COLLECTING COUPONS ON TREES, AND THE COVER TIME OF RANDOM WALKS

Uriel Feige

Abstract. We consider the cover time $E_u[G]$, the expected time it takes a random walk that starts at vertex u to visit all n vertices of a connected graph G. Aleliunas et al introduced the spanning tree argument: for any spanning tree T of the graph G, $E_u[G] \leq W[T]$, where $W[T]$ is the sum of commute times along the edges of T. By refining the spanning tree argument we obtain:

$$E_u[G] \leq \frac{1}{2}(\min_T[W[T]] + \max_{v \in G}[H[u,v] - H[v,u]])$$

where $H[u,v]$ is the hitting time from u to v. We use this bound to show:

1. $\max_G \min_u E_u[G] = (1 + o(1))2n^3/27$. This answers an open question of Aldous.
2. The n-path is the n-vertex tree on which the cover time is maximized. This confirms a conjecture of Brightwell and Winkler.
3. For regular graphs, $E_u[G] < 2n^2$. This improves the leading constant in previously known upper bounds.

We also provide upper bounds on $E_u^+[G]$, the expected time to cover G and return to u.
Key words. Random walks, cover time, coupon collector.
Subject classifications. 05C99, 60J15, 68R10.

1. Introduction

Let G be a connected undirected graph on n vertices. We consider random walks on G, where at each step the random walk moves to a vertex chosen at random with uniform probability from the neighbors of the current vertex. For two vertices $u, v \in G$, the *hitting time* $H[u,v]$ is the expected number of steps it takes a walk that starts at u to reach v, the *commute time* $C[u,v]$ is the expected number of steps that it takes a walk to go from u to v and back to u (that is, $C[u,v] = H[u,v] + H[v,u]$), and the *difference time* (which may be negative) is $D[u,v] = H[u,v] - H[v,u]$.

Let G' be a subgraph of G, and let $u \in G'$. Then the *cover time* $E_u[G']$ denotes the expected number of steps it takes a walk that starts at u to visit all vertices of G', and the *cover and return time* $E_u^+[G']$ denotes the expected number of steps it takes a walk that starts at u to visit all vertices of G', and then return to u. We are interested in proving bounds on these quantities, especially when $G' = G$.

The hitting time between any two vertices can be computed in polynomial time [10]. Computing the cover time of a graph is more difficult. For any graph, the cover time can be estimated with arbitrary degree of accuracy by a random polynomial time algorithm that repeatedly takes random walks and measures their cover times. However, no deterministic polynomial time algorithm is known for computing the cover time, nor for approximating it. The lack of simple characterizations of the cover time also makes it difficult to obtain extremal results (e.g., for which n-vertex graph is the cover time maximal? minimal?).

One approach of upper bounding the cover time is by a "coupon collector" argument. Each time the random walk visits a new vertex it gets a coupon, and the random walk has to collect all $n-1$ coupons. No restrictions are imposed on the order in which coupons are collected. Covering the complete graph K_n is equivalent to the classical coupon collector problem with $n-1$ coupons, giving $E_u[K_n] \simeq n \ln n$. For arbitrary graphs, the coupon collector argument was generalized by Matthews to give the upper bound $E_u[G] \leq \max_{x,y}[H[x,y]] \ln n$ [11]. This bound is particularly useful for graphs that have very fast cover times, such as expander graphs.

A different approach for providing upper bounds on the cover time was introduced by Aleluinas *et al.* [2]. In this approach, the coupons are to be collected in a particular favorable order. A convenient order to consider is one that agrees with Depth First Search of a minimum weight spanning tree of the graph. Let T be a tree defined on the vertices of G (though tree edges need not correspond to edges of G). Define the *weight* of the tree as $W[T] = \sum_{(u,v)\in T} C[u,v]$, where summation is taken over the endpoints of the $n-1$ edges (u,v) of T. It is easy to see that $E_u[G] \leq \min_T[W[T]]$ (and similarly, $E_u^+[G] \leq \min_T[W[T]]$). The spanning tree approach is particularly useful for graphs that have slow cover time, such as the *lollipop* graph (a clique of size $2n/3$ connected to a path of size $n/3$). It was used successfully to provide a tight bound (upto low order terms) of $4n^3/27$ on the cover time of any n-vertex graph [8].

In this paper we study cover problems where neither the coupon collector approach nor the spanning tree approach provide a satisfactory bound. Our research was motivated by a question of Aldous (private communication), namely, how high can $\max_G \min_u E_u[G]$ be?

