

Graph Characteristic from the Gauss-Bonnet Theorem

Hewayda ElGhawalby^{1,2} and Edwin R. Hancock¹

¹ Department of Computer Science, University of York,
YO10 5DD, UK

² Faculty of Engineering, Suez Canal university, Egypt
{howaida, erh}@cs.york.ac.uk

Abstract. In this paper we explore how to characterise graphs using the Gauss-Bonnet theorem. Using the Gaussian curvatures estimated from first-order cycles we compute a global estimate of the Euler index using the Gauss-Bonnet theorem. We commence by embedding the nodes of a graph in a manifold using the heat-kernel mapping. From this mapping we are able to compute the geodesic and Euclidean distance between nodes, and these can be used to estimate the sectional curvatures of edges. Assuming that edges reside on hyper-spheres, we use Gauss's theorem to estimate the Gaussian curvature from the interior angles of geodesic triangles formed by first-order cycles in the graph. From the Gaussian curvatures we make a global estimate of the Euler index of the manifold using the Gauss-Bonnet theorem. Experiments show how the Gaussian curvatures and the Euler characteristics can be used to cluster Delaunay triangulations extracted from real world images.

Keywords: Manifold embedding, Heat kernel, Hausdorff distance, Gaussian curvature, Graph matching.

1 Introduction

The task of characterising the structure of graphs for the purposes of matching, clustering or indexing is a challenging one for which no definitive solution exists. If the characterisation is attempted without recourse to edge weight or node attribute information, then the regularity of many graph structures means that local characteristics are ambiguous, and only global structural information can be used. Examples of the available global characteristics include Cheeger numbers, perimeters, diameters and edge densities. This problem can be particularly severe for k-nearest neighbor graphs, Delaunay triangulations and strongly regular graphs.

Of course, one solution to this problem is to use graph-spectral information. Although the spectrum of a graph (i.e. the set of eigenvalues of the Laplacian matrix) is a global permutation invariant characterisation, the set of eigenvectors are not permutation invariant and can only be used to characterise graphs on a node-by-node basis. Unfortunately, eigenvectors are notoriously unstable,

and small perturbations in graph structure can lead to significant variations in eigenvector directions. Moreover, the estimation of eigenvectors is subject to directional ambiguity.

Another solution to the problem is to make use of spectral geometry [5,2,9]. This allows the nodes of graph to be embedded on a manifold, and the graph to be characterised using the local differential structure of the manifold. Hence, local graph structure can be characterised using invariant quantities such as curvature or torsion [8]. This is the approach adopted by Bai and Hancock [12] who use the heat kernel to construct, embed and characterise the edges of a graph using sectional curvature. This curvature is computed using the difference between the geodesic (edge length) and Euclidean internode distance under the embedding. Of course, the sectional curvature is one of many possible differential invariants that can be computed.

The aim in this paper is therefore to look in greater depth at the diversity of quantities that are furnished by the differential geometry of the manifold. We turn our attention to geodesic triangles. These are formed when a first-order cycle in a graph is embedded. By making a locally hyper-spherical approximation to the geometry of the manifold and the turning angles of the geodesics, we are able to estimate the Gaussian curvature of the triangle. Furthermore, the Gauss-Bonnet theorem allows us to estimate a local Euler characteristic from abutting geodesic triangles. We demonstrate how the estimated curvature information can be used for matching and clustering graphs, extracted from real world image data.

