Hitting Times, Cover Cost, and the Wiener Index of a Tree

Agelos Georgakopoulos∗ Stephan Wagner
October 19, 2012

Abstract

It is well known that the vertices of any graph can be put in a linear preorder so that for random walk on the graph vertices appearing earlier in the preorder are “easier to reach but more difficult to get out of”. We exhibit further such preorders corresponding to various functions related to random walk and study their relationships. These preorders coincide when the graph is a tree, but not necessarily otherwise.

In the case of trees, we prove a simple identity relating hitting times to the Wiener index.

1 Introduction

The Wiener index of a tree $T$ is the sum of the distances of all pairs of vertices of $T$. It has been extensively studied and has found applications in chemistry, communication theory and elsewhere, see [DEG01] and references therein.

The hitting time $H_{rv}$ is the expected number of steps it takes a random walk on $T$ to go from a vertex $r$ to a vertex $v$. We prove a rather surprising connection between the Wiener index and hitting times:

**Theorem 1.** For every tree $T$ and every $r \in V(T)$, we have

$$\sum_{v \in V(T)} (H_{rv} + d(r, v)) = 2W(T) := \sum_{x,y \in V(T)} d(x, y).$$

The sum of the left hand side is dominated by the first subsum $C_r(T) := \sum_{v \in V(T)} H_{rv}$. In [Geo] this sum is dubbed the Cover Cost of $T$, and it is argued that it is related to the Cover Time of a graph, i.e. the expected time for a random walk from $r$ to visit all vertices. Defining the barycentricity $D(r) := \sum_{w \in V(T)} d(r,w)$, the above formula can be rewritten in the following concise form

$$CC(r) + D(r) = 2W(T)$$  \hspace{1cm} (1)

The barycentricity $D(r)$ has been the object of study of several papers in optimization theory []. It is computable in linear time (by a straightforward breadth-first or depth-first search). The same is true for the Wiener index [Dan93], and so we deduce that Cover Cost is computable in linear time.

Similarly to $CC(r)$ we define the the reverse cover cost $RC(r) = \sum_{y \in V(G)} H_{yr}$. We show that

**Theorem 2.** For every tree $T$, the quantity

$$RC(r) + (2n - 1)CC(r) = 4(n - 1)W(T),$$

where $n$ is the number of vertices of $T$, is independent of $r$. Thus a vertex $r$ that maximizes $CC(r)$ minimizes $RC(r)$ and vice versa.

∗Supported by FWF Grant P-24028-N18.
It is well known that the vertices of any graph can be put in a linear preorder \( \leq \) such that vertices appearing earlier in the order are “easier to reach but difficult to get out of”, while vertices appearing later behave the other way round \( [CTW93, \text{Lov93}] \); more precisely, whenever \( x \leq y \) we have \( H_{xy} \leq H_{yx} \). Our next result shows that if the graph is a tree, then this ordering coincides with that of the values of \( RC(x) \), the ordering of the values of \( CC(x) \) reversed, as well as the orderings induced by further functions. We define the weighted barycentricity \( D_\pi \) by \( D_\pi(x) := \sum_{w \in V(T)} d(w)d(r, w) \) and the weighted reverse cover cost \( RC_\pi(x) \) by \( RC_\pi(x) := \sum_{z \in V(T)} d(z)H_{zx} \).

**Corollary 1.1.** For every tree \( T \), and every pair of vertices \( x, y \in V(T) \), the following are equivalent

(i) \( D(x) \leq D(y) \);
(ii) \( D_\pi(x) \leq D_\pi(y) \);
(iii) \( H_{yx} \leq H_{xy} \);
(iv) \( RC_\pi(x) \leq RC_\pi(y) \);
(v) \( RC(x) \leq RC(y) \);
(vi) \( CC(x) \geq CC(y) \).

