Maximum Hitting Time for Random Walks on Graphs

Graham Brightwell, Cambridge University
Peter Winkler, Emory University*

Abstract.

For x and y vertices of a connected graph G, let $T_G(x,y)$ denote the expected time before a random walk starting from x reaches y. We determine, for each $n > 0$, the n-vertex graph G and vertices x and y for which $T_G(x,y)$ is maximized. The extremal graph consists of a clique on $\left\lfloor \frac{2n+1}{3} \right\rfloor$ (or $\left\lceil \frac{2n-2}{3} \right\rceil$) vertices, including x, to which a path on the remaining vertices, ending in y, has been attached; the expected time $T_G(x,y)$ to reach y from x in this graph is approximately $4n^3/27$.

* research supported by ONR grant N00014 85-K-0769

Much attention has recently been focussed on the topic of random walks in graphs: see for instance Aldous [1] and the five subsequent papers in that issue of the *Journal of Theoretical Probability*. In Aleliunas et al. [4], random walks are used to establish the existence of short universal sequences for traversing graphs; in Doyle and Snell [7] they are elegantly associated with electrical networks; in Borre and Meissl [5] they are employed to estimate measurements given by approximate differences. Aldous [2] gives many other contexts in which random walks on graphs arise, and a valuable bibliography [3] compiled by the same author lists many more references on the subject.

A random walk takes place on the vertices of a fixed connected graph G. When the walk is at vertex x, the next step is to one of the neighbors of x, each neighbor chosen with equal probability. The stationary state of the resulting Markov chain is easily computed, (see e.g. [4]), from which we may deduce the simple but remarkable fact that in an infinite such walk, each edge of the graph will be traversed the same proportion of the time; thus, the probability of being at a particular vertex is proportional to its degree. It follows that the *expected return time* for a vertex x, that is, the expected number of steps before a walk commencing at x first returns to x, is equal to twice the number of edges of G divided by the degree of x. Unfortunately, the more general *expected hitting time*, that is, the expected number of steps before a random walk beginning at x first reaches a vertex y, is not so easy to calculate.

Extremal results for random-walk parameters have proven difficult to come by, and expected hitting time is an example. Of course expected hitting time could be as small as 1 (if x has degree 1 and y is its neighbor), but it is not obvious what its maximum value might be in an n-vertex graph. In [8] Lawler proved that this value could not be more than n^3, but (correctly) suspected that the constant factor could be improved. We believe the "maximum hitting time" problem has occurred independently to many people, as it did to us; Lawler apparently heard it first from Paul Erdős.

In this paper we find the n-vertex graphs maximizing the expected hitting time of a vertex y from a fixed starting vertex x. We suspect that our graphs also maximize the *cover time*, i.e., the expected time (from x) to visit all vertices, but have not been able to prove this.

It will be useful at this point to establish some notation. For distinct vertices x and y of a graph G, we define $T_G(x, y)$ to be the expected hitting time of y for a random walk on G starting at x. For W a subset of the set $V(G)$ of vertices of G, let $T_G(x, W)$ be the expected time for a random walk starting from x to reach some vertex of W. It turns out to be convenient for us to consider an altered version of $T_G(x, y)$: for real $M \geq 0$, define $T_G^M(x, y) \equiv T_G(x, y) + Me(G)$, where $e(G)$ denotes the number of edges of the graph G. We shall actually find the graphs G on n vertices (together with distinguished vertices x and y) maximizing $T_G^M(x, y)$ for all real $M \geq 0$. Of course our primary interest is in the case $M = 0$, but in the course of the proofs it will become necessary for us to consider non-zero, even non-integer, M.

For $M \geq 0$, $n \in \mathbb{N}$, we set $T^M(n)$ equal to the maximum of $T_G^M(x, y)$ over all graphs G on n vertices and vertices x, y of G. Call a graph G on n vertices with distinguished vertices x and y (n, M)-*extremal* if $T_G^M(x, y) = T^M(n)$.

