Finding Relations in Polynomial Time

Gilles Caporossi
Ecole Polytechnique
CJP. 6079, Succ. Centre-viUe
Montreal H3C 3A7, Quebec, Canada
e-mail: gillesc@crt.umontreal.ca

Pierre Hansen
GERAD and
Eeole des Hautes Etudes Commerciales
3000, Chemin de la cote-Ste-Catherine
Montreal H3T 2A7, Quebec, Canada
e-mail: pierreh@crt.umontreal.ca

Abstract

Given a set of m observations on n variables, an $0(mn^2)$ algorithm is proposed to find a basis of all affine relations between these variables satisfied by the observations. On a 25 variables example, this new algorithm is 130 000 times faster than the "all subsets" option for linear regression of the SAS package which is a non polynomial alternative. Extension to the cases where squares, ratios, products of pairs of variables or logarithms of such terms appear in the relations is straightforward and remains polynomial. The method is first tested with data for several classical discoveries studied previously by the Bacon programs. Then it is added to the AutoGraphiX system for computer-aided graph theory thus making it entirely automated. To demonstrate the power of the resulting system, five novel relations (or conjectures) in graph theory are found, two of which pertain to mathematical chemistry. Three conjectures involve five invariants, which is more than in most propositions of graph theory. Proofs of two conjectures are also given.

1 Introduction

Given a set of observations of several variables, to find efficiently parsimonious relations that they satisfy is a basic problem in the study of scientific discovery [Langley *et* a/., 1987], [Simon *et* a/., 1997] and, possibly, in its enhancement. More generally, one may seek relations between a subset of the given set of variables. Enumerating all subsets and considering them in turn reduces this problem to the former one, but entails a combinatorial explosion. The purpose of this paper is to show that, for large classes of relations, the latter problem can be solved in polynomial time. More precisely, a polynomial algorithm is proposed for finding a basis of all affine relations between the variables (or powers, ratios, products, or logarithms of them). This algorithm uses basic results of linear algebra, also exploited in data analysis, and in particular in principal component analysis [Pontier *et* a/., 1990]. However, the purpose there, i.e., to explain differences between observations, is the reverse of ours, i.e., to find what properties the observations have in common.

The method is described in the next section. In Section 3 it is applied to several classical discoveries already studied previously by the Bacon programs [Langley *et* al, 1987]. In Section 4, results of its embedding in the AutoGraphiX (AGX) system for computer-aided graph theory are presented. The resulting entirely automated system discovers five novel relations (or conjectures) in graph theory, two of which pertain to mathematical chemistry. Three of these conjectures involve five invariants, which is more than in most propositions of graph theory. Conclusions are drawn in Section 5. Proofs of two among the five conjectures are given in the Appendix.

2 Basic principles

2.1 Description of the algorithm

Consider a set of m observations of n quantitative variables $x_1, x_2, \ldots, x_n$. Let $X_{m\times n} = (x_{ij})$ denote the corresponding data set. Our aim is to find a basis of all affine relations (i.e., relations of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, where the a_j and b are constants) satisfied by the observations. Let $\mu = \frac{1}{m}{}^t X 1_m$ denote the vector of mean values for the columns of X. Then any affine relation between columns of X becomes a linear relation (i.e., a relation of the form $a'_1x'_1 + a'_2x'_2 + \cdots + a'_nx'_m = 0$) between columns of $X' = X - (1_m{}^t\mu)$, the matrix of centered values. Furthermore, all coefficients (except b) are unchanged. Indeed if $x_1 = \sum_{j=2}^{n} c_jx_j + d$ then

$$\mu_1 = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=2}^{n} c_jx_{ij} + d\right) = \sum_{j=2}^{n}\frac{1}{m}\sum_{i=1}^{m} c_jx_{ij} + \sum_{i=1}^{m}\frac{d}{m} = \sum_{j=2}^{n} c_j\mu_j + d$$

and thus

$$x'_1 = x_1 - \mu_1$$

$$= \sum_{j=2}^{n} c_j x_j + d - \left(\sum_{j=2}^{n} c_j \mu_j + d\right)$$
$$= \sum_{j=2}^{n} c_j (x_j - \mu_j) = \sum_{j=2}^{n} c_j x'_j.$$

Centering variables, the first step of our algorithm, thus transforms the problem of finding affine relations into the problem of finding linear ones.

