# Graph Characteristics from the Ihara Zeta Function

Peng Ren, Richard C. Wilson, and Edwin R. Hancock

Department of Computer Science, The University of York, York, YO10 5DD, UK
{pengren,wilson,erh}@cs.york.ac.uk

**Abstract.** This paper shows how to extract permutation invariant graph characteristics from the Ihara zeta function. In a previous paper, we have shown that the Ihara zeta function leads to a polynomial characterization of graph structure, and we have shown empirically that the coefficients of the polynomial can be used as to cluster graphs. The aim in this paper is to take this study further by showing how to select the most significant coefficients and how these can be used to gauge graph similarity. Experiments on real-world datasets reveal that the selected coefficients give results that are significantly better than those obtained with the Laplacian spectrum.

## 1 Introduction

One of the bottlenecks in comparing graphs using graph edit distance[13] is that its computation requires correspondence matches. This problem can be overcome if pattern vectors composed of graph characteristics are used instead, and similarity is measured using the distance between vectors. There are a number of alternative characterizations available. For instance, one could use classical properties of the graph such as diameter, perimeter length or Cheeger number[4] or normalized quantities such as the edge or triangle densities. Several possibilities are offered by spectral graph theory, and these include the Laplacian spectrum[10] or symmetric polynomials computed from the spectral matrix[16].

Recently, however Bai and Hancock have shown that the zeta function of a graph offers an interesting means of characterizing its structure[1]. They have demonstrated that the Rosenberg zeta function[11] is related to the Mellin moments of the heat kernel trace, and have used the moments generated by sampling the Rosenberg zeta function to generate graph feature vectors. The zeta function computed by Bai and Hancock is determined by the Laplacian spectrum. However, in general the zeta function of a graph can be thought of as an analogue of the Riemann zeta function, with prime numbers replaced by prime paths[2]. In fact, the zeta function is a compact representation of information concerning the distribution of paths and path lengths on a graph.

In is this paper, we aim to explore this relationship in a greater depth. We turn to the Ihara zeta function. The Ihara zeta function was first detailed in [7] and [8]. Hashimoto subsequently deduced explicit factorizations for bi-regular bipartite graphs[6]. Bass has generalized Hashimoto's factorization to all finite

N. da Vitora Lobo et al. (Eds.): SSPR&SPR 2008, LNCS 5342, pp. 257–266, 2008.

graphs[2]. In a recent paper we have performed a preliminary study of the Ihara zeta function, and have shown how it can be sampled to give a permutation invariant means of characterizing graphs[12]. However, to be rendered tractable for real world problems the issue of how to avoid sampling the infinities at poles and how to generate stable pattern vectors from the Ihara zeta function must be addressed. We have shown how to overcome this problem by establishing pattern vectors using the polynomial coefficients of the reciprocal Ihara zeta function.

The aim in this paper is to take this work one step further. In particular we aim to study in depth the information conveyed by the individual polynomial coefficients and their relationship to the path length structure of a graph. We show how a reduced set of selected coefficients can be used to efficiently characterize graphs, and that the Euclidean distance between vectors composed of these coefficients can be used to approximate the edit distance between graphs.

## 2 The Ihara Zeta Function

In this section we review the theory of the Ihara zeta function used in our earlier work. The Ihara zeta function of a graph can be denoted in the form of a rational function[2]:

$$Z_G(u) = \left(1-u^2\right)^{\chi(G)} \det\left(\boldsymbol{I} - u\boldsymbol{A} + u^2\boldsymbol{Q}\right)^{-1} \tag{1}$$

Here, $\chi(G)$is the Euler Number of the graph, which is defined as the difference between the vertex number $|V(G)|$ and the edge number $|E(G)|$ of the graph, i.e. $\chi(G) = |V(G)| - |E(G)|$, and $\boldsymbol{A}$ is the adjacency matrix of the graph. The degree matrix $\boldsymbol{D}$ can be generated by placing the column sums as the diagonal elements, while setting the off-diagonal elements zeros. Finally, $\boldsymbol{Q}$ is the difference of the degree matrix $\boldsymbol{D}$ and the identity matrix $\boldsymbol{I}$, i.e. $\boldsymbol{Q} = \boldsymbol{D} - \boldsymbol{I}$.

