Graph Clustering Using Heat Content Invariants

Bai Xiao and Edwin R. Hancock

Department of Computer Science, University of York, York Y01 5DD, UK

Abstract. In this paper, we investigate the use of invariants derived from the heat kernel as a means of clustering graphs. We turn to the heat-content, i.e. the sum of the elements of the heat kernel. The heat content can be expanded as a polynomial in time, and the co-efficients of the polynomial are known to be permutation invariants. We demonstrate how the polynomial co-efficients can be computed from the Laplacian eigensystem. Graph-clustering is performed by applying principal components analysis to vectors constructed from the polynomial co-efficients. We experiment with the resulting algorithm on the COIL database, where it is demonstrated to outperform the use of Laplacian eigenvalues.

1 Introduction

One of the problems that arises in the manipulation of large amounts of graph data is that of embedding graphs in a low dimensional space so that standard machine learning techniques can be used to perform tasks such as clustering. One way of realise this goal is to embed the nodes of a graph on a manifold and to use the geometry of the manifold as a means of characterising the graph. In the mathematics literature, there is a considerable body of work aimed at understanding how graphs can be embedded in manifolds [7]. 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 [12]. A review of methods for efficiently computing distance via embedding is presented in the recent paper of Hjaltason and Samet [5].

The spectrum of the Laplacian matrix has been widely studied in spectral graph theory [4] and has proved to be a versatile mathematical tool that can be put to many practical applications including routing [2], clustering [9] and graph-matching [11]. One of the most important properties of the Laplacian spectrum is its close relationship with the heat equation. The heat equation can be used to specify the flow of information with time across a network or a manifold [10]. 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. Because the heat kernel encapsulates the way in which information flows through the edges of the graph

J.S. Marques et al. (Eds.): IbPRIA 2005, LNCS 3523, pp. 123–130, 2005.

over time, it 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. In a recent paper [13], we have exploited this property and have used heat-kernel embedding for the purposes of graph clustering.

There are a number of different invariants that can be computed from the heat-kernel. Asymptotically for small time, 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 co-efficients of the leading terms in the series are directly related to the geometry of the manifold. For instance, the leading co-efficient is the volume of the manifold, the second co-efficient is related to the Euler characteristic, and the third co-efficient to the Ricci curvature. The zeta-function (i.e. the sum of the eigenvalues raised to a non-integer power) for the Laplacian also contains geometric information. For instance its derivative at the origin is related to the torsion tensor for the manifold. Finally, Colin de Verdiere has shown how to compute geodesic invariants from the Laplacian spectrum [3].

In a recent paper McDonald and Meyers [8] have shown that the heat-content of the heat-kernel is a permutation invariant. The heat content is the sum of the entries of the heat kernel over the nodes of the graph, which may be expanded as a polynomial in time. It is closely related to thre trace of the heat kernel, which is also known to be an invariant. In this paper we show how the co-efficients can be related to the eigenvalues and eigenvectors of the Laplacian. The resulting co-efficients are demonstrated to to outperform the Laplacian spectrum as a means of characterising graph-structure for the purposes of clustering.

2 Heat Kernels on Graphs

In this section, we review the how the heat-kernel is related to the Laplacian eigensystem. 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

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

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

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

where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_{|V|})$ is the diagonal matrix with the ordered eigenvalues $(0 = \lambda_1 < \lambda_2 \leq \lambda_3 ...)$ as elements and $\Phi = (\phi_1|\phi_2|....|\phi_{|V|})$ is the matrix with the correspondingly ordered eigenvectors as columns. Since $\hat{L}$ is symmetric and positive semi-definite, the eigenvalues of the normalised Laplacian are all positive. The eigenvector ϕ_2 associated with the smallest non-zero eigenvalue λ_2 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 heat kernel, i.e. the solution to the heat equation, is a $|V| \times |V|$ matrix found by exponentiating the Laplacian eigenspectrum, i.e. $h_t = \Phi \exp[-\Lambda t] \Phi^T$. For the nodes u and v of the graph G the resulting element is