The equivalence of (i) to (vi) is an easy combinatorial observation, see Section 3. The equivalence of (iii) and (iv) has been proved by Beveridge [Bev09, Proposition 1.1] asserts a weaker statement, but the same proof applies. It is also proved in [CTW93, Lov93] that the vertex minimising \( D \) also minimises \( RC_\pi \).

The above results can be interpreted as follows: there is a function \( D : V(T) \to \mathbb{R}_+ \) the values of which determines the (ordering of the) values of the functions \( CC \) and \( RC \); by Corollary 1.1 knowledge of \( D \) suffices to determine the ordering of the values of \( CC \) and \( RC \), while by Theorems 1 and 2 if in addition we know the constant \( W(T) \) then the exact values of \( CC \) and \( RC \) are determined.

Theorem 3. For every graph \( G \), and every vertex \( x \in V(G) \), we have

\[
CC(x) = mR(x) - \frac{n}{2} R_\pi(x) + K_\pi(G),
\]
\[
RC(x) = mR(x) + \frac{n}{2} R_\pi(x) - K_\pi(G),
\]
\[
RC_\pi(x) = 2mR_\pi(x) - K_{\pi^2}(G), \quad \text{and}
\]
\[
CC_\pi(x) = K_{\pi^2}(G).
\]

\footnote{Beveridge [Bev99, Proposition 1.1] asserts a weaker statement, but the same proof applies. It is also proved there that the vertex minimising \( D \) also minimises \( RC_\pi \).}
Note that unlike trees, the three orderings according to $CC$, $RC$ and $RC_\pi$ are determined by three different functions, namely the functions $D_1 = 2mR(x) - nR_\pi(x)$, $D_2 = 2mR(x) + nR_\pi(x)$ and $D_3 = R_\pi(x)$ respectively (all of which are themselves determined by the two functions $R$ and $R_\pi$). This does not apriori mean that these orderings are different, since there is strong dependence between these functions. We will however construct examples showing that no two of these orderings always coincide.

The fact that $CC_\pi(x)$ is constant was already known; moreover, it can be expressed in terms of the eigenvalues of the matrix $M$ of transition probabilities of $G$ ($m_{ij} = 1/d_i$ if $ij \in E(G)$ and 0 otherwise) as $CC_\pi(x) = 2m \sum_{k=2}^{n} \frac{1}{1 - \lambda_k}$ [Lov93, Formula 3.3.]. Combined with Theorem 3, this yields

$$K_\pi^2(G) = 2m \sum_{k=2}^{n} \frac{1}{1 - \lambda_k}.$$  

Interestingly, a similar formula applies to the Kirchhoff index:

$$K(G) = n \sum_{\lambda \neq 0} \frac{1}{\lambda},$$

the sum being over all nonzero Laplacian eigenvalues of $G$, see [DEG01, Mer89, Moh91].

Finally, using Theorem 1 we are able to find the extremal rooted trees for the Cover Cost $CC(r)$: in Section 3.1 we prove that, for a fixed number of vertices, $CC(r)$ is minimised by the star rooted at a leaf, and maximised by the path rooted at a midpoint. It turns out that the same rooted trees are extremal for Cover Time as well, by theorems of Brightwell & Winkler [BW90] and Feige [Fei97] respectively. Moreover, the same trees are extremal also for the Wiener index [1].

2 Preliminaries

The main tool we will use is the following formula of Tetali [Tet91], expressing hitting times in terms of effective resistances.

$$H_{xy} = \frac{1}{2} \sum_{w \in V(G)} d(w)(r(x,y) + r(w,y) - r(w,x)).$$  (2)

The well-known formula of Chandra et. al. [CRR+89] expressing the commute time $\kappa_{xy} := H_{xy} + H_{yx}$ will also be useful:

$$H_{xy} + H_{yx} = 2mr(x,y).$$  (3)

(The latter formula can immediately be derived from the former.)

