communications in mathematical
match
and in computer chemistry

no. 44, October 2001

MATCDY (44) 305 - 317 (2001)

ISSN 0340 - 6253

GEOMETRY VERSUS TOPOLOGY: OVERLAPPING SPHERES MODEL CORRECTED FOR ANGLES*

Ante Graovac,[a] Alen Orbanić,[b] Tomaž Pisanski[1]

[a]Faculty of Natural Sciences, Mathematics and Education,
University of Split, Nikole Tesle 12, HR-21000 Split, Croatia
The R. Bošković Institute, HR-10002 Zagreb, POB 180, Croatia
graovac@rudjer.irb.hr
[b]University of Ljubljana, FMF, Jadranska 19, SI-1000 Ljubljana, Slovenia
Alen.Orbanic@fmf.uni-lj.si
[c]University of Ljubljana, IMFM, Jadranska 19, SI-1000 Ljubljana, Slovenia
Tomaz.Pisanski@fmf.uni-lj.si

Abstract. All geometrical information is missing in topological, i.e. graph theoretical description of connectivity in molecules. However, a part of molecular geometry could be recovered if some of existing graph drawing algorithms is invoked. For a given molecular graph G on n vertices such an algorithm computes the position of vertices $R(G) = (r_1, r_2, ..., r_n)$ in 3D-space. Here we use the recently introduced NiceGraph Program (NGP) [1] which is a part of the computer program Vega [2].

But on other hand a relationship between topology and geometry of molecules could be reversed, i.e. when molecular connectivity is reconstructed from the molecular geometry data. Recently, we have proposed the Overlapping Spheres (OS) Model as a way to obtain a plausible molecular graph $G' = G'(R)$ after a set of n points R in 3D-space is given.

If for a given graph G its respective 3D-drawing leads to a graph G' identical to G, i.e. if $G'(R) = G$, the graph drawing algorithm is said to be self-consistent; otherwise it is not self-consistent. The consistency of the NGP was tested recently [3] on the selected set of 40 molecular graphs and 30 out of these graphs have passed the consistency test. The discrepancy can be attributed to the fact that the NGP do not take into account the angles during the calculation process.

The aim of the present paper is to improve the OS model by angles considerations. Such a model we call Overlapping Spheres with Angles (OSAN) Model. Two algorithms were developed within the OSAN model and their self-consistency tested. The results obtained show that the OSAN model ensures a rather self-consistent and acceptable framework to reproduce the connectivity of molecules. Considering a test group of graphs used in tests, a success rate of the OS model algorithm is 83% meanwhile a success rate of the first OSAN algorithm reaches 95%.

***Dedicated to Professor Alexandru T. Balaban on the occassion of his 70th birthday.**

INTRODUCTION

Graphs are abstract mathematical objects suitable for describing topology of molecules, namely the connectivity of their constituent atoms. All geometrical information is missing in such a description. A step from topology to geometry is conveniently done by applying some of quantum mechanical models like molecular mechanics and others. The interplay between topology and geometry has been already studied in chemical literature [4,5]. However, a part of molecular geometry is recovered after some of existing graph drawing algorithms is invoked. For a given molecular graph G on n vertices such an algorithm computes the position of vertices $R(G) = (r_1, r_2, ..., r_n)$ in 3D-space. In the present paper we use the recently introduced NiceGraph Program (NGP) [1] which is part of the computer program Vega [2]. In principle the same approach can be used in connection with any other graph drawing algorithm, see [6,7].

But on the other hand a relationship between topology and geometry of molecules could be reversed. A step from geometry to topology is involved in e.g. crystallography where molecular connectivity has to be reconstructed from the molecular geometry data. Recently, we have proposed the overlapping spheres (OS) model as a way to compute a plausible molecular graph $G' = G'(R)$ after a set on n points R in 3D-space is given.

If for a given graph G its respective 3D-drawing R leads to a graph G' identical to G, i.e. if $G'(R) = G$, the graph drawing algorithm is said to be self-consistent; otherwise it is not self-consistent.

The consistency of the NGP was tested recently [3] on the set of 40 molecular graphs describing unbranched and branched chains, cata- and pericondensed benzenoids, non-benzenoids and fullerenes. Out of these 40 examples, 30 have passed the consistency test.

