

Molecular Graph Eigenvectors for Molecular Coordinates

System Demonstration

Patrick W. Fowler
Department of Chemistry
University of Exeter

Tomaz Pisanski
Dept of Theoretical Comp Sci,
IMFM, University of Ljubljana, Slovenia
email: tomaz.pisanski@uni-lj.si

John Shawe-Taylor
Dept of Comp Sci,
Royal Holloway, University of London, England
email: john@dcs.rhbnc.ac.uk

1 Introduction

In [2] the eigenvectors of the adjacency matrix of a molecular graphs have been used to compute molecular coordinates. In [4] the same idea was used to solve a mathematical problem. Here we explain why the largest eigenvectors often give satisfactory results and also explain why it fails for some molecular graphs.

We consider the problem of embedding a graph on n vertices in Euclidean space \mathcal{R}^k , for $k < n$. Typically k would be 3 or 2. By posing the problem as minimising the squared norm of the appropriately weighted distance between adjacent points subject to natural normalising conditions we arrive at a formulation of the problem for which the optimal solution can be simply computed in terms of the eigenvectors of the Laplacian matrix of the (weighted) graph. For the case where the weights are chosen to be unity the solution is independent of the uniform penalty given to non-adjacent vertices.

Let $A(G) = (a_{uv})$ be the adjacency matrix of a simple (positively weighted) n vertex graph G with no loops. Note that u, v are understood to be adjacent iff $a_{uv} > 0$. For non-adjacent vertices $a_{uv} = 0$. Let the diagonal matrix D be given by

$$D_{vv} = d(v) = \sum_{u:(u,v) \in E(G)} a_{uv},$$

the weighted degree of vertex v . The Laplacian matrix is defined to be $Q(G) = Q(A) = D - A$, where $A = A(G)$.

We pose the problem of embedding a graph G as finding a mapping

$$\tau : V(G) \rightarrow \mathcal{R}^k.$$

We will place constraints on this mapping in order to ensure that the representation is natural and hopefully pleasing. We will denote by τ_i the n -dimensional vector formed by taking the i -th coordinate of $\tau(u)$ for all $u \in V(G)$. Thus τ_i is an n -dimensional vector indexed by the vertices of the graph G . Our first requirement is that the centre of gravity of the representation is at the origin. This implies that the vectors τ_i have average entry 0, or $\tau_i \perp \mathbf{j}$, for $i = 1, \dots, k$. The next constraint is that the scaling in all dimensions be similar. This is ensured by requiring that

$$\|\tau_i\|^2 = \sum_{u=1}^n \tau(u)_i^2 = 1.$$

Finally we would like the embedding to retain maximum information about the graph. An example of how information can be lost is given when $\tau_i = \tau_j$ for some $i \neq j$, i.e. τ_i and τ_j are maximally correlated. In this case we have effectively reduced the dimension of the representation by one. Hence maximal information will be represented if the vectors have zero correlation, i.e. $\tau_i \perp \tau_j$, for $i \neq j$. We require adjacent vertices to be close together weighted according to a_{uv} (e.g. for different chemical bond types the value might vary), and require non-adjacent vertices to be far apart. Our definition of the graph drawing problem may therefore be stated as follows.

Problem 1.1 *Graph Drawing of a graph G given by (weighted) adjacency matrix A in \mathcal{R}^k .*

Find a mapping $\tau : V(G) \rightarrow \mathcal{R}^k$, which minimises the following energy function

$$E(\tau) = \sum_{(u,v) \in E(G)} a_{uv} \|\tau(u) - \tau(v)\|_2^2 - \beta \sum_{(u,v) \notin E(G)} \|\tau(u) - \tau(v)\|_2^2,$$

subject to the constraints

$$\begin{aligned} \|\tau_i\| &= 1, & \tau_i &\perp \mathbf{j}, & \text{for } i = 1, \dots, k \\ \tau_i &\perp \tau_j, & & & \text{for } 1 \leq i < j \leq k, \end{aligned}$$

where β is a positive constant controlling the strength of the force driving non-adjacent vertices apart. ■

We are now in a position to state our main result.

Theorem 1.1 *Let G be a connected weighted graph with adjacency matrix A . The graph drawing problem given in Problem 1.1 is solved by taking the weighted graph with adjacency matrix*

$$B_{uv} = \begin{cases} (a_{uv} + \beta) & \text{if } (u, v) \in E(G) \\ 0 & \text{otherwise} \end{cases}$$

and computing the eigenvectors e^1, e^2, \dots, e^n with corresponding eigenvalues $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$ and Laplacian matrix $Q(B)$. An optimal embedding τ is given by $\tau_i = e^{i+1}$, $i = 1, \dots, k$ and the minimal value of $E(\tau)$ is

$$\sum_{\ell=2}^{k+1} \lambda_{\ell} + \beta nk.$$

If $\lambda_{k+1} < \lambda_{k+2}$ then the optimal embedding is unique up to orthogonal transformations in \mathcal{R}^k .

Corollary 1.1 *In the case where the graph is not weighted (i.e. $a_{uv} \in \{0, 1\}$), the optimal embedding does not depend on the parameter β .*

A good example of the kind of image generated by our method is given in Figure 1 which is the embedding generated for the Buckminster fullerene using the 2nd, 3rd and 4th eigenvectors and taking a two dimensional projection.

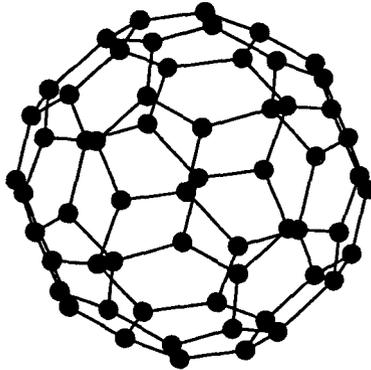


Figure 1: The Buckminster fullerene. The coordinates are determined by the 2nd, 3rd and 4th eigenvector.

In our requirements on the embedding we are forcing the graph to “look spherical”. For graphs with natural excentrical shape our method does not give natural pictures. This problem may also explain why occasionally a better image is created by taking the 2nd, 4th and 5th eigenvectors, [2], [3], rather than the three eigenvectors corresponding to the three smallest non-zero eigenvalues.

It is not clear how the results might be generalised if the norms used are altered, either in the energy function of the accompanying constraints on the vectors τ_i . It may well be that in this case the approach taken in this paper is not applicable and a more standard method of energy minimisation must be applied.

References

- [1] B. Becker and G. Hotz, On The Optimal Layout of Planar Graphs with Fixed Boundary, *SIAM J. Computing* **16**, (1987) 946-972.
- [2] D.E. Manolopoulos, P.W. Fowler, Molecular graphs, point groups, and fullerenes, *J. Chem. Phys.* **96** (1992) 7603-7614.
- [3] T. Pisanski, Recognizing Symmetric Graphs, presented at Graph Drawing '93. Sèvres, France, September 26-29, 1993.
- [4] T. Pisanski and J.S. Shawe-Taylor, Characterising Graph Drawing with Eigenvectors, Departmental Technical Report CSD-TR-93-20, Department of Computer Science, Royal Holloway, University of London, 1993.