Using (2) we can express $CC(x) = \sum_y H_{xy}$ as

$$CC(x) = mR(x) + \frac{1}{2} \sum_{w \in V(G)} d(w) (R(w) - nr(w,x)).$$  (4)

3 Trees

In the case of trees, the effective resistance between two vertices equals their distance. In this section we combine this fact with (2) to obtain some results about trees.

Let us prove, as a warm-up, that the first two inequalities of Theorem 1 are equivalent, i.e. that $D(x) \leq D(y)$ if and only if $D_\pi(x) \leq D_\pi(y)$. We claim that

$$D_\pi(x) = 2D(x) + m,$$  (5)

from which our assertion immediately follows. To show (5), we will check that any edge $e$ has the same contribution to the two sides of the equation, where we think of the contribution of $e$ as the
number of times we add a term $d(x, y)$ such that $e$ lies on the $x$–$y$ path (and thus contributes $1$ unit to the distance). To this end, let $A_x(e)$ be the set of vertices on the same side of $e$ as $x$ and $B_x(e) = V(T) \setminus A_x(e)$ the complement. Then the contribution of $e$ to $D_x(x) := \sum w d(w)d(x, w)$ is, by definition, $\sum_{w \in B_x(e)} d(w)$. By the handshake lemma, the latter sum equals $2|B_x(e)| + 1$ (the $+1$ is due to the endvertex of $e$ in $B_x(e)$). Similarly, the contribution of $e$ to $D(x)$ is $|B_x(e)|$, from which (5) easily follows.

A similar argument will be used in the proof of Theorem 1.

Proof of Theorem 1 We have to prove that

$$CC(x) = 2W(T) - D(x).$$

For a tree we have $m = n - 1$. Moreover, $r(x, y) = d(x, y)$, $R(y) = D(y)$ and $K(T) = W(T)$. Using this, we can rewrite (4) as

$$CC(x) = (n - 1)D(x) + \frac{1}{2} \sum_{w \in V(T)} d(w)(D(w) - nd(x, w)).$$

We now check that any single edge $e$ has the same contribution to this sum and to $2W(T) - D(x)$, which implies our assertion.

To this end, let again $A_x(e)$ be the set of vertices on the same side of $e$ as $x$ and $B_x(e) = V(T) \setminus A_x(e)$ the complement. Then the contribution of $e$ to the above sum is

$$(n - 1)|B_x(e)| + \frac{1}{2} \sum_{w \in A_x(e)} d(w)|B_x(e)| + \frac{1}{2} \sum_{w \in B_x(e)} d(w)(|A_x(e)| - n).$$

Now note that $\sum_{w \in A_x(e)} d(w) = 2|A_x(e)| - 1$ and $\sum_{w \in B_x(e)} d(w) = 2|B_x(e)| - 1$ by the handshake lemma, hence the total contribution is

$$(n - 1)|B_x(e)| + \frac{1}{2}(2|A_x(e)| - 1)|B_x(e)| + \frac{1}{2}(2|B_x(e)| - 1)(|A_x(e)| - n),$$

which simplifies to $(2|A_x(e)| - 1)|B_x(e)|$.

It is also easy to see that

$$W(T) = \sum_e |A_x(e)||B_x(e)|$$

and

$$D(x) = \sum_e |B_x(e)|,$$

from which we conclude that $CC(x) = 2W(T) - D(x)$.  

By (3), the reverse cover cost $RC(x) = \sum_{y \in V(G)} H_{yx}$ is related to the cover cost by

$$RC(x) = 2mR(x) − CC(x).$$

Combined with Theorem 1 this implies that

$$RC(x) = (2n - 1)D(x) - 2W(T),$$

from which Theorem 2 immediately follows.
3.1 The extremal trees

In this section we determine the trees (rooted at a vertex \( r \)) attaining the extremal values of \( CC(r) \).

