Evolving Sets and Their Applications to Local Graph Partitioning

Kwabena Antwi-Boasiako
Advisor: Daniel Spielman
January 3, 2011

1 Motivation

As graphs and networks become increasingly voluminous it becomes more difficult to visualize or reason about them in the same ways one would reason about smaller networks. We would like to be able to detect community structure, and realize other peculiar features of massive networks by studying sparse cuts. This has applications in graph partitioning and solving systems of linear equations quickly [2]. The problem of locally finding a sparse cut in a graph is that of finding a set of vertices with small conductance (i.e. a set of vertices that are connected by very few edges to the rest of the graph) by adaptively exploring a large graph $G$, starting from a specified vertex [1]. For the algorithm to be local, its complexity must be bounded in terms of the size of the set it outputs, with at most a weak dependence on $n$, the number of vertices in $G$ [1].

2 Algorithm

Anderson and Reid [1] provide an algorithm for this problem, EvoCut, that has a local approximation guarantee of $O(\phi^{1/2} \log^{1/2} n)$ and expected work/volume ratio of $O(\phi^{-1/2} \log(n))$, where $\phi$ is the conductance of the cut, $n$ is the number of vertices in the graph and the work/volume ratio is the ratio between the complexity of the algorithm on a given run and the volume of the set output [1]. The main idea behind the algorithm is that of a volume-biased evolving set process (volume-biased ESP) which is a special kind of Markov chain defined on the set of vertices of $G$. 
3 What I will be doing

I will be studying, in detail, the volume-biased evolving set process. I will also be implementing the algorithm *EvoCut* in order to examine its performance on some real-world networks. Anderson and Reid combine *EvoCut* with a technique of Spielman and Teng [2] to yield another algorithm, *EvoPartition*, that has complexity $(m + n\phi^{-1/2}) \cdot O(\text{polylog}(n))$, where $m$ is the number of edges of $G$, and the other variables are as described above. It outputs a set of vertices whose conductance is $O(\phi^{1/2} \log^{1/2} n)$ and whose volume is at least half that of any set with conductance at most $\phi$, where $\phi$ is an input to the algorithm. This algorithm is asymptotically faster by a factor of $\phi^{1/2}$ than any existing algorithm that provides a nontrivial approximation guarantee for the balanced cut problem, although several algorithms provide better approximation guarantees [1]. I will implement *EvoPartition* and investigate how to improve the cuts it produces.

4 Deliverables

These are the items I will produce at the end of the semester:

4.1 Code

The code for the algorithms will be written in C/C++.

4.2 Written report

I will submit a written report that will include a tutorial on the evolving set process, and a summary of my findings on how the algorithms performed on the real-life data. This will include a remark on how accurate the algorithms are in practice, i.e. whether it is worthwhile trading the size of the cuts produced for the speed with which they are produced.

4.3 Presentation

I will present a lecture on my findings to a committee in the Department of Mathematics.
References
