Heat Kernels, Manifolds and Graph Embedding

Xiao Bai and Edwin R. Hancock

Department of Computer Science
University of York, UK

Abstract. In this paper, we investigate the use of heat kernels as a means of embedding graphs in a pattern space. We commence by performing the spectral decomposition on the graph Laplacian. The heat kernel of the graph is found by exponentiating the resulting eigensystem over time. By equating the spectral heat kernel and its Gaussian form we are able to approximate the geodesic distance between nodes on a manifold. We use the resulting pattern of distances to embed the trees in a Euclidean space using multidimensional scaling. The arrangement of points in this space can be used to construct pattern vectors suitable for clustering the graphs. Here we compute a weighted proximity matrix, and from the proximity matrix a Laplacian matrix is computed. We use the eigenvalues of the Laplacian matrix to characterise the distribution of points representing the embedded nodes. Experiments on sets of shock graphs reveal the utility of the method on real-world data.

1 Introduction

One of the problems that arises in the manipulation of large amounts of graph data is that of clustering. Broadly speaking, there are two approaches to the problem. The first of these is to maintain a class prototype, and to cluster by iteratively merging graphs together. The second approach, which avoids the need to maintain a class prototype, is to apply pairwise clustering methods to the edit distance between graphs. Unfortunately, both of these methods involve computing correspondences between nodes, and since this is potentially an NP-hard problem, the computational overheads can be large. An alternative, which does not involve computing explicit correspondences is to embed the nodes of individual graphs in a low dimensional space and to characterise the graph using the distribution of points corresponding to nodes. Central clustering techniques can then be applied to vectors representing the features of the point-distribution associated with the graphs.

In the mathematics literature, there is a considerable body of work aimed at understanding how graphs can be embedded in manifolds [10]. Broadly speaking there are three ways in which the problem has been addressed. First, the graph can be interpolated by a surface whose genus is determined by the number of nodes, edges and faces of the graph. Second, the graph can be interpolated by a hyperbolic surface which has the same pattern of geodesic (internode) distances as the graph [1] [6]. Third, a manifold can be constructed whose triangulation

A. Fred et al. (Eds.): SSPR&SPR 2004, LNCS 3138, pp. 198–206, 2004.

is the simplicial complex of the graph [12, 2]. A review of methods for efficiently computing distance via embedding is presented in the recent paper of Hjaltason and Samet [4].

In the pattern analysis community, there has recently been renewed interest in the use of embedding methods motivated by graph theory. One of the best known of these is ISOMAP [7]. Here a neighborhood ball is used to convert data-points into a graph, and Dijkstra's algorithm is used to compute the shortest(geodesic) distances between nodes. The matrix of geodesic distances is used as input to MDS. The resulting algorithm has been demonstrated to locate well-formed manifolds for a number of complex data-sets. Related algorithms include locally linear embedding which is a variant of PCA that restricts the complexity of the input data using a nearest neighbor graph, and the Laplacian eigenmap that constructs an adjacency weight matrix for the data-points and projects the data onto the principal eigenvectors of the associated Laplacian matrix (the degree matrix minus the weight matrix) [3]. Collectively, these methods are sometimes referred to as manifold learning theory.

One of the most interesting recent developments in this area has been to establish a link between graph-spectra and the geometry of the underlying manifold [5, 8, 9, 11]. Here considerable insight can be achieved through the analysis of the heat kernel of the graph [5, 9]. According to the heat-equation the time derivative of the kernel is determined by the graph Laplacian. The solution to the heat equation is obtained by exponentiating the Laplacian eigensystem over time. The heat kernel encapsulates the way in which information flows through the edges of the graph over time, and is closely related to the path length distribution on the graph. The graph can be viewed as residing on a manifold whose pattern of geodesic distances is characterised by the heat kernel. For short times the heat kernel is determined by the local connectivity or topology of the graph as captured by the Laplacian, while for long-times the solution gauges the global geometry of the manifold on which the graph resides.