1.1. Additional notation. The input graph is denoted by G. Throughout, we assume that G is connected. G can be viewed as a weighted complete graph, where the weight of edge (u, v) is $C[u, v]$. On (a subset of) the vertices of this complete graph we may define a tree T. For vertices $u, v \in T$, let $P[u, v]$ denote the set of directed edges on the simple path leading from u to v along the edges of the tree T. We use the notation $H_T[u, v]$ to define upper bounds on hitting times as can be computed by summing up hitting times along the path of T connecting u and v. Hence $H_T[u, v] = \sum_{(x,y)\in P[u,v]} H[x, y]$, and $C_T[u, v] = H_T[u, v] + H_T[v, u]$. Observe that $H_T[u, v] \geq H[u, v]$, and $C_T[u, v] \geq C[u, v]$. It is interesting to note that there is no reason to define a concept of $D_T[u, v] = H_T[u, v] - H_T[v, u]$. It can be shown that regardless of the tree T on which the path P is defined, $D_T[u, v] = D[u, v]$. This follows from additivity of the difference time.

Lemma 1.1. *For any three vertices, $D[u, v] + D[v, w] = D[u, w]$.*

The proof of Lemma 1.1 appears in [6].

1.2. A new technique and the main result. We provide a coupon collector based refinement to the spanning tree method. As in the spanning tree method, we construct a spanning tree of the graph. But unlike the spanning tree method, we do not enforce a particular order in which the spanning tree has to be traversed. Instead, we use a coupon collector argument, collecting the coupons from parts of the tree that happen to be covered first.

The main contribution of our work is in showing that the above nonstandard coupon collector problem leads to an upper bound that is both simple to express and useful.

Theorem 1.2. *Let G be a connected graph on n vertices. Let T be a tree defined on a subset of the vertices of G, where the edges of T need not be edges of G. Let $E_u[T]$ define the expected time it takes a random walk that starts at $u \in T$ to visit all vertices of T. Let $W[T]$ be the weight of T: the sum of $|T| - 1$ commute times between endpoints of edges of T. Then*

$$E_u[T] \leq \frac{1}{2}(W[T] + \max_{v \in T \setminus \{u\}} [D[u,v]])$$

1.3. Modifications and corollaries. Our techniques apply also to the cover and return time. (Recall the notation of $C_T[u,v]$ from Section 1.1.)

THEOREM 1.3. *Let G be a connected graph on n vertices. Let T be a tree defined on a subset of the vertices of G, and $u \in T$. Then*

$$E_u^+[T] \leq \frac{1}{2}(W[T] + \max_{v \in T}[C_T[u,v]])$$

We conjecture that $C_T[u,v]$ in the above upper bound can be replaced by $C[u,v]$. However, this is of little relevance for our intended applications.

Our original motivation for obtaining Theorem 1.2 was to show that for any graph, there is a "good" starting point from which the expected cover time is at most $W[T]/2$. This follows from the fact that $\min_u \max_v [D[u,v]] \leq 0$ [6]. (In the terminology of [13], a vertex u for which $D[u,v] \leq 0$ for all v is called a *remote* vertex.) Since the minimum weight spanning tree of any connected graph satisfies $W[T] \leq (1+o(1))4n^3/27$ [8], we obtain:

COROLLARY 1.4. *For connected graphs on n vertices,* $\max_G \min_u [E_u[G]] \leq (1+o(1))2n^3/27$.

This answers an open question of Aldous. Using Theorem 1.3 we show:

COROLLARY 1.5. *For connected graphs on n vertices,* $\max_G \min_u [E_u^+[G]] \leq (1+o(1))n^3/9$.

The bounds in Corollaries 1.4 and 1.5 are best possible (up to low order terms), as we shall show in Section 3.1.

For some families of graphs we can show that $\max_{u,v} D[u,v] < cW[T]$, where c is some constant, $0 < c < 1$. For these graphs, we obtain bounds on the cover time that are tighter than those that can be obtained by spanning tree arguments alone. One such family is that of trees. Brightwell and Winkler [3] showed that the n vertex tree with least cover time is the star. They conjectured that the path has maximum cover time. Using Theorem 1.2, we prove this conjecture.

COROLLARY 1.6. *For n-vertex trees T,*

$$\max_T \max_u E_U[T] = \lfloor \frac{5(n-1)^2}{4} \rfloor$$

which is the cover time for a walk that starts at the middle point (or one of two points) on the n-path.

The condition $\max_{u,v} D[u,v] < cW[T]$ holds also for regular graphs. This allows us to improve the known upper bounds on their cover time.

COROLLARY 1.7. *For any regular graph,* $\max_v[E_v[G]] < 2n^2$.

For the cover and return time we obtain:

COROLLARY 1.8. *For any connected d-regular graph, $E_u^+[G] \le n^2(2 + \frac{d-2}{(d+1)^2})$. In particular, $E_u^+[G] \le 25n^2/12$.*

1.4. Asymptotic growth of extremal cover times. In conjunction with previous works of the author [8, 9], we have the following results for connected graphs on n vertices (where $\simeq$ denotes equality up to low order additive terms).
$\max_G \max_u E_u[G] \simeq 4n^3/27$
$\max_G \max_u E_u^+[G] \simeq 4n^3/27$
$\max_G \min_u E_u[G] \simeq 2n^3/27$
$\max_G \min_u E_u^+[G] \simeq 3n^3/27$
$\min_G \max_u E_u[G] \simeq n \ln n$
$\min_G \max_u E_u^+[G] \simeq n \ln n$
$\min_G \min_u E_u[G] \simeq n \ln n$
$\min_G \min_u E_u^+[G] \simeq n \ln n$
The question of which are the extremal graphs for each of the above properties is open.

