On kites, comets, and stars. Sums of eigenvector coefficients in (molecular) graphs.

Rücker, Christoph; Rücker, Getra; Gutman, Ivan

Published in:
Zeitschrift fur Naturforschung - Section A Journal of Physical Sciences

DOI:
10.1515/zna-2002-3-406

Publication date:
2002

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Citation for pulished version (APA):
Rücker, C., Rücker, G., & Gutman, I. (2002). On kites, comets, and stars. Sums of eigenvector coefficients in (molecular) graphs. *Zeitschrift fur Naturforschung - Section A Journal of Physical Sciences*, *57*(3-4), 143-153. https://doi.org/10.1515/zna-2002-3-406

On Kites, Comets, and Stars.
Sums of Eigenvector Coefficients in (Molecular) Graphs

Gerta Rücker, Christoph Rücker[a], and Ivan Gutman[b]

Department of Rehabilitative and Preventative Sports Medicine, University of Freiburg, Hugstetter Str. 55, D-79106 Freiburg

[a] Department of Mathematics, University of Bayreuth, D-95440 Bayreuth

[b] Faculty of Science, University of Kragujevac, P.O.Box 60, YU-34000 Kragujevac, Yugoslavia

Reprint requests to G. R.; Fax: +49 921 55 3385, E-mail: GertaRuecker@aol.com

Z. Naturforsch. **57 a,** 143–153 (2002); received January 22, 2002

Two graph invariants were encountered that form the link between (molecular) walk counts and eigenvalues of graph adjacency matrices. In particular, the absolute value of the sum of coefficients of the first or principal (normalized) eigenvector, s_1, and the analogous quantity s_n, pertaining to the last eigenvector, appear in equations describing some limits (for infinitely long walks) of relative frequencies of several walk counts. Quantity s_1 is interpreted as a measure of *mixedness* of a graph, and s_n, which plays a role for bipartite graphs only, is interpreted as a measure of the *imbalance* of a bipartite graph. Consequently, s_n is maximal for star graphs, while the minimal value of s_n is zero. Mixedness s_1 is maximal for regular graphs. Minimal values of s_1 were found by exhaustive computer search within the sample of all simple connected undirected n-vertex graphs, $n \leq 10$: They are encountered among graphs called *kites*. Within the special sample of tree graphs (searched for $n \leq 20$) so-called *double snakes* have maximal s_1, while the trees with minimal s_1 are so-called *comets*. The behaviour of stars and double snakes can be described by exact equations, while approximate equations for s_1 of kites and comets could be derived that are fully compatible with and allow to predict some pecularities of the results of the computer search. Finally, the discriminating power of s_1, determined within trees and 4-trees (alkanes), was found to be high.

Key words: Molecular Graphs; Walks; Eigenvector Coefficients.

Introduction

When chemists talk about molecular structures and the properties of compounds they often use qualitative and more or less intuitive concepts, such as the complexity of a structure or the diversity of a set of structures. It is natural to ask how such concepts can be rendered quantitative, how something such as the complexity of a structure can be measured. For this purpose (among others) the so-called topological indices (TIs) were introduced [1, 2]. A topological index is a number associated with a graph or a chemical structure and derived therefrom by some well-defined procedure. It is a graph invariant, which means that its numerical value is independent of how a particular graph (structure) is drawn or how its vertices are numbered. Hundreds of TIs have been defined, some purposefully designed, some obtained by mathematical manipulations on already existing definitions, and so it became legitimate to conversely ask for the meaning of a particular TI. This issue is even more urgent if a graph invariant was not constructed by man but was simply uncovered, existing but having gone unnoticed hitherto. In the present work we deal with two such graph invariants which we encountered during our study of walks in molecular graphs.

Definitions

In this work some mathematical properties of two very simple and quite "natural" graph invariants will be investigated, sums of the coefficients of eigenvectors of the adjacency matrix of an n-vertex graph: s_1 is the sum of coefficients of the first (principal) eigenvector, while s_n is the sum of coefficients of the last eigenvector (if unambiguously defined).

Relations between atomic and molecular walk counts, eigenvalues and eigenvector coefficients, and spectral moments have been known for a long time and are described in due detail [3 - 5]. Here the most important formulas are only repeated without proofs. Notations are as in our previous papers, where also the method of proofs can be found. Throughout this paper all graphs are assumed to be connected, which, of course, is an obligatory property of any molecular graph.

Let **A** be the adjacency matrix of a simple connected undirected graph G with n vertices. $\mathbf{A}^k$ then is the k'th power of **A**, its elements are denoted $a_{ij}^{(k)}$. As is well-known, $a_{ij}^{(k)}$ $(i, j = 1, \ldots, n)$ is to be interpreted as the number of walks of length k starting at vertex i and ending at vertex j.

The *atomic walk count* of length k of vertex i, $\mathrm{awc}_k(i)$, sometimes named the extended degree of i of order k, is the i'th row (or column) sum of the matrix $\mathbf{A}^k$,

$$\mathrm{awc}_k(i) = \sum_{j=1}^{n} a_{ij}^{(k)}.$$

The *molecular walk count* of length k, mwc_k, is the sum of all atomic walk counts in $\mathbf{A}^k$,

$$\mathrm{mwc}_k = \sum_{i=1}^{n} \mathrm{awc}_k(i).$$

Further, let us consider for each vertex i the number of its *self-returning walks* of length k, denoted $\mathrm{swc}_k(i)$,

$$\mathrm{swc}_k(i) = a_{ii}^{(k)},$$

and the total number of self-returning walks of length k in the graph (molecule),

$$\mathrm{swc}_k = \sum_{i=1}^{n} a_{ii}^{(k)}.$$

The atomic indices $\mathrm{awc}_k(i)$ and $\mathrm{swc}_k(i)$ may be interpreted as measures of the centrality or involvedness of vertex i within the graph [3, 5 - 7], while the molecular indices mwc_k und swc_k measure a molecule's or graph's complexity [8, 9].

