HOMOMORPHISM POLYNOMIALS OF CHEMICAL GRAPHS

D.M. BERMAN and K.W. HOLLADAY

Department of Mathematics, University of New Orleans, New Orleans, Louisiana 70148, USA

Abstract

We say that a graph G is *homomorphic* to a graph H if there is a mapping p from the vertices of G onto the vertices of H such that $p(u)$ and $p(v)$ are adjacent in H whenever u and v are adjacent in G. The *homomorphism polynomial* of a graph G is a polynomial in two variables that counts the number of homomorphisms of G onto the complete graph of each order. This polynomial can be computed recursively in an analog to the chromatic polynomial. In this paper, we present some results regarding the homomorphism polynomials of the graphs of chemical compounds – in particular, alkane isomers. The coefficients of the homomorphism polynomial can be used to predict the rankings of compounds with respect to several chemical properties. Our results seem to refine those obtained by Randić et al. from path lengths.

A basic problem of theoretical chemistry is to classify chemical compounds with respect to their chemical or physical properties, and to use the classifications to predict the properties of new compounds, the ultimate goal being to do this by computation rather than experimentation. Since the time of Cayley, it has been common to study chemical compounds in terms of graphs constructed with the atoms as vertices and the chemical bonds as edges. A natural application of graph theory is to the study of (constitutional) isomers, i.e. compounds with identical molecular formulae but differing constitution. Indeed, enumerating isomers was one of the first problems of graph theory, and remains one of its most common applications in chemistry.

More recently, there have been efforts to use graph-theoretic parameters to distinguish among isomers and to rank them in order of various chemical and/or physical properties. Ideally, the parameter and the measured values of the property for a few members of the set would provide reasonably accurate extrapolations of

the values for the entire set. Although the eigenvalues of the adjacency matrix of the graph have been extensively used by chemists (see e.g. [1]), there have been numerous attempts to make use of a wider variety of graph properties [2–5] in recent years.

Randić and Wilkins (see [4] and its bibliography) have had some success in using the numbers of paths of length p (the p-path numbers) of the appropriate graphs to rank-order the thermodynamic and physical properties of isomeric alkanes. Table 1 contains some of their data. The 2-path numbers provide a coarse correlation:

Table 1

Comparison of path numbers and homomorphism polynomial for hexanes. The path numbers are the numbers of paths of length 2 and 3. The first homomorphism polynomial number is the coefficient of $x^5 y^6$. It is equal to the number of paths of length 2. The second homomorphism polynomial number is the coefficient of $x^4 y^3$. Qa is the heat of atomization. bp is the boiling point. Qv is the heat of vaporization. Source: ref. [4].

Hexane	Path numbers		Hom. poly.		Qa	bp	Qv
	2	3	1	2			
2,2-MM	7	3	7	13	1801.49	49.7	6.651
2,3-MM	6	4	6	11	1799.63	57.9	6.985
2-M	5	3	5	10	1798.80	60.2	7.160
3-M	5	4	5	9	1798.16	63.5	7.255
n	4	3	4	8	1797.10	68.7	7.555

for different 2-path numbers in an isomer set, there is a parallel change in the properties of the isomers. However, there are a number of cases where two or more isomers from the same set have the same 2-path numbers. Some improvement is obtained by considering as well the 3-path numbers, but this does not remove all of the ambiguities. Moreover, there are some properties that increase with the 2-path number but decrease with the 3-path number.

We propose to use, instead, the homomorphism polynomial, introduced by Bari [6]. We say that a graph G is *homomorphic* to a graph H if there is a mapping p from the vertices of G onto the vertices of H such that $p(u)$ and $p(v)$ are adjacent in H whenever u and v are adjacent in G. (This differs from G and H being isomorphic, in that $p(u)$ and $p(v)$ may be adjacent in H when u and v are not adjacent in G.) In a sense, homomorphism is a generalization of coloring, in that an n-coloring of G can be seen as a homomorphism from G to the complete graph on n vertices. A homomorphism also may be seen as a generalized "folding" of G onto H [7]. It may be this aspect that explains the relevance to chemical properties.