2. Collecting coupons on trees

In this section we prove Theorems 1.2 and 1.3. We shall use Lemma 2.1, which is a simple consequence of the linearity of the expectation. Let A and B denote two events, $A \vee B$ denote the occurrence of one of the events (whichever happens first), $A \wedge B$ the occurrence of both events. Recall that $E[C]$ denotes the expected number of steps until event C occurs.

LEMMA 2.1.
$$E[A] + E[B] = E[A \vee B] + E[A \wedge B]$$

PROOF. By definition, $E[C] = \sum_i iPr[t(C) = i]$. Let $Q_{i,j}$ denote the event that $t(A) = i$ and $t(B) = j$, and let $q_{i,j}$ be the probability that $Q_{i,j}$ occurs. Assume without loss of generality that $i \leq j$. Then $Q_{i,j}$ contributes $iq_{i,j}$ to $E[A]$, $jq_{i,j}$ to $E[B]$, $iq_{i,j}$ to $E[A \vee B]$, and $jq_{i,j}$ to $E[A \wedge B]$. Summing over all values of i and j proves the lemma. □

Now we are ready to prove Theorem 1.2, that
$$E_u[T] \leq \frac{1}{2}(W[T] + \max_{v \in T \setminus \{u\}}[D[u,v]])$$

PROOF. For a graph G, let T be a tree defined on a subset of the vertices of G. That is $V(T) \subset V(G)$. The edges of T have weights induced by the commute times in G between their endpoints. We shall prove Theorem 1.2 for any graph G and any tree T by induction on the number of vertices in T.

For the base case, T contains only two vertices, u and v. Then the left hand side evaluates to $E_u[T] = H[u,v]$, and the right hand side evaluates to
$$\frac{1}{2}(W[T] + D[u,v]) = \frac{1}{2}(C[u,v] + D[u,v]) = H[u,v]$$
and we have exact equality in the statement of the Theorem.

For the inductive step, we make the inductive hypothesis that the theorem holds for any u', and any T' with at most k vertices, and prove for any u and any T with $k+1$ vertices. We distinguish between two cases, according to the degree of u in T.

If the degree of u in T is 1, then by removing u from T we are left with a tree T' with k vertices. Let $y \in V(T')$ denote the neighbor of u in T. We have $E_u[T] \leq H[u,y] + E_y[T']$. The tree T' has only k vertices, so by the inductive hypothesis,
$$E_y[T'] \leq \frac{1}{2}(W[T'] + \max_{v \in T' \setminus \{y\}}[D[y,v]])$$

Using $H[u,y] = \frac{1}{2}(C[u,y] + D[u,y])$, and $W[T] = C[u,y] + W[T']$, and the relation $D[u,v] = D[u,y] + D[y,v]$ (Lemma 1.1), the proof follows.

If the degree of u in T is greater than 1, then we partition T into two subtrees, T_1 and T_2, that intersect only on u. We want a bound on $E_u[T_1 \wedge T_2]$. Let $H[T_1, u]$ denote the expected time it takes a walk that starts at the worst possible vertex of T_1 to hit u, and define $H[T_2, u]$ similarly. Observe that

$$E_u[T_1 \wedge T_2] \leq E_u[T_1 \vee T_2] + \max[H[T_2, u] + E_u[T_1], H[T_1, u] + E_u[T_2]]$$

W.l.o.g., we assume that $H[T_2, u] + E_u[T_1] \geq H[T_1, u] + E_u[T_2]$. Hence

$$E_u[T_1 \wedge T_2] \leq E_u[T_1 \vee T_2] + H[T_2, u] + E_u[T_1]$$

Using Lemma 2.1 to eliminate $E_u[T_1 \vee T_2]$ ($= E_u[T_1] + E_u[T_2] - E_u[T_1 \wedge T_2]$) we obtain:

$$E_u[T_1 \wedge T_2] \leq E_u[T_1] + \frac{1}{2}(E_u[T_2] + H[T_2, u])$$

LEMMA 2.2. *For any tree T defined on the vertices of connected graph G, and for any starting point $u \in T$,*

$$E_u[T] \leq W[T] - H[T, u]$$

PROOF. Let $x \in V(T)$ be the vertex that maximizes $H[T, u]$. That is, $H[T, u] = H[x, u]$. Rather than use $H[x, u]$, we shall use the upper bound $H_T[x, u] \geq H[x, u]$. Since $H_T[x, u]$ is maximized when x is a leaf of T, we may assume that x is a leaf of T. Then there is some Depth First Search order of visiting the vertices of T that starts at u and visits x last. For this particular order, each edge of T is traversed in both directions, except for the path leading from x to u. Hence $E_u[T] \leq W[T] - H_T[x, u]$. □