The discrepancy can be attributed to the fact that the NGP ignores all angles during the calculation process. These angles, as it is known in chemistry, have a rather limited range of values, and in e.g. carbon hydrocarbons the angles between incident edges equal to $180^\circ \pm 1^\circ$, $120^\circ \pm 3^\circ$, and $109.5^\circ \pm 5^\circ$ for sp-, sp^2- and sp^3-hybridized carbons, respectively.

In the original algorithm (OS model) we were given $m=|E(G)|$, expected number of edges. Starting with empty graph on n vertices, we gradually included all possible edges into the graph in the ascending order of their lengths, thereby obtaining a series of graphs $E_n \subseteq G_0 \subseteq G_1 \subseteq ... \subseteq G_k = K_n$, where E_n is an empty graph and K_n a complete graph. If the first graph with m edges in series $G_1,...,G_k$ was equal G, we declared success and otherwise a failure.

METHOD

After the 3D positions r_u of all vertices $u \in V(G)$ of graph G are determined (here by the NGP), we define the distance between a pair u and v of vertices as $d(u,v) = \|r_u - r_v\|$, and then we compute all distances in G. Let $n = |V(G)|$ and let us define a set:

$$S = \{(i, j, d(i, j)) \mid i, j = 1,2,...,n;\ i < j\}.$$

By sorting the triplets in S by the third component in increasing order we obtain (ordered) sequence of triplets $s_1, s_2, ... \in S$. Let us start with $G' = E_n$, a graph containing only n isolated vertices and therefore its edge set $E' = E(G')$ is empty. We go on then over elements $s \in S$ according to the increasing values of their third components ($d(i,j)$), and if for a given $s = (u, v, d(u, v))$ the edge (u,v) is not incident to previous edges it is added to G', and if it is incident with a number of previous edges then all the respective angles of these edges with edge (u,v) are calculated. The angle between two incident edges (i, j) and (i, k) is defined by

$$\theta = \arccos \frac{(r_j - r_i)(r_k - r_i)}{|(r_k - r_i)||(r_k - r_i)|}.$$

Here the improvement with regard to the OS model enters, namely we introduce some minimal allowed (threshold) angle $\theta_{\min}$. We add the edge (u, v) to G' only if its angles with all previous (incident) edges are $\geq \theta_{\min}$. If we go on with traversing until $m = |E(G)|$ edges are obtained in the final graph, we speak of *algorithm 1*. However, it could happen that not all m edges are obtained, or if the obtained edges do not neccessarily reconstruct the original edges of G. If we do not restrict to m edges but we proceed with traversing with respect to a given $\theta_{\min}$, we speak of *algorithm 2*.

In numerical experiments to be described later, the values of $\theta_{\min}$ are varied from 0° to 120°, and their influence on the ability of the final graphs to reproduce the original topology is discussed.

Formal description of OSAN algorithm

Let *A(vertices V, integer m, angle* $\theta_{\min}$*)* be our algorithm where:

V are vertices embedded in Euclidean 3D space;

m exepcted number of edges;

$\theta_{\min}$ a threshold angle.

```
function angle(e,f) calculates an angle between incident edges e and f.

G' = E_n  {an empty graph}
E' = {}
V' = V(G')
Vertices V(G') are enumerated with as 1, 2, ..., n = |V(G')|.
Positions of vertices in 3D space are given with vectors r_i, i = 1,2,...,n.
S = {(i, j, d(i, j)) | i, j = 1,2,...,n}, where d(i, j) = ||r_i - r_j||
(s_i)_{i=1}^{|S|} = elements of S sorted by the third component
for i:= 1 to n(n-1)/2
      s_i = (u_i, v_i, d(u_i, v_i))
      forall edges f in E'
            if (f incident to u_i v_i )
                  if (θ_min > angle(f, u_i v_i))
                        proceed with next i from the
                        beginning of for i:= ... loop
                  endif
            endif
      endforall
      E' = E' ∪ {u_i v_i}
      if(|E'| ≥ m) return G' = (V', E')  {at algorithm 1 only}
endfor
return G' = (V', E')
```

Function `angle` (e, f) calculates the angle that is spanned between two incident edges e and f.

To test both of algorithms we perform a following test. For given graph G we produce its drawing with the NGP algorithm. Then we remove all edges from the drawing and apply our algorithms for a fixed, but otherwise arbitrary threshold angle $\theta_{\min}$ from 0° to 120°.

