A vectorized computer code is developed for the enumeration of walks through the matrix power method for directed graphs. Application of this code to several graphs is considered. It is shown that the coefficients in the generating functions for signed graphs are much smaller in magnitude. It is shown that self-avoiding walks on some graphs can be enumerated as a linear combination of walk GFs of directed paths and rooted-directed paths.

INTRODUCTION

The enumeration of walks on graphs and lattices of chemical interest has numerous applications in many areas of chemistry. Random walk models are useful in the treatment of diffusion, conformations of flexible polymers, among other applications. Walks on graphs are also useful in chemical coding, characterization of graphs, applications of extended Hückel methods to solids through moment generating functions, etc. Randić and co-workers, as well as the present author, have formulated computational techniques for enumerating walks and self-returning walks for ordinary graphs. However, walks on directed graphs, weighted graphs, signed graphs, and graphs containing complex edge weights have not been explored at all. We develop here a vectorized computer code for the enumeration of walks and spectral moments of directed and unsymmetric graphs, in general. The second section describes the computational techniques and codes while the third section comprises results and discussions.

COMPUTATION OF WALKS

A walk on a graph is defined as a sequence of edges that a walker can traverse continuously starting and ending on any vertex. The possibility of repetition of edges is permitted in a random walk. A self-returning walk is a random walk in which one starts and ends the walk in the same vertex. A self-avoiding walk is a walk in which no vertex appears more than once. A walk of length $k$ is a walk consisting of $k$ edges of a graph or lattice in a continuous manner. This is also known as the distance of a walk. Similarly one can define self-avoiding and self-returning walks of length $k$. Suppose $N_k$ is the number of walks of length $k$, the function $W$ shown below, is known as the generating function for the walk

$$W = \sum_{k=0}^{\infty} N_k x^k,$$

where $x$ is a dummy variable and $a^k$ represents walks of length $k$. Similarly one defines the generating function for self-avoiding walk (SWA) and self-returning walk (SRW) as

$$SAW = \sum_{k=0}^{\infty} a^k x^k,$$

$$SRW = \sum_{k=0}^{\infty} M^k x^k,$$

where $N_k$, $a^k$, and $M^k$ are the number of walks of length $k$ from vertex $i$ to vertex $j$.

Note that $N_k$, $a^k$, and $M^k$ are the number of vertices in the graph and that SAW does not have a nonvanishing term beyond $a^{-1}$, since one cannot have a self-avoiding walk of length more than $n-1$ on a graph containing $n$ vertices. Similarly $M^k$ is zero if the graph in question does not have loops. Hence, we omit the constant term in $W$, $SAW$, and $SRW$ in this manuscript.

The adjacency matrix $A$ of an ordinary graph is defined as

$$A = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases}$$

and $i$ and $j$ are connected by an edge.

In general different powers of the adjacency matrix enumerate walks of different lengths. The $i$th element of $A^k$ enumerates the total number of walks of length $k$ between the vertices $i$ and $j$. Consequently the trace of $A^k$ or $trA^k$ is the total number of self-returning walks also known as spectral moments. The sum of all the elements of $A^k$ is $N_k$. In symbols,

$$N_k = \sum_{i=0}^{k} (A^k)_{ii},$$

$$M_k = \sum_{i} (A^k)_{ii} = trA^k,$$

For symmetrical graphs it can be shown that

$$W = P_G(1/x),$$

where $P_G$ is the characteristic polynomial of $A - J$, where $J$ is a matrix in which all entries are units.

The author used his code to compute characteristic polynomials to construct $W$ of graphs before. The coefficients $M_k$ in SRW (self-returning walk generating function) are simply $S_k$, the spectral moments of $A$. $S_k$ is defined as follows

$$S_k = trA^k + \sum_{i} (A^k)_{ii}.$$

The spectral moments $S_k$'s and the coefficients in the characteristic polynomials are related by the following expressions (see, for example, reference (6))

$$S_1 = 0, S_2 = 1, S_3 = 1 + \sum_{i} (A^2)_{ii},$$

Thus the spectral moments $S_k$'s (same as $M_k$) can be recursively obtained using the above relation once the coefficients $C_n$'s in the characteristic polynomial of the graph are determined. All analytical expressions discussed above are valid only for ordinary graphs.

A signed graph is defined as a directed graph for which a signed adjacency matrix $A$ is shown below:

$$A^{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases},$$

and $i$ and $j$ are connected by an edge from $i$ to $j$.

The possibility of nonzero diagonal elements could also be included in the above definition through loops.

