Vertex topological indices and tree expressions, generalizations of continued fractions

Matthew Hudelson

Abstract   We expand on the work of Hosoya to describe a generalization of continued fractions called "tree expressions." Each rooted tree will be shown to correspond to a unique tree expression which can be evaluated as a rational number (not necessarily in lowest terms) whose numerator is equal to the Hosoya index of the entire tree and whose denominator is equal to the tree with the root deleted. In the development, we use $Z(G)$ to define a natural candidate $\zeta(G, v)$ for a "vertex topological index" which is a value applied to each vertex of a graph, rather than a value assigned to the graph overall. Finally, we generalize the notion of tree expression to "labeled tree expressions" that correspond to labeled trees and show that such expressions can be evaluated as quotients of determinants of matrices that resemble adjacency matrices.

Keywords   Hosoya index · Topological index · Z-index

1 Introduction

In 1971, Hosoya inaugurated the study of topological descriptors of graphs in connection with chemical properties. In his initial work, he demonstrated that the number of ways to partition the set of atoms of an alkane into singletons or bonded pairs—a "matching" of the underlying graph in more modern terminology—is highly correlated to the boiling point of the alkane [3]. Hosoya dubbed the number of matchings $Z(G)$ of the graph G its "topological index" Among Hosoya's early observations is the well-known fact that $Z(P_n) = F_{n+1}$, the $(n + 1)$st Fibonacci number [4]. Hosoya was able to generalize this to computing $Z(G)$ where G is a caterpillar by

M. Hudelson (✉)
Department of Mathematics, Washington State University,
P.O. Box 3113, Pullman, WA 99164, USA
e-mail: mhudelson@wsu.edu

using continuants of continued fractions [5–7]. Hosoya and Gutman's most recent paper [8] gives details on relating $Z(G)$ of an underlying caterpillar to the number K of Kekulé structures. Hosoya and Gutman's paper includes an extensive list of other recent applications of $Z(G)$ which we include here as well [2,9–12]. We also note that Hosoya's use of continuants to enumerate $Z(G)$ for caterpillars exactly mirrors tilings of paths by dominos and stacks of squares detailed by Benjamin and Quinn [1]. Insights such as these that connect continued fractions to enumeration problems have led to the results contained in this paper.

For our note, we begin with the observation (made by Hosoya himself and many others) that Hosoya's index can be computed recursively on a graph G by letting u by any vertex and

$$Z(G) = Z(G - u) + \sum_{v \in N(u)} Z((G - u) - v) \tag{1}$$

where $N(u)$ is the set of vertices adjacent to u. Of course to supply initial conditions for the recurrence, we have $Z(G) = 1$ if G is a graph with no edges; this is consistent with the definition of Z as the number of matchings of G. We then perform the simple step of dividing both sides of Eq. 1 by $Z(G - u)$, obtaining

$$\frac{Z(G)}{Z(G - u)} = 1 + \sum_{v \in N(u)} \frac{Z((G - u) - v)}{Z(G - u)}.$$

We define the "vertex topological index"

$$\zeta(G, v) = \frac{Z(G)}{Z(G - u)}$$

and obtain at once a recurrence for ζ:

$$\zeta(G, u) = 1 + \sum_{v \in N(u)} \frac{1}{\zeta(G - u, v)} \tag{2}$$

with initial condition $\zeta(G, u) = 1$ if u has no neighbors in G.

