

A Spectral Generative Model for Graph Structure

Bai Xiao and Edwin R Hancock

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

Abstract. This paper shows how to construct a generative model for graph structure. We commence from a sample of graphs where the correspondences between nodes are unknown *ab initio*. We also work with graphs where there may be structural differences present, i.e. variations in the number of nodes in each graph and the edge-structure. The idea underpinning the method is to embed the nodes of the graphs into a vector space by performing kernel PCA on the heat kernel. The coordinates of the nodes are determined by the eigenvalues and eigenvectors of the Laplacian matrix, together with a time parameter which can be used to scale the embedding. Node correspondences are located by applying Scott and Longuet-Higgins algorithm to the embedded nodes. We capture variations in graph structure using the covariance matrix for corresponding embedded point-positions. We construct a point distribution model for the embedded node positions using the eigenvalues and eigenvectors of the covariance matrix. We show how to use this model to both project individual graphs into the eigenspace of the point-position covariance matrix and how to fit the model to potentially noisy graphs to reconstruct the Laplacian matrix. We illustrate the utility of the resulting method for shape-analysis using data from the COIL database.

1 Introduction

The literature describes a number of attempts aimed at developing probabilistic models for variations in graph-structure. Some of the earliest work was that of Wong, Constant and You [4], who capture the variation in graph-structure using a discretely defined probability distribution. Bagdanov and Worring [3] have overcome some of the computational difficulties associated with this method by using continuous Gaussian distributions. For problems of graph matching Christmas, Kittler and Petrou [1], and Wilson and Hancock [2] have used simple probability distributions to measure the similarity of graphs. There is a considerable body of related literature in the graphical models community concerned with learning the structure of Bayesian networks from data [5].

Recently there has been some research aimed at applying central clustering techniques to cluster graphs. However, rather than characterising them in a statistical manner, a structural characterisation is adopted. For instance, both Lozano and Escolano [7], and Bunke et al. [8] summarize the data using a supergraph. Each sample can be obtained from the super-graph using edit operations. However, the way in which the super-graph is learned or estimated is not statistical in nature. Jain and Wysotzki, adopt a geometric approach which aims to embed graphs in a high dimensional space by means of the Schur-Hadamard inner product [9]. Central clustering methods are then

deployed to learn the class structure of the graphs. The embedding offers the advantage that it is guaranteed to preserve structural information present. Unfortunately, the algorithm does not provide a means of statistically characterising the modes of structural variation encountered.

Hence, the methods described in the literature fall well short of constructing genuine generative models from which explicit graph structures can be sampled. The aim in this paper use ideas from the spectral analysis of graphs to construct a simple and explicit generative model for graph-structure. To this end, we use the heat-kernel embedding to construct a generative model for graph-structure. We use the heat-kernel to map the nodes of a graph to positions in a vector space. Our aim is to construct a statistical model that can account for the distribution of embedded point-positions for corresponding nodes in a sample of graphs. A reference graph is selected, and the correspondences between the nodes of each sample graph and the reference graph are established using the point-matching method of Scott and Longuet-Higgins [6]. We capture variations in graph structure using the covariance matrix for the corresponding embedded point-positions. We construct a point distribution model for the embedded node positions using the eigenvalues and eigenvectors of the covariance matrix. We show how to use this model to both project individual graphs into the eigenspace of the point-position covariance matrix and to fit the model to potentially noisy graphs to reconstruct the Laplacian matrix. We illustrate the utility of the resulting method for shape-analysis. Here we perform experiments on the COIL data-base, and show that the model can be used to both construct pattern spaces for sets of graphs and to cluster graphs.

2 Heat Kernel Embedding

We are interested in using the heat-kernel to embed the nodes of a graph in a vector space. To commence, suppose that the graph under study is denoted by $G = (V, E)$ where V is the set of nodes and $E \subseteq V \times V$ is the set of edges. Since we wish to adopt a graph-spectral approach we introduce the adjacency matrix A for the graph where the elements are

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

We also construct the diagonal degree matrix D , whose elements are given by $D(u, u) = \sum_{v \in V} A(u, v)$. From the degree matrix and the adjacency matrix we construct the Laplacian matrix $L = D - A$, i.e. the degree matrix minus the adjacency matrix. The normalised Laplacian is given by $\hat{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$. The spectral decomposition of the normalised Laplacian matrix is $\hat{L} = \Phi \Lambda \Phi^T$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{|V|})$ is the diagonal matrix with the ordered eigenvalues as elements and $\Phi = (\phi_1 | \phi_2 | \dots | \phi_{|V|})$ is the matrix with the ordered eigenvectors as columns. Since \hat{L} is symmetric and positive semi-definite, the eigenvalues of the normalised Laplacian are all positive. The eigenvector associated with the smallest non-zero eigenvector is referred to as the Fiedler-vector. We are interested in the heat equation associated with the Laplacian, i.e. $\frac{\partial h_t}{\partial t} = -\hat{L} h_t$, where h_t is the heat kernel and t is time. The heat kernel can hence be viewed as describing the flow of information across the edges of the graph

with time. The rate of flow is determined by the Laplacian of the graph. The solution to the heat equation is found by exponentiating the Laplacian eigen-spectrum, i.e. $h_t = \Phi \exp[-t\Lambda]\Phi^T$.

