Introduction
Sentiment analysis is useful for identifying positive or negative feeling in text. This project explores building a sentiment classifier for user-generated product reviews. Instead of just finding positive/negative sentiment, it identifies the star-rating that the review would have received. The classifier takes advantage of labeled and unlabeled reviews using a label propagation technique. A number of unlabeled reviews then selected to be annotated by the crowd to improve the sentiment classifier.

This project aims to answer two questions:
(1) Can the crowd accurately determine sentiment from online reviews?
(2) Are these sentiment classifications useful in this label propagation technique?

Dataset
The dataset is of product reviews from Amazon.[1] The reviews and ratings serve as nodes and labels in the graph, and the similarity between two reviews is the edge weight.

Method
Our sentiment analysis approach uses a graph-based, semi-supervised learning, label propagation algorithm called Modified Adsorption. [2] Label propagation algorithms aims to propagate labels based on how similar a node is with the labeled nodes around it.

Results
We ran experiments on (200, 400) labeled instances and (400, 800) unlabeled instances. Each experiment had (10, 25, 50, 100, 200) of the unlabeled instances annotated by the Crowd[3].

References

Active Learning
We selected nodes to crowdsourced based on a calculated uncertainty score. The uncertainty is indicated by a dummy label.