The aim in this paper is to investigate whether the heat kernel can be used for the purposes of embedding graphs, and in particular trees, on a low dimensional manifold. When the manifold on which the graph resides is locally Euclidean, then the heat kernel may be approximated by a Gaussian function of the geodesic distance between nodes. By equating the spectral and Gaussian forms of the kernel, we can make estimates of the geodesic distances. These distances may then be used to embed the graph in a low-dimensional space. Here we follow the ISOMAP algorithm and use multdimensional scaling to locate a low-distortion embedding of the geodesic distances. Once embedded in this space, we can attempt to extract features that characterise the point-distribution of the embbeded nodes and to use them for the purposes of clustering. To do this we construct a weighted Laplacian matrix for the nodes of the embedded graph by exponentiating the negative squared-distance between nodes. The spectrum of eigenvalues of the Laplacian can be used for the purposes of tree clustering and visualisation.

2 Heat Kernels and Riemannian Manifolds

In this section, we develop a method for approximating the geodesic distance between nodes by exploiting the properties of the heat kernel. To commence, suppose that the graph under study is denoted by $G = (V, E)$ where V is the set of nodes and $E \subseteq V \times V$ is the set of edges. Since we wish to adopt a graph-spectral approach we introduce the adjacency matrix A for the graph where

$$A(u, v) = \begin{cases} 1 & \text{if } (u, v) \in E \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

We also construct the diagonal degree matric D, whose elements are given by $D(u, u) = \sum_{v \in V} A(u, v)$. From the degree matrix and the adjacency matrix we construct the Laplacian matrix $L = D - A$, i.e. the degree matrix minus the adjacency matrix. The normalised Laplacian is given by $\hat{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$. The spectral decomposition of the normalised Laplacian matrix is

$$\hat{L} = \Phi \Lambda \Phi^T = \sum_{i=1}^{|V|} \lambda_i \phi_i \phi_i^T \tag{2}$$

where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_{|V|})$ is the diagonal matrix with the ordered eigenvalues as elements and $\Phi = (\phi_1 | \phi_2 | .... | \phi_{|V|})$ is the matrix with the ordered eigenvectors as columns. Since $\hat{L}$ is symmetric and positive semi-definite, the eigenvalues of the normalised Laplacian fall in the interval $[0, 2]$, i.e. they are all positive. The eigenvector assoicated with the smallest non-zero eigenvector is referred to as the Fiedler-vector. We are interested in the heat equation associated with the Laplacian, i.e.

$$\frac{\partial h_t}{\partial t} = -\hat{L} h_t \tag{3}$$

where h_t is the heat kernel and t is time. The solution is found by exponentiating the Laplacian eigenspectrum, i.e.

$$h_t = \sum_{i=1}^{|V|} \exp[-\lambda_i t] \phi_i \phi_i^T = \Phi \exp[-t\Lambda] \Phi^T \tag{4}$$

The heat kernel is a $|V| \times |V|$ matrix, and for the nodes u and v of the graph G the resulting component is

$$h_t(u, v) = \sum_{i=1}^{|V|} \exp[-\lambda_i t] \phi_i(u) \phi_i(v) \tag{5}$$

When t tends to zero, then $h_t \simeq I - \hat{L}t$, i.e. the kernel depends on the local connectivity structure or topology of the graph. If, on the other hand, t is large, then $h_t \simeq \exp[-t\lambda_m] \phi_m \phi_m^T$, where λ_m is the smallest non-zero eigenvalue and ϕ_m is the assoicated eigevector, i.e. the Fiedler vector. Hence, the large time behaviour is governed by the global structure of the graph.

