CC484: Pattern Recognition **Spring 2018**

Sheet 3: Normailzed Cuts and Similarity Graphs

Abdelrahman Mohamed Abdelnabi, 3466 *May 17*

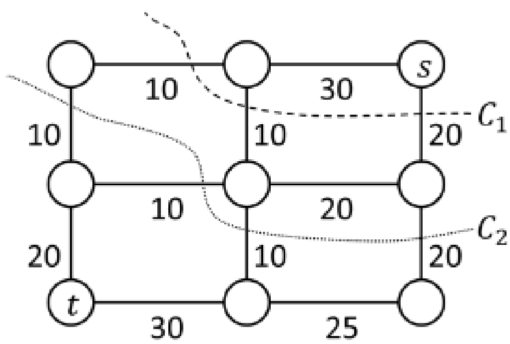
1 Normalized Cut

The cost of a graph partition into two sets A and B is the sum of the weights of edges connecting vertices in A with vertices in B.

$$cut(A,B) = \sum_{i \in A, j \in B} W_{ij}$$

This cost function however, will often isolate vetrices because vertice with a small degree when cut, will have a low cost. To avoid this problem, we use the normalized cut cost funtion, which aims to partition the graph into two nearly equal partitions in size.

$$NCut(A,B) = cut(A,B)(\frac{1}{vol(A)} + \frac{1}{vol(B)})$$



1. Using the un-normalized min cut. The cost of cut C_1 is

$$C_1 = 10 + 10 + 20 = 40$$

   and the cost of C_2 is

$$C_2 = 10 + 10 + 10 + 20 = 50$$

   The un-normalized min cut cost function gives a lower cost to C_1 than to to C_2.

2. When using the normalized cut. $vol(C_i)$ is defined as the sum of all the weights on edges with one end in cluster C_i

$$vol(C_i) = \sum_{v_j \in C_i} d_j$$

For cut C_1, $vol(A) = 100$ and $vol(B) = 280$ and for cluster C_2, $vol(A) = 230$ and $vol(B) = 200$. The cost of C_1 is

$$C_1 = 40(\frac{1}{100} + \frac{1}{280}) = 0.542$$

and the cost of C_2 is

$$C_2 = 50(\frac{1}{230} + \frac{1}{200}) = 0.467$$

The normalized cut cost of cut C_2 is lower than that of C_1 which is what we want for clustering.