**MATCH**
*Communications in Mathematical and in Computer Chemistry*

*MATCH Commun. Math. Comput. Chem.* **93** (2025) 499–507
**ISSN:** 0340–6253
**doi:** 10.46793/match.92-2.499P

# A New Probabilistic Molecular Index

**José Luis Palacios**[a,*]

[a] *Department of Electrical and Computer Engineering. The University of New Mexico. Albuquerque, NM 87131, USA*

jpalacios@unm.edu

(Received May 31, 2024)

**Abstract**

We propose a new molecular index based on a distance on graphs defined through hitting times of random walks on connected graphs. We show its connections to previous probabilistic/electric indices such as the RW index and the Kirchhoff index, compute its value for some families of graphs, and present some open questions.

## 1 Introduction

In what follows, a graph $G = (V, E)$ will be a finite simple connected undirected graph with vertex set $V = \{1, 2, \ldots, n\}$, edge set $E$ and vertex degrees $d_1, d_2, \ldots, d_n$. For all graph theoretical details the reader may consult reference [2].

In Mathematical Chemistry, molecules are modeled using these graphs, where the vertices are the atoms and the atomic bonds are represented by the edges. Many topological indices, or descriptors, i. e., real-valued functions on the domain of all graphs, have been defined with the purpose of capturing physico-chemical properties of the molecules and classifying them according to the values of their indices. One such index is the Kirch-

*Corresponding author.

hoff index defined in [6] as

$$K(G) = \sum_{i<j} R_{ij}, \tag{1}$$

where $R_{ij}$ is the effective resistance between vertices $i$ and $j$ when the graph is thought of as an electrical network, where all the edges have unit resistance.

The simple random walk on $G$ is defined as the Markov chain $\{X_n, n \geq 0\}$ whose state space is $V$ and whose transition probabilities are defined as uniform, from a vertex $i$ to any of its $d_i$ neighboring vertices. The hitting time $T_j$ of the vertex $j$ is defined as the smallest number of jumps needed by the random walk to reach the vertex $j$:

$$T_j = \inf\{n \geq 0 : X_n = j\},$$

and its expected value, when the process is started in state $i$ is denoted by $E_iT_j$. We remark that $E_iT_i = 0$, and this should not be confused with the mean return time to vertex $i$, $E_iT_i^+ = \frac{2|E|}{d_i}$, which involves $T_i^+ = \inf\{n \geq 1 : X_n = i\}$. For facts about hitting times of Markov chains, the reader is referred to [4].

In [7] we showed that there is a close relationship between hitting times and the Kirchhoff index, namely

$$K(G) = \frac{1}{2|E|} \sum_{i<j} (E_iT_j + E_jT_i), \tag{2}$$

so that one can use probabilistic tools and intuitions to this index, in addition to several other fruitful approaches. A good introduction to the relationship between electric networks and random walks on graphs is reference [5].

A recent probabilistic/electrical index was put forward in [3] by Camby et al., the random walk index, defined in the following way: for any pair of vertices $i$ and $j$, a battery is placed between $i$ and $j$ so that a 1 ampere current enters $i$ and exits $j$. This generates a voltage $v_x^{ij}$ on all vertices $x \in V$, and a potential difference on any edge $(x, y)$ given by $v_x^{ij} - v_y^{ij}$.

If the polarity of the battery is inverted, then the potential drop on the edge $(x, y)$ is $v_y^{ij} - v_x^{ij}$, and thus, in order to avoid the dependance on the polarity of the battery, the authors consider the quantity $|v_x^{ij} - v_y^{ij}|$, and they add these quantities over all edges of the graph getting

$$\hat{d}_{ij} = \sum_{(x,y)\in E} |v_x^{ij} - v_y^{ij}|.$$

The authors prove that the function $\hat{d}$ defined on the pairs of vertices $ij$ by the value $\hat{d}_{ij}$ is a metric, and then define the random walk index as

$$RW(G) = \sum_{i<j} \hat{d}_{ij}. \tag{3}$$

Below we will show a relationship between $RW(G)$ and our new probabilistic index.

# 2 The results

We define $HT(G)$, the hitting time index of a graph $G$, as

$$HT(G) = \sum_{i<j} D(i, j),$$

where $D(i, j) = \max\{E_i T_j, E_j T_i\}$. A salient feature of this index is the following

**Proposition 1.** *The function $D$ is a distance on $G$.*

*Proof.* That $D(i, j) \geq 0$ and $D(i, j) = 0$ if and only if $i = j$ is trivial. The triangular inequality perhaps need some explanation. It is clear that for any vertices $i, j, k$ we have

$$E_i T_j \leq E_i T_k + E_k T_j, \tag{4}$$

because when going from $i$ to $j$, we have in general more paths to perform that journey than those that go first to the intermediate vertex $k$, and then proceed to $j$, so the left side of (4) is bounded above by the expected length of the first journey to $k$, plus the expected length of the second journey from $k$ to $j$. Then, clearly

$$E_iT_j \leq D(i,k) + D(k,j), \tag{5}$$

and exchanging the roles of $i$ and $j$

$$E_jT_i \leq D(j,k) + D(k,i) = D(i,k) + D(k,j). \tag{6}$$

Now (5) and (6) imply

$$D(i,j) = \max\{E_iT_j, E_jT_i\} \leq D(i,k) + D(k,j)$$

which completes the proof. ■

The connection between this new index and those with a probabilistic/electric flavor is given in the next

**Proposition 2.** *For any $G$ we have*

$$RW(G) \leq HT(G) \leq 2|E|K(G). \tag{7}$$

*Proof.* The right inequality is easy to prove: $D(i,j) = \max\{E_iT_j, E_jT_i\} \leq E_iT_j + E_jT_i$. Then we use (2).

