Graph characteristics from the heat kernel trace

Bai Xiao [a,*], Edwin R. Hancock [b], Richard C. Wilson [b]

[a] Department of Computer Science School, Beihang University, Beijing 100191, PR China
[b] Department of Computer Science, University of York, York Y010 5DD, UK

ARTICLE INFO

ABSTRACT

Graph structures have been proved important in high level-vision since they can be used to represent structural and relational arrangements of objects in a scene. One of the problems that arises in the analysis of structural abstractions of objects is graph clustering. In this paper, we explore how permutation invariants computed from the trace of the heat kernel can be used to characterize graphs for the purposes of measuring similarity and clustering. The heat kernel is the solution of the heat equation and is a compact representation of the path-length distribution on a graph. The trace of the heat kernel is given by the sum of the Laplacian eigenvalues exponentiated with time. We explore three different approaches to characterizing the heat kernel trace as a function of time. Our first characterization is based on the zeta function, which from the Mellin transform is the moment generating function of the heat kernel trace. Our second characterization is unary and is found by computing the derivative of the zeta function at the origin. The third characterization is derived from the heat content, i.e. the sum of the elements of the heat kernel. We show how the heat content can be expanded as a power series in time, and the coefficients of the series can be computed using the Laplacian spectrum. We explore the use of these characterizations as a means of representing graph structure for the purposes of clustering, and compare them with the use of the Laplacian spectrum. Experiments with the synthetic and real-world databases reveal that each of the three proposed invariants is effective and outperforms the traditional Laplacian spectrum. Moreover, the heat-content invariants appear to consistently give the best results in both synthetic sensitivity studies and on real-world object recognition problems.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Structural information has been shown to perform an important role in object recognition. This is true not only when graph based abstractions of shape are being used [8,51,48], but also when local features are used as object descriptors [49,47]. Recent contributions to this latter endeavor include the "bag of words" representation of spatial relationships [49,27] based on SIFT features [50] and the use of constellations to learn and recognize object classes [47]. However, the use of local graph structures to capture feature arrangement falls well short of using a graph to capture global scene structure. Moreover, graph structures can be used to represent higher level information, and hence are closer to the semantic content of a scene.

One of the most elegant way of characterizing graph structure is to use spectral methods, which make use of the eigenvalues of the adjacency matrix or the Laplacian matrix (the adjacency matrix minus the degree matrix). In fact, the Laplacian spectrum of a graph has found widespread use in computer vision for a number of applications including segmentation [6], routing [7], clustering [14] and indexing [8]. 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 [11]. According to the heat equation, the time derivative of the kernel is determined by the graph Laplacian. The solution to the heat equation is hence 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 over time, it is closely related to the path-length distribution on the graph. An important permutation invariant property of the heat kernel is the trace. This is found by summing and exponentiating the Laplacian eigenvalues with time. Although Lebanon and Lafferty [10] have explored the use of an embedding based on the heat kernel for data clustering, the heat kernel trace has not been thoroughly investigated as a structural representation in computer vision.

The aim of this paper is to overcome this omission by investigating whether the trace of the heat kernel [11] can be used for the purposes of characterizing the properties of graphs. Here we explore

---

* Corresponding author.
  E-mail address: little_point_baixiao@hotmail.com (B. Xiao).

whether the shape of this function can be used to characterize the corresponding graph.

1.1. Related literature

One of the earliest applications of the Laplacian spectrum for pattern analysis was to use the Fiedler vector [1,15,16], i.e. the eigenvector associated with the smallest non-zero eigenvalue, to perform pairwise clustering of data. In addition, several authors have explored the use of the Laplacian and related operators to map data to manifolds in a low dimensional space [13,18–20,9]. The mapping is usually found by expressing the kernel as a Gram matrix, i.e. a matrix formed by the self-inner product of a co-ordinate matrix, and the Young–Householder decomposition allows the co-ordinate matrix to be recovered from the eigenvalues and eigenvectors of the kernel matrix. These methods share the feature of using the spectrum of the Laplacian matrix to map data specified in terms of a proximity matrix to a vector space. For instance in the Laplacian eigenmap [18], the mapping is determined by the raw Laplacian spectrum. The diffusion map [9] of Lafon and Coifman constructs the mapping by raising the Laplacian eigensystem to a negative integer power. This mapping is shown to preserve the distances between nodes under a random walk, or diffusion, on the graph.

The Laplacian eigenvalues [1] can also be used to characterize graphs for the purposes of clustering. For instance Luo et al. [14] have shown how to use the Laplacian spectrum to cluster graphs. Here the ordered eigenvalues are used as a graph feature-vector, and the order of components is invariant to the permutation order of the nodes. Additional permutation invariants that are available include the trace of the Laplacian and its determinant (which is always zero). In subsequent work [23], the same authors showed how to extract a rich family of permutation invariants from a graph by applying symmetric polynomials to the elements of the spectral matrix (the eigenvector matrix premultiplied by the square root of the eigenvalue matrix).

The aim in this paper is to explore whether the heat kernel can also provide useful permutation invariants for the purposes of characterizing graphs. Since the heat kernel admits the addition of time as a parameter, it allows the scale of different substructures to be probed. Again, the simplest invariants that are to hand are the trace and the determinant (non-zero) of the heat kernel. However, for large graphs, the Laplacian can be viewed as a discrete approximation of the Laplace–Beltrami operator [17]. As a result, for such graphs invariant characterizations can also be imported from spectral geometry [11,12]. Here the heat kernel trace has been used to characterize the differential geometry of manifolds using the spectrum of the Laplace–Beltrami operator. From the paramatrix construction of the heat kernel it follows that the trace of the heat kernel admits an asymptotic expansion. Moreover, the coefficients of this expansion are integrals of universal polynomials in the Riemann curvature tensor and its derivatives. However, these coefficients have only been tabulated for low lying values in the expansion [12]. The zeta function (i.e. the sum of the eigenvalues raised to a non-integer power), on the other hand also leads to invariants that are more tractable. As we will show later in this paper, from the Mellin transform it follows that the zeta function is the moment generating function of the heat kernel trace. It hence measures the shape of the heat kernel trace as a function of time. For graphs, the derivative of the zeta function (i.e. its slope) at the origin also yields important information. Simple analysis shows that it is given by the logarithm of the reciprocal of the product of the non-zero Laplacian eigenvalues. It is hence linked to the determinant of the Laplacian.

Another way to characterize the time-dependence of the heat kernel, is to use heat content. The heat content is similar in structure to the heat kernel trace. It is the sum of the complete set of elements of the heat kernel rather than just the sum of its diagonal elements. Gilkey [4] has discussed the heat-content asymptotic in his recent paper. McDonald and Meyers [5] establish that heat content distinguishes some well known examples of isospectral graphs. In a recent paper McDonald and Meyers [24] have shown that the heat content can be expanded purely as a power series in time. The coefficients of the power series are permutation invariants that are linked by Poisson's formula to the eigenvalues and eigenvectors of the graph Laplacian. Moreover, they are more easily computed than the time coefficients in the paramatrix expansion.

1.2. Contribution

As previously stated, the aim in this paper is to explore whether permutation invariants linked to the heat kernel trace can be used for the purposes of characterizing graph structure for the purposes of clustering. Motivated by the literature above, we explore three approaches to the problem. The first of these is to measure the shape of the heat kernel trace by taking moments with respect to time. As noted above, from the Mellin transform the zeta function is the moment generating function of the heat kernel trace. We construct a feature vector whose components are the values of the zeta function with integer argument. The second approach is to explore the derivative of the zeta function at the origin. Also, as noted above, this is related to the product of the non-zero eigenvalues and is thus one of the polynomial invariants studied by Wilson et al. [23]. We hence compare the performance of the zeta-function derivative with the trace of the heat kernel and the set of polynomial invariants. Thirdly, we explore the use of heat content. We expand the heat content as a power series in time, and demonstrate that the coefficients can be related to the eigenvalues and eigenvectors of the Laplacian. Experiments with real-world data taken from the COIL and Caltech–Oxford databases reveal that both the zeta function and its derivative at the origin are useful features for clustering graphs. The zeta function is shown to outperform the Laplacian spectrum as a means of characterizing graph structure.

Although the characterizations used in our study are drawn from the literature on spectral graph theory, we offer a number of novel contributions. First, the representations have not been compared and contrasted before in an experimental setting. Second, our analysis of the heat kernel trace provides links with both the zeta function and the number of spanning trees contained in a graph. We hence use the unified viewpoint furnished by the heat kernel trace to provide links between the different characterizations. This in turn furnishes insights into their meaning and interelationships.

The outline of this paper is as follows. In Section 2 we explain the relationship between the Laplacian eigensystem and the heat kernel trace. Section 3, we develop the three sets of invariants deduced from graph heat kernel which can be used to characterize the graphs. Section 4 performs a sensitivity study on synthetic graphs, and determines the relative sensitivity of the three methods and compares them to the Laplacian spectrum and symmetric polynomials. In Section 5 we turn our attention to real-world data and evaluate the performance of our new characterizations on a number of object recognition and shape analysis problems. Finally, Section 6 offers some conclusions and directions for future research.

2. Heat kernel on graphs

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, which has elements

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

We also construct the diagonal degree matrix D, whose elements are given by $D(u,u) = deg_u = \sum_{v \in V} A(u,v)$. 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 Laplacian has elements

$$L(u,v) = \begin{cases} d_v & \text{if } u = v \\ -1 & \text{if } u \text{ and } v \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The normalized Laplacian is given by $\hat{L} = D^{-1/2} L D^{-1/2}$ and has elements

$$\hat{L}(u,v) = \begin{cases} 1 & \text{if } u = v \text{ and } d_v \neq 0 \\ -\dfrac{1}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The spectral decomposition of the normalized Laplacian matrix is $\hat{L} = \Phi \Lambda \Phi^{\mathrm{T}}$, where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_{|V|})$ is the diagonal matrix with the ordered eigenvalues ($0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{|V|}$) as elements and $\Phi = (\phi_1 | \phi_2 | \cdots | \phi_{|V|})$ is the matrix with the corresponding ordered unit-norm eigenvectors as columns. Since $\hat{L}$ is symmetric and positive semi-definite, the eigenvalues of the normalized Laplacian are all non-negative. The number of zero eigenvalues is the number of connected components in the graph. For a connected graph, there is only one eigenvalue which equals zero. The eigenvector ϕ_2 associated with the smallest non-zero eigenvalue λ_2 is referred to as the Fiedler-vector [1].

We are interested in the heat equation associated with the Laplacian, and this is given by

$$\frac{\partial h_t}{\partial t} = -\hat{L} h_t \tag{4}$$

where h_t is the heat kernel and t is time. The heat kernel is the fundamental solution of the heat equation. It can 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

$$h_t = \mathrm{e}^{-t\hat{L}} \tag{5}$$

We can compute the heat kernel on a graph by exponentiating the Laplacian eigenspectrum (see e.g. [1]), i.e.

$$h_t = \Phi \exp[-\Lambda t] \Phi^{\mathrm{T}} = \sum_{i=1}^{|V|} \exp[-\lambda_i t] \phi_i \phi_i^{\mathrm{T}} \tag{6}$$

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

$$h_t(u,v) = \sum_{i=1}^{|V|} \exp[-\lambda_i t] \phi_i(u) \phi_i(v) \tag{7}$$

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 $h_t \simeq \exp[-\lambda_2 t]\phi_2 \phi_2^{\mathrm{T}}$, 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.
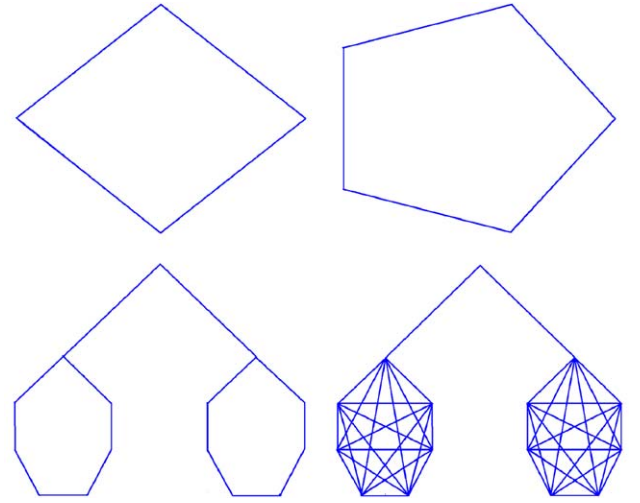

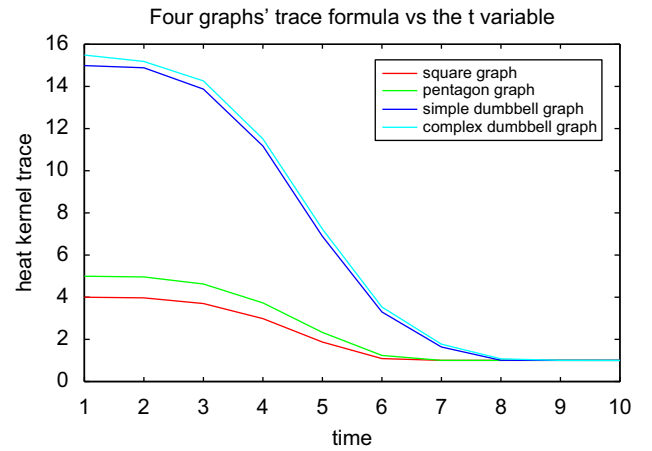Fig. 1. Four graphs used for the heat kernel trace analysis.