We now define a family of graphs which will turn out to include the (n, M)-extremal graphs. For $n, m \in \mathbb{N}$ with $2 \leq m < n$, the n-vertex graph L_n^m consists of a clique on m vertices, including the 'start'-vertex x, and a path of length $t = n - m$ with one end joined to one vertex z of the clique other than x, and the other end of the path being the 'target'-vertex y. See Figure 1. For convenience, we define L_n^n to be the complete graph K_n on n vertices. For $m = n/2$ the graph L_n^m appears in [8] (having been suggested by Erdős) and again in Chandra et al. [6] where it is called a 'lollipop graph', and we shall find it convenient to broaden that term here to cover our two-parameter class: a *lollipop graph* is a graph L_n^m, for $2 \leq m \leq n$.

Figure 1. The lollipop graph L_n^m.

We shall shortly calculate the value of $T_G^M(x, y)$ for $G = L_n^m$ a lollipop graph, but let us see informally why we should expect it to be large. The random walk spends most of its time in the large clique of the graph, taking on average $m - 1$ steps between visits to z. When the walk does hit z, it only steps to the initial vertex of the path with probability $1/m$. Having got as far as the first vertex on the path, the probability that the walk reaches y before returning to z is still only $1/t$. Thus the expected hitting time of y is roughly $m^2 t$. This is maximised, for fixed $n = m + t$, by setting $m = 2n/3$: the value obtained is then about $4n^3/27$. It has been shown several times (Aleliunas et al. [4], Lawler [8]) that $T^M(n) = O(n^3)$: we shall show that the value attained by our lollipop graph is extremal. As M is increased, the (n, M)-extremal lollipop graph has more and more of its vertices in the complete part of the graph, and if $M \geq n-1$, it turns out that the complete graph is (n, M)-extremal.

Our result is as follows.

Theorem.

(i) If $M > n-1$, the complete graph K_n is the only (n, M)-extremal graph.

(ii) If $0 \leq M \leq n-1$, then the lollipop graph L_n^m with $m = \left\lfloor \dfrac{2n+M+1}{3} \right\rfloor$ is (n, M)-extremal. It is the only (n, M)-extremal graph, except when $2n+M+1$ is a multiple of 3, in which case the choice $m = \frac{1}{3}(2n+M-2)$ also yields an (n, M)-extremal lollipop graph.

We define

$$f(n, M) = \begin{cases} T^M_{K_n}(x, y) & (M \geq n-1) \\ T^M_{L_n^m}(x, y) & (M \leq n-1, m = \lfloor \frac{2n+M+1}{3} \rfloor). \end{cases}$$

With this notation, the theorem says that $T_n^M = f(n, M)$. Note that if $M = 0$, then the theorem states that the graph G maximizing $T_G(x, y)$ is a lollipop graph with as nearly as possible $2n/3$ vertices in the complete part.

Before embarking on the proof of the theorem, we put in lemma form the result mentioned above concerning return time. For any vertex x in a graph G, let $T_G(x, x)$ be the expected return time to x, i.e., the average number of steps of a random walk between visits to x.

Lemma 1. *Let x be a vertex of degree $d(x)$ in a graph G. Then $T_G(x, x) = 2e(G)/d(x)$.* □

Proof of the theorem. The first step in the proof is to evaluate the function $f(n, M)$, and to prove that the graphs claimed to be (n, M)-extremal are at least (n, M)-extremal among lollipop graphs (i.e., no n-vertex lollipop graph G, other than those mentioned in the theorem, has $T_G^M(x, y) \geq f(n, M)$).

In the complete graph K_n, the expected time for a random walk starting from x to reach a distinct specified vertex v is just $n-1$. One way to see this is to use Lemma 1: the expected return time from v is exactly n, but a random walk from v consists of a step to some other vertex, which may as well be x, followed by a random walk from x. Thus the expected hitting time of v from x is one less than the expected return time from v.