Hosoya [8] observes that if G has several components $C_1, C_2, \ldots, C_k$ then

$$Z(G) = Z(C_1)Z(C_2) \cdots Z(C_k). \tag{3}$$

This is easy to see as the matchings are independent among the components. From this, we see that if C_1 is the component that contains u, then

$$\begin{aligned} \zeta(G, u) &= \frac{Z(G)}{Z(G - u)} \\ &= \frac{Z(C_1)Z(C_2) \cdots Z(C_k)}{Z(C_1 - u)Z(C_2) \cdots Z(C_k)} \end{aligned}$$

$$= \frac{Z(C_1)}{Z(C_1 - u)}$$
$$= \zeta(C_1, u). \tag{4}$$

2 Ordinary tree expressions

We now restrict our attention to the case where T is a tree with root r. Letting $r_1, r_2, \ldots, r_k$ be the neighbors of r, by deleting r from T, we would obtain a forest each of whose components would be a rooted tree T_i with root r_i. At once, by applying Eqs. 2 and 4 in this situation, we have

Theorem 1 *Given a tree T with root r as well as T_i and r_i as defined above,*

$$\zeta(T, r) = 1 + \sum_i \frac{1}{\zeta(T_i, r_i)}. \tag{5}$$

The initial condition for this recurrence is the single-vertex tree $T = r$ and we already have $\zeta(r, r) = 1$. Alternatively, we could note that r would have no neighbors and the sum in Eq. 4 would be empty.

Any rooted tree (T, r) can now be associated with a "tree expression" $\tau(T, r)$ as generated by the conclusion of Theorem 1. For instance, the tree expression for a five-vertex path P_5 rooted at the central vertex v_3 would be

$$\tau(P_5, v_3) = 1 + \frac{1}{1 + \frac{1}{1}} + \frac{1}{1 + \frac{1}{1}}.$$

More formally, we define a "tree expression" τ and associated rooted tree by the following recursive recipe:

1. The value 1 is a tree expression associated with the single-vertex tree, i.e., $\tau(r, r) = 1$.
2. If $\tau_i, 1 \le i \le k$ are tree expressions then so is $\tau = 1 + \sum_{i=1}^{k} \frac{1}{\tau_i}$. Given that we have associated each rooted tree (T_i, r_i) with τ_i, we associate the rooted tree (T, r) with τ.

We will use the notation $\tau(T, r)$ to represent the tree expression associated with the rooted tree (T, r). Armed with our definition and Theorem 1, we immediately have

Corollary 1 *Given a rooted tree (T, r), $\tau(T, r) = \zeta(T, r)$.*

We would like to conclude that the numerator of $\tau = \tau(T, r)$ yields $Z(T)$, but it's entirely possible that we do not obtain this value if τ is reduced to lowest terms. For instance, if we calculate $\tau(P_5, v_3)$ we obtain the value 2 which does not convey the true value of $Z(P_5) = 8$ or $Z(P_5 - v_3) = 4$. On the other hand, we can circumvent this difficulty simply by not reducing to lowest terms at any time, as we show in the next theorem.

Theorem 2 *If (T, r) is a rooted tree with rooted trees (T_i, r_i) defined as before and if each $\tau(T_i, r_i) = a_i/b_i$ has been evaluated so that $a_i = Z(T_i)$ and $b_i = Z(T_i - r_i)$ then evaluating $\tau(T, r) = a/b$ using the "product common denominator" as opposed to the lowest common denominator yields $a = Z(T)$ and $b = Z(T - r)$.*

Proof We proceed by (strong) mathematical induction on the number of applications of Eq. 5 necessary to complete the evaluation of $\zeta(T, r)$. If this number is zero, then there are no components at all to $T - r$, i.e., $T = r$ is a single vertex. But then $\tau(T, r) = 1 = 1/1$ and $Z(T) = 1$ and $Z(T - r) = 1$ which verifies the basis step. For the inductive step, we assume $a_i = Z(T_i)$ and $b_i = Z(T_i - r_i)$ as in the hypothesis. Now, when faced with evaluating $\tau(T, r) = a/b$, we have

$$
\begin{aligned}
\frac{a}{b} &= 1 + \sum_{i=1}^{k} \frac{1}{\tau(T_i, r_i)} \\
&= 1 + \sum_{i=1}^{k} \frac{b_i}{a_i} \\
&= \frac{a_1 a_2 a_3 \cdots a_k + b_1 a_2 a_3 \cdots a_k + a_1 b_2 a_3 \cdots a_k + \cdots a_1 a_2 a_3 \cdots b_k}{a_1 a_2 a_3 \cdots a_k}
\end{aligned}
$$