Consider then the variance-covariance matrix *V* defined by $V_{n \times n} = {}^t X' X'$.

If the relation

$$x'_j = \sum_{\ell=1, \ell \neq j}^{n} c_\ell x'_\ell \qquad \text{(i)}$$

holds, then

$$v_{jk} = \sum_{i=1}^{m} x'_{ij} x'_{ik} = \sum_{i=1}^{m} \left(\sum_{\ell=1, \ell \neq j}^{n} c_\ell x'_{i\ell} \right) x'_{ik}$$
$$= \sum_{\ell=1, \ell \neq j}^{n} c_\ell \sum_{i=1}^{m} x'_{i\ell} x'_{ik} = \sum_{\ell=1, \ell \neq j}^{n} c_\ell v_{\ell k}$$

which means that if the linear relation (1) holds for the columns of *X'* it also holds for those of *V.* As *V* is symmetric, it still holds for its lines. The second step of our algorithm is thus to compute V.

The third and last step is to diagonalize *V* (with, however, some empty lines if there are relations). This can be done, by Gaussian elimination. In the resulting matrix V, *Dim(Im(V))* lines contain non-zero terms and correspond to independent variables. The remaining n — *Dim(Im(V))* lines contain only zeroes and correspond to dependent variables which may be expressed as linear combinations of the independent ones. These relations form a basis of the null-space of *V.* Using the initial data one can then compute the right-hand sides of the corresponding affine relations.

The case were a dependent variable is expressed as a monomial of several independent ones, instead of an affine expression is easily reduced to the previous one. Indeed, assume

$$x'_j = f \prod_{\ell=1, \ell \neq j}^{n} (x'_\ell)^{e_\ell} \qquad (2)$$

where the *et* and / are constants. Taking logarithms, this relation becomes

$$\log x'_j = \sum_{\ell=1, \ell \neq j}^{n} e_\ell \log(x'_\ell) + \log(f). \qquad (3)$$

So the method just described can be used to find exponents *et* (equal to 0 if x'_ℓ does not appear in the monomial) and the initial data to find /.

The proposed algorithm is illustrated on several examples in Section 3.

2.2 Extension

The two classes of relations described in the previous subsection can be easily and substantially enlarged by computing additional terms from the initial data. The easiest case is to take, in addition to the given variables, all squares and products of pairs of them. Then one uses $n + \frac{n^2+n}{2} = \frac{n^2+3n}{2}$ variables instead of n. The formulas have the form

$$x_j = \sum_{\ell=1, \ell \neq j}^{n} c_\ell x_\ell + \sum_{k=1, k \neq j}^{n} \sum_{\ell=k, \ell \neq j}^{n} c_{k\ell} x_k x_\ell + d.$$

Ratios, higher powers than two or products of three variables may also be considered. If *Xj* is expressed as a monomial, sums or differences of variables may be taken as additional terms.

2.3 Complexity

The first step of the method, computing centered values involves computing μ_i, in 0(mn) time and subtracting μ_j from each x_{ij} again in 0(mn). The second step, computing the variance-covariance matrix V, t a k$\mathbf{0(mn^2)}$ m e . The third step, diagonalizing *V* by Gaussian elimination requires $0(n^3)$ time. In order to avoid spurious relations, one must have *m>n.* Thus the overall complexity is in $\mathbf{0(mn^2)}$ **(or in** $\mathbf{0(n^3)}$**)** if one makes the reasonable assumption that m is 0(n)). If squares and products of pairs of variables are added the complexity rises to $\mathbf{0(n^6)}$. Such problems, with n moderate, are still solvable in a reasonable time.

As an example for comparison, using linear regression allows at most one relation to be found at a time while our method finds a basis of all the relations underlying the data. Furthermore, linear regression needs one variable to be explained by the others, which means that a regression must be computed for each of the n — 1 first variables to be explained by the next ones. To ensure that all relations are found, one cannot escape the "best-subset" criterion for each regression, thus trying all combinations of variables, which has exponential complexity. Treating a 20 variables and 44 observations example with the SAS statistical software needs 10 seconds while our system needs 0.0013 sec (on a Sun Ultra I, 140 MHz). Using 25 variables instead of 20 and still 44 observations, the time required by SAS increases to 4 minutes and a half while our method only needs 0.002 seconds (130 000 times faster).