There are tight connections between the fundamental graph features walks on the one hand and the eigenvalues and eigenvectors of the adjacency matrix on the other. Let $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of **A** and let $\{x_1, \ldots, x_n\}$ be an orthonormal basis of eigenvectors x_i of **A**, where x_i is the eigenvector associated with λ_i. Further, let

$$s_i = \sum_{j=1}^{n} x_{ij} \text{ and } \sigma_i = s_i^2$$

be the sum of coefficients of the i'th eigenvector and its square, respectively. We shall always choose the signs of the coefficients of the eigenvector x_i so that s_i is non-negative.

Obviously, these definitions are unambiguous (except for the sign of s_i) unless the eigenvalue λ_i is degenerate, that is at least for $i = 1$. In the case of degenerate eigenvalues, the eigenvectors are not uniquely determined; for this case it can be shown that the sum of σ_i over the indices i belonging to the same eigenspace is uniquely determined [4, 10]. In the following we demonstrate that s_i and σ_i appear in formulas describing the relations between walk counts, eigenvalues, and eigenvectors.

Walk counts may be described using eigenvalues, eigenvector coefficients and sums of eigenvector coefficients as follows [3]:

$$a_{ij}^{(k)} = \sum_{p=1}^{n} \lambda_p^k x_{pi} x_{pj},$$

$$\mathrm{awc}_k(i) = \sum_{p=1}^{n} \lambda_p^k s_p x_{pi},$$

$$\mathrm{swc}_k(i) = \sum_{p=1}^{n} \lambda_p^k x_{pi}^2,$$

$$\mathrm{swc}_k = \sum_{p=1}^{n} \lambda_p^k = \mathrm{Trace}(\mathbf{A}^k),$$

$$\mathrm{mwc}_k = \sum_{p=1}^{n} \lambda_p^k \sigma_p.$$

These relations are referred to as spectral decomposition [4]. Walk counts primarily depend on the power (length) k. Measuring the centrality of two vertices i and j by $\mathrm{awc}_k(i)$, $\mathrm{awc}_k(j)$ or $\mathrm{swc}_k(i)$, $\mathrm{swc}_k(j)$, one observes in some cases a converging (or not so) oscillation of the relative ranks of particular vertices i and j from one k to the next. Therefore it is appropriate to have k approaching infinity, that is to consider walks of infinite length.

Some Limits of Walk Shares

Cvetković and Gutman [11] defined a sequence $(T_k)_{k\in\mathrm{N}}$ of topological indices

$$T_k = (\mathrm{mwc}_k/n)^{1/k}.$$

T_k^k may be considered as the mean extended connectivity of order k (mean over all vertices). Using the technique of spectral decomposition, these authors showed that the sequence $(T_k)_{k\in\mathrm{N}}$ converges against the principal eigenvalue λ_1:

$$\lim_{k\to\infty}(T_k) = \lambda_1.$$

In this sense λ_1 is the "long term" average degree; Cvetković and Gutman dubbed λ_1 the "dynamic degree". In a less rigorous manner one can view this dynamic degree as follows. If one makes a long random walk in the respective graph, and records the degrees of the vertices encountered, then the average value of these vertex degrees is λ_1.

Consider now the sequence

$$M_k = \mathrm{mwc}_k/\lambda_1^k = \sum_{p=1}^{n}(\lambda_p/\lambda_1)^k\sigma_p$$

and distinguish two cases:

Case 1. $\lambda_1 > |\lambda_n|$, the case of nonbipartite (connected) graphs. Then we have:

$$\lim_{k\to\infty}(M_k) = \sigma_1.$$

This means that mwc_k is of the same order as λ_1^k. This statement is similar to the above result.

Case 2. $\lambda_1 = |\lambda_n|$, i. e. $\lambda_n = -\lambda_1$. This happens in any connected bipartite graph, that is in a tree or in a cyclic graph without odd-membered cycles.

In this case, for k approaching infinity two summands remain, one of which, belonging to the smallest eigenvalue λ_n, has an alternating sign [3]. Therefore we have to consider two partial sequences, one for even, and another for odd k, resulting in

$$\lim_{k\to\infty}(M_{2k-1}) = \sigma_1 - \sigma_n,$$

$$\lim_{k\to\infty}(M_{2k}) = \sigma_1 + \sigma_n,$$

These two limits coincide if and only if $\sigma_n = 0$, as it happens e. g. for even-membered chains [3] and for regular graphs [11]. Both results back the interpretation of λ_1 as dynamic degree, and attract our attention to σ_1 and σ_n. The odd / even case discrimination is typical for the procedure in considering limits of walk counts in bipartite graphs, to be done in the next section.

Table 1. Limits of quotient series of walks, non-bipartite case ($\lambda_1 > |\lambda_n|$).

Enumerator	Denominator			
	swc_k	$\mathrm{awc}_k(i)$	mwc_k	λ_1^k
$\mathrm{swc}_k(i)$	x_{1i}^2	x_{1i}/s_1	x_{1i}^2/σ_1	x_{1i}^2
swc_k	1	–	$1/\sigma_1$	1
$\mathrm{awc}_k(i)$	–	1	x_{1i}/s_1	s_1x_{1i}
mwc_k	–	–	1	σ_1

We now apply spectral decomposition to some sequences of quotients which can be interpreted as relative walk frequencies in a totality of similar walks. The exact procedure was described [3] for the sequence

$$p_k(i) = \mathrm{awc}_k(i)/\mathrm{mwc}_k,$$

where $p_k(i)$ is the relative frequency of walks of length k starting at vertex i among all walks of that length. Similar sequences, constructed analoguously, are

- $\mathrm{swc}_k(i)/\mathrm{swc}_k$, the frequency of self-returning walks of length k of vertex i among all self-returning walks of that length,
- $\mathrm{swc}_k(i)/\mathrm{awc}_k(i)$, the frequency of self-returning walks of length k of vertex i among all walks of that length starting at vertex i,
- $\mathrm{swc}_k(i)/\mathrm{mwc}_k$, the frequency of self-returning walks of length k of vertex i among all walks of that length,
- $\mathrm{swc}_k/\mathrm{mwc}_k$, the frequency of self-returning walks of length k among all walks of that length.

For the transition $k\to\infty$ in the bipartite case it is always necessary to distinguish between odd and even k. Table 1 for (connected) non-bipartite graphs and Table 2 for (connected) bipartite graphs contain the results for the above as well as for other similar quotients. Parts of these results have appeared in [5a] already.

