AI3602 Data Mining: Homework 5

Xiangyuan Xue (521030910387)

1. According to the definition, the modularity $Q_d(C)$ of a community C is specified as

$$Q_d(C) = \frac{\sum_{in}}{m} - \frac{\sum_{tot}^{(in)} \sum_{tot}^{(out)}}{m^2}$$

where m is the number of edges in the graph, $\sum_{in}$ is the summation of edge weights inside the community C, $\sum_{tot}^{(in)}$ is the summation of in-edge weights of the community C, and $\sum_{tot}^{(out)}$ is the summation of out-edge weights of the community C.

To compute the modularity gain $\Delta Q_d(i \to C)$, we should notice that

$$\Delta Q_d(i \to C) = Q_d(\{i\} \cup C) - Q_d(\{i\}) - Q_d(C)$$

Then we can expand the equation step by step as follows

$$\begin{aligned}
\Delta Q_d(i \to C) = & \left(\frac{k_{i,in} + \sum_{in}}{m} - \frac{\left(k_i^{(in)} + \sum_{tot}^{(in)}\right)\left(k_i^{(out)} + \sum_{tot}^{(out)}\right)}{m^2} \right) \\
& - \left(0 - \frac{k_i^{(in)} k_i^{(out)}}{m^2} \right) - \left(\frac{\sum_{in}}{m} - \frac{\sum_{tot}^{(in)} \sum_{tot}^{(out)}}{m^2} \right) \\
= & \frac{k_{i,in}}{m} - \frac{k_i^{(out)} \sum_{tot}^{(in)} + k_i^{(in)} \sum_{tot}^{(out)}}{m^2}
\end{aligned}$$

where $k_{i,in}$ is the summation of edge weights between node i and nodes in community C, $k_i^{(in)}$ is the in-degree of node i, and $k_i^{(out)}$ is the out-degree of node i. Therefore, we can conclude that the modularity gain $\Delta Q_d(i \to C)$ is given by

$$\Delta Q_d(i \to C) = \frac{k_{i,in}}{m} - \frac{k_i^{(out)} \sum_{tot}^{(in)} + k_i^{(in)} \sum_{tot}^{(out)}}{m^2}$$