3 Rediscovering known laws

3.1 Kepler's third law

Kepler's third law is that $p^2/d^3 = ct$ where *p* denotes the period of a planet's orbit and *d* its average distance to the sun. As all those considered in this section, this law was rediscovered by one of the Bacon programs [Langley *et al.*, 1987]. The computing time they use is much larger than with our method. Note, however, that the aim of these programs is different from that pursued here: i.e., to understand the reasoning which led to the discovery

rather than to find an efficient but computer-dependent way to discover the corresponding relation.

A generator was used to produce the following four simulated observations of *p* and *d* (taking units such that ct = I):

$$X = \begin{pmatrix} 9.736442 & 30.380885 \\ 1.371510 & 1.606196 \\ 1.869378 & 2.555911 \\ 2.456725 & 3.850657 \end{pmatrix}$$

or, switching to logarithms,

$$X_{log} = \begin{pmatrix} 2.275875 & 3.413813 \\ 0.315912 & 0.473868 \\ 0.625605 & 0.938408 \\ 0.898829 & 1.348243 \end{pmatrix}.$$

The corresponding variance-covariance matrix is

$$V = \begin{pmatrix} 0.507597 & 0.507597 \\ 0.507597 & 0.507597 \end{pmatrix}.$$

Diagonalization leads to

$$V' = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

which shows that a relation of the form log(p) = $\mathbf{log(d)ct'}$ holds. Then using $\boldsymbol{X_{log}}$, we find *ctf* = 1.5, and an equivalent form of the original relation, p — d^{1*5}. Computing time is too small to be measured.

3.2 Ideal-gas law

The ideal gas law is *PV* = *nRT* where *P is* the pressure (in Pascals), *V* the volume (in cubic meters), n the number of mote and T the temperature (in Kelvins). *R* — 8.32 is the universal ideal gas constant. The generator gave simulated observations of P, V, n and T. These were then converted to the older units, to simulate the situation prevalent at the time of discovery, i.e., before absolute temperatures were used. Then the pressure *p* is measured in atmospheres, the volume *v* in liters and the *t* temperature in Celsius degrees. The conversion rules are *P* = 101325p, *V* = v/1000 and T = t+273.15. Using 25 observations and the extended set of terms including squares and products of pairs of variables led to the relation *nt* = -273.15n + 12.1785pt/ in less than 1/100 of a second. Converting back to the *SI* System units gives *nT* = 0.120192PV, which is an equivalent form of the law presented above.

Note that this relation is even easier to find, using the *SI* System units, if we switch to the logarithms of the values; then an extended data set is not needed.

Sixteen observations are enough and the variance-covariance matrix for these data is

$$V = \begin{pmatrix} 0.506 & 0.049 & -0.076 & 0.631 \\ 0.049 & 0.324 & 0.009 & 0.364 \\ -0.076 & 0.009 & 0.279 & -0.346 \\ 0.631 & 0.364 & -0.346 & 1.343 \end{pmatrix}$$

Diagonalizing this matrix becomes

$$V' = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

which shows that a relation of the form log(T) - log(P) + log(F) - log(n) + *ct* holds. From the logarithms of the data, we find *ct* = 0.120192. So the equivalent form $T = 0.120192PV_n^{-1}$ of the ideal gas law is found instantaneously.

3.3 Newton's gravity law

Newton's Law is that $\boldsymbol{F = G\frac{mM}{D^2}}$ where m and M are the masses of two bodies, *D* the distance between them, *F* their force of attraction and $\boldsymbol{G = 6.67e^{-11}}$ the universal gravitational constant. The problem is formally very similar to the previous one, as one seeks a monomial involving three variables.

The generator provided 25 simulated observations of F, m, *M* and D, then the system obtained from their logarithms the relations $\boldsymbol{F = mMD^{-2}6.67e^{-11}}$ which is equivalent to the law given above. Again the derivation was instantaneous.

In addition to this result, the system gave spurious laws, i.e., m+F = m, *M* + *F* = *M* and D + P = P. It is easy to recognize that they are incorrect, and due to numerical errors. A more careful scaling of units before computation avoids this error.