After running one of these algorithms we obtained the final graph G'. If $G' = G$ we declared a success and otherwise a failure.

In order to determine the optimal values of parameter $\theta_{\min}$ we ran a series of tests on the same data set as in [3]. Given a graph G we were interested in those ranges of parameters $\theta_{\min}$ where algorithm 1 and algorithm 2 produce a success. Angles in those intervals are called a *successful angles*. The aim of the present paper is to find minimal and maximal $\theta_{\min}$ successful angles for the test data set. Our experiments will show that this is not always possible but the OSAN hits more successes than OS model. Having in mind only a plausible character of the NGP results, it is not surprising that the range of acceptable lower and upper limits of threshold angles is far broader and in general different from ranges of allowed angles found in chemistry.

GRAPH SPREAD AND GEOMETRICAL SPREAD

Given a connected graph G the standard graph theoretical distance metrics is defined as: $d(u,v)$ = minimal number of edges on any path between vertices u and v. Maximal distance from vertex v in graph G is defined as: $md(u) = \max_{v \in V(G)} d(u,v)$. A *graph spread* is defined as:

$$\mathrm{GrS}(G) = \frac{\max_{u \in (G)} md(u)}{\min_{u \in (G)} md(u)}.$$

Graph spread is a number between 1 and 2. If for some graph G graph spread is near 1 then the graph is uniformly spread in all directions according to the graph theoretical metrics d. If graph spread is near 2, graph is spread in some directions more than in others.

Given a 3D drawing $R(G) = (r_1, r_2, \ldots, r_n)$ of graph G, a *geometrical spread* is defined as:

$$\mathrm{GeS}(G) = \frac{\max_{1 \le i,j \le n} d(r_i, r_j)}{\min_{1 \le i,j \le n} d(r_i, r_j)},$$

where d stands for standard Euclidean metrics.

To predict a success of both OSAN algorithms, a correlation success of algorithms vs. graph spread (geometrical spread) is studied.

RESULTS

Let us show how restrictions on angles could account for a proper topology, i.e. connectivity of vertices. After the coordinates of vertices are obtained by performing the NGP algorithm on a graph G, OSAN algorithms are performed on the set of the vertices. We have tested the OSAN algorithms on the set of over 40 chemical graphs. As our data set contains the graphs of planar molecules all our computations have been performed in 2D rather than in 3D Euclidean space. However, this restriction is easily lifted.

The Figure 1 shows a number of successes at different minimal threshold angles $\theta_{\min}$ for algorithm 1. When the threshold angle equals 0, algorithm 1 performs like the OS algorithm. Increasing the threshold angle $\theta_{\min}$ a number of sucesses increases up to $\theta_{\min} = 67^{\circ}$. At the greater angles the number of successes decreases. So the value $\theta_{\min} = 67^{\circ}$ is the optimal value for the set of graphs considered. Let us remind that the algorithm reaches success for all

graphs in the test group for some angle θ_{min}, but the angle ranges are quite different. The Figure 2 shows the angle intervals sorted by the minimal threshold angle of a success for the algorithm 1.

The tests on the same group of graphs were also performed with the algorithm 2. As expected, the algorithm 2 is less successful. The Figure 3 shows the number of successes of the algorithm 2 at the different threshold angles θ_{min}. The Figure 4 shows the intervals of the angle θ_{min} for all graphs when the algorithm 2 recovers the original graph. As we see algorithm 2 is successful for some threshold angle θ_{min} in all but 4 graphs. The drawings of these four graphs are depicted in Figure 5. A success of the algorithm 2 strongly depends of the chosen threshold angle θ_{min}. As we can see from the Figure 3, good angles are somewhere between 70° and 90°, and optimal angle is 72°.

