Geometric Characterisation of Graphs

Bai Xiao and Edwin R. Hancock

Department of Computer Science,
University of York, York YO1 5DD, UK.

Abstract. In this paper, we explore whether the geometric properties of the point distribution obtained by embedding the nodes of a graph on a manifold can be used for the purposes of graph clustering. The embedding is performed using the heat-kernel of the graph, computed by exponentiating the Laplacian eigen-system. By equating the spectral heat kernel and its Gaussian form we are able to approximate the Euclidean distance between nodes on the manifold. The difference between the geodesic and Euclidean distances can be used to compute the sectional curvatures associated with the edges of the graph. To characterise the manifold on which the graph resides, we use the normalised histogram of sectional curvatures. By performing PCA on long-vectors representing the histogram bin-contents, we construct a pattern space for sets of graphs. We apply the technique to images from the COIL database, and demonstrate that it leads to well defined graph clusters.

1 Introduction

One of the problems that arises in the manipulation of large amounts of graph data is that of characterising the topological structure of individual graphs. One of the most elegant ways of doing this is to use the spectrum of the Laplacian matrix [7,4]. For instance Shokoufandeh *et al* [10] have used topological spectra to index tree structures, Luo *et al* [3] have used the the spectrum of the adjacency matrix to construct pattern spaces for graphs, and Wilson and Hancock [9] have used algebraic graph theory to construct permutation invariant polynomials from the eigenvectors of the Laplacian matrix. One way of viewing these methods is that of constructing a low-dimensional feature-space the captures the topological structure of the graphs under study.

An interesting alternative to using topological information to characterise graphs is to embed the nodes of a graph on a manifold, and to study the geometry of this manifold. 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]. Third, a manifold can be constructed whose triangulation is the simplicial complex of the graph [2].

However, recently it has been shown that there is a link between the heat kernel of a graph and the geometry of the manifold on which its nodes reside

[12]. The heat kernel is the solution of the heat equation on the graph, and is found by exponentiating the Laplacian eigensystem with 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. Moreover, differential invariants of the manifold can be computed from the heat kernel, and these in turn are related to the Laplacian eigensystem. For instance, the trace of the heat kernel [4] (or the sum of the Laplacian eigenvalues exponentiated with time) can be expanded as a rational polynomial in time, and the coefficients of the leading terms in the series are directly related to the geometry of the manifold. The leading coefficient is the volume of the manifold, the second coefficient is related to the Euler characteristic, and the third coefficient to the Ricci curvature. The zeta-function (i.e. the sum of the eigenvalues raised to a non-integer power) for the Laplacian also conveys geometric information since its derivative at the origin is related to the torsion tensor. This field of study is sometimes referred to as spectral geometry [5,11].

In recent work [13], we have explored how the heat kernel can be used to embed the nodes of a graph on a manifold using a procedure similar to ISOMAP [8]. Here we have performed embedding by applying multidimensional scaling (MDS) to the elements of the heat kernel. In this paper, we aim to explore a more direct approach based on using the heat kernel to characterise the geometry of the manifold on which a graph resides. When the manifold 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 Euclidean distances between nodes. The geodesic distance is given by the floor of the path-length, and this may be computed from the Laplacian spectrum. The difference between the geodesic and Euclidean distances gauges the degree of bending of the manifold, and can be used to estimate the sectional curvatures associated with the edges of the graph. We aim to explore whether sets of graphs can be projected into a pattern space by performing principal components analysis on histograms of sectional curvatures. In this way we aim to characterise the geometry of the manifold in an implicit way.