Fig. 2. Heat kernel trace as a function of t for the four graphs from Fig. 1.

3. Heat kernel invariants

In this section we will describe a number of invariants that can be computed from the heat kernel matrix of the graph. These include the heat kernel trace, the zeta function, heat-content invariants, and the derivative of zeta function at the origin.

3.1. Heat kernel trace

The trace of the heat kernel is the sum of the diagonal elements of the heat kernel matrix of the graph:

$$Z(t) = \mathrm{Tr}[h_t] = \sum_{i=1}^{|V|} \exp[-\lambda_i t] \tag{8}$$

To provide an illustration of the potential utility of the trace formula, in Fig. 1 we show four small graphs with rather different topologies. Fig. 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 have a distinct shape and could form the basis of a useful representation to distinguish graphs. For instance, the more "dumbbell" shaped the graph is the more strongly peaked the trace of the heat kernel at the origin. This is due to the fact that the spectral gap, i.e. the size of λ_2, determines the rate of decay of the trace with time, and this in turn
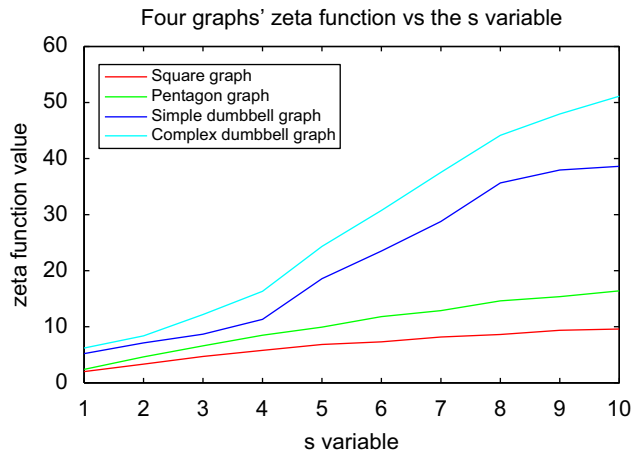
Fig. 3. Zeta function as a function of the variable s for the four graphs from Fig. 1.

is a measure of the degree of separation of the graph into strongly connected subgraphs or "clusters", and this is again due to the fact that for a connected graph, $\lambda_1 = 0$ and as a result

$$Z(t) = 1 + \exp[-\lambda_2 t] + \exp[-\lambda_3 t] + \cdots + \exp[-\lambda_N t]$$

3.2. Zeta function and heat kernel trace moments

The purpose in this section is to use the shape of the heat kernel trace function as a means of characterizing graph structure. Our characterization is found by taking moments of trace function over time.

To commence our development, we consider the zeta function associated with the Laplacian eigenvalues. The zeta function is defined by

$$\zeta(s) = \sum_{\lambda_i \neq 0} \lambda_i^{-s} \tag{9}$$

In other words, it is the result of exponentiating and summing the reciprocal of the non-zero Laplacian eigenvalues. In Fig. 3 we show $\zeta(s)$ as a function of s for the four graphs in Fig. 1. Again the curves are distinct and provide a potential route to distinguish the graphs.

To establish the link between the zeta function and the trace of the heat kernel we make use of the Mellin transform. The Mellin transform of the function $f(t)$ is

$$w(s) = \int_0^\infty t^{s-1} f(t)\, dt \tag{10}$$

Hence, the heat kernel trace has Mellin transform

$$\int_0^\infty \mathrm{Tr}[h_t] t^{s-1}\, dt = \int_0^\infty \sum_{i=1}^{|V|} e^{-\lambda_i t} t^{s-1}\, dt \tag{11}$$

Suppose that the zero eigenvalue $\lambda_1 = 0$ has multiplicity C. Removing this eigenvalue from the sum under the integral, we have

$$\int_0^\infty \mathrm{Tr}[h_t] t^{s-1}\, dt = \int_0^\infty \left[C + \sum_{\lambda_i \neq 0} e^{-\lambda_i t} \right] t^{s-1}\, dt$$

To develop this expression for the heat kernel trace, we note that the gamma function $\Gamma(s)$ is defined to be

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t}\, dt \tag{12}$$

As a result

$$\int_0^\infty t^{s-1} e^{-\lambda_i t}\, dt = \frac{\Gamma(s)}{(-\lambda_i)^s}$$

Hence, we can write

$$\int_0^\infty (\mathrm{Tr}[h_t] - C) t^{s-1}\, dt = \sum_{\lambda_i \neq 0} \frac{\Gamma(s)}{(-\lambda_i)^s}$$

So finally, the zeta function is given by

$$\zeta(s) = \sum_{\lambda_i \neq 0} (-\lambda_i)^{-s} = \frac{1}{\Gamma(s)} \int_0^\infty (\mathrm{Tr}[h_t] - C) t^{s-1}\, dt \tag{13}$$

In other words, the zeta function is related to the Mellin moments of the heat kernel trace. It is hence a way to characterizing the shape of the heat kernel trace with time.

3.3. Zeta-function derivative

The zeta function is also linked to the determinant of the Laplacian. To show this, we re-write the zeta function in terms of the natural exponential with the result

$$\zeta(s) = \sum_{\lambda_i \neq 0} \lambda_i^{-s} = \sum_{\lambda_i \neq 0} \exp[-s \ln \lambda_i] \tag{14}$$

The derivative of the zeta function is given by

$$\zeta'(s) = \sum_{\lambda_i \neq 0} \{-\ln \lambda_i\} \exp[-s \ln \lambda_i] \tag{15}$$

At the origin the derivative takes on the value

$$\zeta'(0) = \sum_{\lambda_i \neq 0} \{-\ln \lambda_i\} = \ln \left\{ \prod_{\lambda_i \neq 0} \left(\frac{1}{\lambda_i} \right) \right\} \tag{16}$$

It is interesting to note that McKay [46] has shown that the derivative of the zeta function at the origin is linked to the number of spanning trees in a graph G through

$$\tau(G) = \frac{\prod_{u \in V} deg_u}{\sum_{u \in V} deg_u} \exp[-\zeta'(0)] \tag{17}$$

As a result, the derivative of the zeta function at the origin is determined by the number of spanning trees in the graph together with the degrees of its vertices. In our experiments, we will use the value of $\zeta'(0)$ as a graph characteristic, and explore whether it can be used to distinguish different graphs.

3.4. Unary attributes with symmetric polynomials

In this section we illustrate the relationship between the derivative of the zeta function at the origin and the symmetric polynomial invariants of Wilson et al. [23].

In their recent paper Wilson et al. [23] have reported a family of permutation invariants that can be computed by applying symmetric polynomials to the elements of the spectral matrix $\sqrt{\Lambda}\Phi$. A symmetric polynomial of n variables $v_1, v_2, \ldots, v_n$ is a function that is unchanged by permutation of its variables. In other words, the symmetric polynomials satisfy

$$f(y_1, y_2, \ldots, y_n) = f(v_1, v_2, \ldots, v_n) \tag{18}$$

where $y_i = v_{\pi(i)}$ and π is an arbitrary permutation of the indices $1, 2, \ldots, n$.

There are two sets of commonly used symmetric polynomials. These are the elementary symmetric polynomials and the power-sum symmetric polynomials. For a set of variables $\{v_1, v_2, \ldots, v_n\}$ the elementary symmetric polynomials are defined as

$$S_1(v_1, \ldots, v_n) = \sum_{i=1}^{n} v_i$$

$$S_2(v_1, \ldots, v_n) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} v_i v_j$$

$$\vdots$$

$$S_r(v_1, \ldots, v_n) = \sum_{i_1 < i_2 < \cdots < i_r} v_{i_1} v_{i_2} \ldots v_{i_r}$$

$$\vdots$$

$$S_n(v_1, \ldots, v_n) = \prod_{i=1}^{n} v_i$$

The output for the n set input variable $\{v_1, v_2, \ldots, v_n\}$ is $S_1, S_2, \ldots, S_n$.

The power-sum symmetric polynomial functions, on the other hand, are defined to be

$$P_1(v_1, \ldots, v_n) = \sum_{i=1}^{n} v_i$$

$$P_2(v_1, \ldots, v_n) = \sum_{i=1}^{n} v_i^2$$

$$\vdots$$

$$P_r(v_1, \ldots, v_n) = \sum_{i=1}^{n} v_i^r$$

$$\vdots$$

$$P_n(v_1, \ldots, v_n) = \sum_{i=1}^{n} v_i^n$$

and also form a basis set over the set of symmetric polynomials. Any polynomial function which is invariant to permutation of the variable indices $(v_i, v_2, \ldots, v_n)$ can be expressed in terms of one of these two sets of polynomials. Moreover, the two sets of polynomials are related to one another by the Newton–Girard formula:

$$S_r = \frac{(-1)^{r+1}}{r} \sum_{k=1}^{r} (-1)^{k+r} P_k S_{r-k} \tag{19}$$

where we have used the shorthand S_r for $S_r(v_1, \ldots, v_n)$ and P_r for $P_r(v_1, \ldots, v_n)$. As a consequence, the elementary symmetric polynomials S_r can be efficiently computed using the power-sum symmetric polynomials. In this paper we will compare our characterizations derived from the heat kernel to the symmetric polynomials of the spectral matrix used by Wilson et al. [23].

There are of course alternative invariants that can be computed from the spectrum of the Laplacian. The most obvious of these is the sum of eigenvalues or the trace of the normalized Laplacian eigenvalue matrix of the graph:

$$S_1(\lambda_1, \lambda_2, \ldots, \lambda_n) = \sum_{i=1}^{n} \lambda_i \tag{20}$$

When the symmetric polynomials take the normalized Laplacian eigenvalues as inputs, then the lowest order polynomial is the sum of the eigenvalues, i.e. the trace of the normalized Laplacian, and
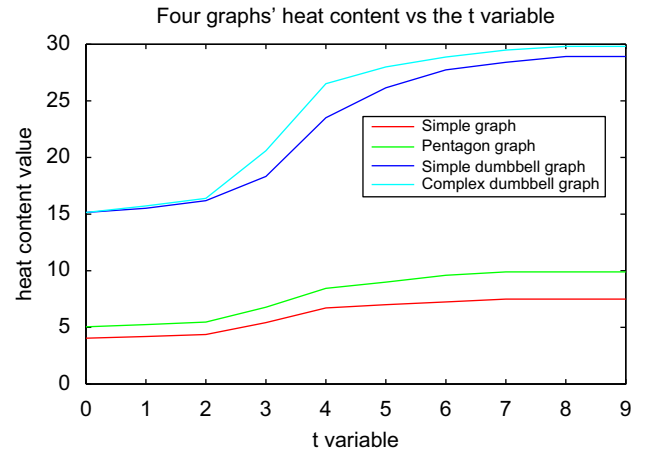

Four graphs' heat content vs the t variable

Fig. 4. Heat content as a function of t for the four graphs from Fig. 1.

the highest order polynomial is the product of non-zero normalized Laplacian eigenvalues. Hence, $S_{|V|}(\lambda_1, \lambda_2, \ldots, \lambda_{|V|}) = \exp[(\zeta'(0))^{-1}]$ where $|V|$ is the number of nodes in the graph. Moreover, in [23], Wilson et al. found it necessary to perform logarithmic squashing of the polynomials to bring them into a convenient dynamic range. This property arises naturally through the analysis of the zeta function.

In our experiments, we will use the symmetric polynomial to construct unary attributes for graph characterization. The inputs for the symmetric polynomial are the columns of the spectral matrix and the eigenvalues of the normalized Laplacian matrix.

