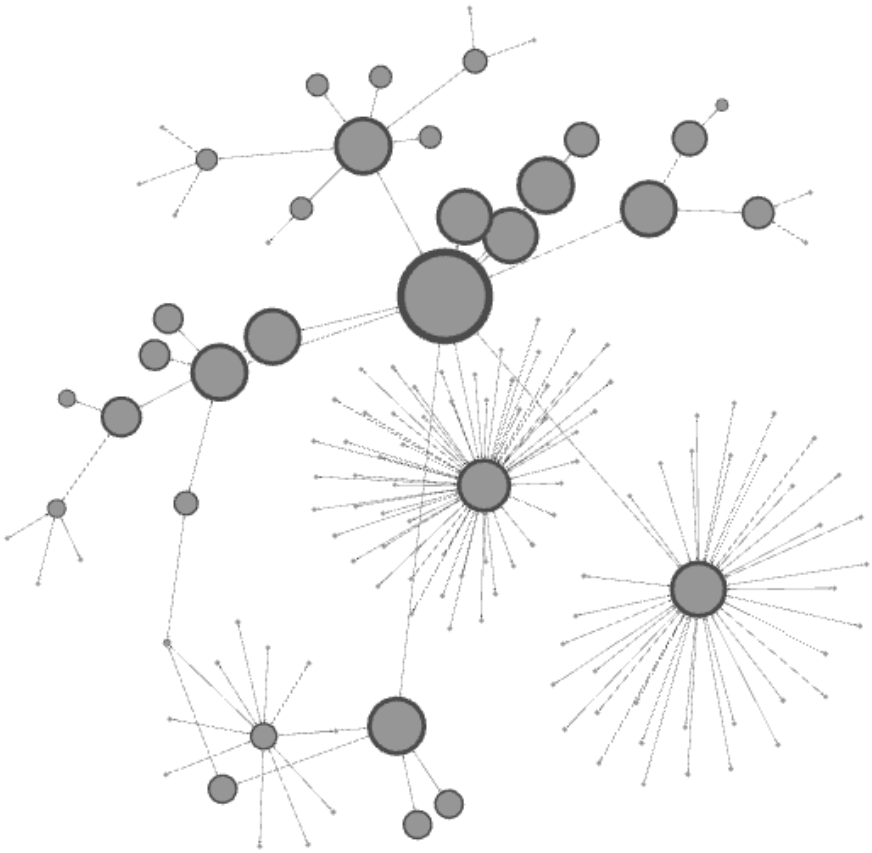
**Question 1**
Seed: "Machine_learning"
$\alpha = 0.3$
$\epsilon = 10^{-5}$
Nodes: 177
Edges: 357



**Question 2**
Seed: “Hipster_%28contemporary_subculture%29”
$\alpha = 0.3$
$\epsilon = 10^{-3}$
Nodes: 118
Edges: 234



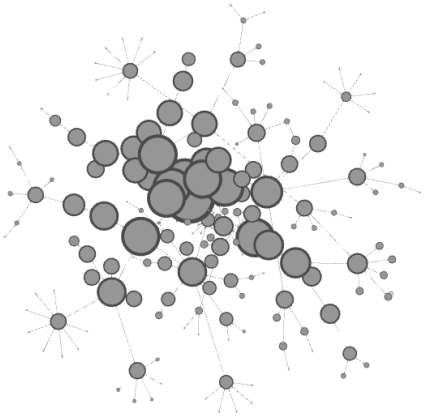
Seed: “Hipster_%28contemporary_subculture%29”
$\alpha = 0.3$
$\epsilon = 2.10^{-4}$
Nodes: 144
Edges: 288



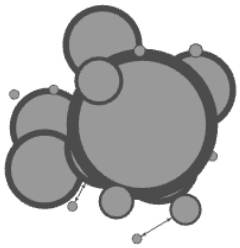
**Question 3**

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", $\alpha = 0.3$, $\epsilon = 10^{-5}$ ).

Standard graph:



Variation 1: increase $\epsilon$ by a factor of 100: new $\epsilon = 10^{-3}$ (alpha remains the same)

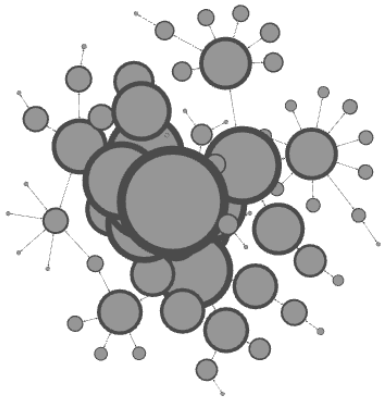


Variation 2: decrease $\epsilon$ by a factor of 100: new $\epsilon = 10^{-7}$ (alpha remains the same)

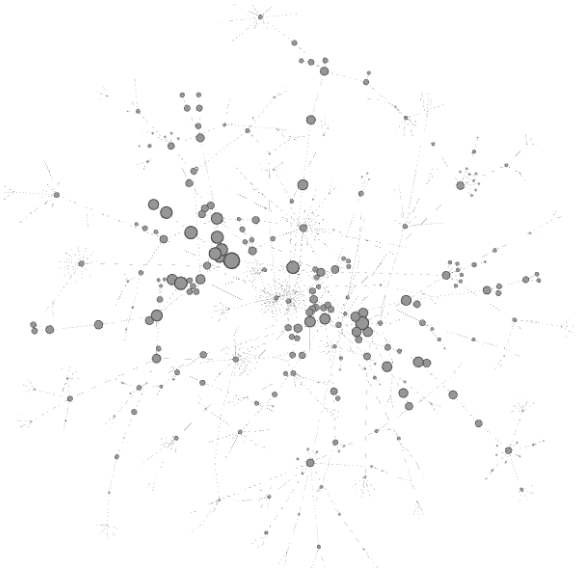
Here we observe that by increasing the value of $\epsilon$, we get a smaller graph and by decreasing the value of $\epsilon$ we get a larger graph. The reason is that $\epsilon$ represents the threshold for which you start pushing residue from a node to its neighbors. If the threshold is small, then 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 $\alpha$:

Variation 3: increase $\alpha$ in by 0.2: new $\alpha = 0.5$ ($\epsilon$ remains the same)



Variation 4: decrease $\alpha$ in by 0.2: new $\alpha = 0.1$ ($\epsilon$ remains the same)



Here we observe that by increasing the value of $\alpha$, we get a smaller graph and by decreasing the value of $\alpha$ we get a larger graph. The reason is that $\alpha$ 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 $\alpha$ 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