For the left inequality, if we set a battery between vertices $i$ and $j$ so that a current of 1 ampere flows between $i$ and $j$, a voltage $v_k^{ij}$ is established at every node $k$ of $V$. We found in [8] an expression for the hitting times of random walks on weighted graphs that in the case of simple graphs reduces to

$$\sum_{k\in V} d_k v_k^{ij} = E_iT_j, \tag{8}$$

and it is immediate to notice that this can be rewritten as

$$\sum_{(x,y)\in E} (v_x^{ij} + v_y^{ij}) = E_iT_j.$$

Now

$$RW(G) = \sum_{i<j} \sum_{(x,y)\in E} |v_x^{ij} - v_y^{ij}| \le \sum_{i<j} \sum_{(x,y)\in E} (v_x^{ij} + v_y^{ij})$$

$$= \sum_{i<j} E_i T_j \le \sum_{i<j} D(i,j) = HT(G) \qquad \blacksquare$$

In [9] we studied highly symmetric graphs for which

$$E_i T_j = E_j T_i, \tag{9}$$

for all $i, j \in G$. It turns out that the family of walk-regular graphs satisfies (9). A graph is walk-regular if the number of $k$-long walks, $k \ge 2$, starting and ending at a vertex $v$ is the same for all $v \in V$. This family contains the families of vertex-transitive, regular edge-transitive and distance regular graphs. For all these graphs we have the following improvement of the right inequality in (7):

**Proposition 3.** *For any walk-regular $G$ we have*

$$HT(G) = |E| K(G). \tag{10}$$

*Proof.* This is due to the fact that $D(i,j) = \max\{E_i T_j, E_j T_i\} = \frac{1}{2}(E_i T_j + E_j T_i)$ . $\blacksquare$

The classical Wiener index (see [10]) is given by the sum of the distances between all vertices. More formally, the distance $d(i,j)$ between vertices $i$ and $j$ is the length of a shortest walk between $i$ and $j$, and the Wiener index is defined as

$$W(G) = \sum_{i<j} d(i,j).$$

The following is a simple observation that yields an interesting relationship

**Proposition 4.** *For any $G$ we have*

$$W(G) \le HT(G) \ . \tag{11}$$

*Proof.* The number of jumps needed for the random walk to reach $j$ starting from $i$ is bounded below by $d(i,j)$. Therefore, so is its expected value, i.e., $E_iT_j \geq d(i,j)$. The same holds for $E_jT_i$. Thus $D(i,j) \geq d(i,j)$. ■

In fact, in [3] they prove something stronger than (11) that, together with proposition 3, can be summarized in the following

**Corollary.** *For every $G$ we have*

$$K(G) \leq W(G) \leq RW(G) \leq HT(G) \leq 2|E|K(G).$$

The relationship between $HT(G)$ and $K(G)$ can be made more precise with the following lower bound

**Proposition 5.** *For any $G$ we have*

$$|E|K(G) \leq HT(G). \tag{12}$$

*Proof.* Use the fact that $D(i,j) = \max\{E_iT_j, E_jT_i\} \geq \frac{1}{2}(E_iT_j + E_jT_i)$. ■

# 3 Some computations and conjectures

Using proposition 3 we have that

$$HT(K_n) = \frac{1}{2}(n-1)^2 n,$$

and

$$HT(C_n) = \frac{1}{6}n^2(n^2-1).$$

Now we will find the $HT$ index of the star graph $S_n$. First, we place a battery between the center vertex $c$ and any leaf vertex $v$ that sends a 1 ampere current from $c$ to $v$. Since the resistance of the edge $cv$ is 1, then the voltage at $c$ and any other leaf other than $v$ is 1, while the voltage at $v$ is 0. Then, using (8),

$$E_cT_v = (n-2) + (n-1) = 2n-3,$$

where the first parenthesis corresponds to $n-2$ vertices with degree 1, and the second parenthesis corresponds to $c$, whose degree is $n-1$. Since obviously $E_vT_c = 1$, we have that $D(c,v) = 2n-3$.