2 Manifold Embeddings

We commence by establishing a matrix representation for graph-structure. Let the graph under study be denoted by $G = (V, E)$ where V is the set of nodes and $E \subseteq V \times V$ is the set of edges. The elements of the adjacency matrix A of the graph G are:

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

To construct the Laplacian matrix of graph G we first establish a diagonal degree matrix D with elements

$$D(u, u) = \sum_{v \in V} A(u, v) = d_u \quad (2)$$

The Laplacian matrix $L = D - A$, is the degree matrix minus the adjacency matrix. Elementwise, we have

$$L(u, v) = \begin{cases} d_u & \text{if } u = v \\ -1 & \text{if } (u, v) \in E \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The normalized Laplacian is given by $\hat{L} = D^{-1/2} L D^{-1/2}$ and has spectral decomposition $\hat{L} = \Phi \Lambda \Phi^T = \sum_{i=1}^{|V|} \lambda_i \phi_i \phi_i^T$ where $|V|$ is the number of nodes,

$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{|V|})$, ($0 < \lambda_1 < \lambda_2 < \dots < \lambda_{|V|}$) is the diagonal matrix with the ordered eigenvalues as elements and $\Phi = (\phi_1|\phi_2|\dots|\phi_{|V|})$ is the matrix with the corresponding ordered eigenvectors as columns. We follow Bai and Hancock [12] and make use of the heat kernel embedding. The heat kernel plays an important role in spectral graph theory. It encapsulates the way in which information flows through the edges of graph over time under the heat equation, and is the solution of the partial differential equation

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

where h_t is the heat kernel and t is the time. The solution is found by exponentiating the Laplacian eigenspectrum as follows

$$h_t = \exp[-\widehat{L}t] = \Phi \exp[-t\Lambda]\Phi^T \tag{5}$$

For the heat kernel, the matrix of embedding coordinates Y (i.e. the matrix whose columns are the vectors of node coordinates) is found by performing the Young-Householder [13] decomposition $h_t = Y^T Y$ as a result the matrix of node embedding coordinates is

$$Y = (y_1|y_2|\dots|y_{|V|}) = \exp[-\frac{1}{2}t\Lambda]\Phi^T \tag{6}$$

where y_u is the coordinate vector for the node u . In the vector space, the Euclidean distance between the nodes u and v of the graph is

$$d_e^2(u, v) = (y_u - y_v)^T (y_u - y_v) = \sum_{i=1}^{|V|} \exp[-\lambda_i t] (\phi_i(u) - \phi_i(v))^2 \tag{7}$$

When a pair of nodes are connected by an edge, then $d_G(u, v) = 1$. We assume that the geodesic between the pair of nodes can be approximated by a segment of circle whose arc-length is the geodesic distance and whose chord-length is the Euclidean distance. If the radius of the circle is $r_s(u, v)$ and the tangent vector to the circle arc undergoes a change in direction of $2\theta(u, v)$, then we have

$$d_g(u, v) = 2r_s(u, v)\theta(u, v) \tag{8}$$

and

$$d_e(u, v) = 2r_s(u, v) \sin \theta(u, v) \tag{9}$$

3 The Gauss-Bonnet Theorem

The Gauss-Bonnet Theorem links the topology and geometry of a surface in an elegant and compact manner. Spivak [10] and Stillwell [11] give accounts of the early history of its development and application.

For a smooth compact oriented Riemannian 2-manifold M , let Δ_g be a triangle on M whose sides are geodesics, i.e. paths of shortest length on the manifold.

Further, let α_1, α_2 and α_3 denote the interior angles of the triangle. According to Gauss's theorem, if the Gaussian curvature K (i.e. the product of the maximum and minimum curvatures at a point on the manifold) is integrated over Δ_g , then

$$\int_{\Delta_g} K dM = \sum_{i=1}^3 \alpha_i - \pi \tag{10}$$

where dM is the Riemannian volume element.

3.1 Geometric Preliminaries

To estimate the Gaussian curvature from the above, we must determine the interior angles α_i of the geodesic triangle. To this end we assume that T is a triangulation of a smooth manifold M , Δ_g be a geodesic triangle on M with angles $\{\alpha_i\}_{i=1}^3$ and geodesic edge lengths $\{d_{gi}\}_{i=1}^3$. Further suppose that Δ_e be the corresponding Euclidean triangle with edge lengths $\{d_{ei}\}_{i=1}^3$ and interior angles $\{\varphi_i\}_{i=1}^3$. We assume that the geodesic index i is a great arc on a sphere with radius r_i , $i = 1, 2, 3$. By averaging over the constituent geodesic edges, we treat the geodesic triangles as residing on a hyper-sphere with radius $r = \frac{1}{3} \sum_{i=1}^3 r_i$. To compute the interior angle of the geodesic triangle, we make use of two formulae. First, from the geometry of the sphere