The powers of adjacency matrices of weighted, directed, and bidirected graphs (different weights for different directions) do not enumerate the "number of walks" in a strict sense. This is because the $ij$ matrix element $A^{ij}$ is given by

$$A^{ij} = \sum_{k} w_{ij} a_{ki} a_{kj} \cdots a_{j_l},$$

where the sum is over all such terms starting with the vertex $i$ and terminating at vertex $j$. Note that if all matrix elements are 0 and 1 for ordinary graphs the above sum gives the number of walks from $i$ to $j$ of the function $f_k$. For a signed graph since some of the matrix product shown above will be negative while the others will be positive and, consequently, $A^{ij}$ gives the net effect of all possible walks from $i$ to $j$ of length $k$. There are some advantages to the evaluation of such generating functions since the coefficients in general tend to be smaller in magnitude. For signed graphs the coefficients of $x$ vanish if $k$ is odd since the sum of contributions of all walks cancel out. Note that the walks of signed and directed graphs are dependent on labeling of vertices, in general.

There are also other advantages in using weighted graphs. For some trees and cyclic graphs a combination of properly weighted directed graphs could be used to enumerate the number of self-avoiding walks as we show here.

In the next section, we shall use the term walk-generating function (WGF) and self-returning walk-generating function (SRWGF) to simply refer to

$$WGF = \sum_{i} a^i x^i,$$

$$SRWGF = \sum_{i} M^i x^i,$$

Thus the coefficients in $WGF$ and $SRWGF$ do not necessarily enumerate the number of walks in the sense they are interpreted for nondirected ordinary graphs. These coefficients measure the net effect of such walks for signed and weighted graphs.

The computation of powers of matrices is efficiently accomplished through the use of vector processors. We developed a vectorized Fortran '77 code and was compiled on an IBM 3000/300 with the vector option and an optimization level = 3. The DO loops were properly organized to achieve maximum vectorization. All matrices were stored as two-dimensional arrays. The original adjacency matrix, the $A^{ij}$ matrix and the $A^k$ matrix were saved in each processor.

RESULTS AND DISCUSSION

Table I compares the results obtained for a series of ordinary and signed graphs containing eight vertices shown in Figure 1. Note that the spectral moments (the number of self-returning walks) differ only in
Table I. WGF and SRWGF for both regular and signed graphs containing eight vertices.a

