Midterm 1 - Q4

Alara Dirik

May 6, 2020

Problem

The "Finding and Evaluating Community Structure in Networks" paper by Newman and Girvan **[1]** introduces modularity Q for the first time, however the description of Q in the paper is not clear. Do some research and then describe it in a better way.

Answer

The paper by Newman and Girvan uses modularity Q as a metric to determine the number of clusters in a given network. Given a network divided into ℓ communities $\{c_1, \ldots, c_\ell\}$, the modularity of this parrticular division is the number of edges within the communities minus the expected number of edges within the communities if the edges were distributed at random **[2]**.

Modularity can be formulated as $Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta\left(c_i, c_j\right)$ where,

- m is the number of edges
- A is the adjacency matrix
- k_i is the degree of node i
- c_i is the community ID of node i

Intuitively, modularity Q lies between $[-1, 1]$ and is positive if the number of edges within groups exceeds the expected number edges. $0.3 < Q < 0.7$ denotes significant community structure, thus modularity Q can be directly optimized to detect communities in a network.

References

[1] Micaleah Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 69 2 Pt 2:026113, 2004.

[2] Mark E. J. Newman. Community detection in networks: Modularity optimization and maximum likelihood are equivalent. *ArXiv*, abs/1606.02319, 2016.