$$\sin\left(\frac{\alpha}{2}\right) = \frac{r}{\sqrt{4r^2 - d_e^2}} \tag{11}$$

Second, when using the double integral to compute the area of a triangle on a sphere

$$\frac{\alpha}{2} = \frac{\pi r^2}{6r^2 - d_e^2} \tag{12}$$

Substituting from (11) and (12) in the Taylor expansion for $\sin(\alpha/2)$, we have

$$\sin(\alpha/2) \simeq (\alpha/2) - \frac{1}{3!}(\alpha/2)^3 \tag{13}$$

$$\frac{\sin(\alpha/2)}{(\alpha/2)} \simeq 1 - \frac{\alpha^2}{24} \tag{14}$$

So we can deduce that

$$\alpha^2 \simeq 24\left(1 - \frac{6r^2 - d_e^2}{\pi r \sqrt{4r^2 - d_e^2}}\right) \tag{15}$$

3.2 The Euler Characteristic

We can now use the Gauss Bonnet Theorem to estimate a global topological characteristic for the manifold. We commence by triangulating M so that each face is a geodesic triangle.

Summing (11) over all the triangular faces gives that the integral of K over all M is 2π times the Euler characteristic of M , i.e.

$$\int_M K dM = 2\pi\chi(M) \tag{16}$$

4 Hausdorff Distance

Once sets of Gaussian curvatures for the geodesic triangles of a graph are to hand, we require a means of comparing different graphs so that we can compute graph similarity. Here to avoid the need to compute explicit node correspondences using the Hausdorff distance between curvature sets. The Hausdorff distance has been used for a number of matching and recognition problems. It provides a means of computing the distance between sets of unordered observations when the correspondences between the individual items are unknown. In its most general setting, the Hausdorff distance is defined between compact sets in a metric space. Given two such sets, we consider for each point in one set the closest point in the second set. Hausdorff distance is the maximum over all these values. More formally, the classical Hausdorff distance (HD) [6] between two finite point sets A and B is given by

$$H(A, B) = \max(h(A, B), h(B, A)) \tag{17}$$

where the directed Hausdorff distance from A to B is defined to be

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{18}$$

and $\|\cdot\|$ is some underlying norm on the points of A and B (e.g., the L2 or Euclidean norm). Regardless of the norm, the Hausdorff metric captures the notion of the worst match between two objects. The computed value represents the largest distance between a point in one set and a point in the other one. Several variations of the Hausdorff distance have been proposed as alternatives to the maximum of the minimum approach in the classical one. These include Hausdorff fraction, Hausdorff quantile [6] and Spatially Coherent Matching [1]. A robust modified Hausdorff distance (MHD) based on the average distance value instead of the maximum value was proposed by Dubuisson and Jain [3], in this sense they defined the directed distance of the MHD as

$$h(A, B) = \frac{1}{N_A} \sum_{a \in A} \min_{b \in B} \|a - b\| \tag{19}$$

For this paper we will consider the classical and the modified Hausdorff distances. Using these ingredients we can describe how Hausdorff distances can be extended to graph-based representations. To commence let us consider two graphs $G_1 = (V_1, E_1, T_1, k_1)$ and $G_2 = (V_2, E_2, T_2, k_2)$, where V_1, V_2 are the sets of nodes, E_1, E_2 the sets of edges, T_1, T_2 are the sets of triangles, and k_1, k_2 the

sets of Gaussian curvatures for geodesic triangles defined in the previous section. We can now write the distances between two graphs as follows:

1) The classic Hausdorff distance (HD) is

$$h_{HD}(G_1, G_2) = \max_{i \in V_1} \min_{j \in V_2} \|k_2(j) - k_1(i)\| \quad (20)$$

2) The modified Hausdorff distance (MHD) is

$$h_{MHD}(G_1, G_2) = \frac{1}{|V_1|} \sum_{i \in V_1} \min_{j \in V_2} \|k_2(j) - k_1(i)\| \quad (21)$$

5 Experiments

In our experiments we use the standard CMU, MOVI and chalet house sequences as our data set [7]. These data sets contain different views of model houses from equally spaced viewing directions. From the house images, corner features are extracted, and Delaunay graphs representing the arrangement of feature points are constructed. Our data consists of ten graphs for each of the three houses. Each node in a Delaunay graph belongs to a first order cycle, and as a result the graph is a triangulation.

To commence, we obtain the manifold embedding for each of the thirty graphs following the steps mentioned in Section 2. We compute the Euclidean distances between the nodes in each graph based on the heat kernel with the values of $t = 10.0, 1.0, 0.1$ and 0.01 . Using the Delaunay graphs to triangulate the manifold, for each geodesic triangle we compute the interior angles deduced using the procedure outlined in Section 3.1. For each triangle we then compute the Gaussian curvature as described in Section 3. We use the Gaussian curvatures of the geodesic triangles as features for the purposes of gauging the similarity of graphs. Using the set of Gaussian curvatures for pairs of graphs, we compute a matrix of Hausdorff distances the thirty different graphs. Here we explore the use of both the classical Hausdorff distance and the modified Hausdorff distance. Finally, we apply multidimensional scaling (MDS) to the distance matrices to embed the individual graphs into a 2D space. Here each graph is represented by a single point. Figure 1 shows the results obtained using the classical Hausdorff distance. The subfigures are ordered from left to right, using $t = 10.0, 1.0, 0.1$ and 0.01 in the heat kernel. Figure 2 shows the corresponding results obtained when the modified Hausdorff distance is used.

Of course, there are alternatives to the heat-kernel embedding, and the procedure for computing Gaussian curvatures can be applied to these too. When the Euclidean distances are computed directly from the Laplacian embedding instead (i.e. $Y = \sqrt{\Lambda} \Phi^T$), then the results shown in Figure 3 are obtained. Here the left-hand subfigure is obtained when using the HD and the right-hand one when using MHD .

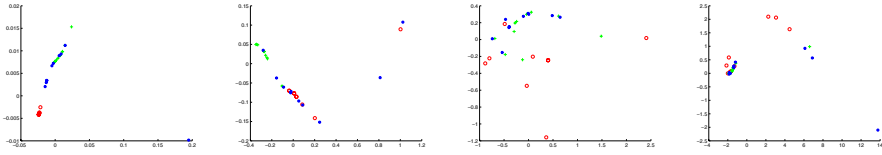


Fig. 1. MDS embedding obtained using HD for the houses data

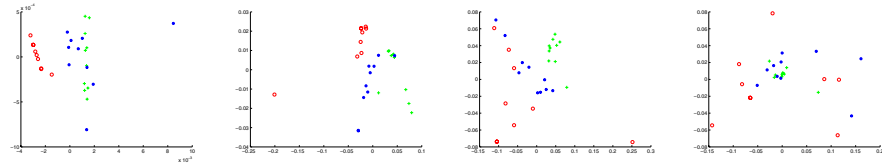


Fig. 2. MDS embedding obtained using MHD for the houses data

To investigate the data in more detail Table 1 shows the rand index for the data as a function of t . This index is computed as follows: We compute the mean for each cluster, we then compute the distance from each point to each mean. If the distance from the correct mean is smaller than those to remaining means, then the classification is correct, if not then classification is incorrect. The rand index is $R = (\# \text{ incorrect}) / (\# \text{ incorrect} + \# \text{ correct})$.