We use the heat kernel to map the nodes of the graph into a vector-space. Let Y be the $|V| \times |V|$ matrix with the vectors of co-ordinates as columns. The vector of co-ordinates for the node indexed u is hence the u^{th} column of Y . The co-ordinate matrix is found by performing the Young-Householder decomposition $h_t = Y^T Y$ on the heat-kernel. Since $h_t = \Phi \exp[-\Lambda t]\Phi^T$, $Y = \exp[-\frac{1}{2}\Lambda t]\Phi^T$. Hence, the co-ordinate vector for the node indexed u is

$$y_u = (\exp[-\frac{1}{2}\lambda_1 t]\phi_1(u), \exp[-\frac{1}{2}\lambda_2 t]\phi_2(u), \dots, \exp[-\frac{1}{2}\lambda_{|V|} t]\phi_{|V|}(u))^T$$

The kernel mapping $\mathcal{M} : V \rightarrow \mathcal{R}^{|V|}$, embeds each node on the graph in a vector space $\mathcal{R}^{|V|}$. The heat kernel $h_t = Y^T Y$ can also be viewed as a Gram matrix, i.e. its elements are scalar products of the embedding co-ordinates. Consequently, the kernel mapping of the nodes of the graph is an isometry. The squared Euclidean distance between the nodes u and v is given by

$$d_E(u, v)^2 = (y_u - y_v)^T (y_u - y_v) = \sum_{i=1}^{|V|} \exp[-\lambda_i t] (\phi_i(u) - \phi_i(v))^2 \quad (2)$$

3 Generative Model

Our aim is to construct a generative model that can be used to represent the statistical variations in a sample of graphs. Let the sample be $T = \{G_1, G_2, \dots, G_k, \dots, G_K\}$ where the k th graph $G_k = (V_k, E_k)$ has node-set V_k and edge-set E_k . The result of performing heat-kernel embedding of the nodes of the k th graph is a matrix of co-ordinates Y_k .

Our aim in this paper is to construct a generative model that can be used to describe the distribution of embedded node co-ordinates for the sample of graphs. Since the graphs contain different numbers of nodes, we truncate the co-ordinate matrices to remove the spatial dimensions corresponding to insignificant eigen-modes of the kernel matrix. Hence, we retain just the first N rows of each co-ordinate matrix. For the graph G_k the truncated node co-ordinate matrix is denoted by \hat{Y}_k .

3.1 Node Correspondences

To construct the generative model, we require correspondences between the nodes of each sample graph and the nodes of a reference structure. Here we take the reference graph to be the graph in the sample with the largest number of nodes. This graph has index $k^* = \arg \max_{G_k \in T} |V_k|$.

To locate the correspondences between the nodes of each sample graph and those of the reference graph, we use the Scott and Longuet-Higgins algorithm. The algorithm uses the distances between the reference graph nodes and the nodes of the sample graph

k to compute an affinity matrix. Let \hat{y}_k^i is the i th column vector of the truncated co-ordinate matrix \hat{Y}_k , i.e. the co-ordinates of the node $i \in V_k$. For the node i of the sample graph G_k and the node j the affinity matrix element is

$$R_{k,k^*}(i, j) = \exp[-\frac{1}{\sigma^2}(\hat{y}_k^i - \hat{y}_{k^*}^j)^T(\hat{y}_k^i - \hat{y}_{k^*}^j)]$$

where σ is a scaling parameter.

