Final project outline

Zhengyu Zou, Kaicheng Guo

April 14, 2023

Project topic: Cheeger's inequality and Spectral Clustering.

Project goal: Presenting a new recommendation algorithm for social networking that requires a lot less user input.

Outline:

1. Warm-up question 30s: If you have a very big graph, how do you disconnect the graph by deleting the minimal number of edges while still keeping the two subgraphs "balanced" (having similar number of vertices). In other words, how to find "good" clusters in a graph (examples on power-point: using minimum edge cut but subgraph not balanced, subgraph balanced but delete a lot of edges, good cluster).

2. Definition of Conductance (Cheeger Constant): The conductance $\phi(G)$ of a graph gives information on the optimal 2-clusters. Define a volume of a subset $S \subset V$ to be the sum of the degrees of all vertices in S:
$$\mathrm{vol}(S) = \sum_{v \in S} \deg(v) \ .$$
Fix $S \subset V$, the conductance of a subset S is given by
$$\phi(S) = \frac{|E(S, V \setminus S)|}{\min(\mathrm{vol}(S), \mathrm{vol}(V \setminus S))} \ .$$
The conductance of the entire graph is simply given by $\min_{S \subset V} \phi(S)$.

3. Naive way to measure conductance and find cluster: One can simply consider every nonempty subset $S \subset V$ as a potential candidate for a cluster. This requires a runtime of $O(2^{|V|})$, which is time inefficient.

4. Cheeger's Inequality to approximate conductance: Consider the Laplacian matrix L_G. Cheeger's Inequality claims that the conductance is bounded by the smallest positive eigenvalue λ_2 of a "normalized Laplacian" $N_G = D^{-1/2} L_G D^{-1/2}$ in the following sense: $\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$.

5. Algorithm: Spectral Partitioning Algorithm
   (a) Find a vector x such that $R(x) \leq O(\lambda^2)$. One can use the power iteration algorithm to find x in linear time, but we do not discuss the details.
   (b) Sort the vertices of V according to x, $x(v_1) \leq x(v_2) \leq \cdots \leq x(v_n)$.
   (c) Return the threshold cut $(v_1, \ldots, v_i, v_{i+1}, \ldots, v_n)$ with the smallest conductance.