Modularity and Heavy-Tailed Degree Distributions

Larry Wilson

Center for Communications Research - La Jolla, 4320 Westerra Ct, San Diego, CA 92114, `larry@ccrwest.org`

Modularity is a measure of the quality of a division of the vertices of a graph into disjoint clusters. Hill climbing on this score is a common method for discovering clusters in graphs. In experiments on LFR graphs with planted clusters, we found that the performance of this clustering is worse for low-degree vertices but quite good for high-degree vertices. The vertices of an LFR graphs have a power-law degree distribution and so there are many more low-degree vertices than high-degree vertices and so the overall clustering performance is poor. One form of the modularity score is

$$\frac{1}{2L}\sum_v\sum_w C_{vw}\left(A_{vw}-\frac{k_v k_w}{2L}\right)$$

where L the number of edges in the graph, k_v the degree of the vertex v, and A and C are adjacency matrices, A for the graph and C for the clusters; that is $C_{vw}=1$ when v and w are in the same cluster and $C_{vw}=0$ otherwise. By loosening the random model from which the modularity score is derived, we recover a variant score in which the "penalty" (that is, the subtracted part in the modularity score) is constant over all pairs of vertices. We find that the same hill climb using this variant modularity performs better at clustering, both for the low-degree vertices and overall.