Quantities σ_1 and σ_n as Graph Invariants

The appearance of s_1 (or $\sigma_1 = s_1^2$) and s_n (or $\sigma_n = s_n^2$) in the above Tables (the latter in the bipartite case only) suggests to look at the behaviour and the meaning of these graph invariants more closely.

Table 2. Limits of quotient series of walks, bipartite case ($\lambda_n = -\lambda_1$).

Enumerator	k	Denominator swc_k	$\mathrm{awc}_k(i)$	mwc_k	λ_1^k
$\mathrm{swc}_k(i)$	even[a]	x_{1i}^2	$2x_{1i}^2/(s_1x_{1i}+s_nx_{ni})$	$2x_{1i}^2/(\sigma_1+\sigma_n)$	$2x_{1i}^2$
swc_k	even[a]	1	–	$2/(\sigma_1+\sigma_n)$	2
$\mathrm{awc}_k(i)$	odd	–	1	$(s_1x_{1i}-s_nx_{ni})/(\sigma_1-\sigma_n)$	$s_1x_{1i}-s_nx_{ni}$
	even	–	1	$(s_1x_{1i}+s_nx_{ni})/(\sigma_1+\sigma_n)$	$s_1x_{1i}+s_nx_{ni}$
mwc_k	odd	–	–	1	$\sigma_1-\sigma_n$
	even	–	–	1	$\sigma_1+\sigma_n$

[a] For (connected) bipartite graphs the coefficients of the first and last eigenvector differ in sign only, thus $x_{1i}^2 = x_{ni}^2$ for all i. Furthermore $\mathrm{swc}_k = 0$ for odd k.

Graph Invariant σ_1 as a Measure of Mixedness of a Graph

The following is observed at least in the non-bipartite case (Table 1): First, a graph with given principal eigenvalue λ_1 contains the more walks, the larger σ_1. Second, the reciprocal σ_1^{-1} measures the share of self returning walks among all walks. Accordingly, a large σ_1 implies a small share of self returning walks, or a high probability that a randomly chosen walk ends at a vertex other than its origin. For these facts we consider σ_1 as a measure of the *mixedness* of a graph. This interpretation is backed by the observation that σ_1 is closely related to the variance of the eigenvector coefficients x_{1i} in an inverse way:

$$\mathrm{Var}(x_{11},\ldots,x_{1n}) = \left[\frac{1}{n-1}\sum x_{1i}^2 - \frac{1}{n}\Big(\sum x_{1i}\Big)^2\right]$$

$$= (n-\sigma_1)/(n(n-1)).$$

Since x_{1i}/s_1, the limit of the sequence $(\mathrm{awc}_k(i)/\mathrm{mwc}_k)_{k\in\mathrm{N}}$, can be interpreted as the contribution of vertex i to the total number of walks (see Table 1), a small variance of the eigenvector coefficients x_{1i} means a rather equal distribution of all walks over individual vertices, that is a high mixedness. As is easily seen, this variance assumes its minimum 0 if all coefficients x_{1i} are mutually equal, wherefore

$$x_{1i} = 1/\sqrt{n}$$

is necessary and thus

$$s_1 = \sum x_{1i} = \sqrt{n} \text{ and } \sigma_1 = n.$$

This is the case if and only if G is a regular graph. Regular graphs therefore are maximally mixed.

There is another aspect of that topic. Let $e = (1,...,1)^{\mathrm{T}}$ be the "space diagonal" in $\mathbb{R}^n$. Then

$$s_1 = x_1^{\mathrm{T}} e$$

is the scalar product of the principal eigenvector and e, and by $||x_1|| = 1$

$$s_1/\sqrt{n} = x_1^{\mathrm{T}} e/\big(||x_1||\,||e||\big) = \cos(x_1, e)$$

is the cosine of the angle between x_1 and e in $\mathbb{R}^n$. Its value is maximal (i. e. 1) if x_1 und e are collinear (if the graph is regular), and small for those graphs whose principal eigenvector is almost perpendicular to e, that is has very different coefficients.

Graph Invariant σ_n as a Measure of the Imbalance of a Bipartite Graph

While the interpretation of σ_1 as a measure of mixedness is cogent for non-bipartite graphs, bipartite graphs present difficulties in the following two facts: First, for calculating a vertex's share of the walks, walks of odd and even lengths have to be distinguished. Second, both limits depend on σ_1 and on σ_n.

In the following we show that σ_n is always smaller than σ_1, and we characterize the graph with maximal σ_n for given n.

Proposition. For (connected) bipartite graphs the following holds:

(i) $\sigma_n < \sigma_1$.
(ii) For fixed n, the graph with maximal σ_n is the star.

Proof: Since the graph is bipartite and connected, $\lambda_n = -\lambda_1$ with nondegenerate λ_n, and $x_{ni}^2 = x_{1i}^2$ for all vertices $i = 1, \ldots, n$, that is, the coefficients of the principal and of the last eigenvector differ by sign if at all, and all coefficients of x_1 have equal sign. Without loss of generality we may assume $x_{1i} > 0$ for $i = 1, ..., n$. Then from the orthogonality of the eigenvectors x_1 and x_n it follows that

$$0 = \sum x_{1i}x_{ni} = \sum_{i\in \mathrm{I}_+} x_{1i}^2 - \sum_{i\in \mathrm{I}_-} x_{1i}^2, \qquad (*)$$

where the vertex sets $\mathrm{I}_+ = \{i : x_{ni} > 0\}$ and $\mathrm{I}_- = \{i : x_{ni} < 0\}$ form that partition of the vertex set $\{1, \ldots, n\}$ induced by the graph's bipartiteness. On the other hand, since x_1 is normalized,

$$1 = \sum_{i=1}^{n} x_{1i}^2 = \sum_{i\in \mathrm{I}_+} x_{1i}^2 + \sum_{i\in \mathrm{I}_-} x_{1i}^2. \qquad (**)$$

By adding (or subtracting, respectively) $(*)$ and $(**)$, we obtain:

$$\sum_{i\in \mathrm{I}_+} x_{1i}^2 = \sum_{i\in \mathrm{I}_-} x_{1i}^2 = \frac{1}{2}. \qquad (***)$$

If we now set

$$s_+ = \sum_{i\in \mathrm{I}_+} x_{1i} \text{ and } s_- = \sum_{i\in \mathrm{I}_-} x_{1i}\ (s_+, s_- > 0),$$

it follows that

$$\sigma_1 = s_1^2 = (s_+ + s_-)^2 \text{ and } \sigma_n = s_n^2 = (s_+ - s_-)^2,$$

and therefrom

$$\sigma_1 - \sigma_n = 4s_+s_- > 0,$$

whereby (i) is proven.

