15-300 Milestone Report

Lichen Zhang

December 11, 2019

1 Major Changes

In my project proposal, my plan was to implement several local clustering algorithm and combine them together with graph convolutional neural network architecture, e.g., Pinterest. As I read the paper, I found that a more critical issue for my application is the notion of graph cluster and corresponding algorithm might not be directly adequate to this task, especially many such local algorithms have the property that, if it does not start at a vertex *in* the cluster, it might even fail to output a result, which is not acceptable for practical application: we need to handle query to any vertex, and output a proper "neighborhood" with respect to that vertex. One direction might be to work on this new notion of neighborhood defined for any vertex. The other direction is to still work on implementing the algorithms. The last possibility is to extend past works on random walks and flow algorithms on computing local clusters.

2 Accomplishment

On one hand, I have installed proper software to do so, for example, the `plato` graph library by Tencent, and `localgraphclustering` library for research purpose. Also, I have access to a working implementation of distributed Pinsage, and large enough dataset. On the other hand, I've read a series of papers on local clustering, from the original `Nibble` algorithm, to more advanced ones using `PageRank` vector and `EvolvingSet`, and `capacity releasing diffusion`. A key building block in these papers are the curve that connects random walk and conductance of a graph developed by Lovasz and Simonovits has been carefully studied and analyzed.

3 Meet Milestone

In my proposal, I was managed to read papers and prepare for implementations, and I have almost finished paper reading part, there are still two or three papers need to be better understood and studied.

4 Surprises

This field of study has a surprisingly short history, and many of them rely on a method developed by Lovasz and Simonovits in 90s. It's tempting to me to adapt some ideas that I have learned with other methods, to develop some new algorithms. For the implementation part, we found a very

nice library, dedicated to research in this area that has already implemented most of algorithms of our interest.

5 Revisions to Milestones

As mentioned above, because of the library I found, the project might change from implementing from scratch to try to adapt their implementations to other applications. Therefore, all "finish the implementation of some algorithms" will be changed to "adapt implementation of some algorithms to Pinterest framework".

6 Resources Needed

Currently, I have everything I need.