Now consider a lollipop graph $G = L_n^m$. Since a walk from x to y must go via z, we have:

$$\begin{aligned} T_G(x, y) &= T_G(x, z) + T_G(z, y) \\ &= m - 1 + T_G(z, y). \end{aligned}$$

If the vertices of the path are labelled consecutively $v_0 = z, v_1, \ldots, v_{t-1}, v_t = y$, where $t = n - m$, then

$$\begin{aligned}
T_G(z,y) &= \sum_{i=1}^{t} T_G(v_{i-1}, v_i) \\
&= \sum_{i=1}^{t} T_{G-\{v_{i+1},\ldots,v_t\}}(v_{i-1}, v_i) \\
&= \sum_{i=1}^{t} \left(T_{G-\{v_{i+1},\ldots,v_t\}}(v_i, v_i) - 1\right) \\
&= \sum_{i=1}^{t} (2e(G - \{v_{i+1},\ldots,v_t\}) - 1) \\
&= 2t\binom{m}{2} + \sum_{i=1}^{t}(2i-1) \\
&= tm(m-1) + t^2.
\end{aligned}$$

Thus

$$T^M_{L^m_n}(x,y) = tm(m-1) + t^2 + m - 1 + M\frac{m(m-1)}{2} + Mt. \qquad (*)$$

We now check that the choice of m indicated in the theorem does indeed maximize the quantity $(*)$, for fixed n and M and $t = n - m$.

We have

$$\begin{aligned}
T^M_{L^m_n} - T^M_{L^{m+1}_n} &= \Big[tm(m-1) - (t-1)(m+1)m\Big] + \Big[t^2 - (t-1)^2\Big] + \Big[m - (m+1)\Big] \\
&\quad + \left[M\frac{m(m-1)}{2} - M\frac{(m+1)m}{2}\right] + \Big[Mt - M(t-1)\Big] \\
&= m(m - 2t + 1) + (2t - 1) - 1 - Mm + M \\
&= m(3m - 2n + 1) + 2n - 2m - 2 - M(m-1) \\
&= (m-1)(3m - 2n - M + 2) \\
&\geq 0 \quad \text{iff} \quad m \geq \frac{1}{3}(2n + M - 2).
\end{aligned}$$

So if $M \geq n - 1$, the complete graph is extremal among lollipop graphs, otherwise the maximum of $(*)$ is given by taking m equal to either $\left\lceil \frac{2n+M-2}{3} \right\rceil$ or $\left\lfloor \frac{2n+M+1}{3} \right\rfloor$.

Thus if $M \geq n - 1$, then $f(n, M) = n - 1 + M\binom{n}{2}$; whereas if $M \leq n - 1$, then

$$f(n,M) = \frac{m(m-1)}{2}(2n - 2m + M) + (n-m)(n-m+M) + m - 1,$$

where m is either of the integer values given above.

This expression is a little too cumbersome for our purposes, and we now find some fairly tight bounds for $f(n,M)$ which do not involve evaluation of integer parts.

Let $m = \dfrac{2n+M+\alpha}{3}$, with $-2 \le \alpha \le 1$. Then

$$
\begin{aligned}
T^M_{Q^m_n} &= \frac{(2n+M+\alpha)(2n+M+\alpha-3)(2n+M-2\alpha)}{54} \\
&\quad + \frac{(n-M-\alpha)(n+2M-\alpha)}{9} + \frac{2n+M+\alpha}{3} - 1 \\
&= \left.\frac{(2n+M)^3 - 3(2n+M)^2}{54} + \frac{(n-M)(n+2M)}{9}\right\} K \\
&\quad + (2n+M)\left[\frac{-\alpha^2+3\alpha-2\alpha^2}{54} - \frac{\alpha}{9} + \frac{1}{3}\right] - \frac{2\alpha^2(\alpha-3)}{54} + \frac{\alpha^2}{9} + \frac{\alpha}{3} - 1 \\
&= K + \frac{(2n+M)}{18}(-\alpha^2-\alpha+6) - \frac{\alpha^2(\alpha-6)}{27} + \frac{\alpha}{3} - 1 \\
&\le K + \frac{(2n+M)}{18}\cdot\frac{25}{4} - \frac{13}{27}.
\end{aligned}
$$

It is easy to see that this quantity is minimized over this range of α at $\alpha = 1$, so

$$K + \frac{2}{9}(2n+M) - \frac{13}{27} \le f(n,M) \le K + \frac{25}{72}(2n+M) - \frac{13}{27}. \qquad (**)$$

We shall also make use of a recurrence satisfied by $f(n,M)$.