3.4 Ohm's law

Ohms' law is that *IL* = *rI* + *v* where I denotes current intensity, *L* inductance, r resistance and *v* tension. The generator provided 25 simulated observations of J, L, r and *v.* Using an extended data set with squares and products of pairs of variables, i.e., 14 terms (or 10 when seeking for expression of a dependent variable as a monomial, because in that case sums of twice the same variable are redundant) the relation *rI* = —*v* + *IL* was found in less than 1/100 of a second.

Note that using only the original observations would not have allowed to find this relation.

4 Finding new relations in graph theory

4.1 The AutoGraphiX system

Several systems have been proposed for computer-added or automated discovery of relations in graph theory, prominent among which are Graph [Cvetkovic *et* a/., 1981); [Cvetkovic and Simie, 1994], and Graffiti [Fajtlowicz, 1988]; [Fajtlowicz, 1998]. They work with *graph invariants,* i.e., quantities such as the number of vertices, number of edges, stability number, chromatic number, radius, diameter, etc., (see [Berge, 1973] for undefined terms) which do not depend on labeling of vertices or edges. Recently, a new system, called AutoGraphiX (AGX) has been developed [Caporossi and Hansen, 1997]. Its basic idea is to consider the problem

of determining extremal graphs for some graph invariant, possibly subject to constraints, and related problem such as refuting conjectures, as problems of combinatorial optimization defined on an infinite family of graphs. Then, powerful heuristics may be brought to bear to find presumably extremal graphs, of moderate size, from which much information may be gathered. More precisely, AGX addresses the following problems:

(a) *Find a graph satisfying given constraints.* Let $i_1(G), i_2(G), \ldots, i_\ell(G)$ denote invariants of a graph G and $p_1, p_2, \ldots, p_\ell$ the values they should take. Consider the problem

$$\min_{G \in \mathcal{G}_n} f(G) = \sum_{k=1}^{\ell} |i_k(G) - p_k|$$

where $\mathcal{G}_n$ denotes the set of graphs with n vertices, and n may be varied. Any graph for which $f(G) = 0$ satisfies the given constraints.

(b) *Find optimal or near optimal values for an invariant $i_0(G)$ subject to constraints.* Consider the problem

$$\min_{G \in \mathcal{G}_n} (\max) f(G) = i_0(G) + M \sum_{k=1}^{\ell} |i_k(G) - p_k|$$

where M is a sufficiently large constant.

(c) *Refute a conjecture,* expressed as $h(G) < g(G)$ where $h(G)$ and $g(G)$ are formulas depending on one or more invariants of G. Consider the problem

$$\min_{G \in \mathcal{G}_n} f(G) = g(G) - h(G).$$

If a graph G for which $f(G) < 0$ is found, the conjecture is refuted.

(d) *Suggest a conjecture.* This can be done in various ways, which use parameterization on the number n of vertices and m of edges of G or other invariants. Surfaces of values for the invariant or formula under study can be displayed graphically, which often suffices to obtain conjectures, and presumably extremal graphs can be studied, which gives further information.

(e) *Suggest a proof.* Checking if presumably extremal graphs found are also obtained with simplified heuristics, such as e.g. local search with a single type of move may suggest ways to prove conjectures for all graphs, or for some classes of them, e.g., by showing that such a move can always be performed on any non-extremal graph and is an improving one.

The heuristics used in AGX fit into the Variable Neighborhood Search metaheuristic framework [Hansen and Mladenovic, 1998]. A local search routine is applied to G, which employs several types of moves: add or remove an edge, displace one or two edges, add or remove a pending vertex, insert a vertex in an edge, etc. Once a locally optimal graph G_1 has been found, one considers a set of neighborhoods $N_1(G_1), N_2(G_1), \ldots, N_{k\max}(G_1)$ increasingly far from G_1 and defined by addition or removal of 1,2,...,kmax edges. A graph G_2 in the neighborhood $N_1(G_1)$, $N_2(G_1)$, etc., is drawn at random and a local search performed leading to a graph G_3. If G_3 has a better value than G_1 the search is recentered there. AGX has been applied to a series of problems of graph theory, some of which pertain to mathematical chemistry [Caporossi and Hansen, 1997]; [Caporossi *et* a/., 1998a]; [Cvetkovic *et* at, 1998]; [Caporossi *et* a/., 1998b]; [Caporossi *et* a/., 1999]. 24 conjectures have been obtained up to now, 11 of which have been proved and none of which has been refuted. In addition, AGX alone refuted 3 conjectures of Graffiti [Caporossi and Hansen, 1997] and, together with some enumeration routines 6 more [Pujol, 1998]. As described, AGX is a tool for computer-assisted discovery. However, addition of a routine implementing the algorithm for finding relations described in Section 2 transforms it into an entirely automated system. Results obtained when doing so are next presented.