3.5. Heat-content invariants

We now turn our attention to heat-content invariants. In a recent paper McDonald and Meyers [24] have described a set of differential invariants that can be derived from the heat content of the heat kernel. The heat content is the sum of the entries of the heat kernel over the nodes of the graph and is given by

$$Q(t) = \sum_{u \in V} \sum_{v \in V} h_t(u, v) = \sum_{u \in V} \sum_{v \in V} \sum_{k=1}^{|V|} \exp[-\lambda_k t] \phi_k(u) \phi_k(v) \tag{21}$$

In Fig. 4, we show the heat content varies with t for the four graphs from Fig. 1. Yet again, the curves have different shapes which depend on whether the graph is "dumbbell" or not. As result heat content can be used as a means for graph characterization.

McDonald and Meyers [24] have shown that the heat content can be expanded as a power series in time, i.e.

$$Q(t) = \sum_{m=0}^{\infty} q_m t^m \tag{22}$$

Using the MacLaurin series $\exp[-\lambda_k t]$ can be expanded as

$$\exp[-\lambda_k t] = 1 + (-\lambda_k)t + \frac{(-\lambda_k)^2 t^2}{2!} + \cdots + \frac{(-\lambda_k)^m t^m}{m!} + \cdots$$

$$= \sum_{m=0}^{\infty} \frac{(-\lambda_k)^m t^m}{m!} \tag{23}$$

Substituting the MacLaurin series with Eq. (21), we have

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

$$= \sum_{m=0}^{\infty} \sum_{u \in V} \sum_{v \in V} \sum_{k=1}^{|V|} \phi_k(u) \phi_k(v) \frac{(-\lambda_k)^m t^m}{m!} \tag{24}$$

As a result the coefficients q_m appearing in the power series expansion of $Q(t)$ are given by

$$q_m = \sum_{k=1}^{|V|} \left\{ \left(\sum_{u \in V} \phi_k(u) \right)^2 \right\} \frac{(-\lambda_k)^m}{m!} \qquad (25)$$

As a result, the time-series coefficient q_m does not depend on the permutation order of the nodes of the graph. Moreover, it is similar in spirit to the permutation invariants described by Wilson et al. in [23], since it involves both eigenvalues and eigenvectors of the normalized Laplacian.

The time-series coefficients q_m are a set of invariants which can be used for graph characterization. Since they depend on both eigenvalues and eigenvectors, they are more likely to be unique than invariants purely based in eigenvalues which are likely to be affected by problems of co-spectrality [31]. We will explore the use of the time-series coefficients for the purpose of graph clustering. To do this

we construct a feature vector $\vec{B} = (q_1, q_2, \ldots, q_k)^T$ from the k leading coefficients of the heat-content power series.

4. Comparing the heat kernel invariants on synthetic data

Before we apply our new methods for characterizing graphs to real world image databases, we first quantitatively investigate their behavior on synthetic data. This study is divided into two parts. We commence by comparing the three new methods with the symmetric polynomial representation, and then evaluate their stability.

4.1. Graph characterization

In the first experiment, we test our methods for graph characterization on synthetic graphs. To this end, we generate five different seed graphs with different structure, i.e. number of nodes and
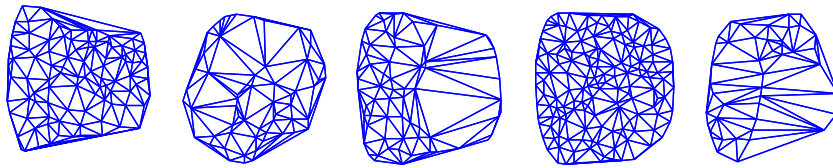

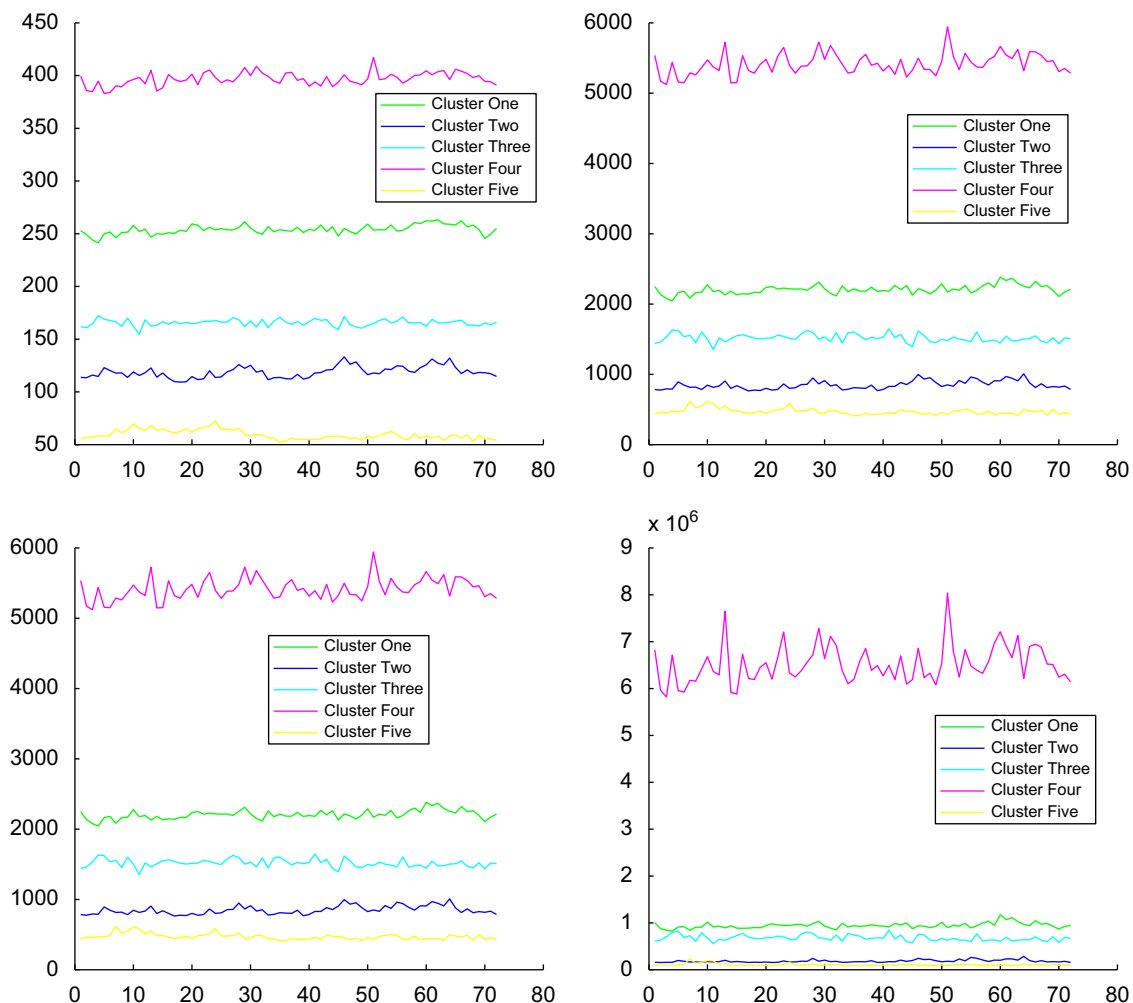**Fig. 5.** Examples of the seed graphs.

Fig. 6. Zeta function $\zeta(s)$ for randomizations of the five seed graphs. From left to right, and top to bottom, $s = 1, 2, 3$ and 4, respectively.

number of edges. We then apply random edit operations to the five seed graphs to simulate the effects of noise. The edit operations are edge deletion, edge insertion and node deletion. The purpose of these operations is to generate five structurally related graph clusters, which we would hope to separate using our methods. In Fig. 5, we show the original seed graphs. For each seed graph, we randomly perform 3–10 edit operations to obtain 70 graph variants.
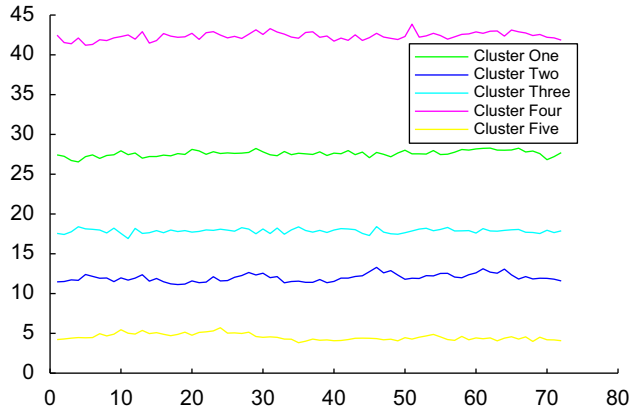

Fig. 7. Derivative of the zeta function at the origin for five clusters of random graphs.

We commence by illustrating the behavior of the zeta function. From left-to-right and top-to-bottom in Fig. 6 we show the values of $\zeta(1)$, $\zeta(2)$, $\zeta(3)$ and $\zeta(4)$ for the 70 randomizations of each seed graph. In each plot the different curves are for the different seed graphs, and the x-axis is the index number of the randomization. The main feature to note is that the curves for the different graph clusters are well separated, and that the individual values of the zeta function do not vary significantly with randomization index. When $s = 1$ and 2, the fluctuations $\zeta(s)$ shown in Fig. 6 are generally small. However, the fluctuation is larger when $s = 4$. This is because when s becomes large the small eigenvalues will play a dominant role in the ζ function.

We now turn our attention to the derivative of the zeta function at the origin. In Fig. 7 we plot the curves $\zeta'(0)$ for the five random graph clusters against randomization index. The curves are well separated for the different objects, and the fluctuations are relatively small.

In Fig. 8, we plot the four leading heat-content invariants coefficients q_1, q_2, q_3 and q_4 separately as a function of the randomization index for the five seed graphs. The coefficients are relatively stable under randomized edit operations.

For the purposes of comparison, in Fig. 9 we plot the mean values of the symmetric polynomials for the five graph clusters, as a function of polynomial order. Here the arguments of the polynomials are the column elements of the spectral matrix $\sqrt{\Lambda}s\Phi$. The different curves in the plot are for the different seed graphs. We select nine different symmetric polynomials for investigation. These are the first
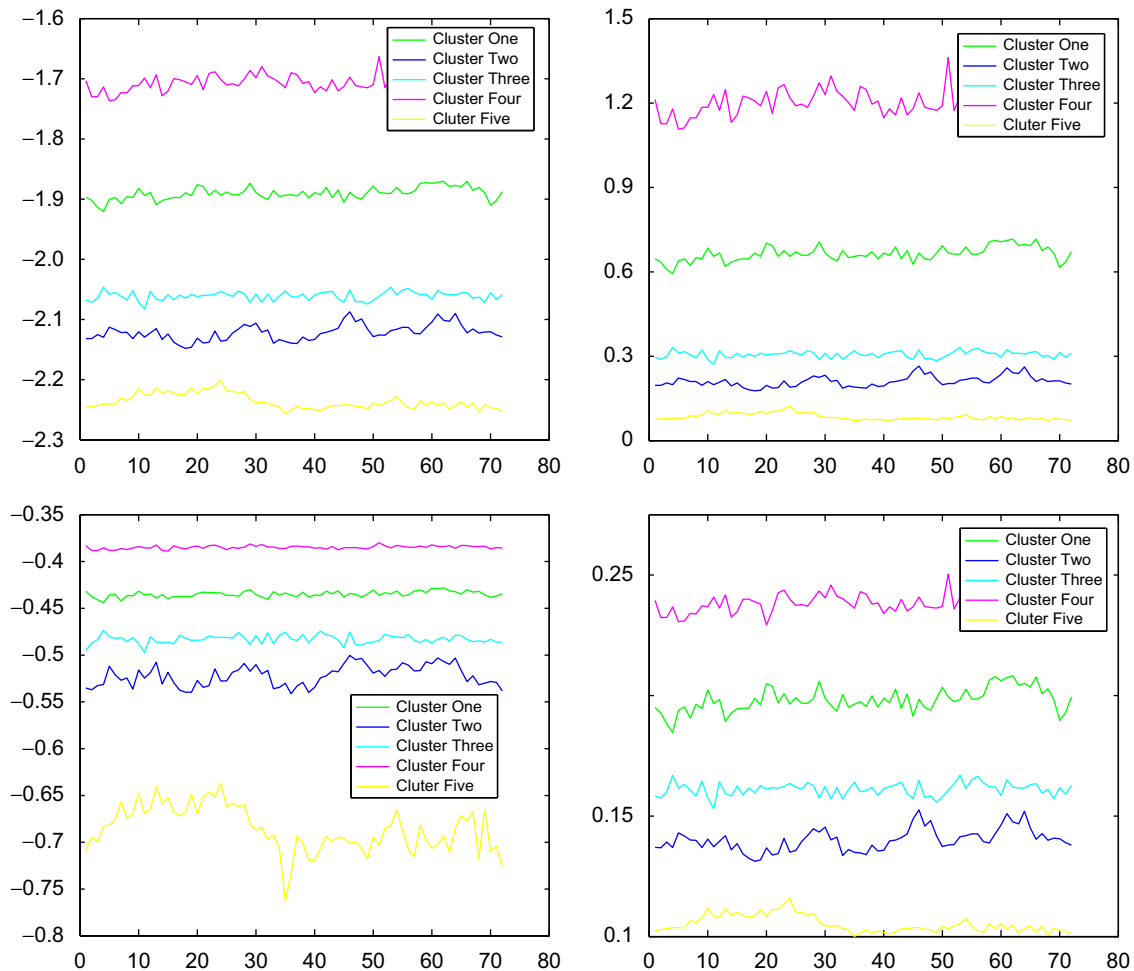