Lemma 2.

(i) If a vertex y has a unique neighbor y' in a graph G, x is another vertex of G and $M \ge 0$, then $T^M_G(x,y) = T^{M+2}_{G-y}(x,y') + M + 1$.

(ii) For $M \le n-1$, $f(n,M) = f(n-1,M+2) + M + 1$.

Proof. (i) We have $T_G(x,y) = T_{G-y}(x,y') + T_G(y',y)$, and further $T_G(y',y) = T_G(y,y) - 1 = 2e(G) - 1 = 2e(G-y) + 1$. Thus $T^M_G(x,y) = T_{G-y}(x,y') + (M+2)e(G-y) + M + 1$, as desired.

(ii) Let G be (n,M)-extremal among n-vertex lollipop graphs. We know that y has a unique neighbor, so from (i) we see that $G-y$ must be $(n-1,M+2)$-extremal among $(n-1)$-vertex lollipop graphs. But we know that $f(n,M)$ is the largest value of $T^M_G(x,y)$ for an n-vertex lollipop graph, and similarly for $f(n-1,M+2)$. □

We now only have to prove that every (n,M)-extremal graph is a lollipop graph. Our strategy will be as follows. First we show that, in every (n,M)-extremal graph G, the neighborhood of y is complete. We then proceed by induction on the number of vertices, considering two major cases. In the case where y has a unique neighbor, we use Lemma 2(i) above. When y has more than one neighbor (but less than $n-2$: the case when y has $n-2$ neighbors is dealt with separately), we contract the neighborhood $R(y)$ of y to get a lower bound on $T^M_G(x,y)$. Finally, we show that this bound is always less than $f(n,M)$.

Lemma 3. *Fix n and M and let G, x and y be chosen so as to maximize $T_G(x,y)+Me(G)$ over n-vertex graphs. Then the neighborhood $R(y)$ of y is complete.*

Proof. Suppose not, and let a and b be non-adjacent neighbors of y in G. Let A be the graph obtained from G by replacing the edge $\{a,y\}$ by $\{a,b\}$, and similarly let B be obtained from G by replacing $\{b,y\}$ by $\{a,b\}$. We claim that

$$T_G(x,y) < \max\left(T_A(x,y), T_B(x,y)\right),$$

which, since G, A and B all have the same number of edges, would contradict the assumption.

Let S denote the set $\{a,b,y\}$ in any of the three graphs G, A or B. The following quantities do not depend on which of the three graphs is under consideration. They are defined symmetrically in a and b.

$T(x,S)$: expected time for a random walk from x to first hit S;

$p(x,a)$: probability (starting from x) that the random walk hits a before any other vertex of S;

$T(a,S)$: expected time for a random walk from a to next hit S, given that no edge inside S is used (i.e., the walk starts out by leaving S);

$p(a,a), p(a,b)$: probability (starting from a) that the random walk hits a (respectively, b) before any other vertex of S, given that no edge inside S is used;

$d(a)$: degree of a, not counting edges inside S.

We see that

$$T(x,y) = T(x,S) + p(x,a)T(a,y) + p(x,b)T(b,y),$$

and therefore it will suffice to prove that in one of A and B, say C,

$$T_C(a,y) > T_G(a,y) \quad \text{and} \quad T_C(b,y) > T_G(b,y).$$

In fact, we may assume without loss of generality that

$$\frac{1+d(b)T(b,S)}{d(b)p(b,y)} \le \frac{1+d(a)T(a,S)}{d(a)p(a,y)};$$

under this assumption, we will show that

$$T_A(a,y) > T_G(a,y) \quad \text{and} \quad T_A(b,y) > T_G(b,y).$$

We now calculate $T_G(a,y)$ from the various transition probabilities. Working in G, we have:

$$T_G(a,y) = \frac{1}{d(a)+1} + \frac{d(a)}{d(a)+1}\Big(T(a,S) + p(a,a)T_G(a,y) + p(a,b)T_G(b,y)\Big)$$

and

$$T_G(b,y) = \frac{1}{d(b)+1} + \frac{d(b)}{d(b)+1}\Big(T(b,S) + p(b,a)T_G(a,y) + p(b,b)T_G(b,y)\Big)$$

so

$$\Big(1 + d(a) - d(a)p(a,a)\Big)T_G(a,y) = 1 + d(a)T(a,S) + d(a)p(a,b)T_G(b,y)$$

and

$$\Big(1 + d(b) - d(b)p(b,b)\Big)T_G(b,y) = 1 + d(b)T(b,S) + d(b)p(b,a)T_G(a,y).$$