$$h_t(u, v) = \sum_{i=1}^{|V|} \exp[-\lambda_i t] \phi_i(u) \phi_i(v) \tag{3}$$

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 $h_t \simeq \exp[-t\lambda_2] \phi_2 \phi_2^T$, where λ_2 is the smallest non-zero eigenvalue and ϕ_2 is the associated eigenvector, i.e. the Fiedler vector. Hence, the large time behavior 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

$$h_t(u, v) = \exp[-t] \sum_{k=1}^{|V|^2} P_k(u, v) \frac{t^k}{k!} \tag{4}$$

Hence, the heat kernel takes the form of a sum of Poisson distributions over the path-length with time as the parameter. The weights associated with the different components are determined by the associated path-length frequency in the graph. As the path-length k becomes large, the Poisson distributions approach a Gaussian, with mean k and variance k.

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

$$P_k(u, v) = \sum_{i=1}^{|V|} (1 - \lambda_i)^k \phi_i(u) \phi_i(v) \tag{5}$$

The geodesic distance between nodes can be found by searching for the smallest value of k for which $P_k(u, v)$ is non zero, i.e. $d_G(u, v) = floor_k P_k(u, v)$.

3 Invariants of the Heat-Kernel

It is well known that the trace of the heat-kernel is invariant to permutations. It is determined by the Laplacian eigenvalues and is given by

$$Z(t) = \sum_{i=1}^{N} \exp[-\lambda_i t] \tag{6}$$

To provide an illustration of the potential utility of the trace-formula, in Figure 1 we show four small graphs with rather different topologies. Figure 2 shows the trace of the heat kernel as a function of t for the different graphs. From the plot it is clear that the curves are distinct and could form the basis of a useful representation to distinguish graphs. For instance, the more bi-partite the graph the more stongly peaked the trace of the heat-kernel at the origin. This is due to the fact the spectal gap, i.e. the size of λ_2, determines the rate of decay of the trace with time, and this in turn is a measure of the degree of separation of the graph into strongly connected subgraphs or “clusters”.

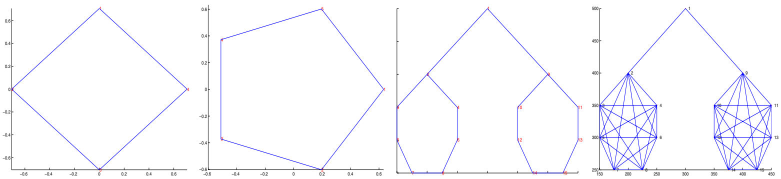

Fig. 1. Four graphs used for heat-kernel trace analysis.

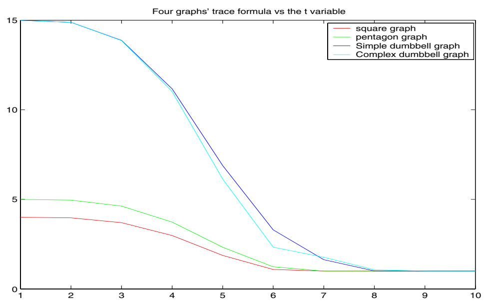

Fig. 2. Heat kernel trace as a function of t for four simple graphs.

Unfortunately, the trace of the heat kernel is limitted use for characterising graphs since for each value of time, it provides only a single scaler attribute. Hence, it must either be sampled with time or a fixed time value selected. However, in a recent paper McDonald and Meyers [8] have shown that the heat-content of the heat-kernel is also an invariant. The heat content is the sum of the entries of the heat kernel over the nodes of the graph and is given by

$$Q(t) = \sum_{u \in V} \sum_{v \in V} h_t(u,v) = \sum_{u \in V} \sum_{v \in V} \sum_{k=1}^{|V|} \exp[-\lambda_k t] \phi_k(u) \phi_k(v) \tag{7}$$

The heat-content can be expanded as a polynomial in time, i.e.

$$Q(t) = \sum_{m=0}^{\infty} q_m t^m \tag{8}$$

By equating the derivatives of the spectral and polynomial forms of the heat content at $t = 0$, the co-efficients are given by

$$q_m = \sum_{i=1}^{|V|} \sum_{u \in V} \sum_{v \in V} \frac{(-\lambda_i)^m}{m!} \phi_i(u) \phi_i(v) \tag{9}$$

In this paper, we will explore the use of the polynomial co-efficients for the purposes of graph-clustering. To do this we construct a vector $\boldsymbol{B} = (q_0, ..., q_5)^T$ from the first six co-efficients of the heat-content polynomial To compare our method with a standard spectral representation we also explore the use of the vector of leading Laplacian eigenvalues $\boldsymbol{B} = (\lambda_2, \lambda_2, ...\lambda_7)^T$ as a feature-vector.