## 3 Polynomial Expression

For md2 graphs, i.e. the graphs with vertex degree at least 2, it is clear-cut that (1) can be rewritten in the form of the reciprocal of a polynomial. However, it is difficult to compute the coefficients of the reciprocal of the Ihara zeta function from (1) in a uniform way, except by resorting to software for symbolic calculation. To efficiently compute these coefficients, it is more convenient to transform the form of the Ihara zeta function in (1) into a concise expression. The Ihara zeta function can also be written in the form of determinant expression[9]

$$Z_G(u) = \frac{1}{\det(\boldsymbol{I} - u\boldsymbol{T})} \tag{2}$$

where $\boldsymbol{T}$ is the Perron-Frobenius Operator[15] on the oriented line graph of the original graph, and is an $2m \times 2m$ square matrix with dimensionality $m$ the number of the edges of the original graph. According to (2), the reciprocal of the Ihara zeta function can be rewritten as:

$$
\begin{aligned}
Z_G^{-1}(u) &= \det(\boldsymbol{I} - u\boldsymbol{T}) \\
&= (u)^{2m}\det\left(\frac{1}{u}\boldsymbol{I} - \boldsymbol{T}\right) \\
&= u^{2m}[c_0\left(\frac{1}{u}\right)^{2m} + c_1\left(\frac{1}{u}\right)^{2m-1} + \cdots + c_{2m-1}\left(\frac{1}{u}\right) + c_{2m}] \\
&= c_0 + c_1 u + \cdots + c_{2m-1}u^{2m-1} + c_{2m}u^{2m}
\end{aligned} \tag{3}
$$

From (3), the coefficients of the reciprocal of the Ihara zeta function can be derived from the coefficients of the characteristic polynomial of the matrix $\boldsymbol{T}$. The calculation of the above coefficients can be converted to a summation of a series of determinants[3]:

$$
c_k = \sum_{\binom{2m}{2m-k}} \begin{vmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,2m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,2m} \\ \cdots & \cdots & \cdots & \cdots \\ b_{2m,1} & b_{2m,2} & \cdots & b_{2m,2m} \end{vmatrix} \tag{4}
$$

There are in total $\binom{2m}{2m-k}$ determinants in the sum. The relevant matrix in each determinant is created by replacing $(2m-k)$ of the $2m$ diagonal elements of the matrix $\boldsymbol{T}$ with -1 and the remaining elements in those corresponding rows and columns with 0.

## 4 Pattern Vectors Using Coefficients as Feature Components

To establish pattern vectors from the Ihara zeta function for the purposes of machine learning, one approach is to consider taking function samples as elements. However, if this strategy is adopted then there is the danger of sampling at poles, and these give rise to infinities. Hence, pattern vectors consisting of function samples are potentially unstable since the distribution of poles is unknown beforehand.

To overcome this problem, we note that the coefficients of the reciprocal of the Ihara zeta function do not give rise to infinities. These coefficients are essentially descriptors of graph structures. As long as $G$ is a simple graph, a) the coefficients $c_3$, $c_4$, and $c_5$ are respectively the negative of twice the number of triangles, squares, and pentagons in $G$, b) $c_6$ is the negative of twice the number of hexagons in $G$ plus 4 times the number of pairs of edge disjoint triangles plus twice the number of pairs of triangles with a common edge, and c) $c_7$ is the negative of twice the number of heptagons in $G$ plus 4 times the number of edge disjoint pairs of one triangle and one square plus twice the number of pairs of one triangle and one square that share a common edge[14]. The highest order coefficient is associated with the number of edges incident to vertex $v_i$, i.e. the node degree $d(v_i)$:

$$
c_{2m} = (-1)^{\chi(G)} \prod_{v_i \in V} (d(v_i) - 1) \tag{5}
$$

The set of coefficients can play the role of pattern vectors for clustering, not only because of their immunity to the distribution of the poles, but also due to their ability to characterize the graph structures.