Let $D(a) = 1 + d(a) - d(a)p(a,a)$ and $U(a) = 1 + d(a)T(a,S)$, with $D(b)$ and $U(b)$ defined symmetrically. Note that our earlier assumption now translates as

$$U(b)d(a)p(a,y) \leq U(a)d(b)p(b,y).$$

We now have

$$T_G(a,y) = \frac{D(b)U(a) + d(a)p(a,b)U(b)}{D(a)D(b) - d(a)d(b)p(a,b)p(b,a)},$$

$$T_G(b,y) = \frac{D(a)U(b) + d(b)p(b,a)U(a)}{D(a)D(b) - d(a)d(b)p(a,b)p(b,a)}.$$

Let us repeat this calculation in the graph A. Here

$$T_A(a,y) = \frac{1 + T_A(b,y)}{d(a)+1} + \frac{d(a)}{d(a)+1}\Big(T(a,S) + p(a,a)T_A(a,y) + p(a,b)T_A(b,y)\Big)$$

and

$$T_A(b,y) = \frac{1}{d(b)+2} + \frac{1 + T_A(a,y)}{d(b)+2} + \frac{d(b)}{d(b)+2}\Big(T(b,S) + p(b,a)T_A(a,y) + p(b,b)T_A(b,y)\Big)$$

so

$$D(a)T_A(a,y) = U(a) + (1 + d(a)p(a,b))T_A(b,y)$$

and

$$(D(b)+1)T_A(b,y) = (U(b)+1) + (1 + d(b)p(b,a))T_A(a,y),$$

and therefore

$$T_A(a,y) = \frac{(D(b)+1)U(a) + (1 + d(a)p(a,b))(U(b)+1)}{D(a)(D(b)+1) - (1 + d(a)p(a,b))(1 + d(b)p(b,a))}$$

and

$$T_A(b,y) = \frac{D(a)(U(b)+1) + (1 + d(b)p(b,a))U(a)}{D(a)(D(b)+1) - (1 + d(a)p(a,b))(1 + d(b)p(b,a))}.$$

To show that $T_A(a,y) - T_G(a,y)$ is positive, we observe that it is a positive multiple of

$$\begin{aligned}
&\Big[(D(b)+1)U(a) + (1+d(a)p(a,b))(U(b)+1)\Big]\Big[D(a)D(b) - d(a)d(b)p(a,b)p(b,a)\Big] \\
&\qquad - \Big[D(b)U(a) + d(a)p(a,b)U(b)\Big]\Big[D(a)(D(b)+1) - (1+d(a)p(a,b))(1+d(b)p(b,a))\Big] \\
&\quad = \Big[U(a)+U(b)+1+d(a)p(a,b)\Big]\Big[D(a)D(b) - d(a)d(b)p(a,b)p(b,a)\Big] \\
&\qquad - \Big[D(b)U(a) + d(a)p(a,b)U(b)\Big]\Big[D(a) - 1 - d(a)p(a,b) - d(b)p(b,a)\Big] \\
&\quad > U(a)\Big[D(b)d(a)p(a,b) - d(a)d(b)p(a,b)p(b,a)\Big] - U(b)d(a)p(a,b)\Big[D(a) - 1 - d(a)p(a,b)\Big] \\
&\quad \geq d(a)p(a,b)\Big[U(a)d(b)(1-p(b,a)-p(b,b))\Big] - U(b)d(a)^2p(a,b)\Big[1-p(a,a)-p(a,b)\Big] \\
&\quad \geq 0,
\end{aligned}$$

using our asymmetrical assumption.