4 Principal Components Analysis

Our aim is to construct a pattern-space for a set of graphs with pattern vectors $\boldsymbol{B}_k$, $k = 1, M$, extracted using heat-content co-efficients. 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 $\mathbf{S} = [\boldsymbol{B}_1 | \boldsymbol{B}_2 | \ldots | \boldsymbol{B}_k | \ldots | \boldsymbol{B}_M]$ 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 $\mathbf{C} = \sum_{i=1}^{M} l_i \boldsymbol{u}_i \boldsymbol{u}_i^T$ on the covariance matrix $\mathbf{C}$, where the l_i are the eigenvalues and the $\boldsymbol{u}_i$ 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 eigenspace is spanned by the s orthogonal vectors $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, .., \boldsymbol{u}_s)$. The individual graphs represented by the vectors $\mathbf{B}_k, k = 1, 2, \ldots, M$ can be projected onto this eigenspace using the formula $\boldsymbol{\mathcal{B}}_k = \mathbf{U}^T \boldsymbol{B}_k$. Hence each graph G_k is represented by an s-component vector $\boldsymbol{\mathcal{B}}_k$ in the eigenspace.

5 Experiments

We have applied our embedding method 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,

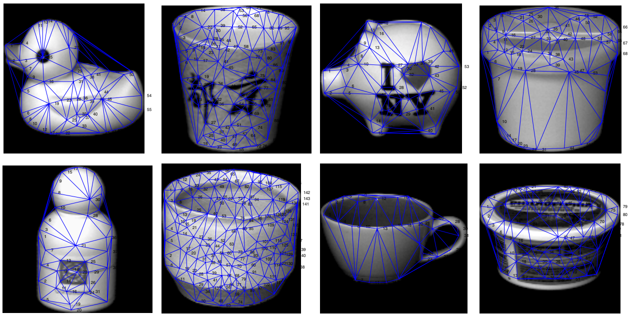

Fig. 3. Eight objects with their Delaunay graphs overlayed.

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 3. 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 embedding procedure has been applied to the resulting graph structures.

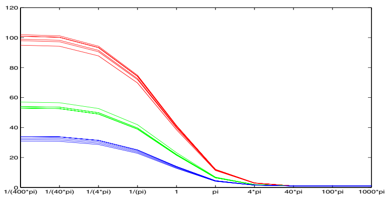

Fig. 4. Heat content as a function of t for 18 COIL graphs.

To commence, we show the heat-content as a function of t for six views of the the second, fifth and seventh objects from the COIL database shown above. From Figure 4 it is clear that objects of the same class trace out curves that are close together. To take this study further, in Figure 5 we plot the six co-efficients q_0, q_1, q_2, q_3, q_4 and q_5 separately as a function of the view number for the eight objects selected from the COIL data-base. The co-efficients are relatively stable with viewpoint. In the left-hand panel of Figure 6 we show the result of performing PCA on the vectors of polynomial co-efficients. For comparison, the right-hand panel in Figure 6 shows the corresponding result when we apply PCA to the vector of leading eigenvalues of the Laplacian matrix $B = (\lambda_2, \lambda_3, ..., \lambda_7)^T$ as the components of a feature vector instead. The main qualitative feature is that the different views of the ten objects are more overlapped than when the heat-content polynomial co-effients are used.

