HW#4 —— Graph Clustering Algorithms

Saket Choudhary

April 28, 2015

Markov Clustering(MCL)

Markov clustering makes use of stochastic flows in a graph. The underlying idea is that a random walk on a dense cluster is likely to stay on the vertices of this cluster before jumping to other cluster. This property is made use by simulating a stochastic flow on the graph such that the flow be inflated wherever the current is strong and downweighted otherwise. This can thus be formulated as a Markov graph. The flows are modelled by calculating successive powers of the matrix.

Restricted Neighborhood Search(RNSC)

RNS clustering tries to minimises a cost function that captures the weight of inter and intra cluster edges. With a random initial value, RNSC iteratively assigns a node to other cluster if that leads to a local minima. Termination can be defined based on number of iterations or convergence of the cost function.

Molecular Complex Detection(MCODE)

Molecular complex detection assigns a weight to each vertex that is proportional to the number of neighbors. Starting from the heaviest vertex it iteratively moves out assigning each vertex to the cluster if it is above a certain threshold.

Comparison

All thress algorithms are parametric and hence the quality of clustering will depend on the choice of appropriate parameters.

Robustness: MCL and RNSC are more robust to noise and missing information in the data, while they are also more robust towards the choice of parameters. Hence in case of missing data MCL or RNSC might be a better choice over MCODE.

Heuristic RNSC and MCODE are heuristic and hence the convergence depends on the initial set of clusters/vertices chosen

References

Brohee, Sylvain, and Jacques Van Helden. "Evaluation of clustering algorithms for protein-protein interaction networks." BMC bioinformatics 7.1 (2006): 488.