

NSF Graduate Research Fellowship Previous Research

Bill March

Opening paragraph: quick summary of things I've worked on/looked into Tie this into interests established before Graph coloring Computational Geometry Fast algorithms Computational biology and chemistry

Hierarchical Clustering and EMST

Context. Euclidean Minimum Spanning Trees are fundamental structures in computational geometry. They have been applied to many things. Additionally, the EMST is equivalent to a hierarchical clustering of the underlying points. This form of clustering is common in biology and cosmology.

The problem has been around for a long time. There are algorithms, but none are great. Now, there are gene microarrays and the SDSS, which scientists might want to apply EMST/hierarchical clustering to. So, we need better scalable algorithms to make this happen.

I approached Alex in Spring 2006 about research opportunities. He suggested it might be possible to apply some of his techniques to the EMST problem. I began working on this problem with him over the summer. Did it myself. Worked with his grad students on learning the other algorithms. Figured out the algorithm, wrote the code, did the experiments on my own.

Details. Computing nearest neighbors - should be able to copy a lot of this from the paper Include a picture of a kd-tree, etc. All the details really exist to set up the proposal.

Results Include some of the better charts

What did I learn? How does it help the proposal?

My role: read and thought, learned some stuffs Learned I wanted something more directly applicable

Submitted EMST paper.