<table>
<thead>
<tr>
<th>Graph (Fig. 1)</th>
<th>Ordinary graph</th>
<th>Signed graph</th>
<th>Ordinary graph</th>
<th>SRWGF</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$14x^2 + 26x + 48x^2$</td>
<td>$-2x^2 + 23x + 4x^2$</td>
<td>$14x^2 + 38x + 16x^2$</td>
<td>$-14x^2 + 38x^2$</td>
</tr>
<tr>
<td></td>
<td>$+ 90x + 198x + 316x^2$</td>
<td>$+ 10x^2$</td>
<td>$+ 90x^2 + 198x^2$</td>
<td>$+ 10x^2$</td>
</tr>
<tr>
<td>II</td>
<td>$14x^2 + 26x + 52x^2$</td>
<td>$-4x^2 + 8x + 24x^2$</td>
<td>$14x^2 + 42x + 140x^2$</td>
<td>$-14x^2 + 42x^2$</td>
</tr>
<tr>
<td></td>
<td>$+ 140x^2 + 160x + 302x^2 + 74x^2 + 1484x^2$</td>
<td>$+ 84x^2$</td>
<td>$+ 140x^2 + 160x + 302x^2 + 74x^2 + 1484x^2$</td>
<td>$+ 84x^2$</td>
</tr>
<tr>
<td>III</td>
<td>$14x^2 + 30x + 60x^2$</td>
<td>$-6x^2 + 10x + 26x^2$</td>
<td>$14x^2 + 46x + 176x^2$</td>
<td>$-14x^2 + 46x^2$</td>
</tr>
<tr>
<td></td>
<td>$+ 12x^2 + 25x^2 + 580x^2 + 1110x^2 + 2308x^2$</td>
<td>$+ 86x^2$</td>
<td>$+ 12x^2 + 25x^2 + 580x^2 + 1110x^2 + 2308x^2$</td>
<td>$+ 86x^2$</td>
</tr>
<tr>
<td>IV</td>
<td>$14x^2 + 30x + 62x^2$</td>
<td>$-6x^2 + 10x + 30x^2$</td>
<td>$14x^2 + 46x + 182x^2$</td>
<td>$-14x^2 + 46x^2$</td>
</tr>
<tr>
<td></td>
<td>$+ 130x^2 + 272x^2 + 570x^2 + 1194x^2 + 2505x^2$</td>
<td>$+ 114x^2$</td>
<td>$+ 130x^2 + 272x^2 + 570x^2 + 1194x^2 + 2505x^2$</td>
<td>$+ 114x^2$</td>
</tr>
<tr>
<td>V</td>
<td>$14x^2 + 32x + 60x^2$</td>
<td>$-8x^2 + 26x + 102x^2$</td>
<td>$14x^2 + 50x + 200x^2$</td>
<td>$-14x^2 + 50x^2$</td>
</tr>
<tr>
<td></td>
<td>$+ 130x^2 + 264x^2 + 660x^2 + 1170x^2 + 2982x^2$</td>
<td>$+ 42x^2$</td>
<td>$+ 130x^2 + 264x^2 + 660x^2 + 1170x^2 + 2982x^2$</td>
<td>$+ 42x^2$</td>
</tr>
<tr>
<td>VI</td>
<td>$14x^2 + 34x + 70x^2$</td>
<td>$-10x^2 + 30x + 70x^2$</td>
<td>$14x^2 + 54x + 242x^2$</td>
<td>$-14x^2 + 54x^2$</td>
</tr>
<tr>
<td></td>
<td>$+ 162x^2 + 362x^2 + 784x^2 + 1660x^2 + 3068x^2$</td>
<td>$+ 112x^2$</td>
<td>$+ 162x^2 + 362x^2 + 784x^2 + 1660x^2 + 3068x^2$</td>
<td>$+ 112x^2$</td>
</tr>
<tr>
<td>VII</td>
<td>$16x + 32x + 62x^2$</td>
<td>$-8x^2 + 16x + 48x^2$</td>
<td>$16x^2 + 48x^2 + 160x^2 + 162x^2 + 544x^2$</td>
<td>$-16x^2 + 48x^2$</td>
</tr>
<tr>
<td></td>
<td>$+ 128x + 256x^2 + 512x^2 + 1024x^2 + 2624x^2$</td>
<td>$+ 160x^2$</td>
<td>$+ 128x + 256x^2 + 512x^2 + 1024x^2 + 2624x^2$</td>
<td>$+ 160x^2$</td>
</tr>
</tbody>
</table>

aFor all signed graphs vertices were labeled sequentially for the largest paths and then branching vertices in the order of increasing vertex numbers they are attached to.

Figure 1. Graphs I-VII containing eight vertices (see Table 1 for walk GPS).

Signs for ordinary and signed graphs Table I. However, the actual walk generating functions differ significantly for ordinary and signed graphs. Note that: a walk from a vertex i to j (i > j) is considered positive while a walk from i to j (i < j) is negative because of weights + 1 and -1, for these two edges. Consequently, the coefficients of all terms with odd powers are zero since there are equal and opposite contributions for every term since $A_j = -A_i$ if i is odd.

Consider a directed chain of length 8 ($L_8$). Label the vertices 1–8 sequentially from left to right. i.e. the edges will be directed from i to j if j > i and only if j > i. This means, for example, the matrix element $a_{ij}$ is 1 but $a_{ji}$ is 0. The remaining adjacency matrix is unsymmetric as expected. The powers of this matrix $A$ enumerate exactly half of the self-avoiding walks (SAW) and in general this is the case for any $L_n$ graph. For the directed $L_n$ graph thus obtained $2W = SAW = 14x^2 + 12x^2 + 10x^2 + 8x^2 + 6x^2 + 4x^2 + 2x^2 + 2x$.

The enumeration of self-avoiding walks can in general be a difficult problem. Randić et al. have formulated an algorithm based on growing paths from nonempty vertices of graphs. This algorithm works reasonably well for graphs which do not contain several rings. For lattice graphs and complex poly cyclic graphs this algorithm can experience a combinatorial explosion. There is in general no polynomial algorithm to compute self avoiding walks. It is hoped that the present study based on matrix powers of directed graphs would lead to finding SAWs as combination of generating functions of directed graphs.
II.

of rooted paths.

WGF, SRWGFs of

Figure 3. A square lattice containing 36 vertices. For the
as a linear combination of the SAW of a path and SAWS

Next we consider a few signed square and honey-
comb lattices. It may be recalled that signed graphs
are bidirected graphs with a weight + 1 for an edge
to it if |< | ≤ 1 if `< |. Note that because of the
introduction of negative weights the coefficients in
WGF enumerate the "net effect" of walks of var-
ious lengths as opposed to the total number of walks
of length k. Hence, the coefficients of odd powers
vanish in WGF.