**Corollary 4.** The minimum value of \( CC(r) \) among all trees of order \( n \geq 2 \) is \( 2n^2 - 6n + 5 \), and it is attained by a star, rooted at one of its leaves.

**Proof.** Given the tree \( T \), it follows from Theorem that the minimum of \( CC(r) \) is achieved when \( D(r) \) attains its maximum. Since \( D(x) \) (as a function of \( x \)) is convex along paths, this maximum can only be attained when \( r \) is a leaf, so we can assume that the root is a leaf in our case. Let \( T' = T \setminus r \) be the rest of \( T \), and let \( r' \) be the unique neighbour of \( r \). Then we have

\[
CC(r) = 2W(T) - D(r) = 2(W(T') + D(r)) - D(r)
\]

\[
= 2W(T') + D(r) = 2W(T') + |T'| + D_r(T').
\]

It is well known [DEG01, EJS76] that the Wiener index is minimized by the star \( S_n \), so \( W(S_n - 1) = (n - 2)^2 \). Moreover, \( D_r(T') \geq |T'| - 1 \) is obvious as well, with equality if and only if \( T' \) is a star and \( r' \) its center. It follows that

\[
CC(r) \geq 2W(S_n - 1) + (n - 1) + (n - 2) = 2(n - 2)^2 + 2n - 3 = 2n^2 - 6n + 5
\]

for every tree \( T \) of order \( n \geq 2 \), with equality if and only if \( T \) is the star \( S_n \) and \( r \) one of its leaves. \( \blacksquare \)

**Corollary 5.** The maximum value of \( CC(r) \) among all trees of order \( n \) is \( (n^3 - n)/3 - [n^2/4] \), and it is obtained by a path, rooted at a midpoint.

**Proof.** Let \( r_1, r_2, \ldots, r_k \) be the neighbours of \( r \) and let \( T_1, T_2, \ldots, T_k \) be the associated branches. Then we have

\[
W(T) = \sum_{i=1}^{k} W(T_i) + \sum_{i=1}^{k} \sum_{j=1}^{k} (D_{r_i}(T_i) + |T_i||T_j|) + D(r),
\]

where the first term accounts for distances between vertices in the same branch, the second term for distances between vertices in different branches, and the last one for distances between the root and other vertices. Moreover,

\[
D(r) = \sum_{i=1}^{k} D_{r_i}(T_i) + |T| - 1.
\]

Therefore,

\[
CC(r) = 2W(T) - D(r) = 2 \sum_{i=1}^{k} W(T_i) + 2 \sum_{i=1}^{k} \sum_{j \neq i}^{k} (D_{r_i}(T_i) + |T_i||T_j|) + \sum_{i=1}^{k} D_{r_i}(T_i) + |T| - 1.
\]

It is known that the Wiener index is maximised by a path [DEG01, EJS76], and it is also easy to see that \( D(r) \) is maximal for a path of which \( r \) is an end. Therefore, \( CC(r) \) increases if we replace each of the branches \( T_i \) by a path with the same number of vertices. This means that we can assume that our tree maximising \( CC(r) \) is a subdivided star and \( r \) its center.

Now assume that \( k > 2 \) and, without loss of generality, that \( |T_1| \leq |T_2| \leq |T_3| \). We claim that if we detach \( T_2 \) from \( r \) and attach it to the last vertex of \( T_1 \), then \( CC(r) \) will increase. To see this, we are going to use the formula

\[
CC(r) = \sum_{e \in E(T)} (2|A_r(e)| - 1)|B_r(e)|
\]
from the proof of Theorem 1. Note that for any edge \( e \) not on \( T_1 \), the sizes of \( A_r(e), B_r(e) \) are not affected by this modification. For an edge \( e \) that does lie on \( T_1 \), its contribution to the sum above changes from \((2A - 1)B\) to \((2(A - t) - 1)(B + t)\), where \( A := |A_r(e)|, B := |B_r(e)| \) (as defined for \( T \) before the modification) and \( t := |T_2| \). The difference between the two expressions is