In order to prove the second statement, we have to show that the difference $|s_+ - s_-|$ is maximal for the star.

By $(***)$ the sums of squares of the eigenvector coefficients over each of the two index sets are constant and equal to $^1/_2$. Therefore the sum s_+ is maximal if all x_{1i} ($i \in \mathrm{I}_+$) are identical, say $x_{1i} = x_+$.

Let n_+ be the cardinality of I_+, n_- the cardinality of I_-. Since $x_1 \not\equiv x_n$, n_+ and n_- are positive and $n = n_+ + n_-$. By $(***)$, we have

$$n_+x_+^2 = \frac{1}{2}, \text{ that is } x_+ = 1/\sqrt{2n_+},$$

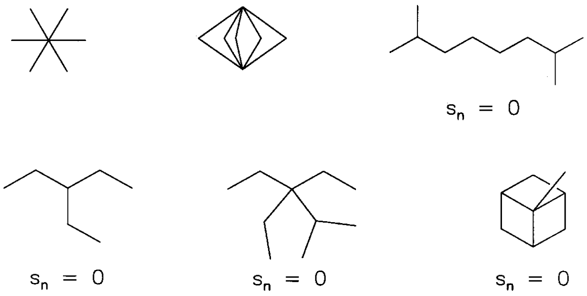

Fig. 1. Some graphs mentioned in the text.

and by definition

$$s_+ = n_+x_+ \text{ and therefore } s_+ = \sqrt{n_+/2}.$$

Thus s_+ is maximal for n_+ as large as possible, and this is the case of $n_+ = n - 1$ and $n_- = 1$. If one index set consists of a single vertex and the second of all other vertices, then the graph is a star (Fig. 1, top left), and the following is true:

$$\sigma_n = (s_+ - s_-)^2 = \left[\sqrt{n-1} - 1\right]^2/2,$$

$$\sigma_1 = (s_+ + s_-)^2 = \left[\sqrt{n-1} + 1\right]^2/2.$$

The proof of the proposition is now complete.

As was shown in the proof, σ_n is the larger the more different are the numbers of vertices of the two vertex sets of a bipartite graph. The more different the numbers of vertices are, the more different are the mean degrees within the two sets. The larger σ_n is, the more different also are the limits of the odd and even walk sequences in Table 2. Therefore it seems natural in a bipartite graph to consider σ_n as a measure of the *imbalance* between both vertex sets and also between walk counts of odd and even length (Fig. 2), or simply of the imbalance of the graph itself. In this sense the star is the most imbalanced among all graphs of a given number of vertices.

What does large imbalance mean? As mentioned above, large imbalance means large differences between the odd and even limits in Table 2, so that, for example, the weight of a certain vertex within the graph, measured by $\mathrm{awc}_k(i)/\mathrm{mwc}_k$, oscillates for increasing k [3].

Where the difference between odd and even limit is marked (as is extremely found for the star), it seems

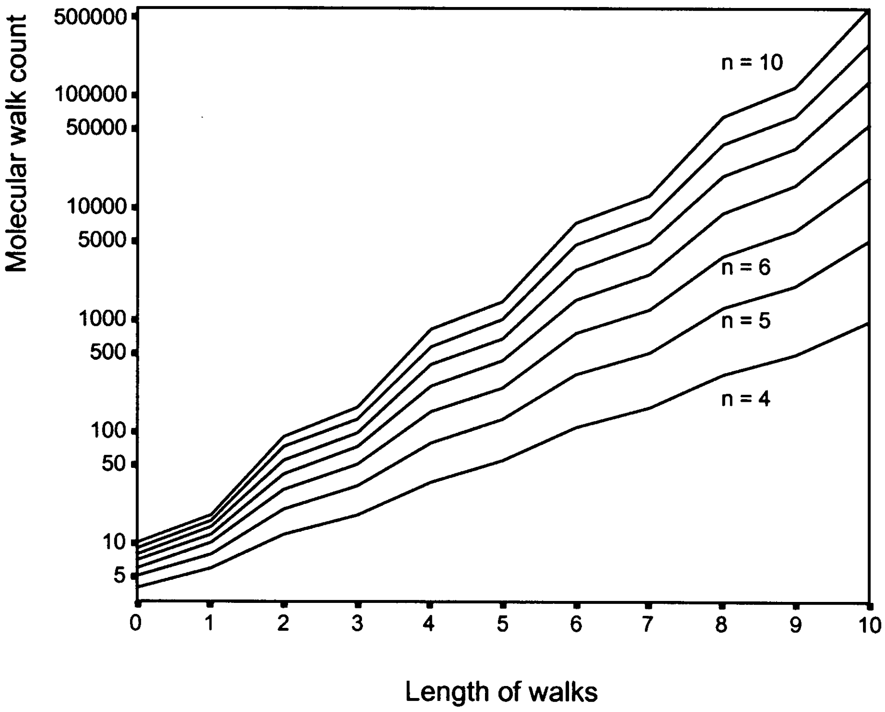

Fig. 2. Molecular walk counts of stars with n vertices (n = 4, ..., 10).

natural to define the share of a vertex among the total number of walks as the average of the odd and even limits, as suggested by us earlier [3]. Let

$$p_{\text{odd}}(i) = \lim_{k\to\infty, k \text{ odd}} (\text{awc}_k(i)/\text{mwc}_k),$$

$$p_{\text{even}}(i) = \lim_{k\to\infty, k \text{ even}} (\text{awc}_k(i)/\text{mwc}_k),$$

and

$$p(i) = [p_{\text{odd}}(i) + p_{\text{even}}(i)]/2.$$

We insert the values of p_{odd} (i) and p_{even} (i) from Table 2, and taking into account $x_{ni}^2 = x_{1i}^2$ we obtain after some transformations

$$p(i) = \frac{x_{1i}}{s_1} S \text{ with } S = \frac{s_1^2 \pm s_1 s_n + s_n^2}{s_1^2 \pm s_1 s_n + s_n^2 \pm \frac{s_n^3}{s_1}}.$$

In this formula, the plus sign holds for vertices i with $x_{ni} = x_{1i}$, while the minus sign holds for those with $x_{ni} = -x_{1i}$. If $|s_n^3| \ll |s_1|$, s_n^3/s_1 may be neglected, and we obtain $S \approx 1$ and so

$$p(i) \approx x_{1i}/s_1,$$

as in the non-bipartite case. If on the other hand $|s_n| > 1$, then this relation is not even approximately valid, as shown by the counterexample of the star.