The *homomorphism polynomial* of a graph G is a polynomial in two variables that counts the number of homomorphisms of G onto the complete graph of each

order. This polynomial can be computed recursively in an exact analogy to the chromatic polynomial. (For other variations, see [8].) The chromatic polynomial can in fact be derived from the homomorphism polynomial: the formula will be given in the proof of theorem 2. Since this derivation is in polynomial time, we see that the basic problem of computing the full homomorphism polynomial for a general graph is of NP-complete difficulty. Our results indicate that non-isomorphic graphs may nonetheless be detected from a small part of the homomorphism poylnomial, which can be computed rather quickly.

To compute the homomorphism polynomial of G, let u and v be any two non-adjacent vertices. We construct from G the two new graphs: G^*uv, in which u and v are replaced by a new vertex v' which is adjacent to each vertex adjacent in G to u or v; and $G + uv$, in which a "pseudo-edge" is added, joining u to v. (We will denote these new graphs as G^* and G^+ when there is no ambiguity.) This process is iterated until G is reduced to graphs each of which is a "pseudo K_p", a p vertex graph in which each pair of vertices is joined by an edge or a pseudo-edge. See fig. 1(c) for an example of this process.

The homomorphism polynomial can then be read from this decomposition: having a term $x^p y^q$ for each resulting pseudo K_p having exactly q pseudo-edges. Note that a graph with n vertices and e edges will have only a single x^n term with a coefficient of 1 and with y an exponent of $C(n, 2) - e$ (where $C(n, k)$ denotes the binomial coefficient). This is because the only pseudo K_n is obtained by the addition of a pseudo-edge between each pair of non-adjacent vertices. In particular, for trees the first term will always be $x^n y^{C(n-1, 2)}$.

We will conventionally write the homomorphism polynomial in decreasing order, first of x exponents and then of y exponents. The first non-trivial term is thus the leading x^{n-1} term. We will rank homomorphism polynomials first by the coefficient of this term. We will show in theorem 1 that for trees this coefficient will equal the 2-path number, so that for this class of graphs representing acyclic chemical compounds, we obtain the same first-order rankings as in [3].

In this paper, we consider properties of alkanes, i.e. acyclic compounds composed entirely of carbon and hydrogen with single bonds. We follow the usual practice [4] of representing alkanes by their carbon backbone (see figs. 1(a) and 1(b) for an example). Thus, the graph model of an n carbon alkane is a tree with n vertices and maximum degree 4.

Table 1 gives a comparison of path numbers and the homomorphism polynomials of hexanes. Notice that the ranking given by the homomorphism polynomial agrees with the orders of the values of the three given properties.

THEOREM 1

Let G be an n vertex tree. The coefficient of the x^{n-1} term having the largest y exponent gives the number of 2-paths in G.
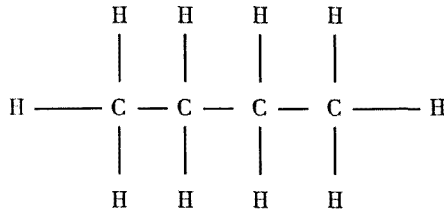
Fig. 1(a). *n*-butane.

Fig. 1(b). *n*-butane.

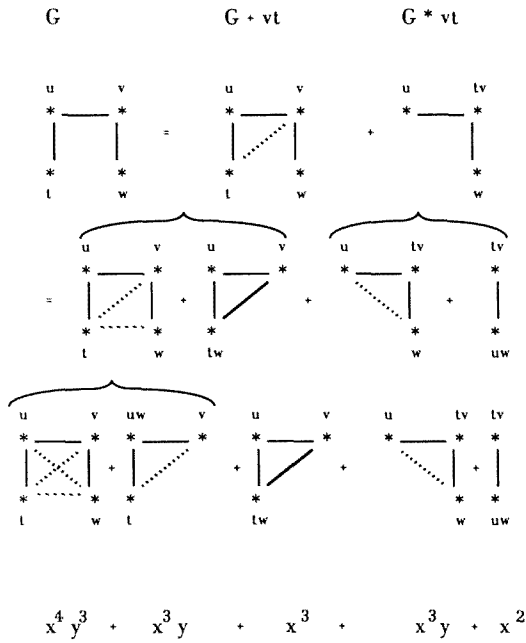$$x^4 y^3 \;+\; x^3 y \;+\; x^3 \;+\; x^3 y \;+\; x^2$$

Fig. 1(c).

Fig. 1. (a) The complete structure graph of *n* butane. (b) The graph of the carbon backbone of this molecule. (c) The computation of the homomorphism polynomial for *n* butane.