\[
(2(A - t) - 1)(B + t) - (2A - 1)B = 2At - 2tB - 2t^2 - t = 2t(A - B) - 2t(t + 1/2),
\]

and this is strictly positive if and only if \( A - B > t + 1/2 \). Clearly, we have \( B \leq |T_1| \) and \( A > |T_2| + |T_3| \), and so \( A - B > |T_2| + |T_3| - |T_1| \geq |T_2| = t \), where we used our assumption about the sizes of the \( T_i \). Since all values are integral, we thus obtain \( A - B > t + 1/2 \) as desired, proving that \( CC(r) \) increases when \( T_2 \) is moved to the end of \( T_1 \).

By iterating the argument, we can assume that \( T \) is a path. The minimum of \( D(r) \) is clearly attained at a midpoint of the path, and the precise value of \( CC(r) \) is easily determined in this case, completing the proof.

4 General formulas

In this section we prove Theorem 3.

We will derive our formulas for \( CC, RC \) and their weighted versions by applying Tetali’s formula. It turns out that this formula is easier to work with when considering differences; we start with calculating some auxiliary quantities: for every vertex \( x \) we have by

\[
CC(x) - RC(x) = \sum_y (H_{xy} - H_{yx})
\]

\[
= \sum_y \frac{1}{2} \sum_w d(w)(r(x, y) + r(w, y) - r(w, x) - r(x, y) - r(w, x) + r(w, y))
\]

\[
= \frac{1}{2} \sum_y \sum_w d(w)(2r(w, y) - 2r(w, x))
\]

\[
= 2K_x(G) - n \sum_w d(w)r(w, x) = 2K_x(G) - nR_x(x). \tag{7}
\]

Moreover, by \( K \) we have

\[
CC(x) + RC(x) = \sum_y \kappa_{xy} = \sum_y 2mr(x, y) = 2mR(x). \tag{8}
\]

Adding \( K \) to \( K \) we obtain \( CC(x) = mR(x) - \frac{r}{2}R_x(x) + K_x(G) \), while subtracting we obtain \( RC(x) = mR(x) + \frac{r}{2}R_x(x) - K_x(G) \). This proves the first part of Theorem 4. For the second part we proceed similarly: we have

\[
CC_x(x) - RC_x(x) = \sum_y d(y)(H_{xy} - H_{yx})
\]

\[
= \sum_y \sum_w d(y)d(w)(r(w, y) - r(w, x))
\]

\[
= 2K_x(G) - 2m \sum_w d(w)r(w, x) = 2K_x(G) - 2mR_x(x).
\]

and

\[
CC_x(x) + RC_x(x) = \sum_y d(y)\kappa_{xy} = \sum_y d(y)2mr(x, y) = 2mR_x(x). \tag{9}
\]
Adding and subtracting as above yields \( CC_π(x) = mR_π(x) + K_π(z(G) - mR_π(x) = K_π(z(G)\) and \( RC_π(x) = mR_π(x) - K_π(z(G) + mR_π(x) = 2mR_π(x) - K_π(z(G)\). This completes the proof of Theorem 3.

The last equality implies that \( RC_π(x) \leq RC_π(y)\) holds if and only if \( R_π(x) \leq R_π(y)\). Using as above it is easy to see that the latter is also equivalent to \( H_yx \leq H_yz\). This shows that the equivalence of inequalities [ii] [iv] of Corollary 1 apply holds for arbitrary graphs if we replace the function \( D_π\) in [ii] by the function \( R_π\). Using the fact that \( R_π = D_π\) in the case of trees, the remarks after Theorem 1 and [iv] completes the proof of Theorem 1.

Except for the aforementioned cases, all other equivalences in Theorem 1 fail for general graphs even if we replace \( D\) and \( D_π\) by their generalisations \( R\) and \( R_π\). To show this we will use the following fact which is interesting on its own right.