Table II compares the WGF and spectral moments
of a signed square lattice (Fig. 3) and the corre-
sponding ordinary square lattice. The ordinary
square lattice is a bidirectional graph with the same weight
+1.0 for both directions. As seen from Table II, the
coefficients of odd powers of the signed lattice
vanish while the coefficients of even powers of the
signed lattice are significantly smaller as these
measure the net effect of taking walks of various
lengths while for ordinary lattices these are the nu-
eral of walks of specified length. The odd coefficients
vanish for signed graphs since the matrix product
terms cancel out. Note that the WGF for signed lattices
is label-dependent. Interested readers can obtain
the labels used for Table II from the author.

The number of "non-self-returning" walks can be
computed using WGF-SRWGF. The coefficients of
various terms in WGF-SRWGF truly enumerate non-
self-returning walks only for regular lattices. Thus the
coefficient of x^n in WGF-SRWGF of the ordinary
lattice in Figure 3 (Table II) is 56 while it is 56 for a signed
lattice.

Table III shows the WGF and SRWGFs of a ho-
neycomb lattice containing 54 vertices (Fig. 4). Note
that since the coefficients of WGF of this lattice grow
astronomically a quadruple precision arithmetic was
invoked. Hence, the execution was slowed down since
vectorization is not possible with the quadruple pre-
cision arithmetic. Yet the CPU time taken was rea-
sonably insignificant for the problem at hand. Again
the general features of the coefficients for the hon-
eycomb lattice follows the square lattice discussed above. The labeling for this lattice can be
obtained from the author.

The code developed above could be potentially
useful for enumerating self-avoiding walks in a faster
manner compared to any known procedure up to
now. The procedure outlined by Randić et al.3 for
finding the number of self-avoiding walks is a non-
polynomial algorithm and thus the CPU time re-
quired grows astronomically for larger complex
graphs containing cycles. However, if matrix power
procedure can be suitably adapted for a combination
of weighted graphs such that the resulting generat-
ing function is the self-avoiding walk polynomial then
it could be very powerful. For example, consider the
tree in Figure 5. If one enumerates the largest path
starting from the vertex labeled 6 one gets L_6. The
SAW of L_6 is given by the matrix powers of the appro-
priate directed graph discussed before

SAW(L_6) = 8x + 6x^2 + 4x^3 + 2x^4.

Now we start with the vertex labeled 6 which was
not visited before and enumerate the largest possible
paths. It can be seen that they are 6-3-4-5 and 6-3-2-1
for the graph in Figure 5. A restricted SAW for

these two graphs can be obtained such that all walks

Figure 4. A honeycomb lattice containing 54 vertices. See Table III for WGF, SRWGFs of regular and signed
lattices.

Figure 5. A brachied tree containing six vertices. The self-avoiding walk GF (SAW) of this tree can be
obtained as a linear combination of the SAW of a path and SAWs
of rooted paths.
al side-chain variables in all conformations, while conformations of the side-chain. We have optimized maximum for the hydroxyl groups for each of the nine

zeta plus polarization basis set and inclusion of core geometries and energies require at least a double-

calculation (~'1¡,~'12) to indicate the conformation with ('1¡ = 187.8° and '12 = 295.3°). We shall use the no-

'1¡ = 187.8° + ~'1¡ and '12 = 295.3° + ~'12. For example,

about '1¡ and '12 with respect to the crystal structure of a large

molecule such as ascorbic acid. The object of this work is to study the

importance of further optimizing the structure of a large molecule, mainly because of its size. The most comprehensive work so far is an STO-3G minimal basis set study without accountig for both the electrostatic interactions and rela-
Plesset level of theory. A complete optimization of all ascorbic acid conformers at the MP2/6-31G* level

Plesset laboratory. SCF calculations using the STO-3G minimal

data is given by

adjacency matrix of this graph. Suppose we call this rooted-self-avoiding walk (RSAW) then

RSAW(L6) = x + 2x² + 2x³.

Since there are two paths (6-3-4-5 and 6-2-1) starting from the vertex 6, SAW of the graph in Figure 5 is given by

SAW (Fig. 5) = SAW (L6) + 2RSAW (L6)

= 10x + 10x² + 8x³ + 2x⁴.