It is interesting to note that the heat kernel is also related to the path length distribution on the graph. If $P_k(u,v)$ is the number of paths of length k between nodes u and v then

$$h_t(u,v) = \exp[-t] \sum_{k=1}^{|V|^2} P_k(u,v) \frac{t^k}{k!} \tag{6}$$

The path-length distribution is itself related to the eigenspectrum of the Laplacian. By equating the derivatives of the spectral and path-length forms of the heat kernel it is straightforward to show that

$$P_k(u,v) = \sum_{i=1}^{|V|} (1-\lambda_i)^k \phi_i(u)\phi_i(v) \tag{7}$$

When the graph is embedded on a manifold in Riemannian space then the pattern of geodesic distances between nodes on the manifold is the same as the path length distribution. However, when the manifold is locally Euclidean, then the heat kernel is approximated by the Gaussian

$$h_t(u,v) = [4\pi t]^{-\frac{n}{2}} \exp[-\frac{1}{4t^2} d(u,v)^2] \tag{8}$$

where $d(u,v)$ is the distance between the nodes u and v on the Euclidean manifold and n is the dimensionality of the space. The aim here is to find an approximation to the geodesic distance between nodes in the embbeding, by equating the spectral and Gaussian forms for the kernel. The result is

$$d(u,v) = 2\sqrt{-t \ln\left\{ (4\pi t)^{\frac{n}{2}} \sum_{i=1}^{|V|} \exp[-\lambda_i t]\phi_i(u)\phi_i(v) \right\}} \tag{9}$$

We can consider the behaviour of this function for large and small values of t. When t is small, making use of the fact that $h_t = I - \hat{L}t$, we have

$$d(u,v) = 2\sqrt{-t\left\{ \frac{n}{2} \ln[4\pi t] + \ln[1 - \hat{L}(u,v)t] \right\}} \tag{10}$$

Hence, the small t behaviour determined by the local topology of the graph. Moreover, since the second term under the-square-root vanishes, the behaviour near $t = 0$ is independant of the structure of the graph. On the other hand, when t is large we can write

$$d(u,v) = 2\sqrt{-t\left\{ \frac{n}{2} \ln[4\pi t] - \lambda_m t + \ln \phi_m(u)\phi_m(v) \right\}} \tag{11}$$

For very large t we have that $d(u,v) \simeq t\sqrt{\lambda_m}$, and hence the effect of local edge-structure is completely smoothed away.

Although the parameter t potentially provides a route to a graph scale-space, here we set $4\pi t = 1$ to simplify the analysis.

3 Multidimensional Scaling

Our aim is to embed the pattern of geodesic distances in a low dimensional space in a manner which minimises the distortion. For this reason we turn to multidimensional scaling(MDS), which is a procedure which allows data specified in terms of a matrix of pairwise distances to be embedded in a Euclidean space. The pairwise geodesic distances between nodes $d(u, v)$ are used as the elements of an $|V| \times |V|$ dissimilarity matrix S, whose elements are defined as follows

$$S(u,v) = \begin{cases} d(u,v) & \text{if } u \neq v \\ 0 & \text{if } u = v \end{cases} \tag{12}$$

In this paper, we use the classical multidimensional scaling method. The first step of MDS is to calculate a matrix T whose element with row r and column c is given by $T(r,c) = -\frac{1}{2}[d(r,c)^2 - \hat{d}(r,.)^2 - \hat{d}(.,c)^2 + \hat{d}(.,.)^2]$, where $\hat{d}(r,.) = \frac{1}{|V|}\sum_{c=1}^{|V|} d(r,c)$ is the average dissimilarity value over the rth row, $\hat{d}_{.c}$ is the dissimilarly defined average value over the cth column and $\hat{d}(.,.) = \frac{1}{|V|^2}\sum_{r=1}^{|V|}\sum_{c=1}^{|V|} d(r,c)$ is the average dissimilarity value over all rows and columns of the dissimilarity matrix T.

We subject the matrix T to an eigenvector analysis to obtain a matrix of embedding co-ordinates X. If the rank of T is $k, k \leq |V|$, then we will have k non-zero eigenvalues. We arrange these k non-zero eigenvalues in descending order, i.e. $l_1 \geq l_2 \geq \ldots \geq l_k > 0$. The corresponding ordered eigenvectors are denoted by $\boldsymbol{u}_i$ where l_i is the ith eigenvalue. The embedding co-ordinate system for the graphs obtained from different views is $X = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_s]$, where $\boldsymbol{f}_i = \sqrt{l_i}\boldsymbol{u}_i$ are the scaled eigenvectors. For the tree-nodes indexed i, the embedded vector of co-ordinates is $\boldsymbol{x}_i = (X_{i,1}, X_{i,2}, ..., X_{i,s})^T$.