Further, $T_A(b,y) - T_G(b,y)$ is a positive multiple of

$$\begin{aligned}
&\Big[D(a)(U(b)+1) + (1+d(b)p(b,a))U(a)\Big]\Big[D(a)D(b) - d(a)d(b)p(a,b)p(b,a)\Big] \\
&\qquad - \Big[D(a)U(b) + d(b)p(b,a)U(a)\Big]\Big[D(a)(D(b)+1) - (1+d(a)p(a,b))(1+d(b)p(b,a))\Big] \\
&\quad = \Big[D(a)+U(a)\Big]\Big[D(a)D(b) - d(a)d(b)p(a,b)p(b,a)\Big] \\
&\qquad - \Big[D(a)U(b) + d(b)p(b,a)U(a)\Big]\Big[D(a) - 1 - d(a)p(a,b) - d(b)p(b,a)\Big] \\
&\quad > D(a)\Big[U(a)d(b)(1-p(b,a)-p(b,b)) - U(b)d(a)(1-p(a,a)-p(a,b))\Big] \\
&\quad \geq 0
\end{aligned}$$

as desired, proving the lemma. □

We have shown that, in any (n,M)-extremal graph, the neighborhood $R(y)$ of y is a clique. In our lollipop graphs, $R(y)$ consists either of y and one other vertex or of the whole graph: the two extreme possibilities.

We need to show that any graph which is not a lollipop graph cannot be (n,M)-extremal: the result will then follow by our calculations on lollipop graphs. Suppose this statement is false, and let G (with distinguished vertices x and y) be an (n,M)-extremal graph, not a lollipop graph, with a minimum number of vertices. It is easy to check that G must have at least 5 vertices.

We suppose first that $r = |R| = 2$, so that y has a unique neighbor v in G. Let $G^- = G - y$. By Lemma 2(i), we have

$$T_G^M(x,y) = T_{G^-}^{M+2}(x,v) + M + 1.$$

If G^- with target-vertex v is not $(n-1, M+2)$-extremal, then we can replace G^- by an $(n-1, M+2)$-extremal graph, and attach y to the target-vertex of that graph, producing a graph showing that G is not (n, M)-extremal, contrary to hypothesis. Thus G^- is $(n-1, M+2)$-extremal, and so a lollipop graph. Hence G is also a lollipop graph, again a contradiction. So $r \geq 3$.

If $r = n$, then G is the lollipop graph K_n, and we are done. We shall also deal with the case $r = n-1$ separately before proceeding with the general case. Suppose then that y is adjacent to all but one other vertex of G: clearly we may as well take x to be this vertex. If x is adjacent to just s vertices of G, a calculation reveals that $T_G(x, y) = n + 1 - \dfrac{2(n-s-2)}{n-1}$, so $T_G^M(x, y) = M\dbinom{n}{2} - M(n-s-1) + n + 1 + \dfrac{2(n-s-2)}{n-1}$. For $M \geq 2$, this is less than $T_{K_n}^M(x, y) = n - 1 + M\binom{n}{2}$; for $M \leq 2$, it is less than $T_{L_n^{n-1}}^M(x, y) = n - 3 + (M+2)\binom{n-1}{2}$.

From now on, we may and shall assume that $3 \leq r \leq n-2$.

The next result is designed to give an upper bound on $T_G(x, y)$ in graphs where the size of $R(y)$ is in this intermediate range. For R a subset of the vertices of G, let G/R denote the "quotient" graph obtained from G by contracting R to a single vertex (also denoted R) and identifying any resulting multiple edges.

Lemma 4. *Let x and y be vertices of a graph G such that $R \equiv R(y)$ is a clique in G, and let $r = |R|$. Then*

$$T_G(x, y) \leq T_{G/R}(x, R) + \frac{4}{r}e(G) - r + 1.$$

Proof. Certainly $T_G(x, R) \leq T_{G/R}(x, R)$, so it suffices to prove that

$$\max_{w \in R} T_G(w, y) \leq \frac{4}{r}e(G) - r + 1.$$

Let w be a vertex in R maximizing $T_G(w, y)$, and suppose the degree of w in G is $r + k - 1$, so w sends k edges out of R. Let G' be the graph formed by removing from G the $\binom{r}{2}$ edges inside R.