Example. Let us consider the star on n vertices. Its central vertex is labeled by number one. It can be shown that for the central vertex

$$p_{\text{odd}}(1) = \frac{1}{2}, \quad p_{\text{even}}(1) = \frac{1}{n},$$

while for vertices 2, . . . , n

$$p_{\text{odd}}(i) = \frac{1}{2(n-1)}, \quad p_{\text{even}}(i) = \frac{1}{n}.$$

Then we have for the central vertex

$$p(1) = \frac{n+2}{4n},$$

and for the outer vertices $i > 1$

$$p(i) = \frac{3n-2}{4n(n-1)}.$$

For $n \to \infty$ we obtain the share of the central vertex $i = 1$ to be $^1/_4$, and therefore that of all other vertices taken together to be $^3/_4$. Considering, on the other hand, the coefficients of the principal eigenvector we see:

$$x_{11} = 1/\sqrt{2}, \quad x_{1i} = 1/\sqrt{2(n-1)} \text{ for } i > 1,$$

that is

$$\frac{s_-}{s_1} = \frac{x_{11}}{s_1} = \left(1 + \sqrt{n-1}\right)^{-1}$$

$$\to 0 \text{ for } n \to \infty \text{ (share of the central vertex)}$$

and

$$\frac{s_+}{s_1} = \frac{(n-1)x_{1i}}{s_1} = \left(1 + 1/\sqrt{n-1}\right)^{-1}$$

$$\to 1 \text{ for } n \to \infty \text{ (share of the outer vertices).}$$

The imbalance of the star graph, measured as s_n, is so marked that the central vertex for large stars (large n) bears a share of $^1/_2$ (odd limit), close to $^1/_4$ (mixed limit), or close to 0 (even limit), depending on the weighting scheme. For the star of Fig. 1 (n = 7), $|s_n|$ is 1,02494. In a somewhat less marked manner one finds the same for other imbalanced graphs such as $K_{2,5}$ (Fig. 1, top middle, $|s_n|$ = 0.58114).

We now seek for those bipartite graphs whose s_n value is minimal. As already mentioned, the minimal value of s_n is zero, found e. g. for chain graphs of even n. This is a special case, to be generalized to such graphs of even n that contain for each vertex from vertex set 1 an equivalent (by symmetry) vertex from vertex set 2. For an example see the even-membered *double snake* in Fig. 1, top right. The corresponding coefficients of eigenvector x_n differ in sign only, cancelling one another and leaving s_n = 0. This fits to the interpretation of s_n as imbalance of a bipartite graph. However, there are bipartite graphs with s_n = 0 outside of this group, even such with odd number of vertices. A few examples are shown in Fig. 1, bottom line.

The two graph invariants s_1 and s_n for bipartite graphs are almost not intercorrelated, as shown in the samples of all connected bipartite graphs with n = 7 (N = 44, r = −0.257) and n = 8 (N = 182, r = −0.304).

Some Properties of σ_1

Though the meaning of σ_1 is not completely clear for very imbalanced graphs such as the stars, we con- tinue to interpret it as a measure of the graph's "mixedness". In this section we consider the distribution of σ_1 within classes of graphs with a fixed vertex count n, and in particular we look for graphs extremal (maximal or minimal) with respect to σ_1 within a class. First we treat the class of all connected graphs on n vertices, then the class of n-vertex trees.

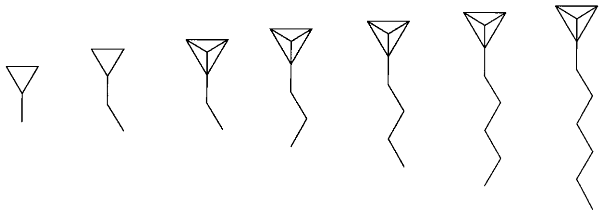

Fig. 3. Graphs with minimal σ_1 for fixed n.

Connected Graphs of Extremal σ_1

As mentioned above, σ_1 is maximal for given n if and only if all degrees are identical, i. e. the graph is regular, independently of its degree. In this case all coefficients of the principal eigenvector are equal and their variance is zero. Examples are the n-cycle, C_n, and the complete graph of n vertices, K_n.

The question for graphs with minimal σ_1 within a class of graphs of constant vertex number is more interesting and less easy to answer. Because

$$\sigma_1 = \left(\sum_{j=1}^{n} x_{1j}\right)^2 \geq \sum_{j=1}^{n} x_{1j}^2 = 1,$$

σ_1 has unity as a lower bound. However, σ_1 could be equal to 1 only if there would be one j_0 with $|x_{1j_0}| = 1$ and $x_{1j} = 0$ for all other j; such a constellation is impossible because $x_{1j} > 0$ for all j.

By a computer search including all simple connected graphs with up to 10 vertices (nearly 12 million graphs) the graphs shown in Fig. 3 were found to have minimal σ_1 within a class of fixed n.