Although the full set of coefficients associated with a graph can be used to construct a pattern vector, only a subset of the coefficients contribute significantly. Some coefficients may be redundant and others may reduce the effectiveness of machine learning algorithms. We thus need to select the subset of salient coefficients, i.e. those that take on distinct values for different classes and exhibit small distinctiveness within class variation. To do this, we compute the between-class scatter $S_b = \sum_{i=1}^{M} N_i(\bar{c}_{k,i} - \bar{c}_k)^2$ and the within-class scatter $S_w = \sum_{i=1}^{M} \sum_{c_{k,i,j} \in C_i} (c_{k,i,j} - \bar{c}_{k,i})^2$ of the individual coefficients, where $\bar{c}_k$ is the mean of all the $c_k$ samples, $\bar{c}_{k,i}$ is the mean of the $c_k$ samples in class $C_i$, $N_i$ is the number of the $c_k$ samples in class $C_i$ and $M$ is the total number of classes. We then use the criterion function $J = (S_b + S_w)/S_w$ to evaluate the performance of individual coefficients. Individual coefficients which give a large contribution to the criterion function are the most significant.

# 5 Experiment

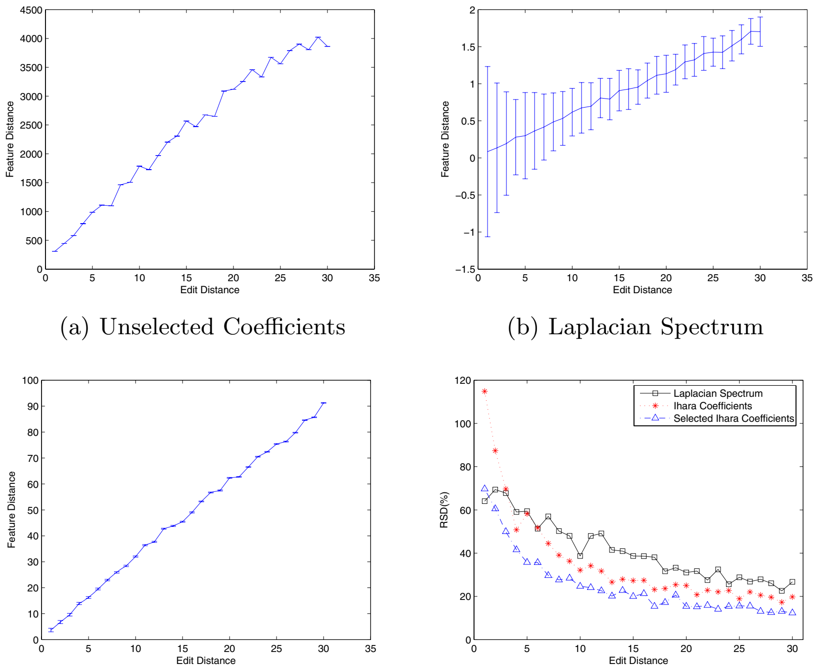
## 5.1 Synthetic Graphs

We commence by investigating the relationship between the edit distance and the Euclidean distance between vectors of Ihara coefficients. We begin with a set of randomly generated md2 graphs. The seed graph for the set has 100 vertices