According to Scott and Longuet-Higgins [10] if R_{k,k^*} is a positive definite $|V_k| \times |V_{k^*}|$ matrix, then the $|V_k| \times |V_{k^*}|$ orthogonal matrix R_{k,k^*}^* that maximises the quantity $Tr[R_{k,k^*}(R_{k,k^*}^*)^T]$ may be found by performing singular value decomposition. To do this they perform the matrix factorisation $R_{k,k^*} = V\Delta U^T$, where V is a $|V_D| \times |V_D|$ orthogonal matrix, U is a $|V_{k^*}| \times |V_{k^*}|$ orthogonal matrix and Δ is a $|V_k| \times |V_{k^*}|$ matrix whose off-diagonal elements $\Delta_{i,j} = 0$ if $i \neq j$ and whose “diagonal” elements $\Delta_{i,i}$ are non-zero. Suppose that E is the matrix obtained from Δ by making the diagonal elements $\Delta_{i,i}$ unity. The matrix R_{k,k^*}^* which maximises $Tr[R_{k,k^*}(R_{k,k^*}^*)^T]$ is $R_{k,k^*}^* = VEU^T$. The element $R_{k,k^*}^*(i, j)$ indicates the strength of association between the node $i \in V_k$ in the graph G_k and the node $j \in V_{k^*}$ in the reference graph. The rows of R_{k,k^*}^* , index the nodes in the graph G_k , and the columns index the nodes of the reference graph G_{k^*} . If $R_{k,k^*}^*(i, j)$ is both the largest element in row i and the largest element in column j then we regard these nodes as being in one-to-one correspondence with one-another. We record the state of correspondence using the matrix C_{k,k^*} . If the pair of nodes (i, j) satisfies the row and column correspondence condition, then we set $C_{k,k^*}(i, j) = 1$, otherwise $C_{k,k^*}(i, j) = 0$.

3.2 Embedded Point Distribution Model

Once we have correspondences to hand, then we can construct the generative model for the set of graphs. To do this we model variations in the positions of the embedded points using a point distribution model. We commence by computing the mean point positions. The matrix of mean-position co-ordinates and the associated covariance matrix are

$$\hat{X} = \frac{1}{T} \sum_{k \in T} C_{k,k^*}^T \hat{Y}_k$$

$$\Sigma = \frac{1}{T} \sum_{k \in T} (C_{k,k^*}^T \hat{Y}_k - \hat{X})(C_{k,k^*}^T \hat{Y}_k - \hat{X})^T$$

To construct the point-distribution model, we perform the eigendecomposition $\Sigma = \Psi \Gamma \Psi^T$ where $\Gamma = diag(\gamma_1, \gamma_2, \dots, \gamma_K)$ is the diagonal matrix of ordered eigenvalues and $\Psi = (\psi_1 | \dots | \psi_K)$ is the matrix with the correspondingly ordered eigenvectors as columns.

We deform the mean-embedded node positions in the directions of the leading eigenvectors of the point-position covariance matrix Σ . Let $\tilde{\Psi}$ be the result of truncating the matrix Ψ after S columns and let b be a parameter-vector of length S . We convert the mean point position matrix with a long vector form. Let $Col_i(\hat{X})$ be the

i th column of the mean-point position matrix \hat{X} . The long vector is given by $\hat{Z} = (Col_1^T(\hat{X}), Col_2^T(\hat{X}), \dots)$. The long vector corresponding to deformed point set position is $\tilde{Z} = \hat{Z} + \tilde{\Psi}b$. The matrix with deformed point position as column is \hat{X} .

An observed configuration of embedded nodes \tilde{Y} may be fitted to the model. To do this the best fit parameters estimated using the least squares procedure

$$b^* = \arg \min_b (\tilde{Y} - \hat{X} - \tilde{\Psi}b)^T (\tilde{Y} - \hat{X} - \tilde{\Psi}b)$$

The best-fit parameter vector is $b^* = \tilde{\Psi}^T (\tilde{Y} - \hat{X})$ and the reconstructed set of embedded point positions is $\tilde{Y}^* = \hat{X} + \tilde{\Psi}b^*$. From the reconstructed point-positions we can recover the Laplacian matrix for the corresponding graph. The heat-kernel for the reconstructed embedded graph is $h_t^* = (\tilde{Y}^*)^T (\tilde{Y}^*) = \exp[-\hat{L}^*t]$ and the Laplacian is hence $\hat{L}^* = -\frac{1}{t} \ln\{(\tilde{Y}^*)^T (\tilde{Y}^*)\}$. From the reconstructed Laplacian we can compute the corresponding adjacency matrix

$$A^* = D - D^{\frac{1}{2}} \hat{L}^* D^{\frac{1}{2}} = D + \frac{1}{t} D^{\frac{1}{2}} \ln\{(\tilde{Y}^*)^T (\tilde{Y}^*)\} D^{\frac{1}{2}}.$$