Corollary 4.1. The cover cost \( CC(r)\) is independent of the starting vertex \( r\) if and only if \( G\) is regular. In this case, we have \( CC(r) = K_π(G) = kK(G)\) where \( k\) is the vertex degree.

Proof. We claim that, for every graph \( G\), and every vertex \( r\) of \( G\), we have

\[
\sum_{z \sim r} (CC(r) - CC(z)) = nd(r) - 2m. 
\]

Note that this claim implies that if \( CC(r)\) is independent of the starting vertex \( r\) then \( G\) is regular, for the left hand side is 0 in that case. To prove [10], we write

\[
\sum_{z \sim r} (CC(r) - CC(z)) = \sum_{z \sim r} \sum_y (H_{ry} - H_zy) 
= \sum_{z \sim r} (H_{ry} - H_zy) + \sum_{z \sim r} (0 - H_zy) 
\]

Now note that for \( y \neq r\) we have \( H_{ry} = 1 + \frac{1}{d(r)} \sum_{z \sim r} H_zy\), for random walk from \( r\) moves to its neighbours \( z\) in its first step. Rearranging this we obtain \( \sum_{z \sim r} (H_{ry} - H_zy) = d(r)\). The return time \( H_{rr}^+\) to \( r\), i.e. the expect time for random walk from \( r\) to reach \( r\) again, is given by \( H_{rr}^+ = 2m/d(r)\). Using this, and an argument similar to the above, we obtain \( \sum_{z \sim r} -H_zy = d(r) - 2m\) Plugging these two equalities into the above sum yields [10].

Suppose, conversely, that \( G\) is \( k\)-regular. By Theorem 3 we have \( CC_π(x) = K_π(G)\). But if \( G\) is \( k\)-regular, then \( CC_π = kCC\) and \( K_π(G) = kK_π(G)\), which yields \( CC(r) = K_π(G)\) as desired.

The second sentence of Corollary 4.1 has been observed by Palacios [Pa10]; the first one might be new.

4.1 Counterexamples to the extension of Corollary 1 to non-trees

We now construct some examples showing that except for the aforementioned equivalence of inequalities [ii] [iv] of Corollary 1, all other equivalences fail for non-trees (with \( D\), \( D_π\) replaced by \( R\), \( R_π\)).

Having seen Corollary 4.1, it easy to construct examples of non-trees in which inequality [vi] of Corollary 1 is not equivalent to any of the others. In the regular graph of Figure 2, for example, the functions \( R_π = 3R\) are non-constant, and by Theorem 1, so are \( RC\) and \( RC_π\).

Our next examples shows that [iv] is not equivalent to any of the other inequalities either: in the graph of Figure 2 we have \( R(x) = R(y)\) but \( R_π(x) \neq R_π(y)\) as the reader will easily check. Combined with Theorem 3 this implies that \( RC_π \neq RC\). Finally, in order to show that [iv] is not equivalent to [iv], it suffices by Theorem 1, to have an example in which \( R(x) - R(y) = R_π(y) - R_π(x) > 0\). The graph of Figure 3 is such an example. For \( k\) large and \( l = k(k+1)/2\) we have \( R(x) - R(y) = l - k\). However, \( R_π(y) - R_π(x)\) is of order \( k\), as the interested reader will be able to check using the fact that when connecting networks in series, the total effective resistance equal the sum of the effective resistances of the subnetworks. The effective resistance between two vertices of a \( k\)-clique is about \( 2/k\).
Figure 1: A regular graph with non-constant $R_x, R, RC$ and $RC_x$.

Figure 2: $R(x) = R(y)$ but $R_\pi(x) \neq R_\pi(y)$.

Figure 3: A graph showing that (v) is not equivalent to (iv). The circle in the left stands for a $k$-vertex clique joined to $x$ by an edge, while an $l$-vertex star is attached to $y$.

References