(a) House Sequences


(b) COIL Datasets

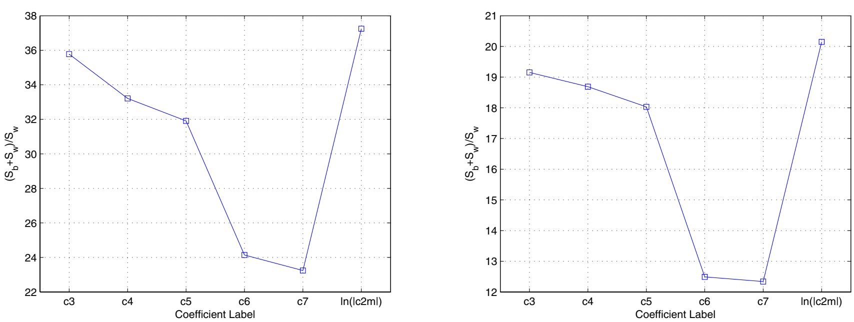
**Fig. 1.** Datasets for Experiments




(a) Unselected Coefficients (b) Laplacian Spectrum

(c) Selected Coefficients (d) Relative Standard Deviation

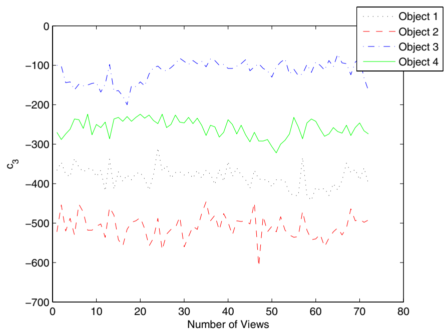
**Fig. 2.** Feature Distance



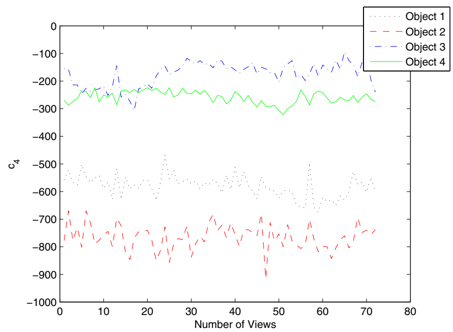

(a) COIL (b) Houses

**Fig. 3.** Criterion Function Value

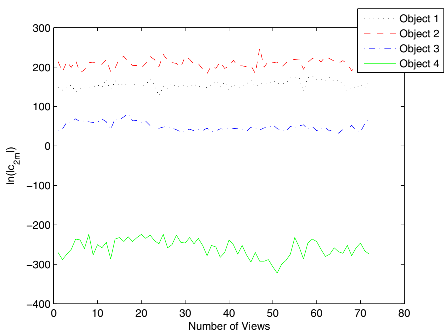
and 300 edges. The remainder of the graphs in the set are obtained by deleting the edges of the seed graph (indexed from 1 to 30). At each level of editing, 100 trials are performed and the edges deleted are chosen randomly, subject to preserving the md2 constraint. We compare the Euclidean distance between the Ihara coefficients with the both the edit distance and the distance between Laplacian spectra. In our experiments, we compute the Ihara coefficients using (4) and