Finally, the similarity of a pair of graphs can be measured using the difference in their best-fit parameter vectors. Since the parameter-vector is just the projection of the corresponding graph into the eigenspace of the model, the difference in parameter vectors is related to the distance between graphs in the eigenspace. Suppose that the graphs G_{k_1} and G_{k_2} have best fit parameter vectors $b_{k_1}^*$ and $b_{k_2}^*$ respectively. The Euclidean distance between the parameter vectors is

$$d^2(k_1, k_2) = (b_{k_1}^* - b_{k_2}^*)^T (b_{k_1}^* - b_{k_2}^*) = (\hat{Y}_{k_1} - \hat{Y}_{k_2})^T \tilde{\Psi} \tilde{\Psi}^T (\hat{Y}_{k_1} - \hat{Y}_{k_2})$$

4 Experiments

In this section we provide some experimental evaluation of our generative model for real-world data. We use the COIL data-base. The data-set contains multiple views of the same object in different poses with respect to the camera. Example images from the data-set are shown in Figures 1. We extract the feature points using the methods of [11]. 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.

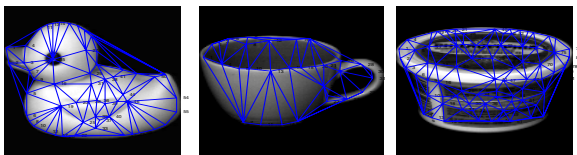


Fig. 1. Three objects from the COIL data-base

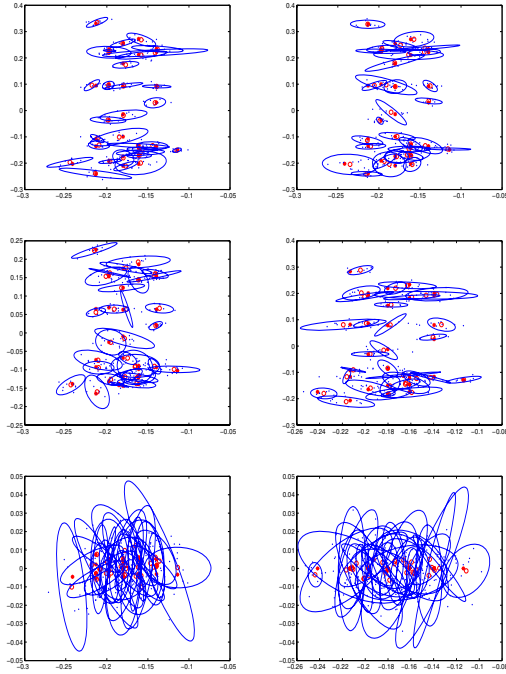


Fig. 2. Embedded point positions and fitted covariance ellipsoids with varying t (from left to right, top to bottom $t = 0.001, 0.01, 0.1, 1, 10, 100$) for the heat kernel

In Figure 2 we show the result of projecting the nodes into the space spanned by the leading two eigenvectors of the heat-kernel. The different panels in the figure are for different values of t , from left to right and top to bottom the t are 0.001, 0.01, 0.1, 1, 10, 100. For this experiment we have taken 15 images from the duck sequence. Each

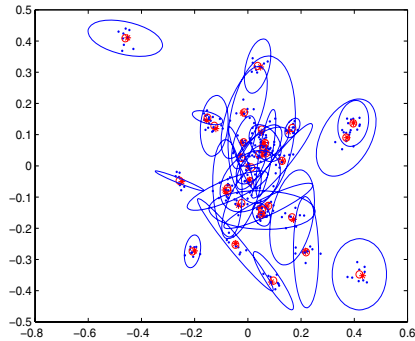


Fig. 3. Embedded point positions and fitted covariance ellipsoids for Laplacian matrix

blue point in the embedding corresponds to a single node of one of the 15 sample graphs. Superimposed on the node-positions as red-points are the locations of the mean node positions. Around each mean node position we have drawn an ellipse. The major and minor axes of the ellipse are in the principal directions of the eigenvectors of the node-position covariance matrix and the lengths of the semi-major axes are the corresponding eigenvalues. There are a number of features to note from this figure. First, for small values of t the embedded points form relatively compact clusters. Second, there is a significant variation in the size and directions of the ellipses. The compactness of the clusters supports the feasibility of our embedding approach and the variation in the ellipses underpins the need for a relatively complex statistical model to describe the distribution of embedded point positions. As the value of t increases then so the overlap of the ellipses increases. For comparison Figure 3 shows the result of repeating the embedding by using the Laplacian spectrum. The node-clusters are more overlapped than those obtained with the heat kernel for small values of t .