Fig. 8. Individual heat-content invariants q_n for five clusters of random graphs (from left to right, and top to bottom, $n = 1, 2, 3$ and 4, respectively).
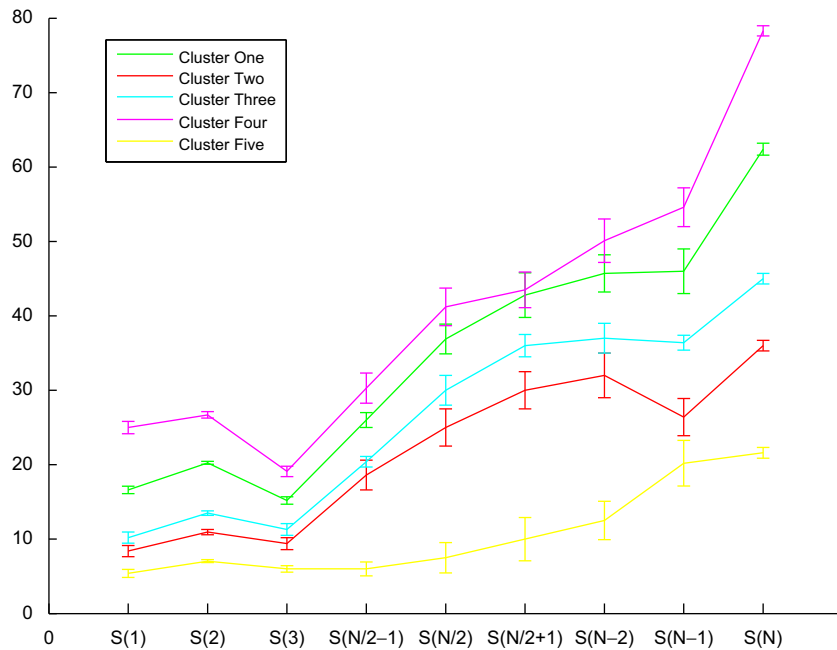
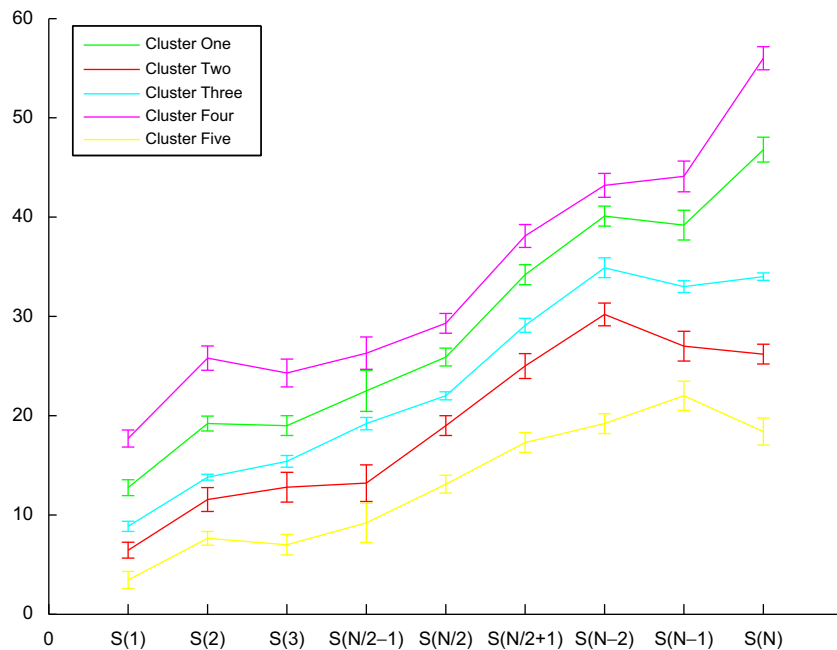**Fig. 9.** Symmetric polynomials using spectral matrix elements for five clusters of random graphs.

Fig. 10. Symmetric polynomials using normalized Laplacian eigenvalues for five clusters of random graphs.

three symmetric polynomials S_1, S_2, S_3, the middle three symmetric polynomials $S_{fix(N/2)-1}$, $S_{fix(N/2)}$, $S_{fix(N/2)+1}$ and last three symmetric polynomials S_{N-2}, S_{N-1}, S_N, where N is the number of node in the graph. The error bars show the standard deviation over the different sets of graphs of the same cluster. Fig. 10 repeats the analysis using the normalized Laplacian spectrum as the arguments of the polynomials.

From Figs. 10 and 9 it is clear that the first three and last three symmetric polynomials computed using the normalized Laplacian eigenvalues and spectral matrix elements as inputs can be used for graph clustering. We also get a better separation of objects when

using the polynomials from the normalized Laplacian spectrum rather than the elements of the spectral matrix.

4.2. Stability evaluation

We now evaluate the stability of the different graph characterization methods. We study six different methods. These are (a) the derivative of zeta function at origin, (b) the heat content invariants, (c) the zeta-function values, (d) the symmetric polynomials with Laplacian eigenvalues as arguments, (e) the symmetric polynomials with spectral matrix elements as arguments and (f) the Laplacian
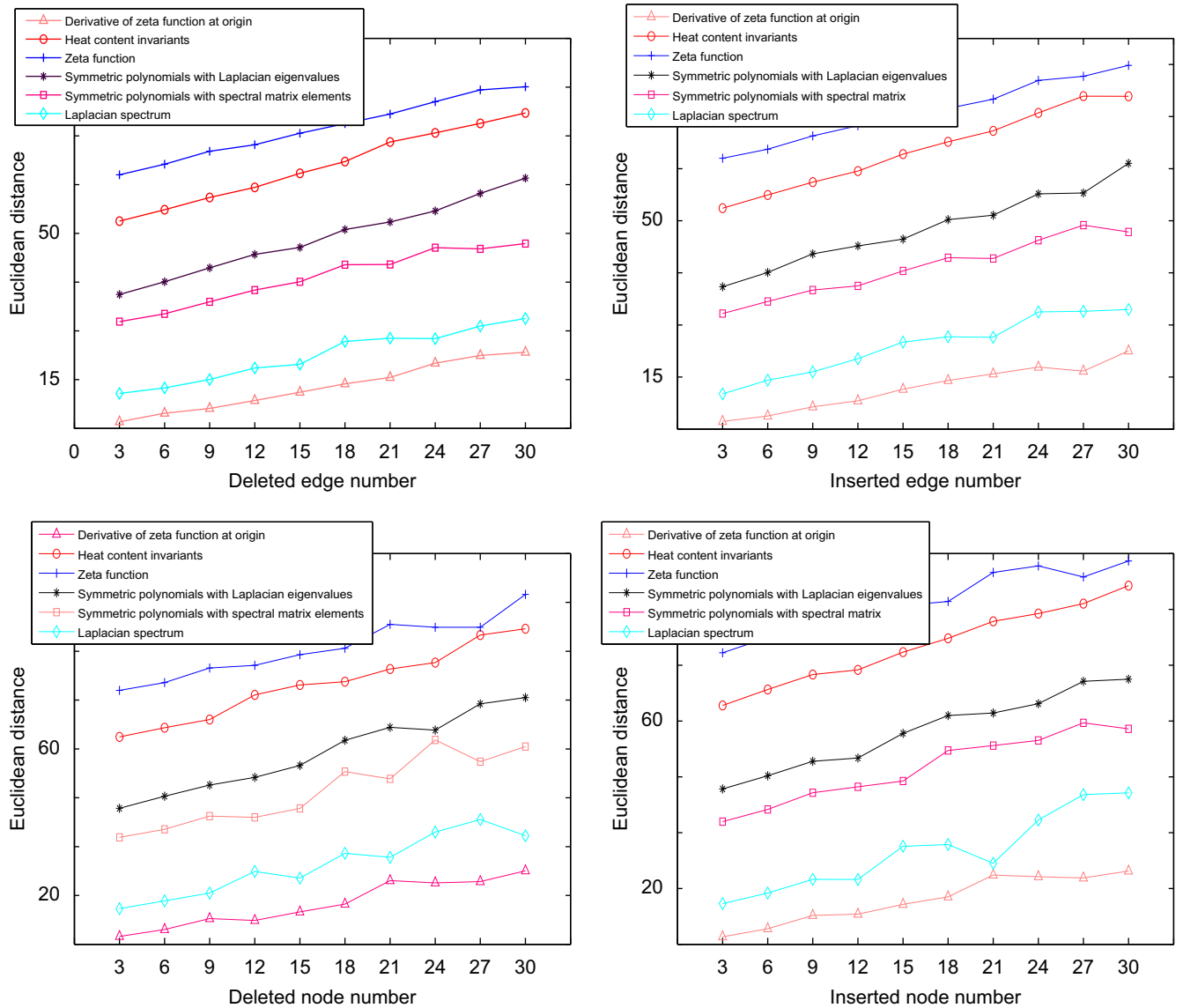
Fig. 11. Euclidean distances from the different graph characterization methods for random graphs (from left to right, and top to bottom, we show edge deletion, edge insertion, node deletion and node insertion, respectively). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

spectrum. We generate a random seed graph with 65 vertices and 170 edges. Modified graph variants are obtained by performing edit operations on the seed graphs. Here we use four different edit operations, namely edge deletion, edge insertion, node deletion and node insertion. We compute the Euclidean distance between the graph feature vector for the original seed graph and the feature vector for its modified variant obtained after the application of the edit operations. For the heat-content invariants, the zeta function values and the Laplacian spectrum, we select the leading six values to construct the required feature vector. For the symmetric polynomials we select the first three and last three values. As a result the feature vectors each have six components. The exception is the derivative of zeta function at origin, which is unary in nature.

By applying the four different edit operations to the seed graph, we obtain four different modified graph variants. Each edit operation is applied between 3 and 30 times to the seed graph. The nodes and edges to be edited are selected at random. For each operation, we perform 30 trials in order to obtain the average distance together with its standard deviation.

In Fig. 11, from left-to-right and top-to-bottom, we show the effects of edge deletion, edge insertion, node deletion and

node insertion. The x-axis in each plot shows the number of edit operations, while the y-axis shows the mean Euclidean distance between the feature vector of the original seed graph and its modified variant after application of the relevant edit operation. In the plots, the different colored lines represent the six different graph characterization methods. The plots show that the five new graph characterizations proposed in this paper result in an approximately linear relationship between the Euclidean distance and the number of edit operations. The Laplacian spectrum, on the other hand, yields a behavior that is less linear.