Proof

From the definition of the homomorphism polynomial we obtain an x^{n-1} term for each pseudo K_{n-1} in the final reduction of G. We will obtain one such from the contraction of each pair u, v of non-adjacent vertices; the other steps in its construction will all be additions of pseudo-edges to the graph $G^\star uv$. The number of pseudo-edges added gives the exponent of y. We will deduce this number by counting the number of actual edges in $G^\star$.

Consider the new vertex u' obtained by contracting u and v. Let w be any other vertex; recall that w is adjacent to u' in $G^\star$ if it is adjacent to u or v in G. Therefore, its degree in $G^\star$ will either be the same as its degree in G or will be one less. The latter will happen if it is adjacent to both u and v, i.e. if u, w, v is a 2-path. Note that there can be only one such w, as G is a tree. Likewise, the degree of u' will either be equal to the sum of the degrees in G of u and v, or else one less if there is a 2-path.

Therefore, the number of actual edges in $G^\star uv$ is either one less than or equal to the number in G, depending on whether or not there is a 2-path from u to v, and hence there will be only two x^{n-1} terms and the coefficient of the one with the larger y exponent will be the number of 2-paths. It is easily computed that the sum of the coefficients of the two x^{n-1} terms can be expressed as the binomial coefficient $C(n-1, 2)$. The possible values of the y exponent will be $C(n-1, 2) - (n-1)$ and $C(n-1, 2) - (n-2)$.

We know of no such simple characterization of the other terms of the homomorphism polynomial, but the proof of theorem 1 gives an indication that the coefficient of the leading x^{n-2} term will also give in some sense a measure of the "compactness" of the structure of the tree. This is therefore the next term we consider. This term can discriminate between pairs of graphs with the same 3-path numbers, and we will present data indicating that this ranking of the homomorphism polynomials agrees with the observed values of several chemical properties (see fig. 1). Note that we do not use the coefficient of the second x^{n-1} term because it conveys no new information. It is just $C(n-1, 2)$ minus the coefficient of the leading x^{n-1} term.

More generally, in a tree the coefficients of the terms with a given exponent p for x will always add up to a constant depending only on p.

THEOREM 2

In the homomorphism polynomial of T_v, a tree with v vertices, the sum of the coefficients of all the terms with x exponent n is equal to the Stirling number of the second kind $S(v-1, n-1)$.

Proof

Define a linear transformation L on the real vector space of polynomials by sending the basis elements x^n to $[x]_n$, the falling factorials of degree n. (The falling factorial is defined by $[x]_n = x(x-1)\ldots(x-n+1)$.) By theorem 2($v$) in Bari [6], the chromatic polynomial $P(G; x)$ of a graph can be obtained from the homomorphism polynomial $h(G; x, y)$ by the formula

$$P(G; x) = L(h(G; x, 1)).\tag{1}$$

The coefficient of x^n in $h(T_v; x, 1)$ is obtained by summing the coefficients of all the terms of $h(T_v; x, y)$ with x exponent n. Let us write this as

$$h(T_v; x, 1) = \sum c_n x^n.\tag{2}$$

It is well known ([9], p. 128) that the chromatic polynomial of T_v is

$$P(T_v; x) = x(x-1)^{v-1}.\tag{3}$$

Substituting (3) into the LHS of (1) and (2) into the RHS, we get

$$x(x-1)^{v-1} = \sum c_n [x]_n.$$

Dividing by x and using the fact that $[x]_n/x = [x-1]_{n-1}$, we obtain

$$(x-1)^{v-1} = \sum c_n [x-1]_{n-1}.$$

This is immediately recognizable as an identity for the Stirling numbers of the second kind, evaluated at $x-1$ (see [10], p. 41). That identity shows that c_n is the Stirling number $S(v-1, n-1)$.