To investigate the role of the number of Laplacian eigenmodes in the reconstruction of the graph-structure we have examined the value of the Froebenius norm $F = \|A - A^*\|$ between the original graph adjacency matrix A and the reconstructed adjacency matrix A^* computed by fitting the generative model. In Figure 4 we show the value of F as a function of the number of eigenmodes used. The different curves in the

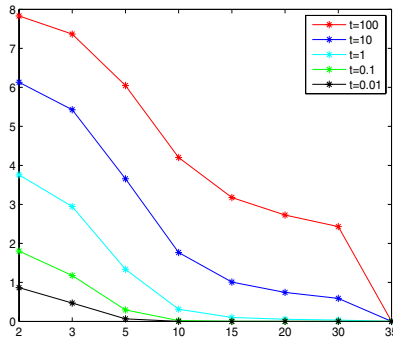


Fig. 4. Froebenius norm as a function of number of eigenmodes

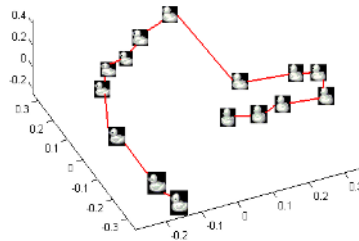


Fig. 5. Eigen-projection of graphs from 15 images in duck sequence

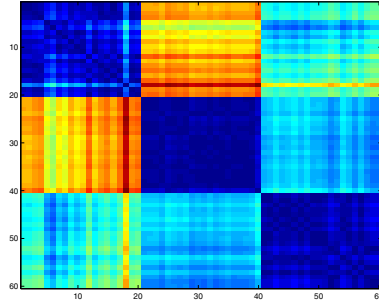


Fig. 6. Distance matrix for fitted parameter vectors

plot are for different values of t . The best reconstructions are obtained with small values of t and an increasing number of eigenmodes.

In Figures 5 we show the result of projecting the embedded node vectors for the graphs extracted from the duck sequence in the COIL data-base onto the eigenvectors of the embedded node position covariance matrix Σ . We have placed a thumbnail image at the location specified by the first three components of the parameter-vector b . The line connecting the thumbnails corresponds to the sequence order of the original images. The main feature to note is that neighboring images in the sequence are close together in the eigenspace.

We have also experimented with the generative model as a means of clustering graphs. In Figure 6 we show the matrix distances between the best fit parameter vectors. The main feature to note is that there is a clear block structure emerges corresponding to the different objects.

5 Conclusions

In this paper we have used the heat-kernel embedding of graphs to construct a generative model for graph structure. The mapping allows nodes of the graphs under study to be embedded as points in a vector-space. The idea underpinning the generative model is to construct a point-distribution model for the positions of the embedded nodes. The required correspondences needed to construct this model are recovered using the Scott and Longuet-Higgins algorithm. The method proves to be effective for computing distances between graphs and also for clustering graphs.

Our future plans revolve around the use of a mixture model to describe the positions of the embedded nodes, and to assess uncertainty in the computation of correspondence.

References

1. W.J. Christmas, J. Kittler, and M. Petrou. Structural matching in computer vision using probabilistic relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):749–764, 1995.
2. R.C. Wilson and E.R. Hancock. Structural matching by discrete relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):634–648, June 1997.

3. A.D. Bagdanov and M. Worring, First Order Gaussian Graphs for Efficient Structure Classification *Pattern Recognition*, **36**, pp. 1311-1324, 2003.
4. A.K.C Wong, J. Constant and M.L. You, Random Graphs *Syntactic and Structural Pattern Recognition*, World Scientific, 1990.
5. D. Heckerman, D. Geiger and D.M. Chickering, Learning Bayesian Networks: The combination of knowledge and statistical data *Machine Learning*, **20**, pp. 197-243, 1995.
6. G.L.Scott,H.C.Longuet-Higgins An algorithm for associating the features of two images *Proceedings of the Royal Society of London*, **244**, pp. 21-26, 1991.
7. M. A. Lozano and F. Escolano, ACM Attributed Graph Clustering for Learning Classes of Images In *Graph Based Representations in Pattern Recognition*, LNCS 2726, pp.247-258, 2003.
8. H.Bunke et al., Graph Clustering Using the Weighted Minimum Common Supergraph. In *Graph Based Representations in Pattern Recognition*, LNCS 2726, pp.235-246, 2003.
9. B. J. Jain and F. Wysotzki, Central Clustering of Attributed Graphs. *Machine Learning*, Vol. 56, pp. 169-207, 2004.
10. G.L.Scott and H.C.Longuet-Higgins, An Algorithm for Associating the Features of two Images. *Proceedings of the Royal Society of London*, Vol. 244, pp. 21-26, 1991.
11. C.G.Harris, and M.J.Stephens, "A Combined Corner and Edge Detector", *Fourth Alvey Vision Conference*, pp. 147-151, 1994.