



Figure 20.7: A weighted graph and its partition into two clusters.

A way to get around this problem is to normalize the cuts by dividing by some measure of each subset A_i . A solution using the volume $\text{vol}(A_i)$ of A_i (for $K = 2$) was proposed and investigated in a seminal paper of Shi and Malik [160]. Subsequently, Yu (in her dissertation [191]) and Yu and Shi [193] extended the method to $K > 2$ clusters. The idea is to minimize the cost function

$$\text{Ncut}(A_1, \dots, A_K) = \sum_{i=1}^K \frac{\text{links}(A_i, \overline{A_i})}{\text{vol}(A_i)} = \sum_{i=1}^K \frac{\text{cut}(A_i)}{\text{vol}(A_i)}.$$

The next step is to express our optimization problem in matrix form, and this can be done in terms of Rayleigh ratios involving the graph Laplacian in the numerators. This theory is very beautiful, but we do not have the space to present it here. The interested reader is referred to Gallier [74].

20.5 Summary

The main concepts and results of this chapter are listed below:

- Directed graphs, undirected graphs.
- Incidence matrices, adjacency matrices.
- Weighted graphs.
- Degree matrix.
- Graph Laplacian (unnormalized).