In Fig. 12, we show the relative deviation, i.e. the standard deviation divided by the mean, for the six graph characterizations under the four different edit operations. This quantity gives an indication of the stability of the different graph characterization methods. From Fig. 12, the zeta-function values, the heat-content invariants and the derivative of zeta function at the origin each give a relatively small relative deviation in all plots. However, it is the heat-content invariants that give the best performance. Moreover, the derivative of the zeta function at the origin outperforms the Laplacian spectrum.
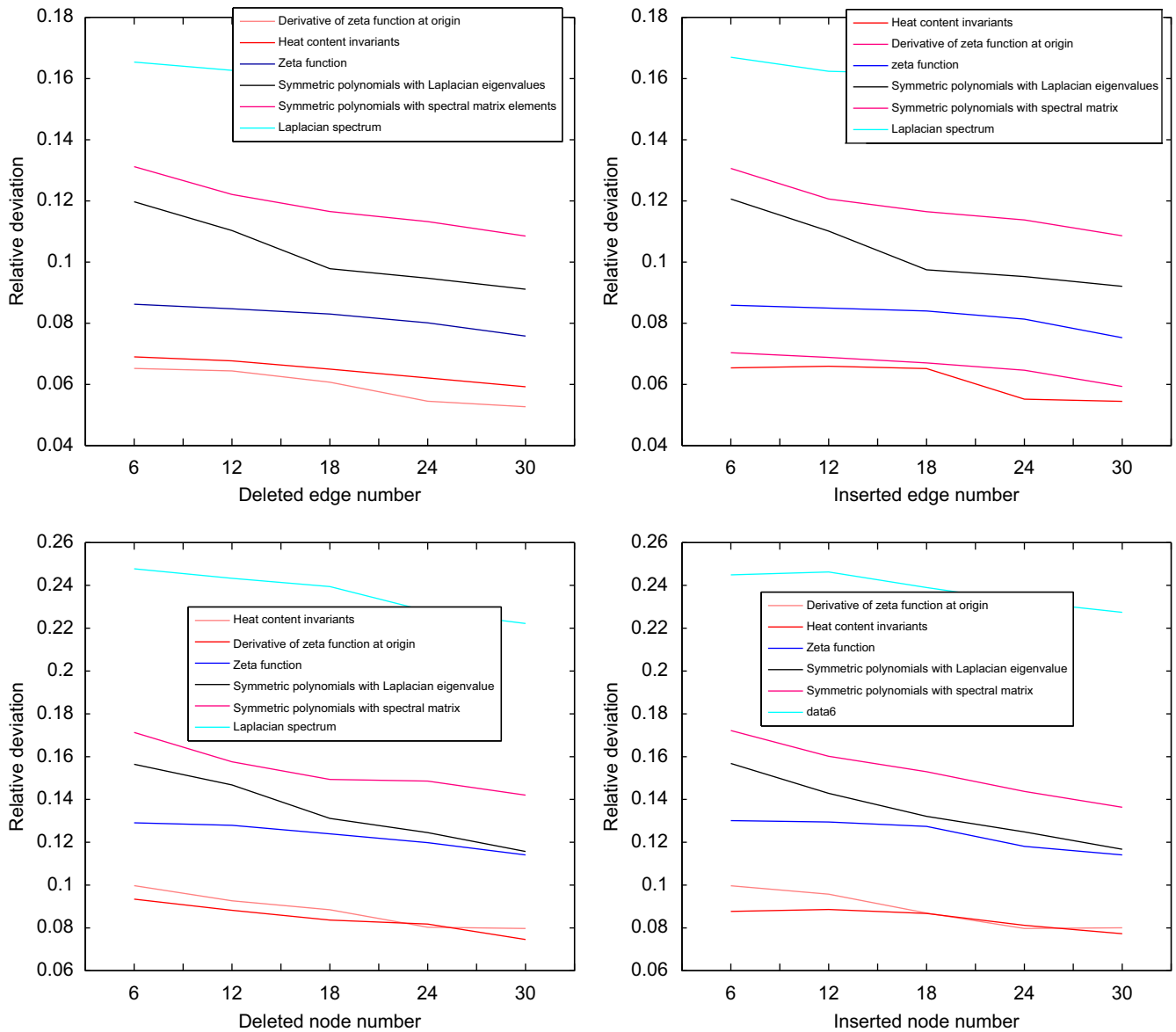
Fig. 12. Relative deviation for the different graph characterization methods for the random graphs (from left to right, and top to bottom, we show edge deletion, edge insertion, node deletion and node insertion, respectively).

5. Experiments on real-world data

In this section, we will apply our methods to real-world image databases. We will explore whether the heat kernel invariants can be used for object recognition. We use three databases with different methods for graph extraction. The first database is COIL-100 [26] which contain different views of a number of 3D objects. The second database is Caltech 256 [36]. The COIL database consists of views of objects against a blank background, while Caltech 256 contains images of objects against a complex background. In the third experiment, we will apply our methods to shape analysis. Here we use a shock tree database which composed of 150 silhouettes of 10 different types of object [32].

5.1. Experiments on the COIL database

We use the COIL 100 database, which consists of a series of 2D views of 3D objects collected at 72 equally spaced viewing directions

on a great circle of the object view-sphere. From each image in each view sequence, we extract corner features [21]. We use the detected feature points as nodes of a graph. The edge relations are obtained by computing the Voronoi tessellations of the feature points, and constructing the region adjacency graph, i.e. the Delaunay graph, of the Voronoi regions. This process is referred to as Delaunay triangulation. An example view can be seen in Fig. 13.

5.1.1. Stability of the characterizations on the coil database

We first choose eight object from the COIL database and explore the stability of the different characterizations with view number. In Fig. 13, superimposed on the images are the detected feature points and their associated Delaunay graphs.

From left-to-right and top-to-bottom in Fig. 14 we show the values of $\zeta(1)$, $\zeta(2)$, $\zeta(3)$ and $\zeta(4)$ as a function of view number for the eight objects. The different curves in the four plots correspond to the different objects. The main feature to note is that the curves for the different objects are well separated, and that the individual values
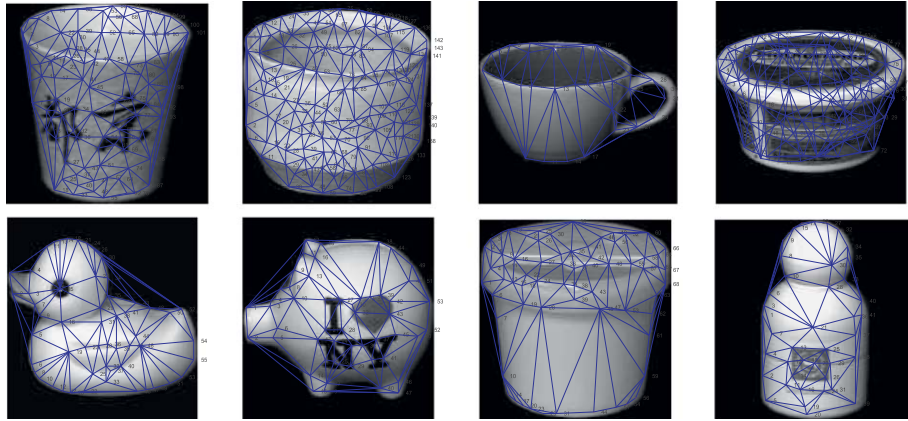
Fig. 13. Eight objects with their Delaunay graphs overlaid (from left to right, and top to bottom, the corresponding object class are denoted as object1, object2,…,object8 as shown in Figs 19 and 20).
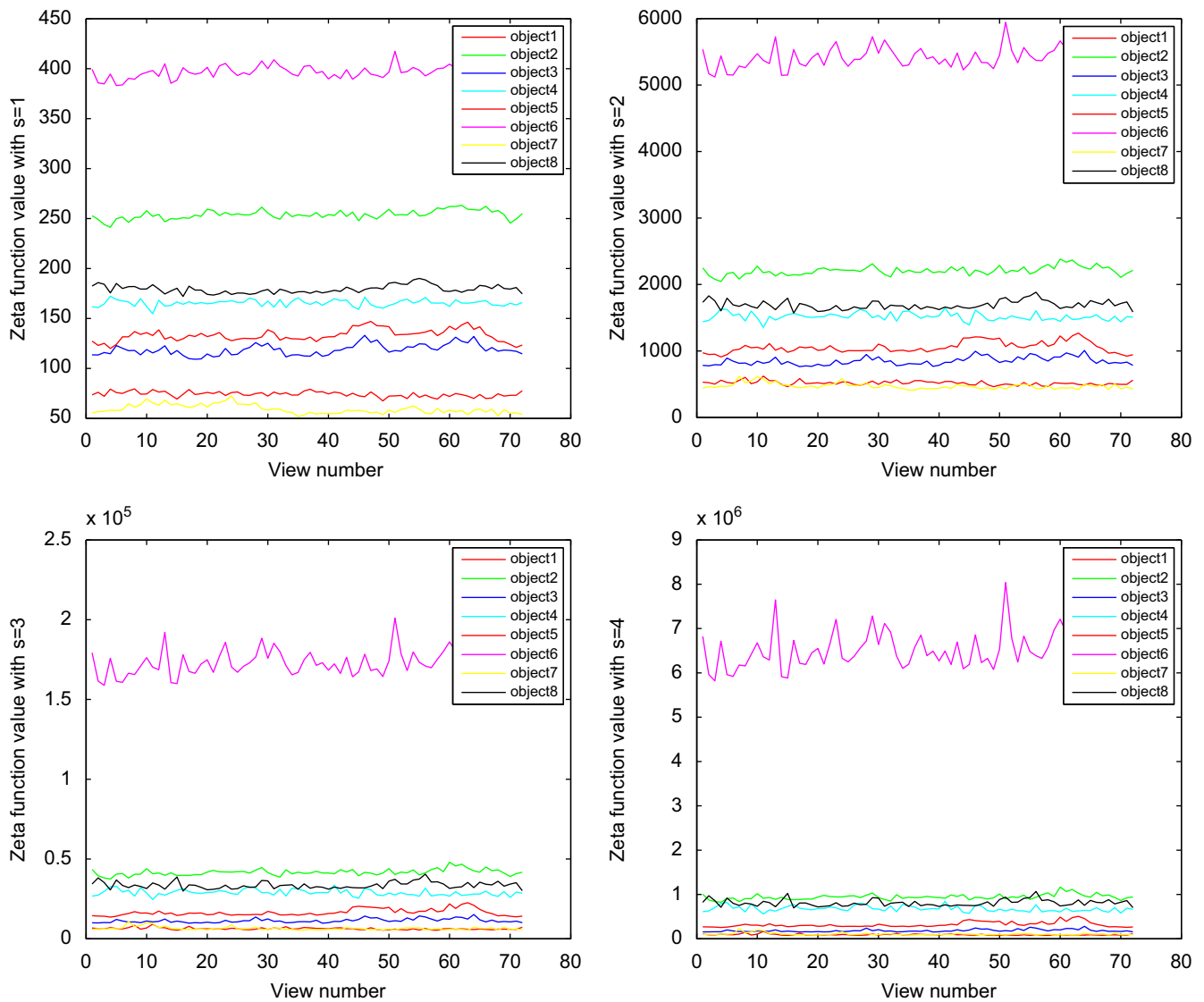

Fig. 14. Zeta function $\zeta(s)$ varying with view number (from left to right, and top to bottom, $s = 1, 2, 3$ and 4, respectively).

of the zeta function do not vary significantly with view number. For small integer input, i.e. $s = 1$ and 2, the fluctuations in the outputs in Fig. 14 are generally small. However, the fluctuation is large when $s = 4$. This is because when s becomes large the small eigenvalues will play a dominant role in the ζ function.

We now turn our attention to the derivative of the zeta function at the origin. In Fig. 15 we show the curves $\zeta'(0)$ as a function of view number for the eight objects from COIL database, respectively. The curves are well separated for the different objects, and the fluctuations are relatively small. In fact in the case of the COIL data the derivative of the zeta function separates the objects better than the individual zeta-function values.

Next we explore the relative stability of the different polynomial characterizations. In Fig. 16 we plot the mean values from the symmetric polynomials over the set of object views as a function of their order. Here the arguments of the polynomials are the column elements of the spectral matrix $\sqrt{\Lambda}\Phi$. The different curves in the plot are for different COIL objects. As with the experiments with synthetic graphs, we use nine different symmetric polynomials. These are the first three symmetric polynomials S_1, S_2, S_3, the middle three symmetric polynomials $S_{fix(N/2)-1}, S_{fix(N/2)}, S_{fix(N/2)+1}$ and last three symmetric polynomials S_{N-2}, S_{N-1}, S_N, where N is the number of node in
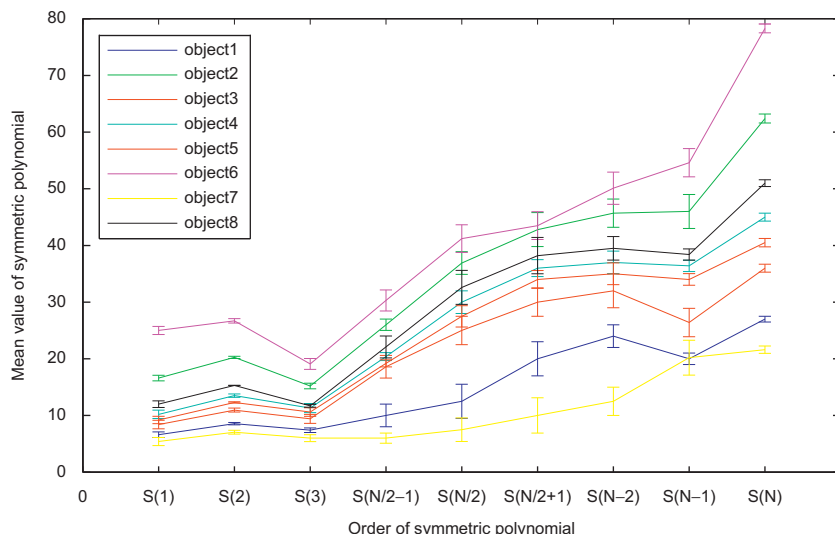
the graph. Again, the error bars show the standard deviation over the different sets of views of the same object. Fig. 10 repeats the analysis using the normalized Laplacian spectrum as the arguments of the polynomials.