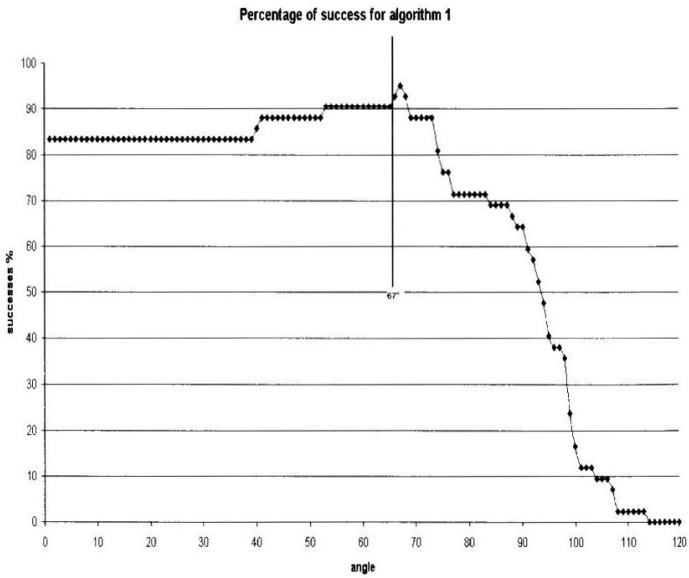


FIGURE 1.Percentage of successes at the different threshold angles for algorithm 1

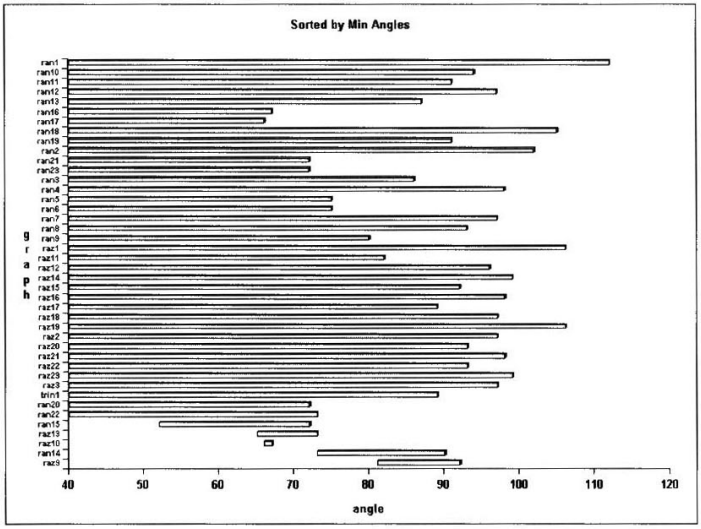


FIGURE 2. Ranges of successful threshold angles for algorithm 1

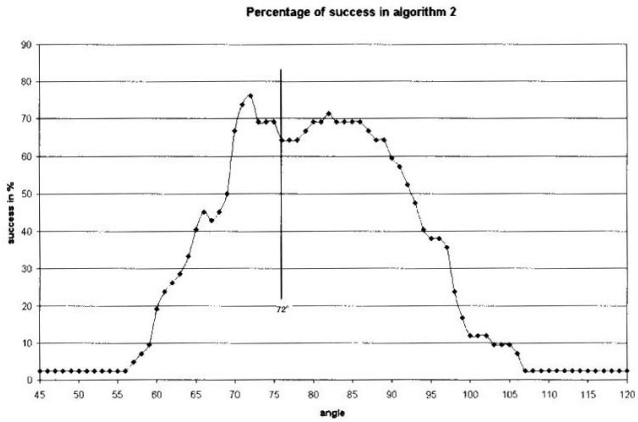


FIGURE 3. Percentage of successes at the different threshold angles for algorithm 2

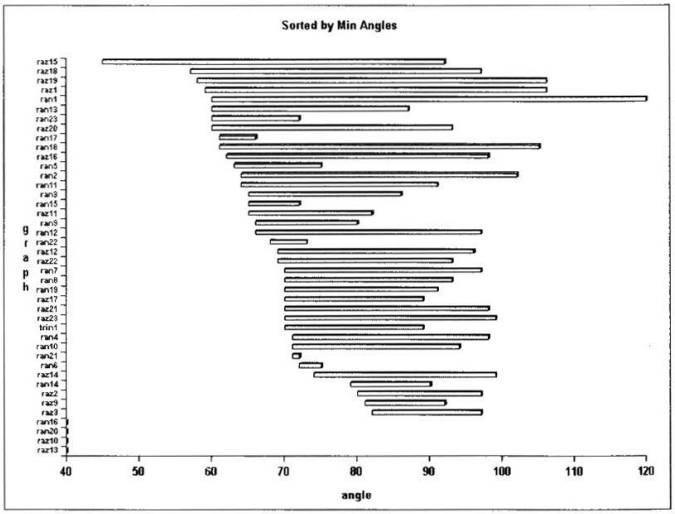


FIGURE 4. Ranges of successful threshold angles for algorithm 1

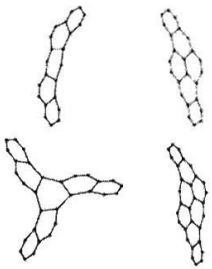


FIGURE 5. Graphs with no success at algorithm 2

Considering the NGP algorithm, it is noticed, that the greater the graph (geometrical) spread the more are the angles incorrect, as we see from e.g. Figure 5. So it is interesting to see if any correlation between the spreads and success of algorithms 1 and 2 (and through this