4.2 A property of color-constrained trees with minimum index

A tree is a bipartite graph, whose vertices may be colored in black and white such that no pair of adjacent vertices have the same color. If the numbers of black and of white vertices are fixed the tree is *color-constrained.* In [Cvetkovic *et* a/., 1998] AGX is used to study color-constrained trees extremal with respect to their index, or largest eigenvalue of their adjacency matrix. Six conjectures are obtained, and four of them proved.

To explore further the extremal trees found, values of the following fifteen invariants were recorded: number n of vertices, number n_1 of pending vertices, number m of edges, diameter D, radius r, stability number a (i.e., maximum number of pairwise non-adjacent vertices), average degree 5, average distance and sum of distances between pairs of vertices, energy (sum of the absolute values of the eigenvalues of the adjacency matrix), maximum degree, largest eigenvalue of the adjacency matrix, Hyperwiener index (sum of the squares of distances between pairs of vertices), Randic index x *see* below), and Chromatic number γ.

Then the algorithm of Section 2 was used to find a basis of affine relations on those invariants. In addition to well-known relations, such as TO = n — 1 and $\gamma = 2$ which are valid for all trees, the unexpected following one was obtained

$$2\alpha - m - n_1 + 2r - D = 0. \tag{4}$$

and the remaining 8 invariants are proved to be linearly independant from any of the considered ones.

This result can also be expressed as follows.

Conjecture 1. *For all color-constrained trees with minimal index*

$$\alpha = \frac{1}{2}(m + n_1 + D - 2r). \tag{5}$$

Furthermore, AGX was used to see whether this conjecture could be extended to all trees. Minimizing and maximizing the left-hand side of (4) then led to the following results.

Conjecture 2* *For all trees*

$$\alpha \leq \frac{1}{2}(m + n_1 + D - 2r). \quad (6)$$

Let $\lfloor a \rfloor$ denote the largest integer smaller or equal to a.

Conjecture 3. *For all trees*

$$\alpha \geq \frac{1}{2}\left(m + n_1 + D - 2r - \left\lfloor \frac{n-2}{2} \right\rfloor\right). \quad (7)$$

Conjecture 2 is proved in the Appendix.

4.3 Two bounds for the connectivity index of chemical graphs

Chemical graphs are extensively used to represent hydrocarbons. Vertices correspond to carbon atoms and edges to bonds between them (hydrogen atoms are usually not represented as their number and bonds are uniquely defined). Consequently the maximum vertex degree is 4. Moreover, chemical graphs are connected. The connectivity index (or Randic index) [Randil, 1975] has been introduced as a measure of molecular branching. It is defined as follows. Let (v_i, v_j) denote an edge of a graph G, *dj* and *dj* the degrees of its end-vertices. Then the *weight* of that edge is $(d_i d_j)^{-1/2}$ and the *connectivity index x{G)* *s the sum of the weights of all edges of *G*. This index is one of the most frequently used molecular structure-descriptors, and the only one to which two books [Kier and Hall, 1976]; [Kier and Hall, 1986] and numerous articles have been devoted.

In [Caporossi *et al.,* 1998b] chemical trees with extremal connectivity index are studied, based on results of AGX, and folly characterized. Applying AGX, with the relation-finding routine, to chemical graphs with presumably minimal connectivity index, did at first give no relations. However, considering the set of local minima among these graphs gave two new relations:

Conjecture 4. *For all chemical graphs*

$$\chi(G) \geq \frac{1}{4}(n_1 + m) \quad (8)$$

Conjecture 5. *For all chemical graphs*

$$\chi(G) \geq \frac{n}{3} + \frac{m}{12}. \quad (9)$$

Conjecture 4 is proved in the Appendix.

5 Concluding remarks

An algorithm for finding automatically from observations of a set of variables a basis of affine relations between these variables, or products, ratios or powers of them, or between logarithms of such terms has been presented. As it is polynomial in the number of variables considered, the combinatorial explosion encountered when each potential relation is considered in turn is avoided.