4 Characterising the Embedded Point Distribution

Once the nodes of a graph have been embedded, we can attempt to characterise the structure of the graph by summarising the distribution of points associated with the nodes. Although there are many alternatives that can be used for this purpose, including statistical moments, here we opt to use a graph-spectral characterisation of the points. To this end, we commence by computing a weighted proximity matrix W with elements

$$W_{i_1,i_2} = \begin{cases} \exp[\frac{-\left\|\boldsymbol{x}_{i_1} - x_{i_2}\right\|_2^2}{2\sigma^2}] & \text{if } \left\|\boldsymbol{x}_{i_1} - \boldsymbol{x}_{i_2}\right\|_2 < r \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where σ is a scale constant and r is the radius of a neigbourhood hypersphere in the embedding space. Unfortunately, the matrix W may have negative eigenvalues. Hence, we turn our attention instead to the Laplacian matrix, since it is positive semi-definite and therefore has positive or zero eigenvalues. The Laplacian matrix is $L_E = W - \Delta$ where Δ is diagonal degree matrix with elements $\Delta(i,i) = \sum_{j \in V} W(i,j)$. The spectral decomposition of the Laplacian matrix is

$L_E = \sum_{i=1}^{n} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T$, where λ_i^k is the ith eigenvalue and $\boldsymbol{e}_i$ is the corresponding eigenvector of the Laplacian matrix $\hat{L}$. Our spectral characterisation of the graph is based on the vector of N leading Laplacian eigenvalues $\boldsymbol{B} = (\lambda_1, ..., \lambda_N)^T$. We can perform pattern analysis on sets of graphs by applying clustering or dimensionality reduction techniques to the the vectors of Laplacian eigenvalues.

Our aim is explore the structure of a set of graphs with pattern vectors B_k, $k = 1, M$. There are a number of ways in which the spectral pattern vectors can be analysed. Here, for the sake of simplicity, we use principal components analysis. We commence by constructing the matrix $\mathbf{V} = [\mathbf{B}_1|\mathbf{B}_2|\ldots|\mathbf{B}_k|\ldots|\mathbf{B}_M]$ with the graph feature vectors as columns. Next, we compute the covariance matrix for the elements of the feature vectors by taking the matrix product $\mathbf{C} = \mathbf{V}\mathbf{V}^T$. We extract the principal components directions by performing the eigendecomposition $\mathbf{C} = \sum_{i=1}^{M} l_i \mathbf{u}_i \mathbf{u}_i^T$ on the covariance matrix $\mathbf{C}$, where the l_i are the eigenvalues and the $\mathbf{u}_i$ are the eigenvectors. We use the first s leading eigenvectors (2 or 3 in practice for visualisation purposes) to represent the graphs extracted from the images. The co-ordinate system of the eigenspace is spanned by the s orthogonal vectors $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, .., \mathbf{u}_s)$. The individual graphs represented by the long vectors $\mathbf{B}_k, k = 1, 2, \ldots, M$ can be projected onto this eigenspace using the formula $\mathbf{x}_k = \mathbf{U}^T\mathbf{B}_k$. Hence each graph G_k is represented by an s-component vector $\mathbf{x}_k$ in the eigenspace.

5 Experiments

In this section we experiment with the application of our clustering algorithm to shock graphs. We tested our algorithm on a database of 150 silhouettes of 10 kinds of objects. A representative view of each object is shown in Figure 1.