(a) $c_3$



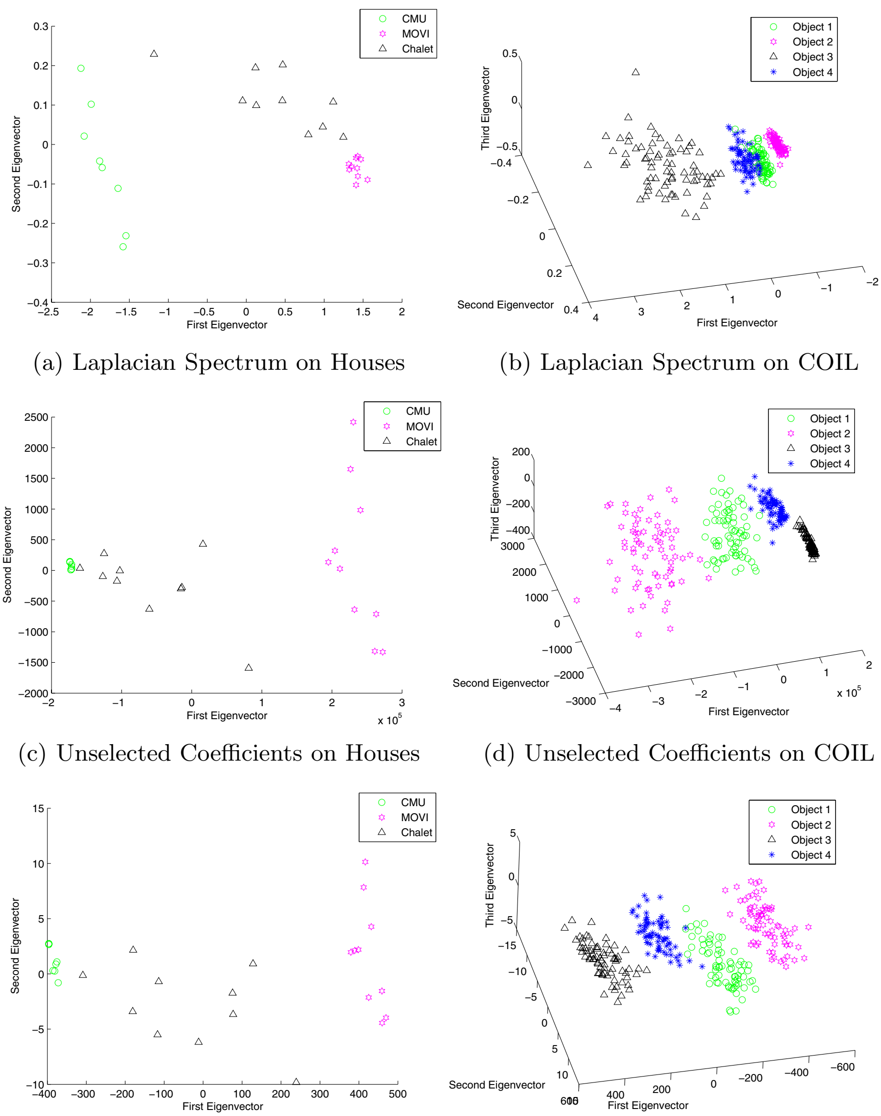

(b) $c_4$




(c) $\ln(c_{2m})$

**Fig. 4.** Coefficients from COIL

(5), constructing a pattern vector in the form $v_G = [c_3 \; c_4 \; c_5 \; c_6 \; c_7 \; \ln(|c_{2m}|)]^T$. The final component of the pattern vector is scaled in a logarithmic manner for the purpose of avoiding high order oscillations. The Lapalacian spectral pattern vectors are formed by taking the second smallest through to the seventh smallest eigenvalues of graph Laplacian as components. The experimental results are shown in Fig.2. Fig.2(a) shows the distance between coefficient pattern vectors from the seed graph and the edited graph as a function of the edit distance, i.e. the number of edges deleted. The relative standard deviation (RSD) is also shown as an error bar. The coefficient distance generally follows the edit distance. Fig.2(b) shows the Laplacian spectral distance as well as the corresponding RSD as a function of the edit distance (but scaled differently to Fig.2(a)). From Fig.2(a) and Fig.2(b) it is clear that the dynamic range of the coefficient feature distance is much larger than that of the spectral distance for corresponding edit distance. Thus the coefficients are more sensitive to graph edits than the Laplacian spectra. To evaluate how reliable the coefficient distance and the spectral distance predict the edit distance, we plot their RSD as a function of edit distance in Fig.2(d). The RSD of the coefficient distance is larger than that of the spectral distance. This means that the Laplacian spectra are more stable than the coefficients under graph edits.