2 Heat Kernels on Graphs

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 the elements are

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

We also construct the diagonal degree matrix 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 $\hat{L} = \Phi \Lambda \Phi^T$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{|V|})$ is the diagonal matrix with the ordered eigenvalues as elements and $\Phi = (\phi_1 | \phi_2 | \dots | \phi_{|V|})$ is the matrix with the ordered eigenvectors as columns. Since $\hat{L}$ is symmetric and positive semi-definite, the eigenvalues of the normalised Laplacian are all positive. The eigenvector associated 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. $\frac{\partial h_t}{\partial t} = -\hat{L} h_t$, where h_t is the heat kernel and t is time. The heat kernel can hence be viewed as describing the flow of information across the edges of the graph with time. The rate of flow is determined by the Laplacian of the graph. The solution to the heat equation is found by exponentiating the Laplacian eigenspectrum, i.e. $h_t = \Phi \exp[-t\Lambda] \Phi^T$. 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 associated eigenvector, i.e. the Fiedler vector. Hence, the large time behavior is governed by the global structure of the graph.

2.1 Geodesic Distance

It is interesting to note that the heat kernel is also related to the path length distribution on the graph. To show this, consider the matrix $P = I - \hat{L}$, where I is the identity matrix. The heat kernel can be rewritten as $h_t = e^{-t(I-P)}$. We can perform a McLaurin expansion on the heat-kernel to re-express it as a polynomial in t . The result of this expansion is

$$h_t = e^{-t} \left(I + tP + \frac{(tP)^2}{2!} + \frac{(tP)^3}{3!} + \dots \right) = e^{-t} \sum_{k=0}^{\infty} P^k \frac{t^k}{k!} \quad (2)$$

The matrix P has elements

$$P(u, v) = \begin{cases} 1 & \text{if } u = v \\ \frac{1}{\sqrt{\deg(u)\deg(v)}} & \text{if } u \neq v \text{ and } (u, v) \in E \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

As a result, we have that

$$P^k(u, v) = \sum_{S_k} \prod_{i=1}^k \frac{1}{\sqrt{\deg(u_i)\deg(u_{i+1})}} \quad (4)$$

where the walk S_k is a sequence of vertices $u_0, \dots, u_k$ of length k such that $u_i = u_{i+1}$ or $(u_i, u_{i+1}) \in E$. Hence, $P^k(u, v)$ is the sum of weights of all walks

of length k joining nodes u and v . In terms of this quantity, the elements of the heat kernel are given by

$$h_t(u, v) = \exp[-t] \sum_{k=0}^{|V|^2} P^k(u, v) \frac{t^k}{k!} \quad (5)$$

We can find a spectral expression for the matrix P^k using the eigen-decomposition of the normalised Laplacian. Writing $P^k = (I - \hat{L})^k = \Phi(I - \Lambda)^k \Phi^T$, the element associated with the nodes u and v is

$$P^k(u, v) = \sum_{i=1}^{|V|} (1 - \lambda_i)^k \phi_i(u) \phi_i(v) \quad (6)$$

The geodesic distance between nodes, i.e. the length of the walk on the graph with the smallest number of connecting edges, can be found by searching for the smallest value of k for which $P^k(u, v)$ is non zero, i.e. $d_G(u, v) = \text{floor}_k P^k(u, v)$.

2.2 Euclidean Distance

Here we are interested in using the heat-kernel to compute an approximate Euclidean distance between nodes. This is the shortest distance between nodes in the vector space in which the manifold resides. Asymptotically with small t , on a Riemannian manifold the heat kernel can be approximated by the so-called paramatrix [11]

$$h_t(u, v) = [4\pi t]^{-\frac{n}{2}} \exp[-\frac{1}{4t} d_G(u, v)^2] \sum_{m=0}^{\infty} b_m(u, v) t^m \quad (7)$$

where $d_G(u, v)$ is the geodesic distance between the nodes u and v on the manifold, n is the dimensionality of the space (taken to be 3 in our later experiments with MDS) and $b_m(u, v)$ real-valued coefficients. When the manifold is locally Euclidean, then only the first term in the polynomial series is significant and the heat kernel is approximated by a Gaussian. Hence, to approximate the Euclidean distance between the embedded nodes we can equate the spectral and Gaussian forms for the kernel, with the result

$$d_E^2(u, v) = -4t \ln \left\{ (4\pi t)^{\frac{n}{2}} \sum_{i=1}^{|V|} \exp[-\lambda_i t] \phi_i(u) \phi_i(v) \right\} \quad (8)$$

