Transition Paths of Karate Club Network

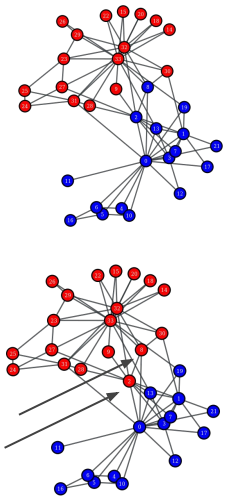
Kirill Konovalov (contribution: 1,2), Esther Visser (contribution: wrote 4,5, helped programming 3), Sepanta Zeighami (contribution: 3)

1 Introduction

In this work we explore two approaches to retrieving the sparsest graph cut based on a real world social network known as the Zachary Karate Club. The network had undergone a fissure based around the conflict of the coach and the president of the club. We compare the performance and suggest new venues to addressing the shortcomings of the chosen methods

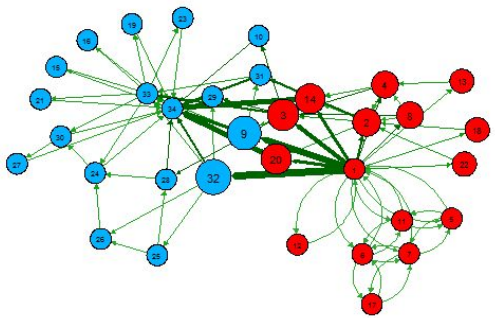
2 Spectral clustering via the Cheeger vector

Spectral decomposition of a normalized graph laplacian of a provides an approximation to graph partitioning. Surprisingly, this method reconstructs the real life partitioning social network of spectacularly well, with only 2 nodes of 34 being labeled incorrectly compared to the ground truth, signified by an arrow on the lower figure.



3 Markov Chains

We used Markov chains to model the social network. The people in the network were modeled as the states of the Markov chain and the probability of moving to another state is uniform among the neighbours. We predict the cut in the social network by checking, starting from a person, whether the probability of reaching the manager is more than the coach or not. The figure below shows the final cut that differs from the original cut by only 1 node. The width and the direction of each edge shows the effective flux and node size shows the transition flux. The directed cycles between two nodes (e.g. 6 and 7) shows that their effective flux in both direction is zero because the probability of starting from them and reaching the manager without reaching the coach is zero. In other words, they will be disconnected from the rest of the graph is the coach is removed.



4 Analysis

Both methods of retrieving the cut in the set of persons predict the actual cut pretty good. The few persons that the prediction predicts wrong are the persons having close/direct contact to both the coach and the president. For predicting those persons right as well, it might be necessary to have a weighted graph. Then both methods can predict the result more precise, because we can see whether persons are closer to the president or the coach not only using other relations they have (which gives more uncertain results).

We can see that for this set the Markov Chain method has performed better in predicting which persons would follow the coach or the president, as it predicts only 1 person wrong, whereas the spectral clustering makes 2 mistakes. However, the method used for the spectral clustering is an approximation, not the exact solution.

5 Conclusion & Future Work

Right now the markov chain method worked better. However, one can easily improve the spectral clustering by finding the optimal cut using the Cheeger vector. However, as this method is NP hard, it does not give fast results. The Markov Chain method works faster, and also gives more kind of information.

For future work, one could try to first find the optimal cut according to the Markov Chain method. After this, one could apply the Cheeger method for cuts that are close to the cut that is found using the Markov Chain, for example cuts that differ by 1 or 2 persons from the cut we already have. This can lead to the optimal solution.

[1] “A global geometric framework for nonlinear dimensionality reduction” Tenenbaum, J.B.; De Silva, V.; & Langford, J.C. Science [2] “Nonlinear dimensionality reduction by locally linear embedding” Roweis, S. & Saul, L. Science 290:2323 (2000) [3] P. Biswas, et al., Semidefinite programming approaches for sensor network localization with noisy distance measurements, IEEE Transactions on Automation Science and Engineering, 3 (2006), pp. 360--371 [4] Gower, J. C. (1966) Some distance … analysis. *Biometrika* **53**, 325–328.