a correlation with consistency of the NGP algorithm). For all graphs, all spreads are calculated and charts of the interval width of the successful angles vs. the spreads for both algorithms are plotted (see Figure 6). On all charts in Figure 6 a linear regression line is drawn. Increasing any spread, the widths of the threshold angle intervals in general shorten, as expected. So both spreads are negatively correlated with the success of OSAN algorithms.

To predict an optimal threshold angle for OSAN algorithms we study a correlation between the geometrical spread and the middle angle of the interval of the successful threshold angles (see Figure 7). Here the corelation is stronger for both algorithms. With linear regression we get the following empirical formulas:

algorithm 1: $\theta_{min} = -10.6\ GeS(G) + 67.7$
algorithm 2: $\theta_{min} = -17{,}7\ GeS(G) + 106{,}1$

At the linear regression for algorithm 2, 4 graphs with no successful angle are omitted. The formulas could be used to predict optimal threshold angle parameters used in OSAN algorithms.

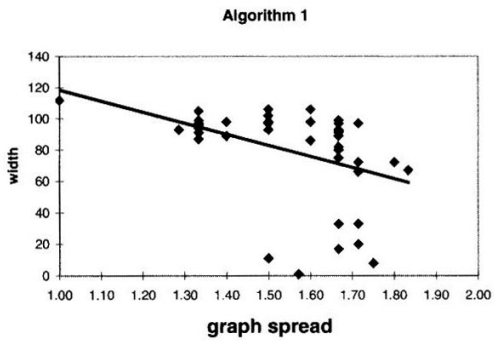
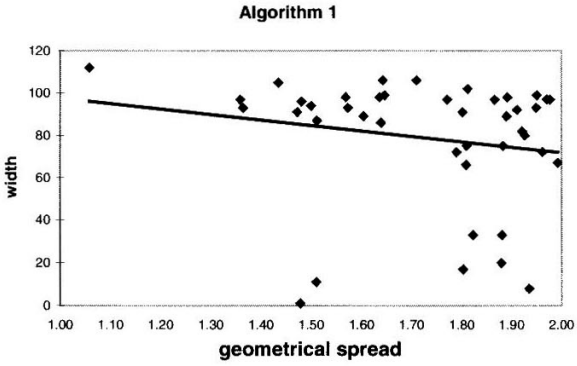


FIGURE 6. Successful threshold angle interval width vs. spreads



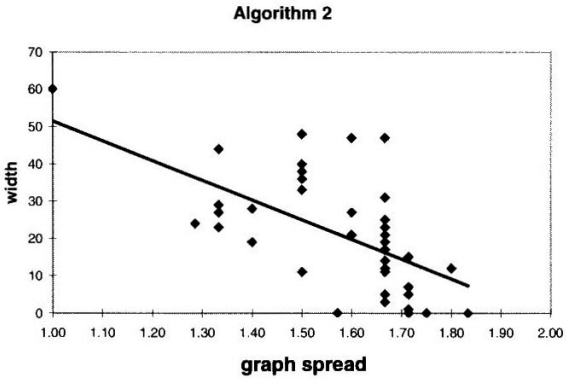


FIGURE 6. (continued)

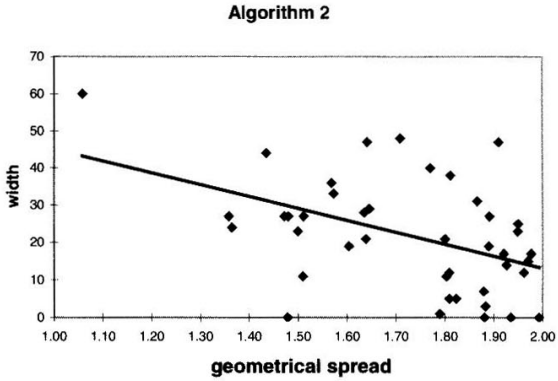


FIGURE 6. (continued)