All graphs in Fig. 3 consist of a head and a tail such that the head is a complete graph on k vertices (k = 3 or 4), and the tail is a chain of the remaining $n - k$ vertices. For obvious reasons we call graphs of this kind *k-kites*. It is plausible that extremely irregular graphs (those of very low mixedness) are formed from a very complex building block (the head) and a very simple building block (the tail). Choosing the correct size of the head k is not so obvious, and for this

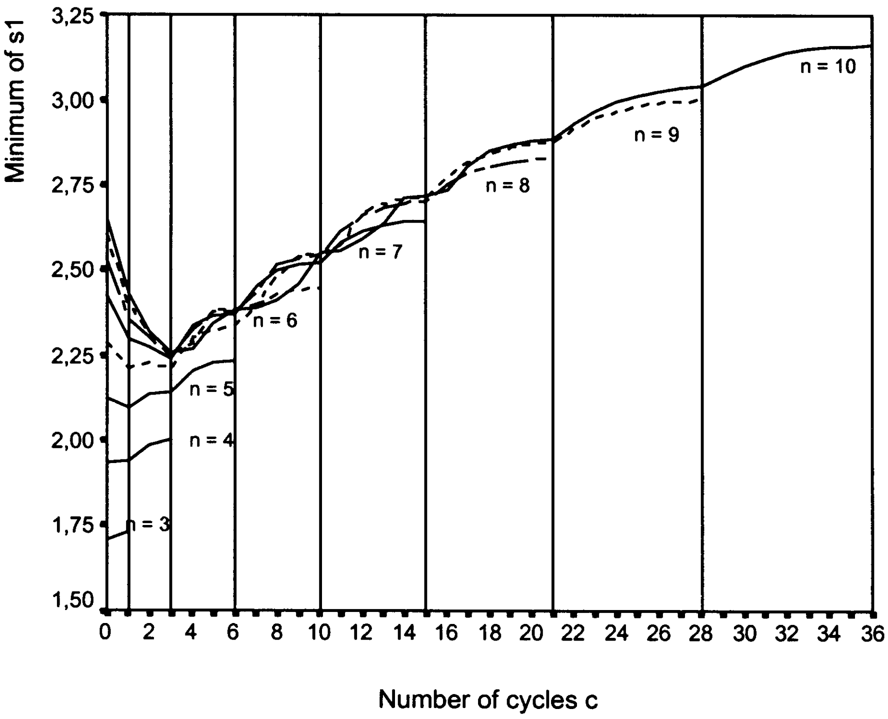

Fig. 4. Minimum of s_1 for given number of vertices n by number of cycles c.

reason we calculated s_1 for several kites of head size between 3 and 8 (see Table 3). Surprisingly, among all kites considered (with the exception of $n \leq 5$), i. e. for n = 6, 7, 8, 9, 10, 11, 12, 20, 50, 99 and 199, the 4-kite was always found to be the one of minimal σ_1.

Further, we empirically examined the minimum of s_1 among simple connected undirected graphs on $n \leq 10$ vertices as a function of the number of cycles c. The above results were confirmed, as illustrated in Figure 4.

For each $n \geq 6$ the minimum value of s_1 is found for c = 3, the cycle number of the 4-kite. For larger c, the minimally mixed graphs tend to have larger s_1 values. The most complex graph in each n-class, i. e. that with maximal number of cycles, is the complete graph K_n with $c = \binom{n-1}{2} = (n-1)(n-2)/2$, which, as a regular graph, clearly has the maximal s_1 value in its class. Note the lack of smoothness of all curves at cycle numbers belonging to complete graphs.

In order to understand the special position of 4-kites we derived approximations for the principal eigenvalue λ_1 and for s_1 of k-kites ($k > 2$) which will be given here without proof.

(Step I) For the principal eigenvalue λ_1 and for $n \to \infty$ the following holds:

$$\lambda_1 \approx \frac{1}{2}\left[k - 3 + (k-1)\frac{\sqrt{k+2}}{\sqrt{k-2}}\right].$$

Approximate values obtained using this formula are very close to the exact ones for n = 12 already, as shown in Table 4. Moreover, we can see that

$$\lim_{k\to\infty} (\lambda_1(k)/(k-1)) = 1,$$

and so λ_1 is of order $k - 1$, as is also seen in Table 4.

(Step II) All vertices within the kite's head are equivalent by symmetry, with the exception of the vertex bearing the tail. Therefore the eigenvector coefficients belonging to these vertices are identical, say x. For high values of n the following approximation holds:

$$x^2 \approx \{k - 1 + (\lambda_1 - k + 2)^2 + [\lambda_1(\lambda_1 - k + 2) - (k-1)]^2\}^{-1}.$$

Table 3. s_1 for kites by vertex number n, head size k, and number of cycles c (row minimum is printed in bold).

n	k: 3 c: 1	4 3	5 6	6 10	7 15	8 21
3	1.73205					
4	**1.93892**	2.00000				
5	**2.08516**	2.13935	2.23607			
6	2.21247	**2.20857**	2.33511	2.44949		
7	2.29803	**2.23865**	2.36947	2.52392	2.64575	
8	2.35839	**2.25064**	2.37959	2.54345	2.70415	2.82843
9	2.39966	**2.25520**	2.38237	2.54781	2.71642	2.87579
10	2.42713	**2.25690**	2.38311	2.54873	2.71863	2.88405
11	2.44504	**2.25753**	2.38330	2.54893	2.71901	2.88531
12	2.45654	**2.25776**	2.38336	2.54897	2.71908	2.88549
20	2.47556	**2.25789**	2.38337	2.54898	2.71909	2.88552
50	2.47598	**2.25789**	2.38337	2.54898	2.71909	2.88552
99	2.47598	**2.25789**	2.38337	2.54898	2.71909	2.88552
199	2.47598	**2.25789**	2.38337	2.54898	2.71909	2.88552
*	2.58230	**2.26434**	2.38470	2.54941	2.71927	2.88561

* Approximated (step IV).

Table 4. λ_1 for kites by vertex number n, head size k, and number of cycles c.

n	k: 3 c: 1	4 3	5 6	6 10	7 15	8 21
3	2.00000					
4	2.17009	3.00000				
5	2.21432	3.08613	4.00000			
6	2.22833	3.09651	4.05137	5.00000		
7	2.23321	3.09787	4.05480	5.03404	6.00000	
8	2.23499	3.09805	4.05503	5.03547	6.02420	7.00000
9	2.23566	3.09807	4.05505	5.03553	6.02490	7.01809
10	2.23591	3.09808	4.05505	5.03553	6.02492	7.01847
11	2.23601	3.09808	4.05505	5.03553	6.02492	7.01848
12	2.23605	3.09808	4.05505	5.03553	6.02492	7.01848
20	2.23607	3.09808	4.05505	5.03553	6.02492	7.01848
50	2.23607	3.09808	4.05505	5.03553	6.02492	7.01848
99	2.23607	3.09808	4.05505	5.03553	6.02492	7.01848
199	2.23607	3.09808	4.05505	5.03553	6.02492	7.01848
*	2.23607	3.09808	4.05505	5.03553	6.02492	7.01848

* Infinity (step I).