(a) Laplacian Spectrum on Houses
(b) Laplacian Spectrum on COIL
(c) Unselected Coefficients on Houses
(d) Unselected Coefficients on COIL
(e) Selected Coefficients on Houses
(f) Selected Coefficients on COIL

**Fig. 5.** Clustering Performance

## 5.2 Visual Clustering

We apply the pattern vectors composed of Ihara coefficients to two graph datasets used previously in the work of Bai and Hancock. The first set of graphs are extracted from three sequences of images of model houses (referred to as the CMU, MOVI and chalet sequences in Fig.1(a)). The second set of graphs are extracted

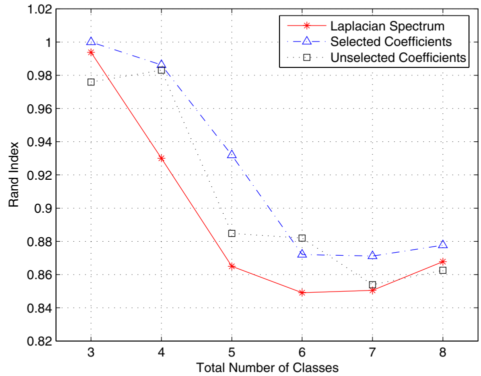



**Fig. 6.** Rand Index

from images of the objects in COIL database (Fig.1(b)). We first extract corner points using the Harris detector. Then we establish Delaunay graphs based on these corner points as nodes. The pattern vectors are formed as explained in Section 5.1. We perform PCA on the pattern vectors to embed them into a 3-dimensional space. We then locate the clusters using $K$-means method.

**Table 1.** Rand Indices

| | (a) | (b) | (c) | (d) | (e) | (f) | (g) |
|---|---|---|---|---|---|---|---|
| House | 0.82 | 0.71 | 0.83 | 0.73 | 0.73 | 0.83 | 0.40 |
| COIL | 0.99 | 0.84 | 0.98 | 0.79 | 0.79 | 0.80 | 0.66 |

Table 1 gives the Rand indices obtained when clustering is attempted using different features. The methods are (a) the Laplacian spectrum (second smallest eigenvalue through to the seventh samllest eignevalues), (b) the Rosenberg zeta function used by Bai and Hancock sampled at the integer values 1 to 6, (c) the Ihara zeta function sampled over the range from 0.001 to 0.006 with interval 0.001, (d) the Ihara zeta function sampled over the range from 0.01 to 0.06 with interval 0.01, (e) the Ihara zeta function sampled over the range from 0.11 to 0.16 with interval 0.01 and (f) the Ihara zeta function sampled over the range from 0.1 to 0.6 with interval 0.1. The clusters are located using the $K$-means method. In Table 1 we can see that the proposed method outperforms the Rosenberg zeta function, and is comparable with the Laplacian spectrum. The final three columns in Table 1 indicate the performance of pattern vectors constructed by sampling the Ihara zeta function. If the samples are appropriately chosen, then this method is comparable with the use of the coefficients. However, when function is sampled in the proximity of poles, the clustering performance deteriorates.

Finally, we explore which combination of coefficients gives the best performance. To do this, we select the coefficients according to the criterion introduced in Section 4.2. We select 3 objects for each of which 10 sample images are used as training data to compute the criterion function value. Fig.3 shows the criterion function values for the coefficients extracted from the two datasets. The first and last coefficients offer more discrimination than the middle ones. This is because the remaining coefficients provide no significantly increase in information over $c_3$, $c_4$ and $c_{2m}$ since they are determined by the number of triangles and squares in the graph and the node degrees. This information can be fully represented by $c_3$, $c_4$ and $c_{2m}$. Based on this feature selection analysis we work with the pattern vector $v_{Gs} = [c_3\ c_4\ \ln(|c_{2m}|)]^T$. The three components of $v_{Gs}$ extracted from the first 4 objects in the COIL dataset are shown in Fig.4 as a function of view numbers. Each line in a plot represents the coefficient extracted from one object. The lines in each plot are well separated thus indicating that the three coefficients are sufficient to distinguish different object classes. Fig.2(c) shows the feature distance computed using the three selected coefficients as a function of the edit distance. Compared with Fig.2(a), the selected coefficients are more linear with the edit distance than the unselected coefficients. From Fig.2(d), it is clear that although the unselected coefficients are less stable than the graph spectra, the selected coefficients offer best stability. We apply the pattern vectors of a) Laplacian spectra, b) unselected Ihara coefficients and c) selected Ihara coefficients to both the house sequences and the first four objects in COIL dataset. Fig.4 shows the clustering results obtained by performing PCA on the pattern vectors. The selected coefficients outperform both the Laplacian spectra and the unselected coefficients,and give clusters with better separation.