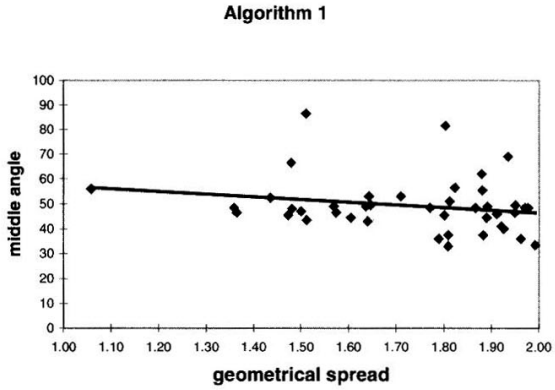


FIGURE 7. Middle angles of successful angle intervals vs. geometrical spreads

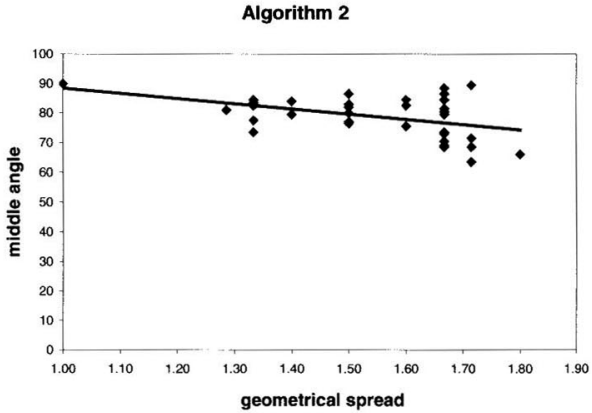


FIGURE 7. (continued)

CONCLUSION

From the data of 40 chemical graphs it is infered that the threshold angle has optimal value $\theta_{opt} = 70^\circ \pm 3^\circ$. At this value the algorithm 1 recognizes 90% of graphs while the algorithm 2 recognizes 70% of sample. The algorithm 1 has superior recognition rate compared with the OS model. The algorithm 2 has lower recognition rate than the OS model, but it does not need additional parameter about expected number of edges.

The OSAN model can be used at setting parameters in the NGP algorithm so the greater consistency could be reached. A representative test group of graphs could be chosen and successes at different threshold angles could be measured at the different NGP parameters. A similar tests like these made on planar molecular graphs could be done for molecular graphs in 3D space. (e.g. fullerenes).

The OSAN model includes the OS model and it is more powerful and successful. As drawings produced by the NGP algorithm more or less ignore the information about chemical

angles it is belived that success would be much more greater if the drawings were more like in actual molecular geometry.

The OS model and the OSAN model algorithms could be used for testing a consistency of other molecular graph drawing algorthms, see [6,7].

Acknowledgements

Supported in part by Ministrstvo za znanost in tehnologije Republike Slovenije, proj no. J1-8901 and J1-8549.

REFERENCES

[1] T. Pisanski, Bor Plestenjak, and Ante Graovac: NiceGraph Program and Its Applications in Chemistry, *Croat. Chem. Acta* **1995**, *68*, 283-292.

[2] Vega Version 0.5, Quick Reference Manual and Vega Graph Gallery (T. Pisanski, ed.), IMFM, Ljubljana, http://vega.ijp.si

[3] T. Pisanski, Marko Razinger and Ante Graovac: Geometry versus Topology: Testing Self-Consistency of the NiceGraph Program, *Croat. Chem. Acta* **1996**, 69, 827-836.

[4] A.T. Balaban, Ed., From Chemical Topology to Three-Dimensional Geometry, *Plenum Press*, New York, 1997; IX + 420 pp.

[5] Erica Flapan, When Topology Meets Chemistry-An Introduction to Molecular and Topological Chirality, *Cambridge University Press & Mathematical Association of America*, 2000; paperback, 0-521-66254-0, 24,95 USD.

[6] Matjaž Kaufman, Tomaž Pisanski, Dragan Lukman, Branko Borštnik and Ante Graovac, Graph-drawing algorithms geometries versus molecular mechanics in fullerenes, *Chemical Physics Letters* **1996**, *259*, 420-424.

[7] T. Pisanski, John Shawe-Taylor, Characterizing Graph Drawings with Eigenvectors, *Journal of Chemical Information and Computer Sciences*, **2000**, *40*, 567-571