(Step III) It can be shown that the principal eigenvector coefficient sum s_1 can be calculated exactly (i. e., by using the exact value of x) as

$$s_1 = \frac{[(k-2)\lambda_1 - 1]x - x_0}{\lambda_1 - 2},$$

where x_0 is the coefficient belonging to the terminal vertex in the chain (maximal distance from the kite's head). For large n, x_0 is nearly zero.

(Step IV) Inserting approximation (I) for λ_1 and approximation (II) for x into (III) and neglecting x_0 leads to an approximation for s_1 for $n \to \infty$.

The s_1 values so estimated are in good accordance with those obtained by computer, at least for $k > 4$ (see Table 3). The foremost result is that s_1 and σ_1 converge for $n \to \infty$, in other words they are independent of n or the tail's length $n-k$. The approximation further predicts that for large values of n (i. e., starting with n = 5), the smallest s_1 or σ_1 is that for k = 4, as was observed. Note that these considerations do not imply any statement on other classes of graphs (non-kites).

We examined the statistical distribution of s_1 for all 853 distinct connected simple graphs with n = 7. As mentioned above, all values were found to lie between 2.23865 (4-kite) and $\sqrt{7}$ = 2.64575 (regular graphs), with mean 2.53611 and standard deviation 0.06845.

Trees of Maximal σ_1

While among cyclic graphs σ_1 is maximal for regular graphs, a tree obviously cannot be regular, with exception of the trivial trees of n = 1 or 2. A candidate for a tree with rather equal distribution of eigenvector coefficients, i. e., of large σ_1, might be the chain (path) graph of n vertices. It is, however, known that in path graphs the interior vertices are associated with large eigenvector coefficients, the exterior ones with very small coefficients [4]. The variance of the coefficients therefore is considerable, the chain's mixedness is low. It was therefore tempting to equalize eigenvector coefficients by attaching short branches at both ends of a chain (Fig. 1, top right). This modification was more successful than initially expected, in that all graphs obtained in this manner (we call them *double snakes*) were found to have no more than two distinct principal eigenvector coefficients each, i. e., their principal eigenvector is

$$(\frac{1}{2}, \frac{1}{2}, 1, 1, ..., 1, 1, \frac{1}{2}, \frac{1}{2})^{\mathrm{T}}$$

with a normalizing factor of $(n-3)^{-0.5}$, associated with the eigenvalue λ_1 = 2. From this we obtain

$$\sigma_1 = \frac{(n-2)^2}{n-3}.$$

From this formula the variance of the eigenvector coefficients can be calculated [4]. It approaches zero for increasing n ("almost all coefficients are equal").

Table 5. s_1 for comets by vertex number n and central vertex degree k (row minimum is printed in bold).

n	k: 2 (n-Alkanes)	3 (Snakes)	4	5	6	7	8
2	1.41421						
3	1.70711						
4	1.94650	**1.93185**					
5	2.15470	2.13099	**2.12132**				
6	2.34190	2.31281	**2.28550**	2.28825			
7	2.51367	2.48138	**2.42522**	2.42540	2.43916		
8	2.67347	2.63927	2.54201	**2.52910**	2.55595	2.57794	
9	2.82360	2.78829	2.63779	**2.60256**	2.63427	2.67920	2.70711
10	2.96569	2.92979	2.71480	**2.65182**	2.68182	2.74010	2.79628
11	3.10095	3.06480	2.77550	**2.68344**	2.70871	2.77256	2.84498
12	3.23032	3.19414	2.82248	**2.70307**	2.72320	2.78860	2.86826
20	4.11807	4.08437	2.94710	**2.73167**	2.73855	2.80249	2.88604
50	6.42751	6.40269	2.95680	**2.73205**	2.73861	2.80252	2.88605
99	9.00242	8.98391	2.95680	**2.73205**	2.73861	2.80252	2.88605
199	12.73213	12.71876	2.95680	**2.73205**	2.73861	2.80252	2.88605
$*$			2.95680	**2.73205**	2.73861	2.80252	2.88605

* Approximated (step IV).

Table 6. λ_1 for comets by vertex number n and central vertex degree k.

n	k: 2 (n-Alkanes)	3 (Snakes)	4	5	6	7	8
2	1.00000						
3	1.41421						
4	1.61803	1.73205					
5	1.73205	1.84776	2.00000				
6	1.80194	1.90211	2.07431	2.23607			
7	1.84776	1.93185	2.10100	2.28825	2.44949		
8	1.87939	1.94986	2.11199	2.30278	2.48849	2.64575	
9	1.90211	1.96157	2.11688	2.30725	2.49721	2.67624	2.82843
10	1.91899	1.96962	2.11917	2.30869	2.49931	2.68190	2.85308
11	1.93185	1.97538	2.12026	2.30917	2.49983	2.68301	2.85697
12	1.94188	1.97964	2.12080	2.30932	2.49996	2.68327	2.85761
20	1.97766	1.99317	2.12132	2.30940	2.50000	2.68328	2.85774
50	1.99621	1.99897	2.12132	2.30940	2.50000	2.68328	2.85774
99	1.99901	1.99974	2.12132	2.30940	2.50000	2.68328	2.85774
199	1.99975	1.99994	2.12132	2.30940	2.50000	2.68328	2.85774
$*$	2.00000	2.00000	2.12132	2.30940	2.50000	2.68328	2.85774

* Infinity (step I).

In this sense double snakes seem to be the "most regular trees", and the double snake of infinite length is the most regular among these.

Trees of Minimal σ_1

Here, as in the case of general graphs, we concentrate on graphs consisting of a heavy head and a long tail. In this case also we first undertook a complete computer search within all trees of a particular n up to n = 20. The result of the search, i. e., the tree with minimal σ_1 in the respective class, is shown in Figure 5. We call trees of this kind, consisting of a star with degree k and a chain, *k-comets*. Table 5 contains s_1 values of several comets, Table 6 the corresponding principal eigenvalues λ_1. The central vertex of the minimal σ_1 comet has degree 4 up to n = 7, from n = 8 onwards the degree is 5, as is seen in Table 5 for n = 11, 12, 20, 50, 99, 199.