We now observe that

$$
\begin{aligned}
b &= a_1 a_2 \cdots a_k \\
&= Z(T_1) Z(T_2) \cdots Z(T_k) \\
&= Z(T - r)
\end{aligned}
$$

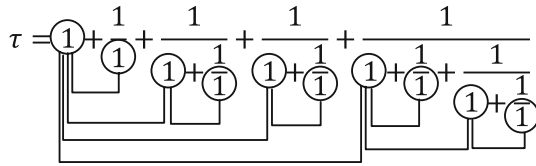which follows from Eq. 3. Finally, we have from Corollary 2 that

$$
\frac{a}{b} = \zeta(T, r) = \frac{Z(T)}{Z(T - r)}
$$

which, together with $b = Z(T - r)$, implies $a = Z(T)$ as desired. $\square$

We can see this in action in our example

$$
\begin{aligned}
\tau(P_5, v_3) &= 1 + \frac{1}{1 + \frac{1}{1}} + \frac{1}{1 + \frac{1}{1}} \\
&= 1 + \frac{1}{2} + \frac{1}{2} \\
&= \frac{4}{4} + \frac{2}{4} + \frac{2}{4} = \frac{8}{4}
\end{aligned}
$$

Fig. 1 The rooted graph (T, v) produced from the leading 1s of τ

We have seen how to go from rooted trees to tree expressions. The reverse direction of constructing a tree from a given tree expression has an intriguing graphical interpretation. Suppose we have a tree expression $\tau = 1 + \sum_i \frac{1}{\tau_i}$. The initial "1" in the expression $1 + \sum_i \frac{1}{\tau_i}$ will be called the "leading 1"; each of the τ_i expressions will in turn have leading 1s. We construct the rooted tree (T, r) recursively. If $\tau = 1$, we declare T to be the single vertex r. Otherwise, we assume that each tree expression τ_i has led to the construction of the rooted tree (T_i, r_i). The leading 1 of τ is declared to be the root r and we join r to each of the vertices r_i, forming the tree T. The construction of (T, r) from τ is precisely the "unraveling" of the construction of τ from (T, r) and so we have a one-to-one correspondence between rooted trees and tree expressions.

One means of picturing the construction of (T, r) from τ is simply to circle all of the leading 1s and then joining any leading 1 to the leading 1s of expressions at the 'next level under the same fraction bar.' This is demonstrated in Fig. 1.

3 Labeled tree expressions

We can generalize tree expressions to be applied to trees whose vertices are assigned numerical labels by replacing the 1s that appear in denominators with the labels. More specifically, suppose (T, v_1) is a rooted tree with vertices $v_0, v_1, \ldots, v_n$ and vertex i has label x_i. We then define its *labeled tree expression* $\lambda(T, v_0)$ by the recursive recipe

1. If the vertex set of T is $\{v_0\}$, then $\lambda(T, v_0) = x_0$.
2. If the root is adjacent to vertices $v_{i_1}, v_{i_2}, \ldots, v_{i_k}$ and that C_j is the component of $T - v_0$ that contains v_{i_j}, then $\lambda(T, v_0) = x_0 + \sum_j 1/\lambda(C_j, v_{i_j})$.

