Abstract

In this dissertation, we propose a spectral distance to measure the difference between graphs in the sense of their structures. For two graphs $G(V, E)$ and $G'(V', E')$ with different size N and N' respectively, their eigenvalues of normalized Laplacian matrices can be written as:

$$0 = \lambda_0 \leq \lambda_1 \leq \ldots \lambda_{N-1} \leq 2 \text{ and } 0 = \lambda'_0 \leq \lambda'_1 \leq \ldots \lambda'_{N'-1} \leq 2.$$

We define the spectral distance $D(G, G')$ as follows:

$$D(G, G') = \int |\rho_G(x) - \rho_{G'}(x)| dx,$$

where $\rho_G(x) = \frac{1}{N} \sum_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\lambda_i)^2}{2\sigma^2}}$ is the eigenvalues density for graph G and σ here is the bandwidth. This distance is bounded in $[0, 2]$ and proved to be a pseudometric on the space of graphs.

For some particular graph classes, we prove that the distance between two graphs belonging to this class becomes zero, when their sizes tend to be infinity. Generally, it is difficult to destroy the structure of one large graph by finite edit steps. On the one hand we prove that the distance between two large graphs belonging to the same class and only differing in finitely many edit operations is equal to $\mathcal{O}(1/N)$. On the other hand we show that the distance between two graphs lying in different classes is bigger than a certain value that is independent of the graph sizes. Moreover, while computing the graph edit distance is NP-complete, the time complexity for calculating the spectral distance can be reduced to $\mathcal{O}(N^2)$.

Based on the relationship between the evolutionary distance and the spectral distance, we employ the spectral distance to predict the evolutionary relationship among graphs. Compared with other distances on graphs, our method yields good results in accurately reconstructing graphs.

One of the most important applications of our spectral distance is that we can use it to predict the evolutionary relation among networks stemming from the real databases. In biology, the protein-protein interaction networks and the domain co-occurrence networks are two popular types of networks with a functional perspective in organisms. We generate networks of differing species from different databases. By calculating the quartet distance, the spectral distance can produce similar or more accurate phylogenetic trees than the ones obtained by other graph distances, when we compare them to expert classifications. In principle, the reconstruction using our spectral distance can be carried out on any group of networks without labeled vertices information. As another example, we apply our method to linguistic networks. It also gives a relative correct relationship among different language systems.