2^{nd} Milestone Report for 15-400, Spring 2017

Apoorva Bhagwat

Major Changes :

Instead of correlation clustering, we have decided to focus on stability notions for a related problem that has received attention recently. It's the problem of finding endogenously formed communities, which is quite similar to correlation clustering, except clusters are allowed to overlap (a point can belong to multiple communities). Avrim has pointed me to two recent papers on this problem.

Accomplishments So Far :

I have been reading about approximation algorithms for correlation clustering using natural LP and SDP relaxations of the constraints, and other material recommended by Avrim, such as work by Konstantin and Yury Makarychev on approximation algorithms for clustering.

In the last meeting with Avrim, we came up with four high level questions that are worth exploring, based on work by [Balcan et al] (link) and [Arora et al] (link) :

1. Consider a noisy model with a 'ground truth' set of sparse communities, where edges are deleted randomly. Is it possible to recover the true communities? (this corresponds to a small value of the parameter Θ in Balcan et al's approach)
2. Looking at community detection in the context of Bilu-Linial stability (I am particularly interested in this question).
3. See if techniques from Balcan et al's model is applicable to Arora et al's model.
4. Consider community detection as clustering in the stochastic block model with overlapping blocks. Tensor methods that are used for non-overlapping stochastic block models may be applicable here (this seems unfeasible for this semester since I do not know much about these models).

Meeting The Milestone :

I did not meet the milestone set for Jan 30th, because we decided to focus on a slightly different problem, so we still need to narrow down a concrete problem to work on. We came up with a list of potential questions in our meeting, and I'm currently trying to formulate a problem based on the approaches in [Balcan et al] and [Arora et al].

Surprises :

I haven't met with any unpleasant surprises so far. The literature on different techniques

in approximation algorithms is vast, and so the process of getting acquainted with the material and coming up with interesting questions is slow, but that was expected.

Revisions To 15-400 Milestones :

The milestones are still similar to my original milestones, except that I will be studying a variation of the problem I had originally proposed.

Resources Needed :

I haven't felt the need for any more non-trivial resources so far.