In the case of a path $P = v_0 v_1 \ldots v_n$ rooted at v_0, we obtain as its labeled tree expression the continued fraction

$$\lambda(P, v_0) = x_0 + \cfrac{1}{x_1 + \cfrac{1}{x_2 + \cfrac{1}{\ddots \cfrac{}{x_{n-1} + \cfrac{1}{x_n}}}}}.$$

Hosoya and Gutman [8] point out that

$$\lambda(P, v_0) = \frac{\det M(x_0, x_1, x_2, \ldots, x_n)}{\det M(x_1, x_2, \ldots, x_n)}$$

where

$$M(d_1, d_2, d_3, \ldots, d_k) = \begin{pmatrix} d_1 & 1 & & & & & \\ -1 & d_2 & 1 & & & & \\ & -1 & d_3 & 1 & & & \\ & & -1 & d_4 & & & \\ & & & & \ddots & & \\ & & & & & d_{k-1} & 1 \\ & & & & & -1 & d_k \end{pmatrix}$$

denotes the $k \times k$ tridiagonal matrix whose diagonal entries are $d_1, \ldots, d_k$, whose entries just above the diagonal are all 1 and whose entries just below the diagonal are all -1.

We show that in fact this observation can be generalized to labeled tree expressions. If G is a labeled graph where vertex v_r is assigned label x_r, we define the matrix $B(G)$ by the formula

$$B(G)_{ij} = \begin{cases} x_i, & i = j; \\ 1, & i < j, v_i \text{ adjacent to } v_j; \\ -1, & i > j, v_i \text{ adjacent to } v_j; \\ 0, & i \neq j, v_i \text{ not adjacent to } v_j. \end{cases}$$

In short, this matrix has the vertex labels along the diagonal, takes on the values of $A(G)$, the adjacency matrix of G above the diagonal and takes on the negatives of the values of $A(G)$ below the diagonal.

As an example of this more general situation, if (T, v_0) is the rooted tree that produces the tree expression τ in Fig. 1,

$$\tau(T, v_0) = 1 + \frac{1}{1} + \frac{1}{1 + \dfrac{1}{1}} + \frac{1}{1 + \dfrac{1}{1}} + \frac{1}{1 + \dfrac{1}{1} + \dfrac{1}{1 + \dfrac{1}{1}}},$$

then

$$\lambda(T, v_0) = x_0 + \frac{1}{x_1} + \frac{1}{x_2 + \dfrac{1}{x_3}} + \frac{1}{x_4 + \dfrac{1}{x_5}} + \frac{1}{x_6 + \dfrac{1}{x_7} + \dfrac{1}{x_8 + \dfrac{1}{x_9}}},$$

and (suppressing the zero elements),

$$B(T) = \begin{pmatrix} x_0 & 1 & 1 & & 1 & & 1 & & & \\ -1 & x_1 & & & & & & & & \\ -1 & & x_2 & 1 & & & & & & \\ & & -1 & x_3 & & & & & & \\ -1 & & & & x_4 & 1 & & & & \\ & & & & -1 & x_5 & & & & \\ -1 & & & & & & x_6 & 1 & 1 & \\ & & & & & & -1 & x_7 & & \\ & & & & & & -1 & & x_8 & 1 \\ & & & & & & & & -1 & x_9 \end{pmatrix}$$

We make some preliminary observations that will assist in the proof of our next main result that generalizes the Hosoya and Gutman observation, linking $\lambda(T, r)$ and the determinant of B-matrices. Given a tree T, we observe that if we expand the determinant of $B(T)$ by the usual signed permutation sum

$$\det B(T) = \sum_{\sigma \in S_n} (-1)^{\operatorname{sgn}(\sigma)} \prod_{i=1}^{n} (B(T))_{i,\sigma(i)}$$

then an x_j value occurs in a term if and only if $\sigma(j) = j$. Also, since T has no cycles of length 3 or more, no permutation with a cycle of length 3 or more can make a nonzero contribution to $\det B(T)$. Thus, if $B(T)_{j,k}$ is in any term $\prod_{i=1}^{n}(B(T))_{i,\sigma(i)}$ that makes a nonzero contribution to $\det B(T)$ then so is $B(T)_{k,j}$.