3 Geometric Properties of the Manifold

In this paper our aim is to explore whether the geometric properties of the embedding can be used for the purposes of characterising and clustering graphs. To do this, we can make numerical estimates of the sectional curvature between pairs of nodes. The sectional curvature is determined by the degree to

which the geodesic bends away from the Euclidean chord. Hence for a geodesic on the manifold, the sectional curvature can be estimated easily if the Euclidean and geodesic distances are known. Suppose that the geodesic can be locally approximated by a circle. Let the geodesic distance between the pair of points u and v be $d_G(u, v)$ and the corresponding Euclidean distance be $d_E(u, v)$. Further let the radius of curvature of the approximating circle be $r_s(u, v)$ and suppose that the tangent-vector to the manifold undergoes a change in direction of $2\theta_{u,v}$ as we move along a connecting circle between the two points. In terms of the angle $\theta_{u,v}$, the geodesic distance, i.e. the distance traversed along the circular arc, is $d_G(u, v) = 2r_s(u, v)\theta_{u,v}$, and as a result we have that $\theta_{u,v} = d_G(u, v)/2r_s(u, v)$. The Euclidean distance, on the other hand, is given by $d_E(u, v) = 2r_s(u, v)\sin\theta_{u,v}$, and can be approximated using the McLaurin series

$$d_E(u, v) = 2r_s(u, v)\left(\theta_{u,v} - \frac{1}{6}\theta_{u,v}^3 + \dots\right)$$

Substituting for $\theta_{u,v}$ obtained from the geodesic distance, we have

$$d_E(u, v) = d_g(u, v) - \frac{d_g(u, v)^3}{24r_s^2(u, v)}$$

Solving the above equation for the radius of curvature, the sectional curvature of the geodesic connecting the nodes u and v is approximately

$$k_s(u, v) = \frac{1}{r_s(u, v)} = \frac{2\sqrt{6}(d_G(u, v) - d_E(u, v))^{\frac{1}{2}}}{d_G(u, v)^{\frac{3}{2}}} \quad (9)$$

Since for an edge of the graph, we have that $d_G(u, v) = 1$, the squared sectional curvature associated with an embedded edge is $k_e^2(u, v) = 24(1 - d_E(u, v))$. As a result we can construct the squared sectional curvature matrix $k_e^2 = 24(A + \ln[(4\pi t)^{2nt}\Phi \exp[-4\Lambda t^2]\Phi^T])$.

To characterise the geometry of the graph embedding we construct a histogram of sectional curvatures. The sectional curvatures are assigned to m bins and the normalised contents of the j th bin is denoted by $h(j)$. The feature vector for the graph is constructed from the normalised bin-contents and $\mathbf{B} = (h(1), h(2), \dots, h(m))^T$.

For the purposes of comparison, we will compare the results of using sectional curvature histograms with a number of alternative representations. The first of these is the Laplacian spectrum. Here we use the leading m eigenvalues of the normalised Laplacian $\hat{L}$ to construct the feature-vector $\mathbf{B} = (\lambda_1, \dots, \lambda_m)^T$. We have also explored using the normalised histogram of geodesic distances computed from the path length distribution $P^k(u, v)$.

Our aim is explore the structure of a set of graphs with pattern vectors $\mathbf{B}_k$, $k = 1, M$ extracted using sectional curvature histograms. There are a number of ways in which the graph pattern vectors can be analysed. Here, for the sake of simplicity, we use principal components analysis (PCA). We commence

by constructing the matrix $\mathbf{S} = [\mathbf{B}_1|\mathbf{B}_2|\dots|\mathbf{B}_k|\dots|\mathbf{B}_M]$ 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 $\mathbf{C} = \mathbf{S}\mathbf{S}^T$. 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 (3 in practice for visualisation purposes) to represent the graphs extracted from the images. The co-ordinate system of the eigen-space is spanned by the s orthogonal vectors $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_s)$. The individual graphs represented by the vectors $\mathbf{B}_k, k = 1, 2, \dots, M$ can be projected onto this eigen-space using the formula $\mathbf{B}_k = \mathbf{U}^T \mathbf{B}_k$. Hence each graph G_k is represented by an s -component vector $\mathbf{B}_k$ in the eigen-space.