Applying Lemma 2.2 to T_2, we obtain $E_u[T_2] + H[T_2, u] \leq W[T_2]$. Using the inductive hypothesis $E_u[T_1] \leq \frac{1}{2}(W[T_1] + \max_{x \in T_1} D[u, x])$, and the identity $W[T] = W[T_1] + W[T_2]$, the proof of Theorem 1.2 follows. □

We now prove Theorem 1.3, that for any connected graph G,

$$E_u^+[T] \leq \frac{1}{2}(W[T] + \max_{v \in T}[C_T[u, v]])$$

The proof is simpler than that of Theorem 1.2, and is given in less detail.

PROOF. By induction on the size of T. If T contains only one edge, we have exact equality in the statement of the lemma. Assume the theorem holds for any u', and any T' with at most k vertices, and prove for T with $k+1$ vertices. We distinguish between two cases, according to the degree of u in T.

If the degree of u in T is 1, let T' be the tree that remains by removing u from T, and let $y \in V(T)$ denote the neighbor of u in T. Then we have

$E_u^+[T] \le C[u,y] + E_y^+[T']$. The tree T' has only k vertices, hence $E_y^+[T'] \le \frac{1}{2}(W[T'] + \max_{v \in T'}[C_{T'}[y,v]])$. Using $W[T] = C[u,y] + W[T']$, $C[u,y] = C_T[u,y]$, and $C_T[u,v] = C_T[u,y] + C_{T'}[y,v]$, the proof follows.

If the degree of u in T is greater than 1, then we partition T into two subtrees, T_1 and T_2, that intersect only on u. Note that each such subtree contains at most k vertices, and so it obeys the inductive hypothesis. We want a bound on $E_u^+[T_1 \wedge T_2]$. Observe that

$$E_u^+[T_1 \wedge T_2] \le E_u^+[T_1 \vee T_2] + \max[E_u^+[T_1], E_u^+[T_2]]$$

Assuming w.l.o.g. that $E_u^+[T_1] \ge E_u^+[T_2]$, and using Lemma 2.1 to eliminate $E_u^+[T_1 \vee T_2]$ we obtain:

$$E_u^+[T_1 \wedge T_2] \le E_u^+[T_1] + \frac{1}{2} E_u^+[T_2]$$

Using $E_u^+[T_1] \le \frac{1}{2}(W[T_1] + \max_{x \in T_1} C_T[u,x])$, and $E_u^+[T_2] \le W[T_2]$, the proof follows. □

3. Applications

We present three applications of our bounds. The first uses the fact that for any connected graph G, there exists a spanning tree T of weight $W[T] \le (1+o(1))4n^3/27$ [8]. The second uses the fact that if G is a tree, then $W[G] = W[T] = 2(n-1)^2$. The third uses the fact that for connected d-regular graphs, any spanning tree T has weight $W[T] < 3dn^2/(d+1)$ [5]. In all three cases, we obtain bounds on $D[u,v]$ (or $C_T[u,v]$) that are smaller than the corresponding values of $W[T]$, and get bounds on the cover time that are stronger than those that one obtains by using $W[T]$ alone.

We shall often need to compute hitting times and commute times, both for the purpose of obtaining our upper bounds, and for providing examples that show that they are tight. The analogy between random walks and resistance in electrical networks [7] proves helpful in this respect (for those who have better intuition for electrical resistance than for the behavior of random walks). An undirected graph can be viewed as an electrical network in which each edge represents a resistor of 1 ohm. The *effective resistance* $R[u,v]$ between two vertices is the voltage that develops between them if they are connected to a current source of 1 amp (by Ohm's law). Effective resistance is monotone (lowering the resistance of a resistor in the network cannot increase the effective resistance), obeys the triangle inequality, is additive for series connections, and its inverse is additive for parallel connection. The following useful identity is proved in [4].

LEMMA 3.1. *For any connected graph with m edges and any two vertices, the commute times and effective resistance are related by $C[u,v] = 2mR[u,v]$.*

By standard properties of electrical resistance, for any edge (u,v) of the graph, $R[u,v] \leq 1$, with equality if and only if (u,v) is a cut edge of the graph. It follows from Lemma 3.1 that if (u,v) is a cut edge of the graph, then $C[u,v] = 2m$ (for a more direct proof, see [2]).

COROLLARY 3.2. *For any cut edge (u,v), $H[u,v] = 2m' - 1$, where m' is the number of edges that remain in the graph if all edges on v's side of the cut are removed.*

PROOF. Removing all edges from v's side of the cut does not change $H[u,v]$. However it does make $H[v,u] = 1$. Subtracting this hitting time from the commute time, which after removing the edges is $2m'$, proves the corollary. □

3.1. Where to start a random walk. We use Theorem 1.2 to prove Corollary 1.4, that for any connected graph there is a vertex u with $E_u[G] \leq (1+o(1))2n^3/27$.

