Proposition 20.7. *Let $G = (V, W)$ be a weighted graph. The number c of connected components $K_1, \ldots, K_c$ of the underlying graph of G is equal to the dimension of the nullspace of both L_{sym} and L_{rw}, which is equal to the multiplicity of the eigenvalue 0. Furthermore, the nullspace of L_{rw} has a basis consisting of indicator vectors of the connected components of G, that is, vectors $(f_1, \ldots, f_m)$ such that $f_j = 1$ iff $v_j \in K_i$ and $f_j = 0$ otherwise. For L_{sym}, a basis of the nullpace is obtained by multiplying the above basis of the nullspace of L_{rw} by $D^{1/2}$.*

A particularly interesting application of graph Laplacians is graph clustering.

20.4 Graph Clustering Using Normalized Cuts

In order to explain this problem we need some definitions.

Definition 20.20. Given any subset of nodes $A \subseteq V$, we define the *volume* $\mathrm{vol}(A)$ of A as the sum of the weights of all edges adjacent to nodes in A:

$$\mathrm{vol}(A) = \sum_{v_i \in A} \sum_{j=1}^{m} w_{i\,j}.$$

Given any two subsets $A, B \subseteq V$ (not necessarily distinct), we define $\mathrm{links}(A, B)$ by

$$\mathrm{links}(A, B) = \sum_{v_i \in A, v_j \in B} w_{i\,j}.$$

The quantity $\mathrm{links}(A, \overline{A}) = \mathrm{links}(\overline{A}, A)$ (where $\overline{A} = V - A$ denotes the complement of A in V) measures how many links escape from A (and $\overline{A}$). We define the *cut* of A as

$$\mathrm{cut}(A) = \mathrm{links}(A, \overline{A}).$$

The notion of volume is illustrated in Figure 20.5 and the notions of cut is illustrated in Figure 20.6.

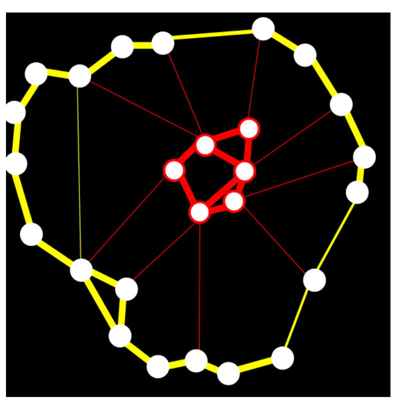

Figure 20.5: Volume of a set of nodes.