We also note that if $G - v$ has components $C_1, C_2, \ldots, C_k$ then there is a permutation matrix P such that $P^{-1}B(G - v)P$ is a block diagonal matrix, each block being of the form $B(C_i)$. This is tantamount to renumbering the vertices so that for each $1 \leq i \leq k-1$, the vertices of component C_i have numbers below those of C_{i+1}. Thus, $\det B(G - v) = \prod_{i=1}^{k} \det B(C_i)$.

We are now ready to prove our result.

Theorem 3 *If (T, r) is a labeled tree then*

$$\lambda(T, r) = \frac{\det B(T)}{\det B(T - r)}$$

where $B(T - r)$ is the specific matrix obtained by removing the row and column corresponding to r. We adopt the convention that $\det B(T - r) = 1$ to make sense of the case when T is the graph with the single vertex r.

Proof The proof is by (strong) induction on the number of vertices n in T. The basis step is when T has the single vertex r with label x. In this case, $\lambda(T, r) = x$ and $B(T) = [x]$ by definition. With the convention that $\det B(T - r) = 1$, the basis step is easily verified. For the inductive step, suppose the vertices are $v_0 = r, v_1, \ldots, v_{n-1}$. Suppose further that v_i is labeled x_i and suppose $r = v_0$ is adjacent to vertices $v_{i_1}, v_{i_2}, \ldots, v_{i_k}$ where v_i is the root of component C_i of the graph $T - r$. If we expand

det $B(T)$ along "row 0", corresponding to $r = v_0$, we obtain

$$\det B(T) = x_0 \det B(T - r) + \sum_j \det B(T - r - v_{i_j}).$$

The second expression arises from each occurrence of a $B(T)_{0,i_j} = 1$. From our preliminary observations, $B(T)_{i_j,0} = -1$ must also appear in the term. These two terms form a transposition and so contribute a second factor of -1 to the product forming this term in the determinant; the remaining factor is $\det B(T - r - v_{i_j})$.

We now divide each side of our formula for $\det B(T)$ by $\det B(T - r)$, obtaining

$$\frac{\det B(T)}{\det B(T - r)} = x_0 + \sum_j \frac{\det B(T - r - v_{i_j})}{\det B(T - r)}.$$

Finally, if C_j is the component of $T - r$ that contains v_{i_j} then by our preliminary observations,

$$\frac{\det B(T - r - v_{i_j})}{\det B(T - r)} = \frac{\det B(C_j - v_{i_j})}{\det B(C_j)}$$

$$= \frac{1}{\lambda(C_j, v_{i_j})}$$

by the inductive hypothesis. Therefore,

$$\frac{\det B(T)}{\det B(T - r)} = x_0 + \sum_j \frac{1}{\lambda(C_j, v_{i_j})} = \lambda(T, r)$$

by the recursive definition of $\lambda(T, r)$ and we are done.   □

We finish with the following result that links $\det B(G)$ to matchings in the more general case when G is a graph with no even cycles.

Theorem 4 *Let G be a graph with no even cycles and let $P(G) = \det B(G)$, treated as a polynomial in the x_i. Then the coefficient of $x_{i_1} x_{i_2} \cdots x_{i_k}$ of $P(G)$ enumerates the perfect matchings of the graph G' induced by the vertices whose indices are not in $\{i_1, i_2, \ldots, i_k\}$.*

Proof Again, we analyze the determinant by means of the usual formula involving permutations $\sigma \in S_n$:

$$P(G) = \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma)} \prod_{i=1}^{n} (B(G))_{i,\sigma(i)}. \qquad (6)$$

The only permutations that contribute to the desired coefficient are those that have each of $i_1, i_2, \ldots, i_k$ as their only fixed points. Let σ be such a permutation and suppose that σ has a cycle of length at least 3. Let $C = (a_1, a_2, \ldots, a_q)$ be the cycle

of any length at least 3 with the lowest entry such that each $(B(G))_{a_i,a_{i+1}}$ along with $(B(G))_{a_q,a_1}$ is either 1 or -1. Since G has no even cycles, q must be odd.