Figs. 16 and 17 support our findings on synthetic data. It is clear that the first three and last three symmetric polynomials computed using the normalized Laplacian eigenvalues and spectral matrix elements as inputs can be used for object separation. We also get a better separation of objects when using the polynomials from the normalized Laplacian spectrum rather than the elements of the spectral matrix.

We now turn our attention to the invariants computed from the heat content characterization described in Section 3.5. In Fig. 18, we plot the four leading heat-content invariants co-efficients q_1, q_2, q_3 and q_4 separately as a function of the view number for the eight objects selected from the COIL database. The co-efficients are relatively stable with viewpoint.

5.1.2. Clustering
Having established the stability of the different characterizations under variation in viewing direction, we now explore whether thay can be used to cluster the graphs for objects of the same class. To this end, in Fig. 19, we show the result of performing principal component analysis (PCA) on the feature vectors of permutation invariants. In Fig. 19, from left to right, we show the result with the leading zeta-function values, the symmetric polynomials with Laplacian eigenvalues, and the heat-content invariants. The different objects are denoted by points of a different color. For comparison, we repeat the experiments for symmetric polynomials with spectral matrix elements and Laplacian eigenvalues. The results are given in Fig. 20. For this unsupervised process, we have computed the Rand index [25] for the different methods. The Rand index is defined as

$$R_I = \frac{C}{C + W} \tag{26}$$

where C is the number of agreements and W is the number of disagreements in cluster assignment. If two objects are in the same class (i.e. ground truth cluster) and belong to the same experimentally extracted cluster, then this counts as an agreement. If, on the other hand, two objects are in the same class, but appear in different experimentally extracted clusters, then this counts as a disagreement. The index is hence the fraction of views of a particular object that
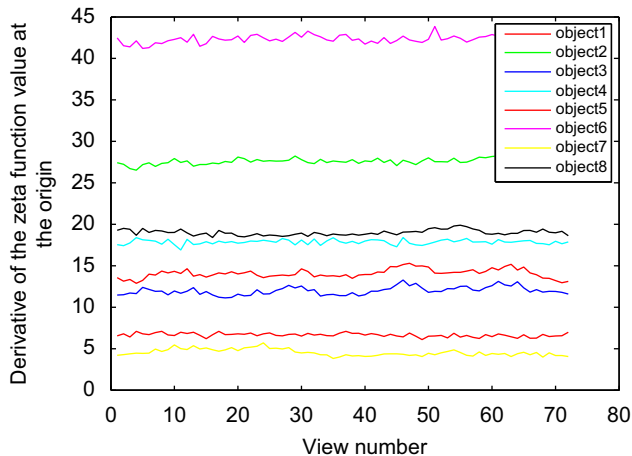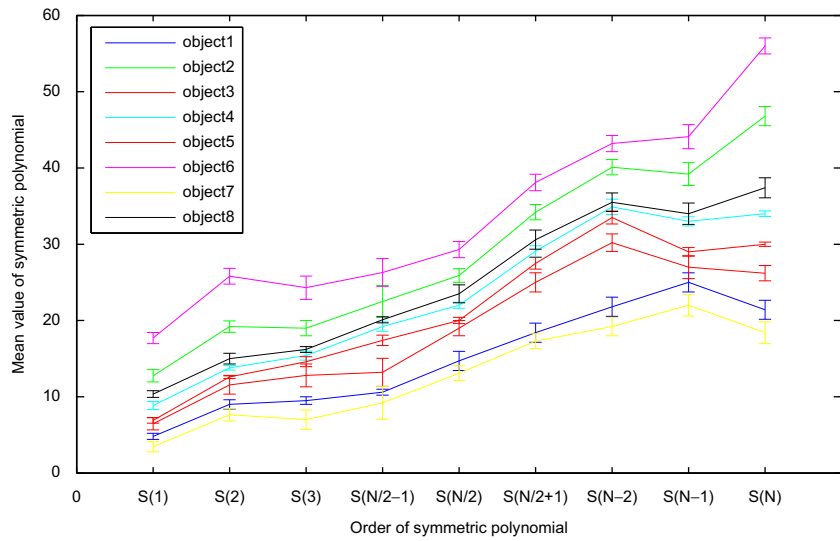

Fig. 15. Derivative of the zeta function at the origin for the COIL database.

Fig. 16. Symmetric polynomials using spectral matrix elements for the COIL database.

Fig. 17. Symmetric polynomials using normalized Laplacian eigenvalues for the COIL database.
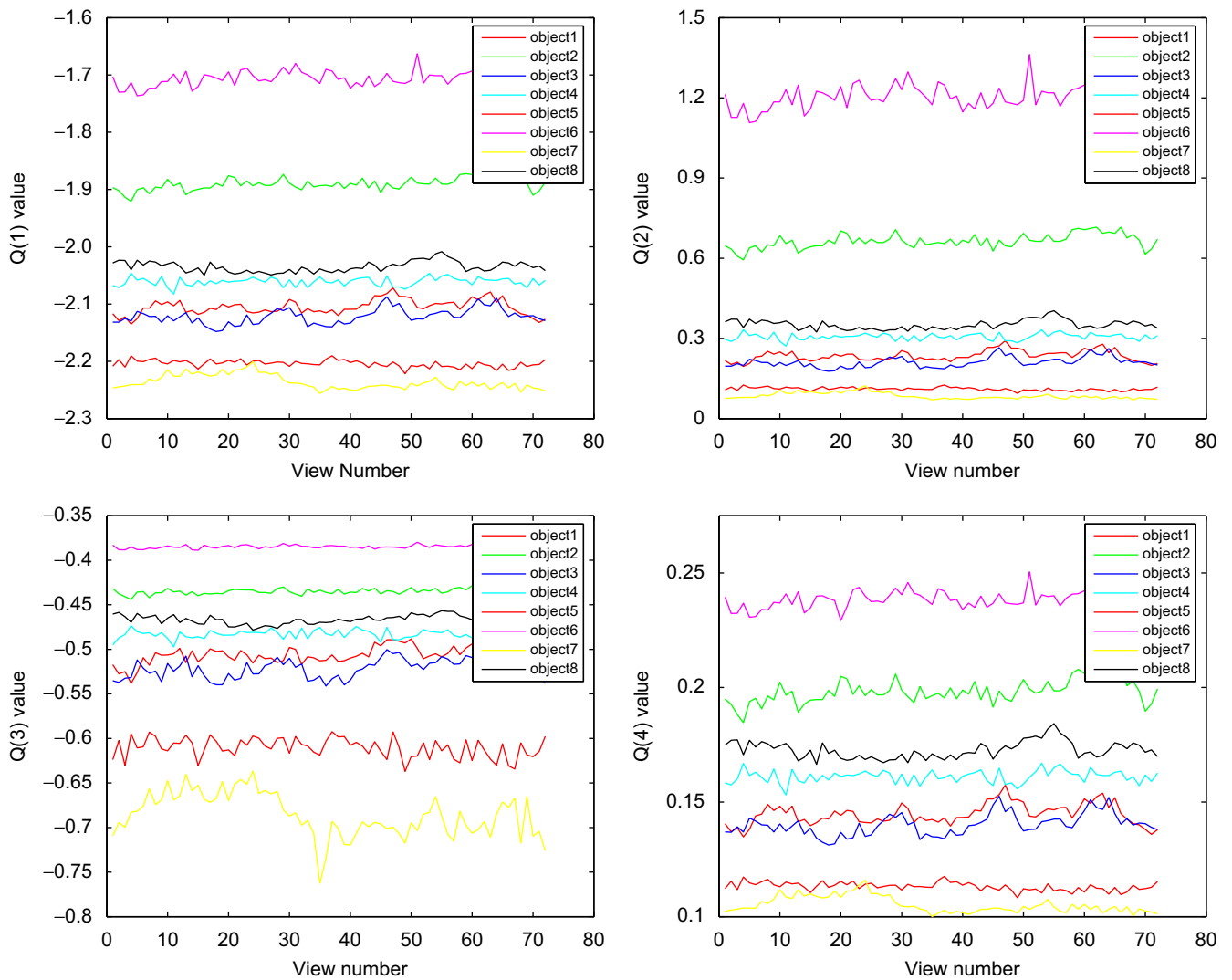

Fig. 18. Individual heat-content invariants as a function of view number.
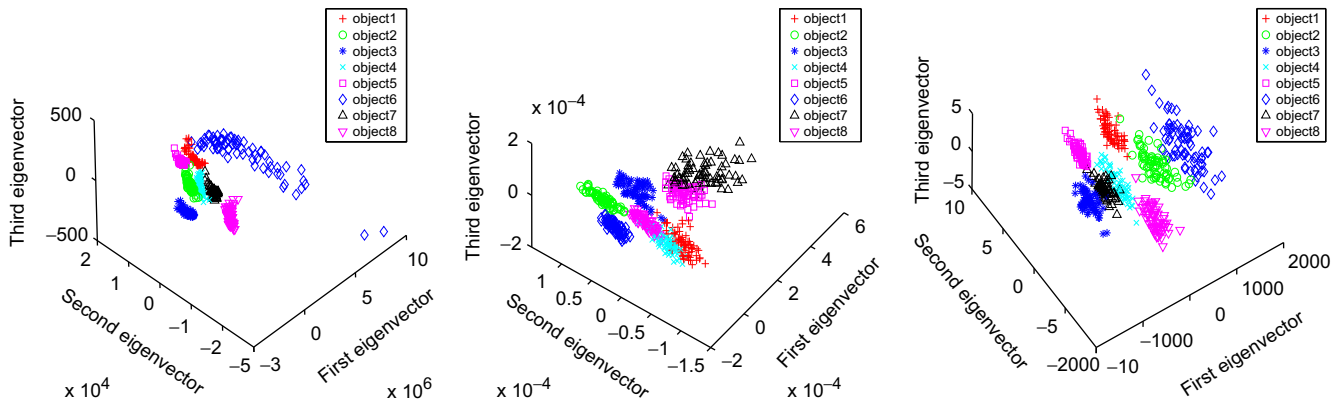
Fig. 19. Principal component analysis results for the heat kernel invariants for the COIL database: zeta-function values (left), symmetric polynomials using normalized Laplacian eigenvalues (middle) and heat-content invariants (right).
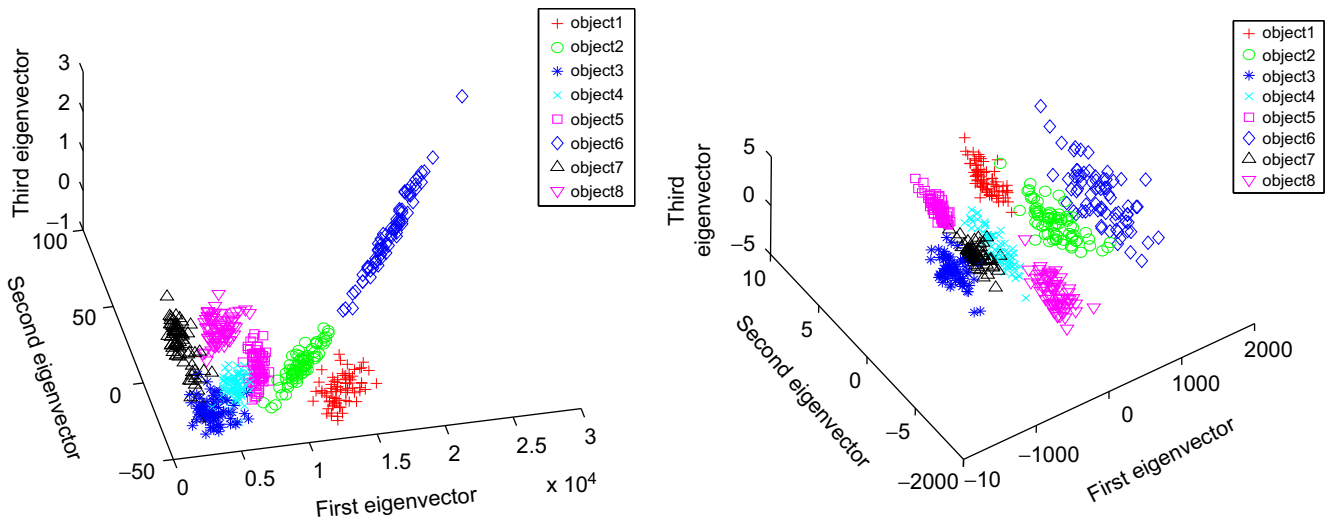

Fig. 20. Principal component analysis results for the traditional spectral methods on the COIL database: symmetric polynomials using spectral matrix elements (left) and Laplacian spectrum (right).

are closer to an object of the same class than to one of a different class. The Rand index takes a value in the interval [0,1], where 1 corresponds to a perfect clustering.