Once rankings are obtained, a natural question arises concerning the extent to which the qualitative results can be extended to quantitative predictions about the ranked properties. We have preliminary results using linear regression that show good correlations between homomorphism polynomial coefficients and physical constants such as heat of atomization and heat of formation. Figure 2 and table 2 show the results for the hexanes, a case where the value of 13 for 2, 2-MM gives a good linear fit to the data. Here, the plausible value of 12 would give a correct ranking but not as good a fit. When interpreting the regression results in table 2, it should be remembered that the homomorphism polynomial coefficients are giving counts of certain objects and hence must be small integers. The results shown are about as good as small integers can give. This same discretization effect also appears in fig. 3, which gives the results for heptanes. The regression results for heptanes in table 3 are reason-
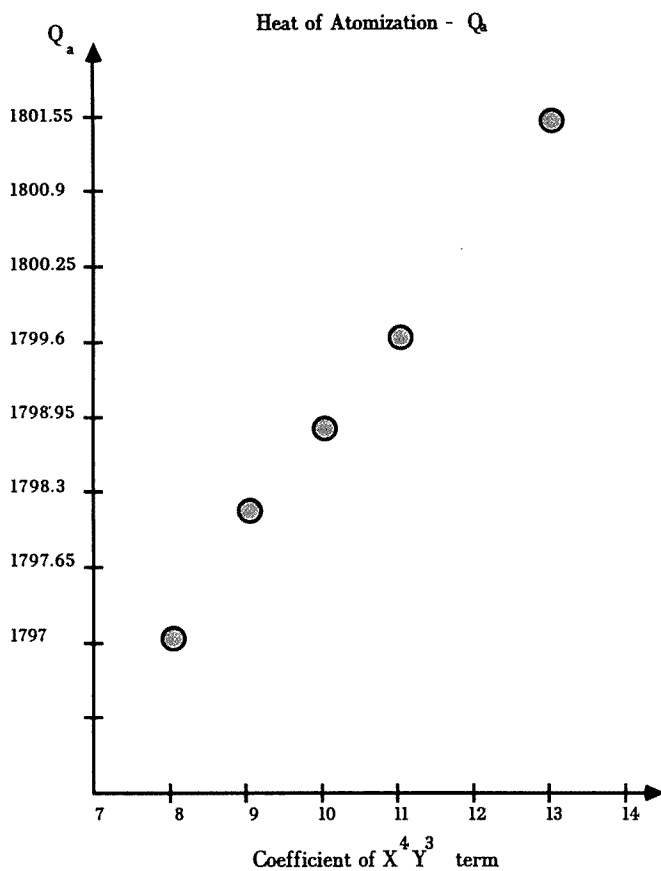
Heat of Atomization - Q_a

Fig. 2. Graph of heat of atomization for the hexanes versus the coefficient of the $X^4 Y^3$ term of the homomorphism polynomial of the structure graph of the hexane. See table 2 for the actual numbers.

Table 2

Regression results for hexanes and heat of atomization

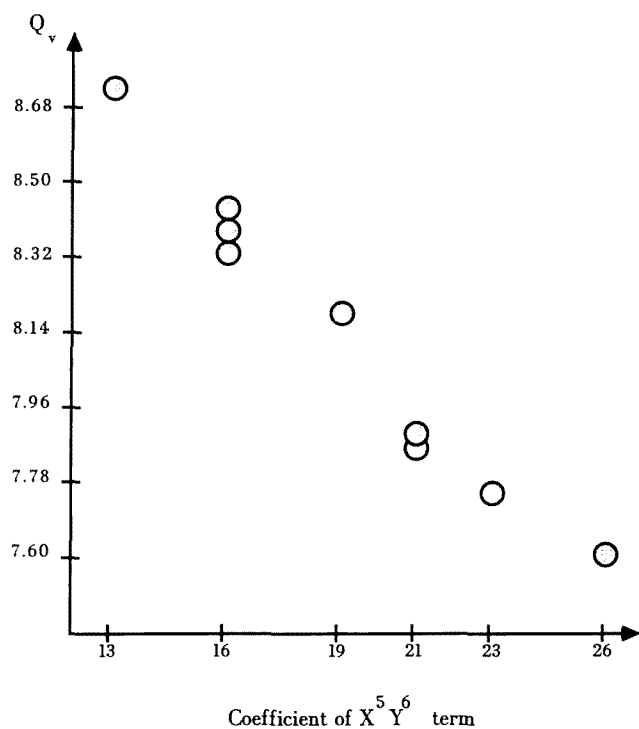| Hexane	Hom. poly. coefficient	Qa	Least squares predicted	Residual
2,2-MM	13	1801.49	1801.440	0.051
2,3-MM	11	1799.63	1799.723	− 0.093
2-M	10	1798.80	1798.864	− 0.064
3-M	9	1798.16	1798.006	0.154
n	8	1797.10	1797.148	− 0.048

Fig. 3. Graph of heat of vaporization for the heptanes versus the coefficient of the $X^5 Y^6$ term of the homomorphism polynomial for the heptane. See table 3 for the actual numbers.