PROOF. By Lemma 1.1, it follows that there exist a vertex u such that for any vertex v, $D[u,v] \leq 0$ (as proved in [6]). From Theorem 1.2, we obtain that for this particular u and any tree T that contains u, $E_u[G]] \leq \frac{1}{2}W[T]$. The proof follows by using the bound $W[T] \leq (1+o(1))4n^3/27$ for the minimum weight spanning tree of G [8]. □

The bound in Corollary 1.4 is best possible. Consider the *grape* graph composed of a path of length $n/3$, and a clique of size $2n/3$ hanging in the middle of the path. The graph has $(1+o(1))2n^2/9$ edges. The effective resistance between u and v, the two endpoints of the path, is $n/3$. From Lemma 3.1, $C[u,v] = (1+o(1))4n^3/27$. By symmetry, $H[u,v] = H[v,u] = C[u,v]/2$. Hence no matter which of the two endpoints is hit first, $2n^3/27 + o(n^3)$ lower bounds the expected time remaining until the graph is covered.

We use Theorem 1.3 to prove Corollary 1.5, that for any connected graph there is a vertex u with $E_u^+[G] \leq (1+o(1))n^3/9$.

PROOF. For graph G, construct minimum spanning tree T, using only edges of G. By [8], it has weight at most $(1+o(1))4n^3/27$. Observe that between any two adjacent vertices on T, $C_T[x,y] = C[x,y] < n^2$. Let z and w be a pair of vertices that maximize $C_T[z,w]$. Then we choose a vertex u along the path connecting z and w on T such that $C_T[u,z] \le \frac{1}{2}(C_T[z,w]+n^2)$, and $C_T[u,w] \le \frac{1}{2}(C_T[z,w]+n^2)$. Such a vertex u must exist, and satisfies $\max_{v\in G} C_T[u,v] \le \frac{1}{2}(W[T]+n^2)$. Hence $E_u^+[G] \le 3W[T]/4 + n^2/4 \le (1+o(1))3n^3/27$. □

The above bound is best possible.

LEMMA 3.3. *For the grape graph (a path of length $n/3$, and a clique of size $2n/3$ hanging in the middle),* $\min_v[E_v[grape]] \ge (1+o(1))n^3/9$.

PROOF. Let x and y be the two endpoints of the path. $E_v^+[\{x,y\}]$ provides a lower bound on $E_v^+[G]$. We shall first show that $E_u[\{x,y\}] \ge (1+o(1))n^3/9$, where u is the midpoint of the path.

By Lemma 3.1, $C[u,x] = C[u,y] \simeq 2n^3/27$. Using symmetry arguments, $E_u^+[\{x,y\}] = E_u^+[\{x\} \vee \{y\}] + E_u^+[\{x\}]$. To compute $E_u^+[\{x\} \vee \{y\}]$, identify the two endpoints x and y (that is, the path is replaced by a cycle of length $n/3$), and name the merged vertex z. Then $E_u^+[\{x\} \vee \{y\}] = H[u,z] + H[x,u]$. By series/parallel connection principles, $R[u,z] = n/12$. Hence $C[u,z] = (1+o(1))n^3/27$. The hitting time from z to u remains unchanged if we remove the clique (which is connected only to u), and hence is $O(n^2)$. It follows that $H[u,z] = (1+o(1))n^3/27$. Hence $E_u^+[\{x\} \vee \{y\}] \ge (1+o(1))n^3/27$, and from $E_u^+[\{x\}] = C[u,x]$ we obtain that $E_u^+[\{x,y\}] \ge n^3/9 + o(n^3)$.

Now consider any starting vertex v. If v is in the clique, $E_v^+[\{x,y\}] \ge E_u^+[\{x,y\}]$, as the random walk must pass through u on its way to x and y, and return through u on its way back to v. So assume that v is on the path, between u and x, distance k away from u, where $1 \le k \le n/6$. We proceed by a case analysis.

1. x is hit before u. Then the expected time that remains is $H[x,y] + H[y,v] \simeq 2n^3/27 + 4kn^2/9$.
2. u is hit before x. The expected time that remains will be $E_u^+[\{x,y\}]$ with some additive correcting factor, that may be either positive or negative. We have two subcases.
   (a) x is hit before y. Then after y is hit and the return to u, the walk still needs to return to v, paying $H[u,v]$.
   (b) y is hit before x. Then after x was hit, the walk did not need to return all the way to u. It could have stopped at v. Hence we overpaid $H[v,u]$.

Because both subcases are equally likely, the expected time that remains is $E_u^+[\{x,y\}] + D[u,v]/2$. From Corollary 3.2, $D[u,v] \simeq 4kn^2/9$.