Since q is odd, the number of the values $(B(G))_{a_i,a_{i+1}}$ or $(B(G))_{a_q,a_1}$ that equal -1 must have the opposite parity of the number of values that equal 1. If we let σ' be the permutation where C is reversed but otherwise $\sigma'(i) = \sigma(i)$, then we are interested in the values $(B(G))_{a_{i+1},a_i}$ along with $(B(G))_{a_1,a_q}$. Here, the parities of the number of -1 values and the number of 1 values are reversed. We conclude that the contributions of σ and σ' to $P(G)$ eliminate each other. Such permutations come in pairs (thanks to the definition of C) and so the only permutations that can contribute to the desired coefficient without being cancelled by another permutation are those that fix $i_1, i_2, \ldots, i_k$ and transpose all of the remaining entries in pairs. Each transposition (a, b), $a < b$, contributes to the product in Eq. 1 one factor of -1 in $(-1)^{\text{sign}(\sigma)}$, a factor of 1 from $(B(G))_{a,b}$, and a factor of -1 from $(B(G))_{b,a}$. In summary, each $(B(G))_{i_j,i_j} = x_{i_j}$ and each transposition contributes two factors of -1 to the product, and so the product $(-1)^{\text{sign}(\sigma)} \prod_{i=1}^{n} (B(G))_{i,\sigma(i)} = x_{i_1} x_{i_2} \ldots x_{i_k}$. Finally, each transposition corresponds to an edge in G' and so σ corresponds to a perfect matching of G', which yields the result. $\square$

We close by noting the following corollaries.

Corollary 2 *Given a graph G with no even cycles (this includes trees), the Hosoya index $Z(G)$ is $P(G)(1, 1, \ldots, 1)$ and the number of perfect matchings is $P(G)$ $(0, 0, \ldots, 0)$.*

Corollary 3 *Given a graph G with no even cycles and a vertex v,*

$$\zeta(G, v) = \frac{P(G)(1, 1, \ldots, 1)}{P(G - v)(1, 1, \ldots, 1)}.$$

For such graphs, this provides a polynomial-time formula for computing $\zeta(G, v)$.

References

1. A. Benjamin, J. Quinn, *Proofs that Really Count. The Art of Combinatorial Proof* (Mathematical Association of America, Washington, DC, 2003)
2. G.G. Cash, J. Math. Chem. **37**, 117–125 (2005)
3. H. Hosoya, Bull. Chem. Soc. Jpn. **44**, 2332–2339 (1971)
4. H. Hosoya, Fibonacci Q. **11**, 255–266 (1973)
5. H. Hosoya, Continuant, caterpillar, and topological index Z. Fastest algorithm for degrading a continued fraction. Nat. Sci. Rep. Ochanomizu Univ. **58**(1), 15–28 (2007)
6. H. Hosoya, Important mathematical structures of the topological index Z for tree graphs. J. Chem. Inf. Model. **47**, 744–750 (2007)
7. H. Hosoya, Continuant, caterpillar, and topological index Z. Novel identities involving Fibonacci, Lucas, and generalized Fibonacci numbers. Nat. Sci. Rep. Ochanomizu Univ. **58**(2), 11–20 (2008)

8. H. Hosoya, I. Gutman, Kekulé structures of hexagonal chains—some unusual connections. J. Math. Chem. **44**, 559–568 (2008)
9. H. Liu, X. Yan, Z. Yan, M.J. Zhou, MATCH Commun. Math. Comput. Chem. **57**, 235 (2007)
10. X. Pan, C. Yang, M.J. Zhou, MATCH Commun. Math. Comput. Chem. **57**, 371 (2007)
11. S.G. Wagner, MATCH Commun. Math. Comput. Chem. **57**, 221 (2007)
12. A. Yu, X. Lv, The Merrifield–Simmons indices and Hosoya indices of trees with k pendant vertices. J. Math. Chem. **41**, 33–43 (2007)