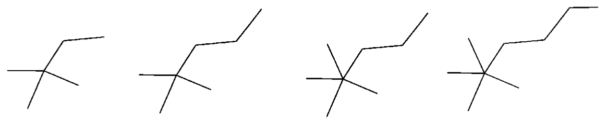

Fig. 5. Trees with minimal σ_1 for fixed n.

As in the case of kites, we derived formulas for comets. Let n be the number of vertices and k the degree of the central vertex.

(Step I) The following approximation of the principal eigenvalue λ_1 holds for $k > 2$ and for $n \to \infty$:

$$\lambda_1 \approx \frac{k-1}{\sqrt{k-2}}. \qquad (1)$$

Exact λ_1 values for the comets depicted in Fig. 5 are found in Table 6.

(Step II) Since the primary vertices in the comet's head are all equivalent by symmetry, all the eigenvector coefficients associated with them have to be identical, say x. By considering the eigenvector equations, for large n a formula including all coefficients of the principal eigenvector can be derived over a geometrical series as a function of x, and thence finally

$$x^2 \approx \frac{k-3}{2(k-1)(k-2)}, \; (k > 3, n \text{ large}). \qquad (2)$$

(Step III) It can be shown that for the principal eigenvector coefficient sum s_1 the following holds:

$$s_1 = \frac{x[(k-2)\lambda_1 - (k-1)] - x_0}{\lambda_1 - 2}, \; (k > 3). \qquad (3)$$

Table 7. Discriminating power of s_1 for trees and alkanes.

n	# Trees	# Distinct s_1 values	Reso-lution	# Alkanes	# Distinct s_1 values	Reso-lution
6	6	6	1	5	5	1
7	11	11	1	9	9	1
8	23	23	1	18	18	1
9	47	47	1	35	35	1
10	106	106	1	75	75	1
11	235	235	1	159	159	1
12	551	550	0.998	355	354	0.997
13	1301	1297	0.997	802	798	0.995
14	3159	3153	0.998	1858	1853	0.997
15	7741	7722	0.998	4347	4332	0.997
16	19320	19257	0.997	10359	10311	0.995
17	48629	48475	0.997	24894	24781	0.995
18	123867	123494	0.997	60523	60262	0.996
19	317955	316953	0.997	148284	147627	0.996
20	823065	820567	0.997	366319	364788	0.996

(Step IV) Inserting approximations (1) and (2) for λ_1 and x, respectively, into (3) and neglecting x_0 yields as limit for s_1 for $n \to \infty$:

$$s_1 = \frac{\sqrt{k-1}\sqrt{k-3}\left[\sqrt{k-2}-1\right]}{\sqrt{2}\left[k-1-2\sqrt{k-2}\right]}, \; (k > 3). \quad (4)$$

These limits are approached rather rapidly for the comets (for n = 50 the error is less than 0.000005). In particular (4) reveals that the minimum s_1 value for $n > 7$ is in fact found for k = 5, again (as for the kites) independently of the tail's length. Thus for $n > 7$ the 5-comet is, somewhat surprisingly, the least mixed comet. (These considerations do not state anything about trees of different type, i. e., non-comets.) 2- and 3-comets (*n*-alkanes and so-called *snakes*) are special cases, in that for their more regular structures their s_1 values are considerably higher than those of other comets.

Discriminating Power of s_1 for Trees and for Alkanes

The discriminating power of s_1 was determined for simple tree graphs and for 4-trees (alkane graphs). Eigenvector coefficients and their sums were calculated as double-precision numbers, for comparison of s_1 values ten decimal places were used, values were compared within each class of constant n. Results are shown in Table 7. First degeneracies appear within trees and alkanes at n = 12 (dodecanes), as is the case with Balaban's index J [12]. However, there are fewer degeneracies for s_1 than for J, e. g. for the 355 topologically distinct alkanes of n = 12 there are 349 distinct J values and 354 distinct s_1 values.

[1] a) N. Trinajstić, Chemical Graph Theory, second edition 1992, Boca Raton, Florida. b) A. T. Balaban, Ed., From Chemical Topology to Three-Dimensional Geometry, New York 1997. c) J. Devillers and A. T. Balaban, Topological Indices and Related Descriptors in QSAR and QSPR, Amsterdam 1999.

[2] a) R. Todeschini and V. Consonni, Handbook of Molecular Descriptors, New York 2000. b) M. V. Diudea, Ed., QSPR / QSAR Studies by Molecular Descriptors, Huntington, N.Y. 2001.

[3] C. Rücker and G. Rücker, J. Chem. Inf. Comput. Sci. **34**, 534 (1994).

[4] I. Gutman, C. Rücker, and G. Rücker, J. Chem. Inf. Comput. Sci. **41**, 739 (2001).

[5] a) D. Bonchev, W. A. Seitz, and E. Gordeeva, J. Chem. Inf. Comput. Sci. **35**, 237 (1995). b) D. Bonchev and L. B. Kier, J. Math. Chem. **9**, 75 (1992). c) D. Bonchev, L. B. Kier, and O. Mekenyan, Int. J. Quantum Chem. **46**, 635 (1993).

[6] G. Rücker and C. Rücker, J. Chem. Inf. Comput. Sci. **33**, 683 (1993).

[7] a) M. Randić, J. Comput. Chem. **1**, 386 (1980). b) M. Randić, W. L. Woodworth, and A. Graovac, Int. J. Quantum Chem. **24**, 435 (1983).

[8] G. Rücker and C. Rücker, J. Chem. Inf. Comput. Sci. **40**, 99 (2000).

[9] S. Nikolić, N. Trinajstić, I. M. Tolić, C. Rücker, and G. Rücker, in D. Bonchev and D. H. Rouvray, Eds., Complexity in Chemistry, Mathematical Chemistry series, vol. 7, in press.

[10] C. Rücker and G. Rücker, J. Math. Chem. **9**, 207 (1992).

[11] D. M. Cvetković and I. Gutman, Croat. Chem. Acta **49**, 115 (1977).

[12] A. T. Balaban, Chem. Phys. Lett. **89**, 399 (1982) .