We use the K-nearest neighbor algorithm to locate clusters. If over half of the neighbors belong to the same category as the ground truth object class, the corresponding graph is correctly clustered. In our experiment, we set K to 5. The Rand index for zeta ζ function characterization clustering is 0.96. For the symmetric polynomial clustering we use the first and last three symmetric polynomial values $S_1, S_2, S_3, S_{N-2}, S_{N-1}, S_N$ as feature vector for graph clustering. The Rand index value for the symmetric polynomial using Laplacian eigenvalues as inputs is 0.93 and the Rand index value for the symmetric polynomial using spectral matrix elements as inputs is 0.89. The Rand index for the heat-content coefficients clustering is 0.98, for zeta-function values is 0.96 and for the Laplacian eigenvalues it is 0.78. This experiment shows that the heat kernel invariants proposed in the paper can be used as a means of graph clustering.

We proceed to experiment with our different characterizations on the entire COIL 100 database. Here we use supervised learning methods. For each class of object, we select 30 graphs randomly from the relevant class as training data. For comparability, we construct

the feature vectors of the same length (10 in our experiments) from the different graph characterizations. We have used several standard supervised learning algorithms, generative mixture models (GMM) [37], support vector machines (SVM) [38] and linear discriminant analysis (LDA) [34]. In Table 1, we give the classification accuracy obtained with different graph characterization methods and different learning methods. The heat kernel invariants give the best overall performance, irrespective of the learning algorithm used. The feature vector of zeta-function values gives performance that is comparable to the symmetric polynomials. Overall the best results are obtained with the heat-content invariants with the GMM classification method.

5.2. Experiments on the Caltech 256 database

In this section, we apply our methods to the Caltech 256 database [36]. As noted above, when compared with the COIL database the Caltech database is considerably more complex. The database contains images of different examples of the same kind of object. The images are taken from different view points and contain different

Table 1
Classification performance for the COIL 100 database.

	Zeta-function values (%)	Heat-content invariants (%)	Symmetric polynomials with Laplacian eigenvalues (%)	Symmetric polynomials with spectral matrix (%)	Tradition Laplacian spectrum (%)
LDA	93	97	94	90	83
GMM	94	97	94	91	84
SVM	92	96	94	91	83

Fig. 21. Some example objects selected from the Caltech 256 database.

backgrounds corresponding to different environments. The corresponding graphs thus contain occlusion and clutter. Some example views of some selected objects are shown in Fig. 21. For this database, we extract relational graphs in which the nodes denote primitive image features and the links denote Gestalt relationships between the image features. We use line patterns as our primitives (i.e. nodes) [29]. We use Gestalt relationships [39,30] that reflect human perceptual organization between line primitives. For each pair of lines we compute measures of proximity, continuity and parallelism. We then combine these Gestalt relations to generate a the relational graph along the lines suggested in [40,22].

For two line segments e_i and e_j, the proximity measure $E_{Proximity}(e_i, e_j)$ is defined as

$$E_{Proximity}(e_i, e_j) = 1 - \min\left(1, \frac{d_{gap}}{\min(l_i, l_j)}\right)$$

where l_i is the length of the line segment e_i and d_{gap} is the minimum distance between the end points of two line segments. The quantity $E_{Continuity}(e_i, e_j)$ is defined as

$$E_{Continuity}(e_i, e_j) = \min\left(\frac{l_i + l_j}{l_i + l_j + d_{gap}}, \cos^2(\theta)\right)$$

The parallelism between e_i and e_j is determined by difference in length between the constituting edge segments, l_o, the length of the symmetry axis, l_{sym}, the average width, μ_w, and variation, σ_w, of the parallel strip, and the minimum fit error, Err_{sym}, of a constant curvature segment to the symmetry axis [40]:

$$E_{Parallelism}(e_i, e_j) = \min\left(\frac{\mu_w}{\mu_w + \sigma_w}, \frac{l_{sym}}{l_{sym} + Err_{sym}}, \frac{l_{sym}}{l_{sym} + l_o}\right)$$

The final weight $w(e_i, e_j)$ between two line segments e_i and e_j is given by

$$w(e_i, e_j) = E_{Proximity}(e_i, e_j) \times (E_{Continuity}(e_i, e_j) + E_{Parallelism}(e_i, e_j))$$

as suggested by Mohan and Nevatia in [40].

With the relational graph representation in hand, we perform object recognition with the Caltech 256 database. In Fig. 22, we show the performance rate as a function of number of classes. As before, in this experiment we randomly select 30 graphs from each class for training, and then use the Gaussian mixture model (GMM) as our learning method. Fig. 22 shows that with a suitable relational graph
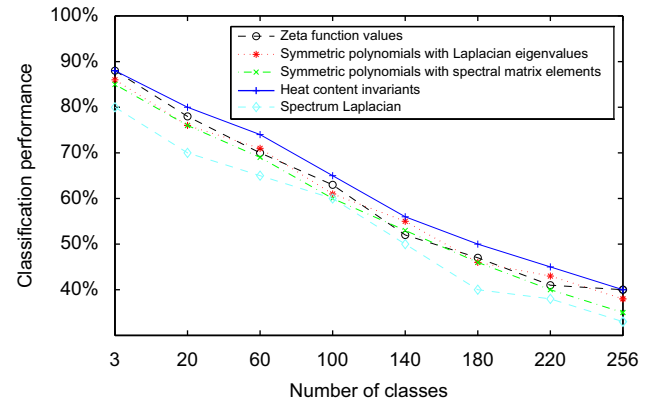
Fig. 22. Experimental results on the Caltech 256 database.

representation in hand, then our graph characterization methods can deliver reasonable classification results when compared with alternative methods [36]. The classification rate is high (over 80%) with no more than 20 objects classes. The performance drops when more object classes are added. This can be alleviated by incorporating more sophisticated learning methods i.e. semi-supervised GMM [43] into our future work. Again, the heat kernel invariants give thre best overall performance.

5.3. Applications on shape analysis

We have also explored the application of our new graph characterization methods to shape analysis. Here graph based methods have provided a powerful means of characterizing the underlying variations in shape due to deformation, articulation or natural class variability [41]. We have applied our methods to shock trees. This is a tree structure based representation of the differential structure of the boundary of a 2D shape. It is obtained by locating the shape skeleton, and examining the differential behavior of the radius of the bitangent circle from the skeleton to the object boundary, as the skeleton is traversed [42]. Recently, the shock tree extraction process has been further improved by Torsello and Hancock [32]. Firstly, they have shown how to compute the skeleton more reliably in the proximity of high curvature detail [44]. Secondly, they have shown
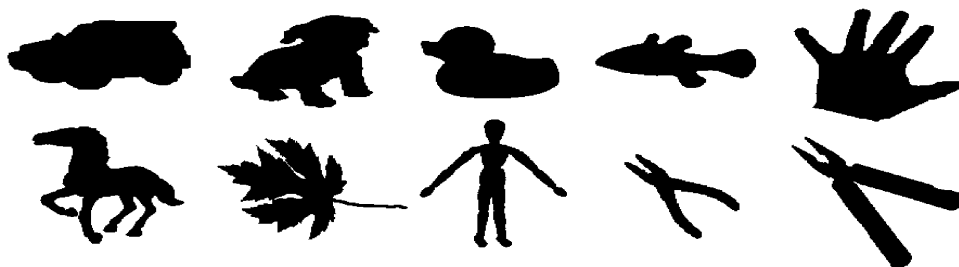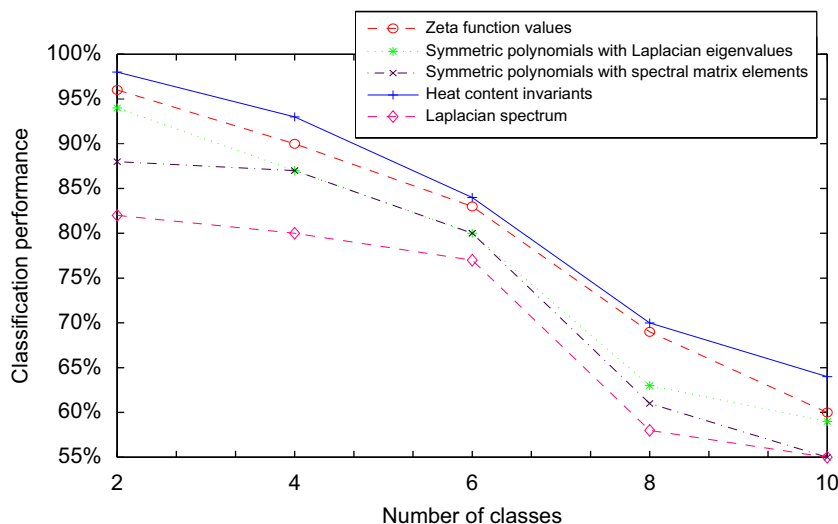
Fig. 23. Sample views of shape silhouettes.

Fig. 24. Shape silhouette classification performance as a function of the number of silhouettes.
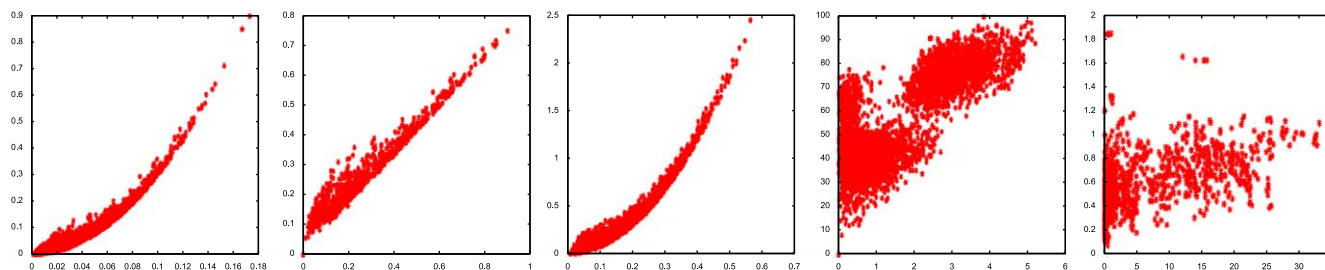

Fig. 25. Scatter plot of the feature vector distance for the different characterizations as a function of edit distance (*y*-axis)—from left to right, first (zeta-function values), second (heat-content invariants), third (symmetric polynomials with Laplacian eigenvalues), fourth (symmetric polynomials with spectral matrix) and the fifth (the Laplacian spectrum).

how to determine the similarity of skeletal branches using a simple measure computed from the distance map from the skeleton to the boundary [45].

In this paper we use the methods described in [32] to extract a tree representations from the silhouettes of 2D shape. A representative view of each object studied is shown in Fig. 23.

In the database, there are over 15 shapes from 10 classes. For each shape we can construct a tree representation of the skeleton. Using the Laplacian spectrum of the tree, we compute the different invariants described earlier. A GMM (Gaussian mixture model) is applied to the feature vectors constructed from the invariants for the purposes of classification. In Fig. 24, we plot the proportion of shapes from the database correctly classified as the number of shape classes

is increased. All five methods explored decrease in classification accuracy as the size of the database increases. The best performance is given by the heat-content invariants, closely followed by the feature vector of zeta-function values. The poorest performance is given by the spectrum of Laplacian eigenvalues.

