

CS328 Modularity maximisation with graph sparsification

Ayon Biswas

1 Problem Definition

In this proposal, we consider the problem of identifying communities using modularity maximisation[1]. We shall investigate the impact of different graph sparsification methods on module identification. The following are the questions we want to answer:

- How is the result of clustering on the sparsified graph affected in comparison to the original graph?
- What is the speed-up of the clustering method on graphs with different sparsification methods?
- Which graph sparsification method gives best results for modularity maximisation?

2 Algorithm Description

We shall first select edges of the graph using a strategy such as random edge sampling[2], G-spar, L-spar[3] or spectral sparsification[4]. We shall find communities in the sparsified graph using Clauset-Newman-Moore greedy modularity maximization algorithm[5]. We will then investigate the trade-off between speed-up of the clustering algorithm versus the quality of the results.

3 Datasets

We will use the following SNAP datasets[6]:

- Information networks - Wiki (<https://snap.stanford.edu/data/wikispeedia.html>) and Flickr (<https://snap.stanford.edu/data/web-flickr.html>)
- Social networks - Twitter (<https://snap.stanford.edu/data/ego-Twitter.html>) and Facebook (<https://snap.stanford.edu/data/ego-Facebook.html>)

References

- [1] M E Newman. Modularity and community structure in networks. *Proc Natl Acad Sci U S A*, 103(23):8577–8582, June 2006.
- [2] David Ron Karger. *Random sampling in graph optimization problems*. PhD thesis, 1995.
- [3] Venu Satuluri, Srinivasan Parthasarathy, and Yiye Ruan. Local graph sparsification for scalable clustering. 2011.
- [4] Joshua Batson, Daniel A Spielman, Nikhil Srivastava, and Shang-Hua Teng. Spectral {Sparsification} of {Graphs}: {Theory} and {Algorithms}. *Commun. ACM*, 2013.
- [5] Aaron Clauset, M. E.J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 2004.
- [6] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014.