We next consider the expected time from w to y given that the walk starts out from w by leaving R. This is at most $T_{G'}(w, R) + \max_{u \in R} T_G(u, y) \leq T_{G'}(w, w) + T_G(w, y)$, and Lemma 1 tells us that $T_{G'}(w, w) = 2e(G')/k = 2\left(e(G) - \binom{r}{2}\right)/k$.

Thus

$$T_G(w, y) \leq \frac{1}{r+k-1}\left[1 + \sum_{\substack{u \in R \\ u \neq z, y}} (T_G(u, y) + 1) + k\left(\frac{2\left(e(G) - \binom{r}{2}\right)}{k} + T_G(w, y)\right)\right]. \qquad (***)$$

What is $T_G(y,y)$? On the one hand, by Lemma 1 it is $2e(G)/(r-1)$: on the other it is $\frac{1}{r-1}\sum_{\substack{u\in R\\ u\neq y}}(T_G(u,y)+1)$. So

$$\sum_{\substack{u\in R\\ u\neq w,y}}(T_G(u,y)+1)=2e(G)-T_G(w,y)-1,$$

and hence $(***)$ yields:

$$(r+k-1)T_G(w,y)\leq 1+2e(G)-T_G(w,y)-1+2e(G)-r(r-1)+kT_G(w,y),$$

and therefore $T_G(w,y)\leq \frac{4e(G)}{r}-(r-1)$, as claimed. □

From Lemma 4, we see that

$$\begin{aligned}T_G^M(x,y) &\leq T_{G/R}(x,R)+(M+4/r)e(G)-r+1\\ &\leq T_{G/R}(x,R)+(M+4/r)\left[e(G/R)+\binom{r}{2}+(n-r)(r-2)\right]-r+1,\end{aligned}$$

since contracting R removes at most $\binom{r}{2}$ edges inside r and at most $(n-r)(r-2)$ edges from $V-R$ to R (since y has no neighbors outside R).

Therefore

$$\begin{aligned}T_G^M(x,y) &\leq T^{M+4/r}(n-r+1)+(M+4/r)\left(\binom{r}{2}+(n-r)(r-2)\right)-r+1\\ &\leq f(n-r+1,M+4/r)+(M+4/r)\left(\binom{r}{2}+(n-r)(r-2)\right)-r+1,\end{aligned}$$

by the minimality of G, and it is now sufficient to prove that this quantity is less than $f(n,M)$.

Let

$$\Delta=\Delta(n,M,r)=f(n,M)-f(n-r+1,M+4/r)-(M+4/r)\left(\binom{r}{2}+(n-r)(r-2)\right)+r-1.$$

We shall show that $\Delta(n,M,r)>0$ for $0\leq M$ and $3\leq r\leq n-2$.

The formula for $f(n,M)$ depends on whether the proposed (n,M)-extremal graph is the complete graph or not. Thus we have three cases to consider.

Case 1. $M\geq n-1$. In this case, $f(n,M)=n-1+M\binom{n}{2}$ and $f(n-r+1,M+4/r)=n-r+(M+4/r)\binom{n-r+1}{2}$, and thus

$$\begin{aligned}\Delta(n,M,r) &= 2(r-1)+(M+4/r)\left[\binom{n}{2}-\binom{n-r+1}{2}-\binom{r}{2}-(n-r)(r-2)\right]-\frac{4}{r}\binom{n}{2}\\ &= 2(r-1)+(M+4/r)(n-r)-2n(n-1)/r\\ &\geq 2(r-1)+(n-1+4/r)(n-r)-2n(n-1)/r \quad \equiv Y(n,r).\end{aligned}$$

The quantity $rY(n,r) = 2r^2 - 6r + r(n-1)(n-r) - 2n^2 + 6n$ is 0 if $n = r$, and we now show that it is increasing in n for each fixed $r \geq 3$. Indeed

$$\begin{aligned}\frac{\partial}{\partial n} rY(n,r) &= r(2n - r - 1) - 4n + 6 \\ &\geq 3(2n-4) - 4n + 6 \\ &= 2n - 6 > 0\end{aligned}$$

for $n > r \geq 3$. Thus $rY(n,r) > 0$ for $n > r \geq 3$, and hence $\Delta(n, M, r) > 0$.