Table 1. A rand index vs. t

	lap	t=10	t=1.0	t=0.1	t=0.01
HD	0.2667	0.2333	0.3000	0.5333	0.5333
MHD	0.1333	0.1667	0.0333	0.1333	0.4000

In previous work [4], we used the sectional curvature associated with the edges of a graph under the heat kernel embedding as a feature for the purposes of gauging the similarity of graphs, and hence clustering them. Table 2 shows the rand index for the results obtained as a function of t .

A comparison shows that the curvature attributes associated with the edges give a slightly better clusters than those obtained using the attributes derived

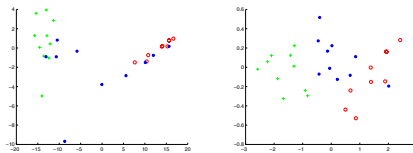


Fig. 3. MDS embedding obtained using curvatures computed from the Laplacian eigenmap and distances computed using HD (left) and MHD (right) for houses data

Table 2. A rand index vs. t . induced when the sectional curvature of the edges was used.

	lap	t=10	t=1.0	t=0.1	t=0.01
HD	0.0.2333	0.1000	0.1667	0.4333	0.0333
MHD	0.1000	0.1333	0.2333	0.1333	0.0333

Table 3. A rand index vs. t . induced when the sectional curvature of the edges was used.

lap	t=10	t=1.0	t=0.1	t=0.01
0.0.3333	0.4667	0.2333	0.3000	0.4000

from the triangles. To take this study one step further we have constructed normalised histograms of the Gaussian curvature associated with the geodesic triangles on each graph. The histograms can be treated as vectors, and can be used as input to principal components analysis. Using the leading two principal component projections we repeat the Rand index computation, The error deduced is shown in Table 3, which ranges between 0.23 and 0.47. Next, we consider the Euler characteristic whose computation was detailed in Section 3.2. In Table 4 we list the mean and variance for each group of graphs (for each house).

Table 4. The mean and variance for the Euler characteristic of the manifold embedding of each graph

		lap	t=10	t=1.0	t=0.1	t=0.01
1st house	Mean	0.0217	-0.1725	2.2361	1.2835	5.1177
	Variance	0.0099	0.0001	0.1830	4.8437	1.7395
2nd house	Mean	0.0850	-0.7514	4.0221	49.1613	8.3172
	Variance	0.1344	0.0003	0.1077	12.4022	10.2801
3rd house	Mean	0.0545	-0.3970	3.5168	16.1415	6.1704
	Variance	2.4059	0.0120	0.8116	150.9340	17.5878

Finally, we investigate how the Gaussian curvatures of the geodesic triangles are distributed over the Delaunay graph. Figures 4, 5 and 6 show distribution for sample embeddings computed from the heat kernel at ($t=0.1$).

From the sequence it is clear that the Gaussian curvature distribution over the different views of each house is stable, moreover it moves smoothly from positive (elliptical) to negative (hyperbolic) regions. This suggests that the arrangement of triangles and their Gaussian curvatures could be used as the basic of a matching algorithm.

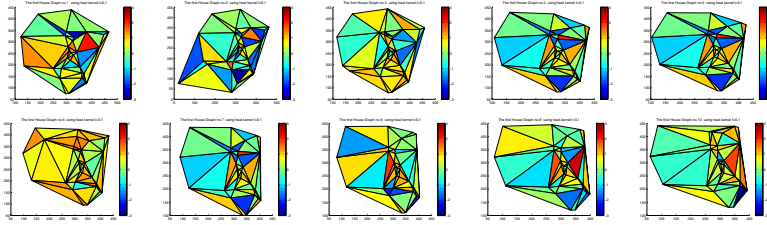


Fig. 4. The distribution of the Gaussian curvatures of the geodesic triangles for the ten graphs of the 1st house