We then confine our attention to the COIL dataset, and evaluate the clustering performance obtained with different numbers of object classes. After performing PCA on the pattern vectors, we locate the clusters using the $K$-means method and calculate the Rand index. The Rand index for each pattern vector is plotted as a function of class number in Fig.4. The selected coefficients gives the best performance and the Laplacian spectra the poorest performance.

## 6 Conclusion

In this paper, we show how to construct pattern vectors for graph characterization using the Ihara zeta function. We compute polynomial coefficients using the reciprocal of the Ihara zeta function. We construct pattern vectors by performing feature selection on the coefficients. Unlike the samples of the Ihara zeta function, the coefficients are not prone to singularities due to poles. The method outperforms the Laplacian spectrum both in terms of stability and in clustering performance on md2 graphs.

## References

1. Bai, X., Hancock, E.R.: Recent results on heat kernel embedding of graphs. In: Brun, L., Vento, M. (eds.) GbRPR 2005. LNCS, vol. 3434, pp. 373–382. Springer, Heidelberg (2005)
2. Bass, H.: The Ihara-Selberg zeta function of a tree lattice. International Journal of Mathematics 6(3), 717–797 (1992)

3. Brooks, B.P.: The coefficients of the characteristic polynomial in terms of the eigenvalues and the elements of an n×n matrix. Applied Mathematics letters 19(6), 511–515 (2006)
4. Chung, F.K.: Spectral graph theory. American Mathematical Society (1997)
5. Haemers, W.H., Spence, E.: Enumeration of cospectral graphs. European Journal of Combinatorics 25(2), 199–211 (2004)
6. Hashimoto, K.: Zeta functions of finite graphs and representations of p-adic groups. Advanced Study of Pure Mathematics 15, 211–280 (1989)
7. Ihara, Y.: Discrete subgroups of PL(2, $k_{\varphi}$). In: Proceeding Symposium of Pure Mathematics, pp. 272–278 (1965)
8. Ihara, Y.: On discrete subgroups of the two by two projective linear group over p-adic fields. Journal of Mathematics Society Japan 18, 219–235 (1996)
9. Kotani, M., Sunada, T.: Zeta functions of finite graphs. Journal of Mathematics University of Tokyo 7(1), 7–25 (2000)
10. Luo, B., Wilson, R.C., Hancock, E.R.: Spectral embedding of graphs. Pattern Recognition 36(10), 2213–2223 (2003)
11. Rosenberg, S.: The Laplacian on a Riemannian manifold. Cambridge University Press, Cambridge (2002)
12. Ren, P., Wilson, R.C., Hancock, E.R.: Pattern vectors from the Ihara zeta function. In: IEEE International Conference on Pattern Recognition (submitted, 2008)
13. Sanfeliu, A., Fu, K.S.: A distance measure between attributed relational graphs for pattern recognition. IEEE Transanctions on Systems, Man, and Cybernetics 13, 353–362 (1983)
14. Scott, G., Storm, C.K.: The coefficients of the Ihara zeta function. Involve - A Journal of Mathematics 1(2), 217–233 (2008)
15. Storm, C.K.: The zeta function of a hypergraph. Electronic Journal of Combinatorics 13(1) (2006)
16. Wilson, R.C., Luo, B., Hancock, E.R.: Pattern vectors from algebraic graph theory. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(7), 1112–1124 (2005)