The above result can be easily verified. Although the above procedure for complex cyclic graphs can get complicated it may be a better and efficient alternative to compute the SAW of a given graph as a linear combination of SAWs of L6s and rooted paths. Such advancement and extension of these procedures to complex cyclic graphs could be the topic of future investigations.

CONCLUSION

In this investigation we developed a vectorized computer code in FORTRAN 77 to enumerate the walk generating functions (WGF) and self-returning walk generating functions (SRWGF) for directed graphs, signed graphs and weighted codes. The code was

applied to several graphs and lattices. It was also shown that the self-avoiding walks on some graphs can be enumerated as a linear combination of WGFs of directed maximal paths and rooted-directed maximal paths.

References


15. T. Atwater Laboratory, Middletown, Connecticut 06457

Ab Initio Study of Ascorbic Acid Conformations

Mohammad A. Al-Laham and G.A. Petersen* Department of Chemistry, Wesleyan University, Hall-Atwater Laboratory, Middletown, Connecticut 06457

Paul Haske Department of Molecular Biology and Biochemistry and Department of Chemistry, Wesleyan University, Hall-Atwater Laboratory, Middletown, Connecticut 06457

Received 13 March 1990; accepted 6 July 1990

The STO-3G optimized structures of nine different staggered conformers of ascorbic acid are presented. The largest energy difference between the nine local minima is 5.1 kcal/mol. Comparison of the relative energies of the fully optimized structures of ascorbic acid conformers with those of nonoptimized conformers shows that full optimization is essential to obtain meaningful results. However, optimization of the ring structure is almost independent of optimization of the side-chain structure. One of the STO-3G optimized gas phase conformers is very close to the X-ray structure of the crystal.

INTRODUCTION

Ascorbic acid or Vitamin C (Fig. 1) has been a molecule of great biochemical interest since it was isolated and described in 1928. The crystal structure was determined in 1968 by Hvoslef using X-ray crystallography and neutron diffraction. The theoretical work has been done on ascorbic acid, mainly because of its size. The most comprehensive work so far is an STO-3G minimal basis set study without geometry optimization by Carlson, Cable and Pedersen. Other work includes semiempirical calculations on free radicals derived from ascorbic acid and α-hydroxytertronic acid to help determine the structure and UV spectrum of ascorbic acid, and to study the π-electrons in the ring. Ab initio SCF calculations on α-hydroxytertronic acid have been used as a model for the electronic structure of ascorbic acid. The object of this work is to study the importance of fully optimizing the structure of a large molecule such as ascorbic acid, by comparing the fully optimized results from this work with the previous nonoptimized results of Carlson, Cable and Pedersen. In this article, we consider the separate optimization of the α-hydroxytertronic acid (ring) and the 1,2-diol sidechain, and we compare the results with those obtained when we fully optimize the molecule as one unit, in order to study the separability of calculations on large systems.

Accurate ab initio calculations of conformational geometries and energies require at least a double-

zeta plus polarization basis set and inclusion of cor-

relation effects on at least the second-order Möller-Plesset level of theory. A complete optimization of all ascorbic acid conformers at the MP2/6-31G* level of theory is not practical at this time. On the other hand, SCF calculations using the STO-3G minimal basis set give qualitatively correct results through accounting for both the electrostatic interactions and the Pauli repulsive interactions between electrons (the RMS error in the STO-3G rotational barriers for ethane, methanol, and acetaldehyde is 0.52 kcal/mol). This level of calculation also allows us to compare our optimized results with the nonoptimized results of Carlson, Cable, and Pedersen. The local minima on the ascorbic acid potential energy surface are related by rotations about the two dihedral angles, τ1 (C3-C4-C5-C6 in Figure 1) and τ2 (C4-C5-C6-O6 in Figure 1). Carlson, Cable, and Pedersen investigated the nine staggered conformations obtained by rotations of 0°, 120°, and 240° about τ1 and τ2 with respect to the crystal structure (τ1 = 187.8° and τ2 = 295.3°). We shall use the notation (Δτ1,Δτ2) to indicate the conformation with τ1 = 187.8° + Δτ1 and τ2 = 295.3° + Δτ2. For example, the (0,0) conformer is the STO-3G optimized con-

former corresponding to the crystal structure. We have not performed a similar systematic investiga-
tion of the possible conformations of the hydroxyl groups because of the large number of conformers (324) that would need to be considered. In order to be consistent with the earlier nonoptimized study with which we wished to compare our results, we instead limited our search to finding the global minimum for the hydroxyl groups for each of the nine conformations of the side-chain. We have optimized all side-chain variables in all conformations, while


COI 0192-8651/91/0113-112$04.00