Homework 6
Rob Schwartz

Exercise 1

The equation to calculate a clustering coefficient is as follows:

$$\eta(v) = \frac{2*(\text{\# of neighbor nodes' edges to other neighbors})}{\text{degree}*(\text{degree} - 1)}$$

The clustering coefficient, roughly, predicts how likely a neighboring node of *v* is to connect to another neighboring node of *v*.

In this exercise, the node's neighbors are represented by *z*, and the density of each community (G1 having two communities and G2 having one community) is represented by λ.

We know that *v* has the same number of neighboring nodes in both graphs. Given how we calculate η(v) (see first formula), we can acknowledge that the denominator will be the same for both *v1* and *v2*. Therefore, any difference in η(v) will be due to changes in the formula's numerator (# of neighbor nodes' edges to other neighbors).

We now must determine how the number of communities to which *v* belongs influences the aforementioned denominator. In other words, how does community membership influence how often *v*'s neighbors are neighbors with one another?

Well, because density in each community is the same, the nodes in each community are equally likely to be neighbors with other nodes in that respective community. Naturally, this means that a community with more nodes (a larger community) will have more total connected nodes than a smaller community if those two communities have the same density.

This fact is important. Larger communities with density λ have more connected nodes than smaller communities with density λ. In G1, there are two separate communities, each of which necessarily hold at least one of *v*'s neighbors. You can adjust the number of neighbors in each community however you like, but you cannot change that there will be fewer of *v*'s neighbors in each of G1's communities than there will be in G2's single community.

In a community with more of *v*'s neighbor nodes, then, more of those neighboring nodes will be connected than there would be in a community with fewer of *v*'s neighboring nodes. Even in edge cases, such as when λ = 1 (i.e. when all nodes in a community are connected), there will be more of *v*'s neighbors connected because there are simply more of *v*'s neighbors in the community.

We can return to the numerator of the clustering coefficient equation. With our new knowledge, we can assert that communities with more of *v*'s neighboring nodes will have a greater number of edges connecting *v*'s neighbors than in communities with fewer of *v*'s neighbors. Therefore, a

vertex belonging to a community that has more of that vertex's neighboring nodes (i.e. G2) will have a greater clustering coefficient than a community with fewer of that vertex's neighboring nodes (both communities in G1).

Therefore, 2 is correct. For any choice of λ and z, it holds $\eta 2(v) \geq \eta 1(v)$.