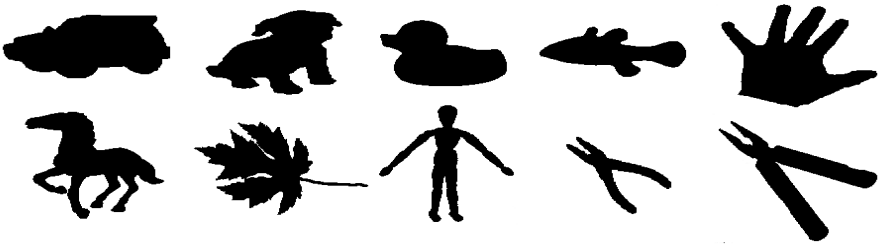

Fig. 1. Sample views of the 10 objects

In our experiments, we will compare the results obtained by using our algorithm with those obtained by using direct spectral analysis of trees. The direct spectral analysis of trees commences by first constructing the Laplacian matrix $\hat{L}$ for the tree. Then we use the leading Laplacian eigenvalues λ_i of the matrix $\hat{L}$ to construct the spectral feature vector $\boldsymbol{B} = (\lambda_1, ..., \lambda_N)^T$. After the spectral feature vectors have been extracted from the trees, we apply PCA (principal component analysis).

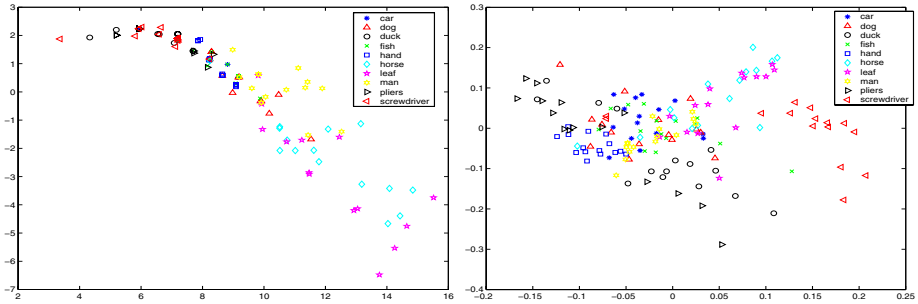

Fig. 2. Direct spectral analysis (left) and heat-kernel analysis (right)

Our first experiment compares our algorithm with the direct spectral analysis of the trees from the entire database of 150 shock trees. In Figure 2 the left-hand panel shows the result of the direct spectral analysis of the shock trees, while the right-hand panel is the result of applying our heat kernel analysis. There is a legend in the top left-hand corner of each plot that explains the shape correspondence of each of the symbols. There are a number of points that can be drawn from these plots. First, in the case of the direct spectral analysis the data distribute themselves along a trajectory in the embedding space. This may be attributable to the problem of co-spectrality of the trees. However, after the heat kernel analysis is performed, the trees distribute themselves over the 2D space. Moreover, in the case of the direct spectral analysis the different shapes are interspersed along the trajectory. It is hence not possible to allocate the shapes to reliably assign shapes to classes on the basis of their position in the plots. The possible exception is that the horses and leafs are separated at the bottom right hand corner of the plot. In the case of the heat kernel analysis, the trees could be better separated. Although there is considerable overlap near the center of the plot, it appears that there is scope for separating the screwdrivers, pliers and leafs.

In our second experiment, we have repeated the procedure above for a smaller database which contains only three representative shapes. The three shapes used for test are the hands, the leafs and the men. For each shape there are 15 different views corresponding to different viewing directions. The left-hand panel of Figure 3 shows the results of direct spectral analysis, while the right-hand panel shows the result of heat kernel analysis. In the case of the direct spectral analysis, the shapes are poorly separated. In the case of the heat kernel analysis, there is good separation.

To further investigate this three-class data, in Figure 4 the two panels show the distances $d_T(k_1, k_2) = (\boldsymbol{B}_{k_1} - \boldsymbol{B}_{k_2})^T(\boldsymbol{B}_{k_1} - \boldsymbol{B}_{k_2})$ between the vectors of eigenvalues for the trees indexed $k_1$and k_2. The left-hand panel is for the direct spectral analysis of the trees and the right-hand panel is for the spectral vectors extracted by performing the heat-kernel embedding. Here the classes emerge as clear blocks in the distance matrix for the heat-kernel embedding, while for the direct spectral analysis the block structure is more confused.