Moreover, adding a routine implementing this algorithm to the system AGX led to an entirely automated system for finding relations in graph theory. Its use is illustrated by the derivation of five novel conjectures, two of which pertain to mathematical chemistry. Many more applications are expected in the near future.

A Proofs of two conjectures

Theorem 1 (Conjecture 2). *For all trees with stability number* α_1, m *edges,* n_i *pending vertices, diameter D and radius r*

$$\alpha \leq \frac{1}{2}(m + n_1 + D - 2r).$$

Proof. Recall that in a tree $D = 2r$ if D is even and $D = 2r - 1$ if JO is odd. Hence, *D* - 2r = *-(D* mod (2)). Assume first that *D* is even. Define the *star* of a vertex V_i to be the set of edges incident with v_i,. Clearly, stars of vertices of a stable set are disjoint. Consider a maximum stable set S with α_1 pending vertices and α_2 inner vertices. Then, $m \geq \alpha_1 + 2\alpha_2 = 2\alpha - \alpha_1 \geq 2\alpha - n_1$ and $\alpha \leq \frac{1}{2}(m + n_1)$

Assume then that *D* is odd. Consider a path *P* of length *D.* Either only one end vertex of *P* belongs to *S* or there is a pair of consecutive vertices v_i and v_j of *P* which do not belong to 5, and the edge (v_i, v_j) does not belong to the star of any vertex of 5. In both cases $m \geq 2\alpha - n_1 + 1$ and $\alpha \leq \frac{1}{2}(m + n_1 - 1)$. □

Theorem 2 (Conjecture 4). *For all chemical graphs G with connectivity index* x(G), n_1 *pending vertices and m edges*

$$\chi(G) \geq \frac{1}{4}(n_1 + m).$$

Proof. Let n_i, $i = 1, \ldots, 4$ denote the numbers of vertices of degree i of G and y_{ij}, $i, j = 1, \ldots, 4$; $i \leq j$ denote the numbers of edges of G with end vertices of degree i and j. Then consider the 6 equations:

$$\begin{aligned} n_1 + n_2 + n_3 + n_4 &= n \\ n_1 + 2n_2 + 3n_3 + 4n_4 &= 2m \\ y_{12} + y_{13} + y_{14} &= n_1 \\ y_{12} + 2y_{22} + y_{23} + y_{24} &= n_2 \\ y_{13} + y_{23} + 2y_{33} + y_{34} &= n_3 \\ y_{14} + y_{24} + y_{34} + 2y_{44} &= n_4 \end{aligned}$$

obtained by counting vertices and degrees. These equations are clearly independent. Solving for $n_1, n_2, n_3, n_4, y_{14}$ and y_{44} and substituting y_{14} and y_{44} in the objective function

$$\min \chi(G) = \sum_{i=1}^{4} \sum_{j=i}^{4} \frac{1}{\sqrt{ij}} x_{ij} \qquad (10)$$

leads to the relation

$$\min \chi(G) = \frac{1}{4}n_1 + \frac{1}{4}m + \left(\frac{1}{\sqrt{2}} - \frac{1}{2}\right) y_{12} + \left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right) y_{13} + \frac{1}{4} y_{22} + \left(\frac{1}{\sqrt{6}} - \frac{1}{4}\right) y_{23} + \left(\frac{1}{2\sqrt{2}} - \frac{1}{4}\right) y_{24} + \frac{1}{12} y_{33} + \left(\frac{1}{2\sqrt{3}} - \frac{1}{4}\right) y_{34}. \qquad (11)$$

As $y_{12}, y_{13}, y_{22}, y_{23}, y_{24}, y_{33}$ and y_{34} are non-negative and have positive coefficients in (11) the result follows. □

Conjecture 5 can be proved in a similar way.

Aknowledgments

The authors thank Francois Bellavance and Robert Latour from Ecole des Hautes Etudes Commerciales de Montreal for their advices on statistical methods, and Ivan Gutman from university of Kragujevac (Yougoslavia) for discussions about AGX and its results. This work was supported by NSERC grant OGPOO 39682, FCAR grant 32EQ 1048 and a grant from CETAI-HEC.

References

[Berge, 1973] Claude Berge. *Graphs and Hypergraphs.* Wiley, New York, 1973.