We pursue the analysis further in Fig. 25. Here we show scatter plots of the Euclidean distances between feature vectors for the permuation invariants versus tree edit distance. Here the shock tree edit distance is computed using the method described in [32]. In Fig. 25 the five scatter plots, from left to right, are for the zeta-function values, the heat-content invariants, the symmetric polynomials with Laplacian eigenvalues, the symmetric polynomials with spectral matrix elements, and, the Laplacian spectrum.

There is a clear regression trend for the first three of these, i.e. the zeta-function values, the heat-content invariants and the symmetric polynomials with Laplacian eigenvalues. For these the invariant features can proxy for the computation of tree edit distance, and this offers an advantage since no node correspondences are needed. However, for the remaining two there is no such trend.

6. Conclusion

In this paper we have explored the use of several invariants computed from the heat kernel to characterize graphs. First we have explored the use of the zeta function as a means of characterizing the shape of the heat kernel trace. Using the Mellin transform, the zeta function is linked to the moment generating function of the heat kernel trace. We have shown that the derivative of the zeta function at the origin is related to the determinant of the Laplacian matrix. The slope of the zeta function at the origin is related to the product of the non-zero Laplacian eigenvalues, and is hence an invariant to the permutation of the node order of the graph. We have further investigated alternative unary permutation invariants computed by applying symmetric polynomials to the spectral matrix. Finally, we have shown how the use of heat content can lead to a series of invariants. In common with the symmetric polynomials, the heat-content coefficients are permutation invariants that depend on both the eigenvalues and eigenvectors of the Laplacian matrix.

We have experimented with the three newly proposed graph characterizations on synthetic and real-world databases. The main conclusions from this study are as follows. First, the most effective characterization is the set of heat-content invariants. Second, each of the three new characterizations outperforms the symmetric polynomials. The poorest performance is achieved with the Laplacian spectrum. The derivative of the zeta function at the origin, which is a unary attribute gives similar performance to the symmetric polynomials. We have applied our characterizations to object recognition. By using existing learning algorithms i.e. GMM (Gaussain mixture model), we have developed a technique which is comparable in performance to existing methods reported in the literature.

There are a number of ways in which the work reported in this paper can be extended.

First, we aim to explore more depth the links with spectral geometry, in particular the Selberg zeta function [2]. A good place to start is with the work of Stark and Terras [3]. A second line of investigation will be to explore in more depth the link between the heat kernel invariants and symmetric polynomials. Finally, we intend to explore different application domains. For instance, we will explore whether the invariants described in this paper can be used to characterize molecular structures and small world networks [35].

References

[1] F.R.K.Chung, Spectral Graph Theory, American Mathematical Society, 1997.
[2] D.A. Hejhal, The Selberg Trace Formula for PSL(2,R), Springer, Berlin, 1983.
[3] H.M. Stark, A.A. Terras, Zeta functions of finite graphs and coverings, Adv. Math. 121 (1996) 126–165.
[4] P. Gilkey, in: D. Krupka, D. Saudners, (Eds.) The Spectral Geometry of Operators of Dirac and Laplace Type, 2008, pp. 287–324.
[5] P. Mcdonald, R. Meyers, Isospectral polygons, planar graphs and heat content, Proc. Amer. Math. Soc. 131 (2003) 3589–3599.
[6] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Pattern Anal. Mach. Intelli. 22 (2000) 888–905.
[7] J.E. Atkins, E.G. Bowman, B. Hendrickson, A spectral algorithm for seriation and the consecutive ones problem, SIAM J. Comput. 28 (1998) 297–310.
[8] A. Shokoufandeh, S. Dickinson, K. Siddiqi, S. Zucker, Indexing using a spectral encoding of topological structure, in: IEEE Conference on Computer Vision and Pattern Recognition, 1999, pp. 491–497.
[9] R.R. Coifman, S. Lafon. Diffusion maps. Appl. Comput. Harmonic Anal. (2004).
[10] J. Lafferty, G. Lebanon, Diffusion kernels on statistical manifolds, J. Mach. Learn. Res. 6 (2005) 129–163.
[11] S.T. Yau, R.M. Schoen, Differential Geometry, Science Publication, 1988.
[12] P.B. Gilkey, Invariance Theory, the Heat Equation, and the Atiyah-Singer Index Theorem, Perish Inc., 1984.
[13] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[14] B. Luo, R.C. Wilson, E.R. Hancock, Spectral embedding of graphs, Pattern Recognition 36 (2003) 2213–2230.
[15] B. Mohar, Laplace eigenvalues of graphs—a Survey, Discrete Math. 109 (1992) 171–183.
[16] L. Lovaz, Random walks on graphs: a survey, Combinatorics, Paul Erds is eighty 2 (1996) 353–397.
[17] M. Hein, J.Y. Audibert, U. von Luxburg, From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacian, in: 18th Annual Conference on Learning Theory, 2005, pp. 470–485.
[18] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (2003) 1373–1396.
[19] J.B. Tenenbaum, V.D. Silva, J.C. Langford, A global geometric framework for non-linear dimensionality reduction, Science 290 (2000) 586–591.
[20] X. He, P. Niyogi, Locality preserving projections, in: NIPS03.
[21] C.G. Harris, M.J. Stephens, A combined corner and edge detector, in: Fourth Alvey Vision Conference, 1994, pp. 147–151.
[22] S. Sarkar, K.L. Boyer, Quantitative measures of change based on feature organization: eigenvalues and eigenvectors, Comput. Vision Image Understanding (1998) 110–136.
[23] R. Wilson, B. Luo, E.R. Hancock, Pattern vectors from algebraic graph theory, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 2220–2237.
[24] P. Mcdonald, R. Meyers, Diffusions on graphs, poisson problems and spectral geometry, Trans. Am. Math. Soc. 354 (2002) 5111–5136.
[25] W.M. Rand, Objective criteria for the evaluation of clustering, J. Am. Statist. Assoc. 66 (1971) 846–850.
[26] S.A. Nene, S.K. Nayar, H. Murase, Columbia object image library (COIL 100), Columbia University, 1996.
[27] R. Fergus, P. Perona, A. Zisserman, Object class recognition by unsupervised scale-invariant learning, IEEE Trans. Pattern Anal. Mach. Intell. 2 (2003) 264–271.
[29] B. Huet, E.R. Hancock, Line pattern retrieval using relational histograms, IEEE Trans. Pattern Anal. Mach. Intell. 21 (1999) 1363–1370.
[30] S. Sarkar, K.L. Boyer, Perceptual organization in computer vision, IEEE Trans. Syst. Man Cybernet. 23 (1999) 382–399.
[31] D. Emms, E.R. Hancock, S. Severini, R.C. Wilson, A matrix representation of graphs and its spectrum as a graph invariant, Electron. J. Combin. 13 (2006) 140–154.
[32] A. Torsello, A. Robles-Kelly, E.R. Hancock, Discovering shape classes using tree edit-distance and pairwise clustering, Int. J. Comput. Vision 72 (3) (2007) 259–285.
[34] R.O. Duda, P.E. Hart, D.H. Stork, Pattern Classification, Wiley Intersceince, New York, 2000.
[35] I. Duraka, A.-L. Barabasi, S.J. Zhou, T.C. Germann, P.S. Lomdahl, A.R. Bishop, Molecular dynamics investigation of the surface stress distribution in a Si/Ge quantum dot superlattice, Phys. Rev. 60 (1999) 2150–2153.
[36] G. Griffin, A. Holub, P. Perona, Caltech-256 Object Category Dataset, Technical Report, California Institute of Technology.
[37] M.A.F. Figueiredo, A.K. Jain, Unsupervised learning of finite mixture models, IEEE Pattern Anal. Mach. Intell. 24 (3) (2002) 381–396.
[38] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[39] K. Koffka, Principles of Gestalt Psychology, Harcourt, New York, 1935.
[40] R. Mohan, R. Nevatia, Erceptual organization for scene segmentation and description, IEEE Pattern Anal. Mach. Intell. 14 (1992) 616–635.
[41] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 509–522.
[42] B.B. Kimia, A.R. Tannenbaum, S.W. Zucker, Shapes, shocks, and deforamtions, Int. J. Comput. Vision 15 (1995) 189–224.
[43] S. Vempala, G. Wang, A spectral algorithm for learning mixture models, J. Comput. Syst. Sci. 68 (2004) 841–860.
[44] A. Torsello, E.R. Hancock, Correcting curvature density effects in the Hamilton–Jacobi skeleton, IEEE Trans. Image Process. 15 (2006) 877–891.
[45] A. Torsello, E.R. Hancock, A skeletal measure of 2D shape similarity, Comput. Vision Image Understanding 95 (2004) 1–29.
[46] B.D. McKay, Spanning trees in regular graphs, Eur. J. Combin. 4 149–160.
[47] R. Fergus, P. Perona, A. Zisserman, Object class recognition by unsupervised scale-invariant learning, IEEE Comput. Vision Pattern Recognition 2 (2003) 264–271.
[48] M. Demirci, R. Leuken, R. Veltkamp, Matching shape graphs, in: Proceedings in Applied Mathematics and Mechanics, 2008.
[49] F. Li, P. Perona, A Bayesian hierarchical model for learning natural scene categories, IEEE Comput. Vision Pattern Recognition 2 (2005) 524–531.
[50] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110.
[51] A. Lobay, D.A. Forsyth, Shape from texture without boundaries, Int. J. Comput. Vision 67 (1) (2006) 71–91.

About the Author—Bai Xiao received the B.Eng. degree in Computer Science from Beihang Univeristy of China in 2001. From 2002 to 2006, he was a Ph.D. student at Computer Science Department, University of York, UK under the supervision of Professor Edwin R. Hancock. From September 2006 to December 2008, he was a Research Officer (Fellow, Scientist) at Computer Science Department, University of Bath. He is now an Associate Professor at Computer Science School, Beihang University (BUAA). His current research interests include graph spectral analysis, image segmentation and recognition, statistical pattern recognition, and feature extraction. He has published more than 20 papers in journals and refreed conferences. He has been a programme committee member for many national and international conferences.

About the Author—Edwin R. Hancock studied physics as an undergraduate at the University of Durham and graduated with honors in 1977. He remained at Durham to complete the Ph.D. degree in the area of high-energy physics in 1981. Following this, he worked for 10 years as a researcher in the fields of high-energy nuclear physics and pattern recognition at the Rutherford-Appleton Laboratory (now the Central Research Laboratory of the Research Councils). In 1991, he moved to the University of York as a lecturer in the Department of Computer Science. He was promoted to senior lecturer in 1997 and to reader in 1998. In 1998, he was appointed to a chair in computer vision. Professor Hancock now leads a group of some 15 faculty, research staff, and Ph.D. students working in the areas of computer vision and pattern recognition. His main research interests are in the use of optimization and probabilistic methods for high and intermediate level vision. He is also interested in the methodology of structural and statistical pattern recognition. He is currently working on graph-matching, shape-from-X, image databases, and statistical learning theory. He has published more than 130 journal papers and 450 refereed conference publications. He was awarded the Pattern Recognition Society medal in 1991 and an outstanding paper award in 1997 by the journal Pattern Recognition. In 1998, he became a fellow of the International Association for Pattern Recognition. He has been a member of the editorial boards of the journals IEEE Transactions on Pattern Analysis and Machine Intelligence and Pattern Recognition and Computer Vision and Image Understanding. He has also been a guest editor for special editions of the journals Image and Vision Computing and Pattern Recognition. In 2008 the University of Durham awarded him the D.Sc. degree for his published work.

About the Author—RICHARD WILSON received the B.A. degree in Physics from the University of Oxford in 1992. In 1996, he completed a Ph.D. degree at the University of York in the area of pattern recognition. From 1996 to 1998, he worked as a Research Associate at the University of York. After a period of postdoctoral research, he was awarded an Advanced Research Fellowship in 1998. In 2003, he took up a lecturing post and he is now a Reader in the Department of Computer Science at the University of York. He has published some 100 papers in journals, edited books and refereed conferences. He received an outstanding paper award in the 1997 Pattern Recognition Society awards and has won the best paper prize in ACCV 2002. He is currently an Associate Editor of the journal Pattern Recognition. His research interests are in statistical and structural pattern recognition, graph methods for computer vision, high-level vision and scene understanding. He is a member of the IEEE computer society.