By standard properties of random walks on paths, the probability of case 1 is $k/(n/6)$ and the probability of case 2 is $(n/6-k)/(n/6)$. Taking the weighted average of both cases we obtain that $E_v^+[\{x,y\}] \simeq n^3/9 + 4k^2n/3$. $\Box$

3.2. Trees. Trees on n vertices have $n-1$ edges, and each edge is a cut edge, with effective resistance 1 between its endpoints. This makes Lemma 3.1 and Corollary 3.2 easy to apply. In particular, the commute time between any two endpoints of an edge is $2(n-1)$.

Of all n-vertex trees, the expected commute time is maximized for the end points of the n-path, where it is $2(n-1)^2$. Since the commute time is always bounded by $W[T]$, and $W[T] = 2(n-1)^2$ for any tree, it follows from Theorem 1.3 that the n-path is the tree with maximum cover and return time.

Brightwell and Winkler [3] conjectured that of all trees, the n-path also maximizes the expected cover time, for a walk that starts at the mid-point of the path (or one of the two midpoints, for even n). We prove this conjecture.

LEMMA 3.4. *For any n-vertex tree,*

$$\max_{u,v}[D[u,v]] \le \lfloor \frac{(n-1)^2}{2} \rfloor$$

PROOF. Let T be an arbitrary n-vertex tree, and consider the vertices that maximize $D[u,v]$. Let ℓ be the number of edges along the path connecting u and v. Then $C[u,v] = 2\ell(n-1)$, $H[v,u] \ge \ell^2$, and $D[u,v] \le 2\ell(n-1) - 2\ell^2$. This is maximized when $\ell = (n-1)/2$ for odd n, and $\ell = n/2$ for even n. $\Box$

From Theorem 1.2 we obtain Corollary 1.6, that for any n-vertex tree T,

$$\max_u E_u[T] \le \lfloor \frac{5(n-1)^2}{4} \rfloor$$

This is exactly the cover time for the n-path. The n-path is the *unique* n-vertex tree with this cover time. This follows from the fact that for a tree to be extremal, we must have for the starting point u and any leaf v

$$D[u,v] = \lfloor \frac{(n-1)^2}{2} \rfloor$$

(this follows from requiring that all inequalities are tight in the proof of Theorem 1.2).

3.3. Regular graphs. The diameter bound for regular graphs can be used in order to bound the maximum commute time.

LEMMA 3.5. *For any connected d-regular graph, the diameter (number of edges in the shortest path between two most distant vertices) is smaller than* $\frac{3n}{d+1} - 1$.

PROOF. (The proof presented is a slight refinement of a proof given in [1].) Let v_0 and v_ℓ be the vertices that are furthest apart, and let $v_1, v_2, \ldots, v_{\ell-1}$ be the shortest path connecting them. Let V_{path} be the set of $\ell + 1$ vertices on this path, and let $\bar{V}_{path}$ be the set of $n - \ell - 1$ vertices that are not on this path. Since the degree of each vertex is d, and the path contains ℓ edges, then there are $d(\ell + 1) - 2\ell$ edges connecting V_{path} and $\bar{V}_{path}$. Each vertex in $\bar{V}_{path}$ is connected to at most 3 vertices in V_{path}, since otherwise it would serve as a shortcut contradicting the shortest path property. Hence

$$3(n - \ell - 1) \geq d(\ell + 1) - 2\ell$$

implying the lemma. □

We shall use the following lemma, whose proof appears in [5].

LEMMA 3.6. *For any connected d-regular graph G on n vertices and two adjacent vertices u and v, let $\delta[u, v] = R[u, v] - 2/(d + 1)$. Then $\delta[u, v] \geq 0$, and $\sum_{(u,v)} \delta[u, v] = n/(d + 1) - 1$. In particular, the weight of any spanning tree is at most $nd(2(n - 1)/(d + 1) + n/(d + 1) - 1) < 3dn^2/(d + 1)$.*

COROLLARY 3.7. *Let G be connected d-regular graph. Let P be the shortest path connecting vertices x and y. Then $C_P[x, y] \leq (n^2 d(d + 7)/(d + 1)^2 - nd)$, where commute times are measured relative to the path P (as in Sect. 1.1).*

PROOF. By Lemma 3.6, for any set of ℓ edges, their total effective resistance is at most $\frac{n}{d+1} - 1 + \frac{2\ell}{d+1}$. Substitute an upper bound for ℓ from Lemma 3.5 and use Lemma 3.1. □

LEMMA 3.8. *For any connected d-regular graph and vertex u, there exists a spanning tree T of weight $W[T] < 3n^2 d/(d + 1)$, and with maximum commute time* $\max_v [C_T[u, v]] < n^2 d(d + 7)/(d + 1)^2$.