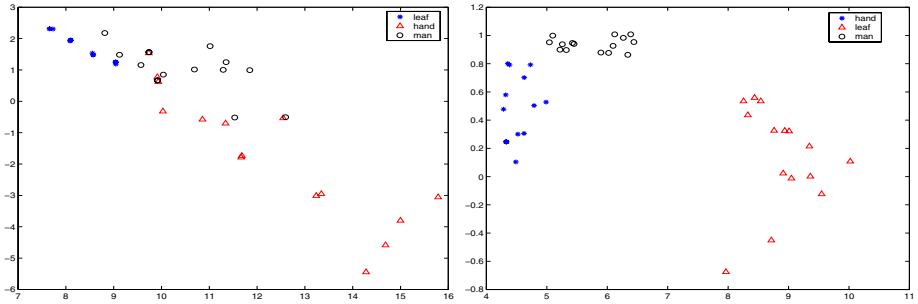

Fig. 3. Direct spectral analysis (left) and heat-kernel analysis (right)

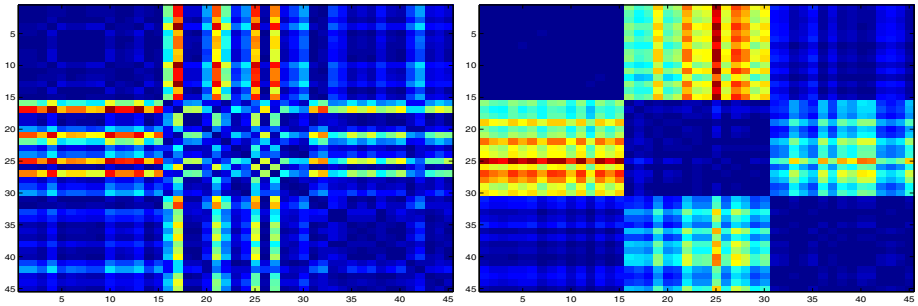

Fig. 4. Distance matrices for direct spectral analysis (left) and heat kernel analysis (right)

6 Conclusion and Future Work

In this paper we have explored how the use of heat kernels can lead to a measure of geodesic distance that can be used for the purposes of embedding graphs in low dimensional Euclidean spaces. The distance measure is found by equating the spectral and Gaussian forms of the heat kernel. We show how MDS can be used to embed the the distances, and how a spectral characterisation of the embedded graphs can be used for graph-clustering. We experiment with the method on sets of shock trees. Here the characterisation which results from the geodesic analysis is better than that obtained from the raw spectral features of the graphs. There are clearly a numbert of ways in which the work reported in this paper can be extended. For instance, it would be interesting to study the controlled effects of varying the time parameter, and to see if this leads to a natural definition of "scale-space" for the graphs.

References

1. A.D. Alexandrov and V.A. Zalgaller. Intrinsic geometry of surfaces. *Transl. Math. Monographs*, 15, 1967.
2. A. Ranicki. Algebraic l-theory and topological manifolds. *Cambridge University Press*, 1992.

3. M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Neural Information Processing Systems*, 14:634–640, 2002.
4. G.R. Hjaltason and H. Samet. Properties of embedding methods for similarity searching in metric spaces. *PAMI*, 25:530–549, 2003.
5. A. Grigor'yan. Heat kernels on manifolds, graphs and fractals. *preprint*, 2003.
6. H. Busemann. The geometry of geodesics. *Academic Press*, 1955.
7. J.B. Tenenbaum, V.D. Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:586–591, 2000.
8. J. Lafferty and G. Lebanon. Diffusion kernels on statistical manifolds. *CMU preprint*, 2004.
9. T. Coulhon M. Barlow and A. Grigor'yan. Manifolds and graphs with slow heat kernel decay. *Imperial College preprint*, 2000.
10. N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some its algorithmic application. *Combinatorica*, 15:215–245, 1995.
11. A.J. Smola and R. Kondor. Kernels and regularisation of graphs. 2004.
12. S. Weinberger. Review of algebraic l-theory and topological manifolds by a.ranicki. *BAMS*, 33:93–99, 1996.