Major Changes:
There is a now a possibility that I will be working with Andy Yang and Lichen Zhang (both in 15300) towards a final product. Our goals are very different though and if there is any collaboration at all, it will just be trading our thoughts and at the most, at the end, it may culminate in publishing a paper together where we combine our results. I very hesitant on this because Lichen’s work, as he has told me, involves improving some graph clustering subroutine in an existing ML system; I have no experience and no interest in ML and I want to stick to trying to improve the CRD algorithm from a theoretical point of view. Most likely, we won’t end up doing anything at all together and we’ll just end up talking about our thoughts on various graph clustering algorithms.

What I have accomplished so far:
I have been mostly reading and thinking about the CRD algorithm. Andy has reached out and found the code of the algorithm and passed along the Github to me. I have spent a large portion of my time reading on other graph clustering methods such as the classic random walks (Andersen and Lang)/paint spilling methods. Gary has told me to look into the idea of spectral rounding; while I understand the idea on a high level, I need to work a bit more to learn the math.

Meeting my Milestone
Here are the tasks that I set out to do:

- Be well read in the the background information of CRD and other related algorithms
- Reading about related concepts such as the random walk local cluster algorithm, isoperimetric cuts and the Cheeger inequality
- Build intuition on the Push-Relabel algorithm
- Contact Di Wang, the author of the paper that introduced the CRD process, for his code and implementation of the CRD algorithm.

I have done all the tasks that I have set out to do except parts of Task 2. I have yet to really figure out the exact math behind the Cheeger inequality; my linear algebra skills need some brushing up.

Surprises:
I’m just surprised at how useful graph clustering is in the real world.

Revisions to Milestones
My January 27th milestone remains as it is. For the break though, I was looking forward to really working out the math behind the CRD algorithm so I can have a better clue of how to move forwards. I have a feeling that the CRD algorithm is more complicated than it needs to be. The idea that it encapsulated shouldn’t be that hard to replicate with a similar but more easily expressed algorithm.

Resources Needed

It turns out that I might need to run CRD on some pretty large graphs. I think I’ll be using AWS/EC2. I have talked to Laxman Dhulipala, a PhD student here; he has a lot of experience working with large graphs. Other than that, nothing else.