PROOF. Use Breadth First Search starting at u to construct a spanning tree. Then the bound on the commute time follows from Corollary 3.7, and the bound on the weight of the tree follows from Lemma 3.6. □

Now, from Theorem 1.3, we obtain Corollary 1.8, that for any connected d-regular graph, the maximum expected cover and return time is less than $n^2(2 + \frac{d-2}{(d+1)^2})$. Our bound obtains its maximum $25n^2/12$ at $d = 5$.

A d-regular *necklace* (for $d \geq 3$) is constructed from roughly $n/(d+1)$ cliques, each on $(d+1)$ vertices, and two cliques on $(d+2)$ vertices. The cliques are arranged on a line, with the two larger cliques at the two ends of the line. Any two adjacent cliques are connected by an edge. Within each one of the inner cliques, there are now two vertices of degree $d+1$. The edge between these two vertices is removed, to make the graph d-regular. For the outer cliques, all degrees are $d+1$ except for one vertex of degree $d+2$. If d is odd, edges can be removed from the outer cliques so as to make them d regular as well.

The regular graph whose expected cover and return time is conjectured to be worst is the 3-regular necklace, for which the bound is $n^2(\frac{3}{2} + o(1))$. For large d, our bound is $n^2(2 + o(1))$, whereas the d-regular necklace has a cover and return time of $n^2(1 + o(1))$ (for odd d). We remark that our upper bound is not sensitive to the existence of a constant number of vertices of degree $d \pm 1$ (except for low order additive terms). In this respect it is nearly tight, for the path.

To bound the cover time, we bound $\max_{u,v}[D[u,v]]$ over all d-regular graphs, and then use Theorem 1.2.

LEMMA 3.9. *For any connected d-regular graph,* $\max_{u,v}[D[u,v]] \leq (n+d-2)^2/2$.

PROOF. Tetali [12] shows that $D[u,v] = \sum_z d_z(R[v,z] - R[u,z])$, where d_z is the degree of vertex z. For d-regular graphs, $d_z = d$. Hence $D[u,v] = d\sum_z(R[v,z] - R[u,z])$. By the triangle inequality for resistances, $R[v,z] - R[u,z] \leq R[u,v]$. For convenience of notation, we let R denote $R[u,v]$.

We partition $V(G)$ into three groups of vertices, A, B, and C. Then we will upper bound $\sum_z(R[v,z] - R[u,z])$ as $\sum_{z\in A}(R[v,z] - R[u,z]) + \sum_{z\in B}(R[v,z] - R[u,z]) + \sum_{z\in C}(R[v,z] - R[u,z])$.

To describe the partition, consider the shortest path P from v to u. Let $\ell + 1$ denote the number of vertices on the path, let p_i for $0 \leq i \leq \ell$ denote the vertices on the path, where $p_0 = v$ and $p_\ell = u$. Group A contains the vertices a_i for $0 \leq i \leq \lfloor R \rfloor$ as follows. Vertex a_i is chosen as the first vertex on the path P for which $i-1 < R[v, a_i] \leq i$. In particular, $a_0 = v$, and all a_i are distinct. To see that all a_i are well defined, observe that $R[v, p_\ell] = R$ and $R[v, p_{i+1}] \leq R[v, p_i] + 1$, where the last inequality follows from $R[p_i, p_{i+1}] \leq 1$ and the triangle inequality for resistances.

By the triangle inequality for resistances, $R[u, a_i] + R[v, a_i] \geq R$. Hence $R[u, a_i] \geq R - i$. It follows that $\sum_{z \in A}(R[v, z] - R[u, z]) \leq 0$.

Group B contains vertices that are not in A, and are neighbors of a_i for even i. Specifically, for each such a_i, we choose $d-2$ neighbors, making sure that neither a_{i-1} nor a_{i+1} are chosen. All vertices chosen this way are distinct. Suppose to the contrary that vertex $b \notin A$ is both a neighbor of a_i and a neighbor of a_j, where both i and j are even. Then the distance between a_i and a_j is at most 2. By the shortest path property of path P, it follows that $j = i + 2$. Now because $b \neq a_{i+1}$ there are two parallel paths of length two between a_i and a_{i+2}. Hence $R[a_i, a_{i+2}] \leq 1$, contradicting our construction with $R[a_i, a_{i+2}] > 1$.

Let b_{ij}, for $0 \leq i \leq \lfloor R/2 \rfloor$ and $1 \leq j \leq d-2$, denote the jth neighbor of a_{2i} chosen to be in B. Then by the triangle inequality for resistances, $R[v, b_{ij}] \leq 2i+1$ and $R[u, b_{ij}] \geq R - 2i - 1$. It follows that $\sum_{z \in B}(R[v, z] - R[u, z]) \leq R(d-2)$.

Group C contains all remaining vertices. Since $|A| \geq R$ and $|B| \geq R(d-2)/2$, $|C| \leq n - R - R(d-2)/2 = n - Rd/2$. For vertex $c \in C$ we use the inequality $R[v, c] - R[u, c] \leq R$ to conclude that $\sum_{z \in C}(R[v, z] - R[u, z]) \leq R(n - Rd/2)$.