4 Experiments

We have applied our geometric technique to images from the COIL data-base. The data-base contains views of 3D objects under controlled viewer and lighting conditions. For each object in the data-base there are 72 equally spaced views, which are obtained as the camera circumscribes the object. We study the images from eight example objects. A sample view of each object is shown in Figure 1. For each image of each object we extract feature points using the method of [6]. We have extracted graphs from the images by computing the Voronoi tessellations of the feature-points, and constructing the region adjacency graph, i.e. the Delaunay triangulation, of the Voronoi regions. Our technique has been applied to the resulting graph structures.

In Figure 2 we show example histograms for the twenty views for four of the objects in the COIL database. Here the histograms are stacked behind each other, and are ordered by increasing view number. There are a number of conclusions that can be drawn from the histograms. First, the histograms for the same object are relatively stable with view number. Second, the histograms for

Fig. 1. Eight objects with their Delaunay graphs overlayed

Fig. 2. Sets of histograms for different views of 4 different COIL objects

Fig. 3. Clustering results using the histograms of the sectional curvature (left), the Laplacian spectrum (centre) and the geodesic distance histogram (right).

the different objects have different shapes. The plots shown were obtained with $t = 0.001$.

The results of applying PCA to the vectors of histogram bin contents are shown in the right-hand panel of Figure 3. We obtain well defined clusters of objects. For comparison, the centre panel of Figure 3 shows the corresponding result when spectral clustering is used, and the right-hand panel that obtained using the histogram of geodesic distances. The main qualitative feature is that the different views of the ten objects are more overlapped than when the sectional curvature histograms are used. For all three clustering methods, we have computed the Rand index. For the curvature histograms the maximum rand index is 0.91 (at $t=0.001$), for the spectral method it is 0.57, and for the geodesic

distance histograms it is 0.72. Hence the results extracted using the curvature histograms appear to give the best clusters.

5 Conclusion and Future Work

In this paper we have explored how the use of the heat kernel can lead to a geometric characterisation of the structure of a graph. Specifically, we show how a measure of the difference between the geodesic and Euclidean distances between nodes can be used to compute sectional curvature, and this can be used for the purposes of gauging the differential structure of the manifold on which the graph resides. There are clearly a number of ways in which the work reported in this paper can be extended. For instance, it would be interesting to investigate if the distances and curvatures could be used to aid the process of visualising or drawing graphs, or for graph matching.

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. R. C. Wilson B. Luo and E.R. Hancock. Spectral embedding of graphs. *Pattern Recognition*, 36:2213–2223, 2003.
4. F.R.K.Chung. Spectral graph theory. *American Mathematical Society*, 1997.
5. P. B. Gilkey. Invariance theory, the heat equation, and the atiyah-singer index theorem. *Publish or Perish Inc.*, 1984.
6. C.G. Harris and M.J. Stephens. A combined corner and edge detector. *Fourth Alvey Vision Conference*, pages 147–151, 1994.
7. H.Sachs, D.M.Cvetkovic, and M.Doob. Spectra of graphs. *Academic Press*, 1980.
8. J.B.Tenenbaum, V.D.Silva, and J.C.Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:586–591, 2000.
9. B. Luo R. C. Wilson and E.R. Hancock. Pattern vectors from algebraic graph theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:to appear, 2005.
10. A. Shokoufandeh, S. Dickinson, K. Siddiqi, and S. Zucker. Indexing using a spectral encoding of topological structure. *CVPR*, pages 491–497, 1999.
11. S.Rosenberg. The laplacian on a Riemannian manifold. *Cambridge University Press*, 2002.
12. S.T.Yau and R.M.Schoen. Differential geometry. *Science Publication*, 1988.
13. X.Bai and E.R.Hancock. Heat kernels, manifolds and graph embedding. *Structural,Syntactic, and Statistical Pattern Recognition*, pages 198–206, 2004.