Also, if we place a battery sending a current of 1 ampere between any two leaves $v$ and $w$, since the effective resistance between these two vertices is equal to 2, then we must have that the voltages are equal to 2 in the vertex $v$, equal to 1 in the vertex $c$ (and in all leaves other than $v$ and $w$) and equal to 0 in the vertex $w$. Using (8) we conclude that

$$E_vT_w = 2 + (n-3) + (n-1) = 2(n-1).$$

Finally we get

$$HT(S_n) = (n-1)(2n-3) + \binom{n-1}{2}2(n-1) = (n-1)(n^2-n-1).$$

The computation of $HT(P_n)$, for the path graph $P_n$ on $n$ vertices, is more involved. First we compute $E_kT_{k+1}$, for any $1 \leq k \leq n-1$, by placing a battery between the vertices $k$ and $k+1$ so that the current flowing from $k$ to $k+1$ is equal to 1. Then it is clear that all vertices prior to and including 1 have voltage equal to 1, and all those following and including $k+1$ have voltage 0. Therefore, by (8) we get

$$E_kT_{k+1} = 2k-1.$$

And thus, for $i<j$, we have

$$E_iT_j = \sum_{k=i}^{j-1}(2k-1) = (i+j-2)(j-i). \tag{13}$$

A similar argument for hitting times in the opposite direction yields

$$E_jT_i = (2n-i-j)(j-i). \tag{14}$$

Then, in view of (13) and (14) we have that $E_iT_j \geq E_jT_i$, i.e., $D(i,j) =$

$E_i T_j$, if and only if $i + j \geq n + 1$. This fact allows us to write, finally that

$$HT(P_n) = \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \sum_{m=k+1}^{n-k+1} E_m T_k + \sum_{k=\lfloor \frac{n}{2} \rfloor + 2}^{n} \sum_{m=n-k+2}^{k-1} E_m T_k.$$

Thus, for example, $HT(K_6) = 75$, $HT(S_6) = 145$, $HT(C_6) = 210$, and $HT(P_6) = 223$. We should point out that, as opposed to the $RW$ index, our $HT$ index discriminates between the graphs $C_n$ and $P_n$ and between $S_n$ and $K_n$.

Inequalities (7) and (12) imply that $HT(G) = \Theta(|E|K(G))$, though this fact does not help in finding graphs that maximize and minimize $HT(G)$ because the interplay between $|E|$ and $K(G)$ is not simple: the largest values for $|E|$ and $K(G)$ are $\Theta(n^2)$ and $\Theta(n^3)$ respectively, but they don't occur at the same graph; similarly, the smallest values for $|E|$ and $K(G)$ are both linear, but there is no graph where these occur simultaneously. The complete graph $K_n$ has the smallest value for $K(G) = n - 1$, and the largest value for $|E|$; also, for all $i, j$, it seems to have the smallest values for $D(i, j) = n - 1$, and so, even though its $|E|$-value is large, it seems to be a good candidate to minimize $HT(G)$. On the other hand, the path $P_n$ maximizes $K(G)$, though its $|E|$-value is only linear, so that $HT(P_n) = \Theta(n^4)$, with the constant of the leading term equal to $\frac{1}{6}$, and this order of magnitude seems to be the largest possible for the $HT$ index.

If we turn our attention at the largest values for hitting times, these occur for the lollipop graph, i.e., a complete graph on roughly $\frac{2n}{3}$ vertices, attached to a path made with the remaining vertices, that attains the maximal value $\frac{4}{27}n^3$ from any vertex of the complete part to the endpoint of the path, as shown in [1], and since this occurs for about two thirds of the vertices, the largest $HT$ can be for a lollipop graph is also $\Theta(n^4)$, but with a constant smaller than that of $HT(P_n)$.

By brute force, it can be verified for $n \leq 6$ that the graphs that maximize and minimize $HT(G)$ are indeed, $P_n$ and $K_n$, respectively. It would nice to settle these conjectures for general $n$.

# References


[1] G. Brightwell, P. Winkler, Maximum hitting time for random walks on graphs, *Rand. Struct. Algor.* **1** (1990) 263–276.

[2] F. Buckley, F. Harary, *Distance in Graphs*, Addison-Wesley, Redwood City, 1990.

[3] E. Camby, G. Caporossi, M. H. M. Paiva, M. E. V. Segatto, Expected distance based on random walks, *J. Math. Chem.* **56** (2018) 618–629.

[4] P. G. Doyle, J. L. Snell, *Random Walks and Electrical Networks*, Math. Assoc. America, Washington, 1984.

[5] M. Grinstead, J. L. Snell, *Introduction to Probability*, AMS, Providence, 1997.

[6] D. J. Klein, M. Randić, Resistance distance, *J. Math Chem.* **12** (1993) 81–95.

[7] J. L. Palacios, Resistance distance in graphs and random walks, *Int. J. Quantum Chem.* **81** (2001) 29–33.

[8] J. L. Palacios, E. Gómez, M. Del Río, Hitting times of walks on graphs through voltages, *J. Prob.* **2014** (2014) #852481.

[9] J. L. Palacios, G. Markowsky, Kemeny's constant and the Kirchhoff index for the cluster of highly symmetric graphs, *Appl. Math. Comput.* **406** (2021) #126283.

[10] H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.