Altogether, $\sum_z (R[v, z] - R[u, z]) \leq R(n - Rd/2) + R(d-2)$, which is maximized for $R = (n + d - 2)/d$, giving $\sum_z (R[v, z] - R[u, z]) \leq (n + d - 2)^2/2d$. Hence $D[u, v] \leq (n + d - 2)^2/2$. □

Corollary 3.10. *For any connected d-regular graph and for any vertex u, $E_u[G] \leq 3n^2 d/2(d+1) + \min[(n + d - 2)^2/2\ ,\ n^2 d(d+7)/(d+1)^2 - nd]/2$.*

Proof. Follows from Theorem 1.2. The term $3n^2 d/2(d+1)$ is obtained from Lemma 3.6. The term $(n + d - 2)^2/2$ is obtained from Lemma 3.9. The term $n^2 d(d+7)/(d+1)^2 - nd$ is obtained from Corollary 3.7 (the commute time bounds the difference time from above). □

Corollary 3.10 gives two upper bounds on the cover time of regular graphs. When n is large relative to d, it is better to use the first upper bound, namely, $3n^2 d/2(d+1) + (n + d - 2)^2/4$. In particular, when $n > \frac{d-2}{\sqrt{2}-1}$, the first upper bound is smaller than $2n^2$. When n is small relative to d, it is better to use the second upper bound, namely, $3n^2 d/2(d+1) + (n^2 d(d+7)/(d+1)^2 - nd)/2$.

In particular, when $n < \frac{d(d+1)^2}{2(d-2)}$, the second upper bound is smaller than $2n^2$. Observe that for every $d > 2$, $\frac{d-2}{\sqrt{2}-1} < \frac{d(d+1)^2}{2(d-2)}$. Hence for any $d > 2$, at least one of the two upper bounds is smaller than $2n^2$. Hence we obtain Corollary 1.7, that for any regular graph, $\max_v[E_v[G]] < 2n^2$. (The proof of Corollary 1.7 in cases that the graph is either 1-regular or 2-regular is straightforward, and is left to the reader.) For a walk starting at the middle of the 3-regular necklace, the cover time is $15n^2/16$ (up to low order terms). This is conjectured to be the regular graph that is most difficult to cover [1].

Acknowledgements

The author is a Yigal Alon Fellow, and the incumbent of the Joseph and Celia Reskin Career Development Chair.

References

[1] D. Aldous, J. Fill, *Reversible Markov chains and random walks on graphs.* Book in preparation. Draft available via homepage http://www.stat.berkeley.edu/users/aldous.

[2] R. Aleliunas, R. M. Karp, R. J. Lipton, L. Lovász, C. Rackoff, Random walks, universal traversal sequences, and the complexity of maze problems. In *Proceedings of the 20th Annual Symposium on Foundations of Computer Science*, pages 218–223, San Juan, Puerto Rico, October 1979.

[3] G. Brightwell, P. Winkler, Extremal Cover Times for Random Walks on Trees. *Journal of Graph Theory*, 14(5):547–554, 1990.

[4] A. Chandra, P. Raghavan, L. Ruzzo, R. Smolensky, P. Tiwari, The electrical resistance of a graph captures its commute and cover times. In *Proceedings of the 21st Annual ACM Symposium on the Theory of Computing*, 1989, 574–586.

[5] D. Coppersmith, U. Feige, J. Shearer, Random Walks on Regular and Irregular Graphs. *SIAM J. on Discrete Math.*, 9(2):301–308, May 1996.

[6] D. Coppersmith, P. Tetali, P. Winkler, Collisions among random walks on a graph. *SIAM J. on Discrete Math.*, 6(3):363–374, August 1993.

[7] P. Doyle, J. Snell, *Random walks and electrical networks.* The Mathematical Association of America, Washington D.C., 1984.

[8] U. Feige, A Tight Upper Bound on the Cover Time for Random Walks on Graphs. *Random Structures and Algorithms*, Vol. 6, No. 1 (1995) 51–54.

[9] U. Feige, A Tight Lower Bound on the Cover Time for Random Walks on Graphs. *Random Structures and Algorithms*, Vol. 6, No. 4 (1995) 433–438.

[10] J. Kemeny and J. Snell, *Finite Markov Chains.* Springer Verlag, 1983.

[11] P.C. Matthews, Covering Problems for Brownian Motion on Spheres. *Ann. Probab.*, 16:189-199, 1988.

[12] P. Tetali, Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4:101-109, 1991.

[13] P. Tetali, P. Winkler, Simultaneous reversible Markov chains. In *Combinatorics, Paul Erdős is Eighty*, Vol. 1, Keszthely, Hungary (1993), 433-451.

Manuscript received April, 1996

Uriel Feige
Department of Applied Mathematics
The Weizmann Institute
Rehovot, Israel.
feige@wisdom.weizmann.ac.il.