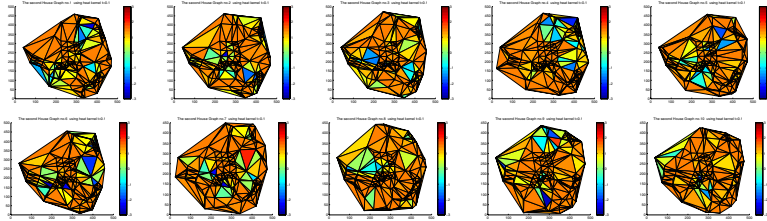


Fig. 5. The distribution of the Gaussian curvatures of the geodesic triangles for the ten graphs of the 2nd house

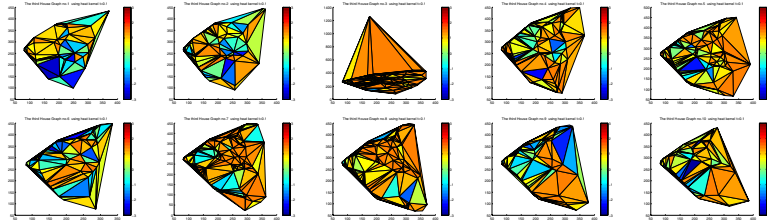


Fig. 6. The distribution of the Gaussian curvatures of the geodesic triangles for the ten graphs of the 3rd house

6 Conclusion and Future Work

We have shown how to compute the Gaussian curvatures of geodesic triangles and the Euler characteristic for graphs embedded on a manifold using the heat kernel. Experiments show that these two quantities provide effective features for graph clustering, but perform marginally worse than the sectional curvatures of the edges. However, they do lead to a deeper representation of the differential structure of embedded graphs. As future work we have a number of different lines of research in mind. First we aim to explore whether our characterisation can lead to the construction of Grassmanian manifolds for graphs. Second, we

aim to explore whether the arrangement of geodesic triangles and their Gaussian curvature can be used for attributed graph matching.

References

1. Boykov, Y., Huttenlocher, D.: A new bayesian framework for object recognition. In: Proceeding of IEEE Computer Society Conference on CVPR, vol. 2, pp. 517–523 (1999)
2. Brooks, R.: *Geometry, Spectral theory, Groups, and Dynamics*. Amer. Mathematical Society (2005)
3. Dubuisson, M., Jain, A.: A modified hausdorff distance for object matching, pp. 566–568 (1994)
4. ElGhawalby, H., Hancock, E.R.: Measuring graph similarity using spectral geometry. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2008. LNCS, vol. 5112, pp. 517–526. Springer, Heidelberg (2008)
5. Gilikey, P.B.: *Invariance Theory, The heat equation, and the Atiyah-Singer index Theorem*. Publish or Parish Inc., USA (1984)
6. Huttenlocher, D., Klanderman, G., Rucklidge, W.: Comparing images using the hausdorff distance. *IEEE. Trans. Pattern Anal. Mach. Intell.* 15, 850–863 (1993)
7. Luo, B., Wilson, R.C., Hancock, E.R.: Spectral embedding of graphs. *Pattern Recognition* 36, 2213–2230 (2003)
8. Rosenberg, S.: *The Laplacian on a Riemannian manifold*. Cambridge University Press, Cambridge (2002)
9. Saucan, E., Appleboim, E.: Curvature based clustering for DNA microarray data analysis. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3523, pp. 405–412. Springer, Heidelberg (2005)
10. Spivak, M.: *A Comprehensive Introduction to Differential Geometry*, 2nd edn., vol. 1-5. Publish or Parish, Houston (1979)
11. Stillwell, J.: *Mathematics and its History*. Springer, New York (1974)
12. Xiao, B., Hancock, E.R.: Heat kernel, riemannian manifolds and graph embedding. In: Fred, A., Caelli, T.M., Duin, R.P.W., Campilho, A.C., de Ridder, D. (eds.) SSPR&SPR 2004. LNCS, vol. 3138, pp. 198–206. Springer, Heidelberg (2004)
13. Young, G., Householder, A.S.: Discussion of a set of points in terms of their mutual distances. *Psychometrika* 3, 19–22 (1938)