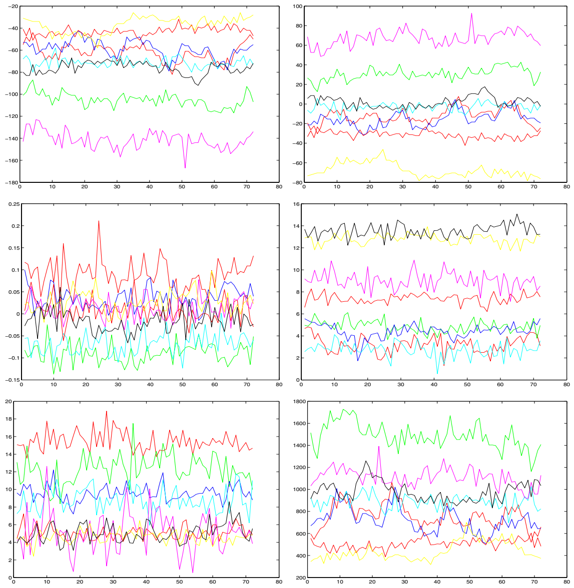

Fig. 5. Individual heat-content co-efficients as a function of view number.

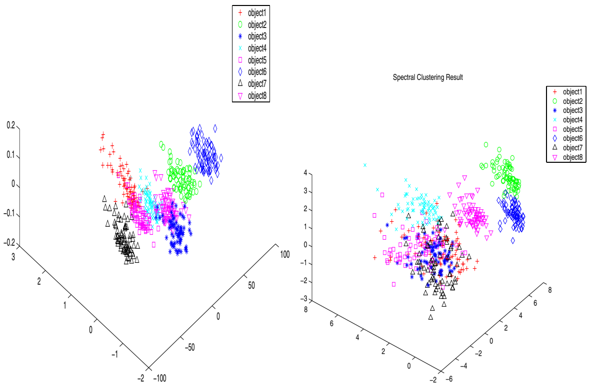

Fig. 6. Applying PCA to the heat-content differential co-efficients (left) and Laplacian spectrum (right).

To investigate the behavior of the two methods in a more quantitative way, we have computed the Rand index for the different objects. The Rand index is defined as $R_I = \frac{C}{C+W}$ where C is the number of "agreements" and W is the number of "disagreements" in cluster assignment. The index is hence the fraction of views of a particular class that are closer to an object of the same class than to one of another class. For the heat-content co-efficients, the Rand index is 0.88 while for the Laplacian eigenvalues it is 0.58.

6 Conclusion and Future Work

In this paper we have explored how the use of heat-content can lead to a series of invariants that can be used for the purposes of clustering. There are clearly a number of ways in which the work reported in this paper can be extended. These include the use of features which have a direct geometrical meaning such as the Euler characteristic, the torsion of the mean and Gaussian curvatures of the manifold.

References

1. A.D.Alexandrov and V.A.Zalgaller. Intrinsic geometry of surfaces. *Transl.Math. Monographs*, 15, 1967.
2. J. E. Atkins, E. G. Bowman, and B. Hendrickson. A spectral algorithm for seriation and the consecutive ones problem. *SIAM J. Comput.*, 28:297–310, 1998.
3. Colin de Verdiere. Spectra of graphs. *Math of France*, 4, 1998.
4. F.R.K.Chung. Spectral graph theory. *American Mathematical Society*, 1997.
5. G.R.Hjaltason and H.Samet. Properties of embedding methods for similarity searching in metric spaces. *PAMI*, 25:530–549, 2003.
6. C.G. Harris and M.J. Stephens. A combined corner and edge detector. *Fourth Alvey Vision Conference*, pages 147–151, 1994.
7. N.Linial, E.London, and Y.Rabinovich. The geometry of graphs and some its algorithmic application. *Combinatorica*, 15:215–245, 1995.
8. P.Mcdonald and R.Meyers. Diffusions on graphs, poisson problems and spectral geometry. *Transactions on Amercian Mathematical Society*, 354:5111–5136, 2002.
9. Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE PAMI*, 22:888–905, 2000.
10. S.T.Yau and R.M.Schoen. Differential geometry. *Science Publication*, 1988.
11. S.Umeyama. An eigen decomposition approach to weighted graph matching problems. *IEEE PAMI*, 10:695–703, 1988.
12. S.Weinberger. Review of algebraic l-theory and topological manifolds by a.ranicki. *BAMS*, 33:93–99, 1996.
13. X.Bai and E.R.Hancock. Heat kernels, manifolds and graph embedding. *IAPR Workshop on Structural,Syntactic, and Statistical Pattern Recognition, Lecture Notes in Computer Science 3138*, pages 198–206, 2004.