Table 3

Regression results for heptanes and heat of vaporization. The values for heat of vaporization are from [4]. The homomorphism polynomial coefficient 1 is the coefficient of $X^5 Y^6$ and the coefficient 2 is the coefficient of $X^5 Y^5$

Heptane	Hom. poly. 1	2	Qv	Least squares predicted	Residual
n	13	36	8.739	8.6404	0.0986
2-M	16	37	8.325	8.3487	− 0.0237
3-M	16	35	8.391	8.4030	− 0.0120
3-E	16	33	8.421	8.4573	− 0.0363
2,4-MM	21	34	7.872	7.9892	− 0.1172
2,3-MM	19	34	8.191	8.1656	0.0254
2,2-MM	23	36	7.764	7.7585	0.0055
3,3-MM	21	36	7.901	7.9349	− 0.0339
2,2,3-MMM	26	33	7.669	7.5754	0.0936
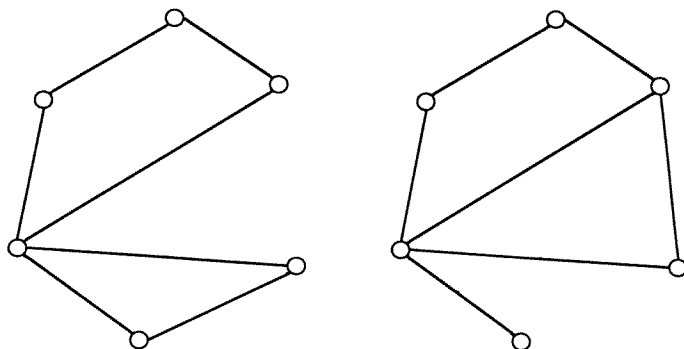
$$x^6 y^8 + 2x^5 y^5 + 4x^5 y^4 + 2x^5 y^3 + 7x^4 y^2 + 8x^4 y + 6x^3$$

Fig. 4. The above two non-isomorphic graphs have the same homomorphism polynomial. The common homomorphism polynomial is displayed below the graphs. This example was found by M. Albertson.

able and provide further evidence that quantitative results are within the reach of our methods. Once we have more data, we will examine the regression coefficients themselves to see if they can be predicted.

It is well known [11] that there exist families of non-isomorphic graphs having the same eigenvalues. Slater [12] has disproved Randić's conjecture that non-isomorphic trees must be distinguishable by their numbers of paths of each length. We have an example due to Albertson of non-isomorphic graphs with identical homomorphism polynomials (fig. 4). Girse [13] has asked whether the homomorphism polynomial can always distinguish trees. Even if it can not distinguish all trees, it would be important to know whether it can distinguish all graphs in the chemically important class of trees with maximum degree 4.

Acknowledgements

References

[1]   D.H. Rouvray, C.R. Acad. Sci. Paris Ser. C 274(1967)1561.
[2]   J.L. Fox, C & EN (1983)45.
[3]   M. Randić, *Int. J. Quant. Chem.: Quantum Biology Symp. 5* (Wiley, New York, 1978) pp. 245–255.

[4]   M. Randić and C.L. Wilkins, J. Phys. Chem. 83, 11(1979)1525.

[5]   M. Randić, SIAM J. Alg. and Disc. Meth. 6(1985)145.

[6]   R.A. Bari, J. Comb. Inf. & Sys. Sci. 7(1982)56.

[7]   C.R. Cook and A.B. Evans, *Proc. 10th S.E. Conf. Comb., Graph Theory, and Computing,* Congressus Numerantium XXIII (1979) pp. 305 – 314.

[8]   D. Joyce, Discr. Math. 50(1984)51.

[9]   J.A. Bondy and U.S.R. Murty, *Graph Theory with Applications* (North-Holland, New York, 1976).

[10] C. Berge, *Principles of Combinatorics* (Academic Press, New York, 1971).

[11] A.J. Schwenk, in: *New Directions in the Theory of Graphs,* ed. F. Harary (Academic Press, New York, 1973) pp. 275 – 307.

[12] P.J. Slater, J. Graph Theory 6(1982)89.

[13] R.D. Girse, in press.