[Caporossi and Hansen, 1997] Gilles Caporossi and Pierre Hansen. Variable Neighborhood Search for Extremal Graphs. 1. The AutoGraphiX System. Les Cahiers du GERAD G-97-41, Ecole des Hautes Etudes Commerciales, Montreal, June 1997, revised September 1997. To appear in *Discrete Mathematics.*

[Caporossi *et al.,* 1998a] Gilles Caporossi, Dragos Cvetkovid, Ivan Gutman, and Pierre Hansen. Variable Neighborhood Search for Extremal Graphs. 2. Finding Graphs with Extremal Energy. Les Cahiers du GERAD G-98-44, Ecole des Hautes Etudes Commerciales, Montreal, August 1998. To appear in *Journal of Chemical Information and Computer Sciences.*

[Caporossi *et al.,* 1998b] Gilles Caporossi, Ivan Gutman, and Pierre Hansen. Variable Neighborhood Search for Extremal Graphs. 4. Chemical Trees with Extremal Connectivity Index. To appear in *Computers and Chemistry.*

[Caporossi *et al.,* 1999] Gilles Caporossi, Andrey Dobrynin, Ivan Gutman, and Pierre Hansen. Trees with Palindromic Hosoya Polynomials. Les Cahiers du GERAD G-99-10, Ecole des Hautes Etudes Commerciales, Montreal, January 1999.

[Cvetkovic and Simie, 1994] Dragos Cvetkovic and Slobodan Simil. Graph Theoretical Results Obtained by the Support of the Expert System "Graph". *Bulletin de l'Acadimie Serbe des Sciences et des Arts,* T CVn:19-41,1994.

[Cvetkovid *et al.,* 1981] Dragos Cvetkovic, L. Kraus, and Slobodan Simic. Discussing Graph Theory with a Computer. I. Implementation of Graph Theoretic Algorithms. *Publ. Elektrotehn. Fak. Ser. Mat. Fiz.,* 716-734, University of Beograd, 1981.

[Cvetkovic *et al.,* 1998] Dragos Cvetkovic, Slobodan Simic, Gilles Caporossi, and Pierre Hansen. Variable Neighborhood Search for Extremal Graphs. 3. On the Largest Eigenvalue of Color-constrained Trees. Les Cahiers du GERAD G-98-66, Ecole des Hautes Etudes Commerciales, Montreal, December 1998.

[Fajtlowicz, 1988] Siemion Fajtlowicz. On Conjectures of Graffiti. *Discrete Mathematics,* 72:113-118,1988.

[Fajtlowicz, 1998] Siemion Fajtlowicz. Written on the Wall. Version 05-1998, 1998. Regularly updated file accessible via email from siemionOmath.uh.edu.

[Hansen and Mladenovic, 1998] Pierre Hansen and Nenad Mladenovic. An Introduction to Variable Neighborhood Search. In I.H. Osman, S. Voss, S. Martello and C. Roucairol, editors, *Meta-Heuristics: Advances and Trends in Local Search Paradigms for Optimization,* pages 433-458. Kluwer, Boston, 1998.

[Kier and Hall, 1976] Lemont B. Kier and Lowell H. Hall. *Molecular Connectivity in Chemistry and Drug Research.* Academic Press, New York, 1976.

[Kier and Hall, 1986] Lemont B. Kier and Lowell H. Hall. *Molecular Connectivity in Structure-activity Analysis.* Wiley, New York, 1986.

[Langley *et al.,* 1987] Pat Langley, Herbert A. Simon, Gary L. Bradshaw, and Jan M. Zytkow. *Scientific Discovery. Computational Exploration of the Creative Process.* MIT Press, Cambridge, Massachusetts, 1987.

[Pontier *et al.,* 1990] Jacques Pontier, Anne-Beatrice Dufour, and Myriam Normand. *Le modile Euclidien en analyse des donnees.* Editions de l'Universite Libre de Bruxelles, Bruxelles, 1990.

[Pujol, 1998] Florian Pujol. *Etude de conjectures en thiorie des graphes.* Rapport de stage ISIMA, directed by Pierre Hansen, Montreal, 1998.

[Randic, 1975] Milan Randic. On Characterization of Molecular Brandling. *Journal of the American Chemical Society,* 97:6609-6615,1975.

[Simon *et al.,* 1997] Herbert A. Simon, Raul E. Valdes-Perfe, and Derek H. Sleeman. Scientific Discovery and Simplicity of Method. *Artificial Intelligence,* 91:177-181,1997.