Case 2. $n-1 \geq M \geq n-r-4/r$. In this range, the complete graph K_{n-r+1} is $(n-r+1, M+4/r)$-extremal among lollipop graphs, so, as in Case 1,

$$\Delta(n, M, r) = f(n, M) - n + 2r - 1 - (M + 4/r)\left(\binom{n}{2} - n + r\right).$$

We now use Lemma 2 ($f(n,M) = f(n-1, M+2) + M + 1$ for $M \leq n-1$) repeatedly. In each range $n - 3(j-1) - 1 \geq M \geq n - 3j - 1$ ($j \geq 1$), we apply this lemma j times to arrive at:

$$f(n, M) = f(n-j, M+2j) + jM + j^2 = (n-j-1) + (M+2j)\binom{n-j}{2} + jM + j^2.$$

Thus in each range,

$$\Delta(n, M, r) = A_j + M\left[\binom{n-j}{2} + j - \binom{n}{2} + n - r\right],$$

where A_j is independent of M.

The quantity in square brackets is negative for each $j \geq 1$, so $\Delta(n, M, r)$ decreases with M over each range, and so over the entire range of M covered by this Case.

Case 3. $M \leq n - r - 4/r$. In this final range, both values of f are given by non-complete lollipop graphs. It seems that we need to estimate these values of f fairly precisely.

Lemma 5. *If $M + 4/r \leq n - r$, and $3 \leq r \leq n - 2$, then*

$$f(n, M) - f(n-r+1, M+4/r) > \frac{1}{9}(2n+M)(2n+M-2r+4/r)(r-1-2/r) + M/2.$$

Proof. Recall (**), which states that, for $M \leq n-1$,

$$f(n, M) = \frac{(2n+M)^3 - 3(2n+M)^2}{54} + \frac{(n-M)(n+2M)}{9} + g(n, M),$$

where

$$\frac{2}{9}(2n+M) - \frac{13}{27} \le g(n,M) \le \frac{25}{72}(2n+M) - \frac{13}{27}.$$

We show first that

$$M/2 < W \equiv \frac{(n-M)(n+2M)}{9} + g(n,M) - \\ \frac{(n-M-r+1-4/r)(n+2M-r+1+8/r)}{9} - g(n-r+1, M+4/r).$$

We have:

$$\begin{aligned}
9W &\ge (n-M)(r-1-8/r) + (n+2M)(r-1+4/r) - (r-1-8/r)(r-1+4/r) \\
&\quad + 9\left(\frac{2}{9}(2n+M) - \frac{13}{27}\right) - 9\left(\frac{25}{72}(2n+M) - \frac{13}{27}\right) \\
&= n(2r-2-4/r+4-25/4) + M(r-1+16/r+2-25/8) - (r-1-8/r)(r-1+4/r) \\
&\ge n(2r-4/r-17/4) + 5M - (r-1-8/r)(r-1-4/r).
\end{aligned}$$

For $r=3$, each term is positive, and so $9W > 5M$. For $r \ge 4$:

$$9W - 5M \ge n(r-1-2/r) - (r-1-2/r)^2 > 0.$$

Thus $W > 5M/9 \ge M/2$, as required.

So it is now sufficient to prove that

$$h(2n+M) - h(2n+M-2r+2+4/r) \ge \frac{1}{9}(2n+M)(2n+M-2r+4/r)(r-1-2/r),$$

where

$$h(a) = \frac{a^3 - 3a^2}{54}.$$

Setting $U = 2n - M$ and $t = 2r - 2 - 4/r$, we see that this is equivalent to proving:

$$h(U) - h(U-t) \ge \frac{1}{18}Ut(U-t-2).$$

This is straightforward enough, since

$$\begin{aligned}
54[h(U) - h(U-t)] &= U^3 - U^3 + 3U^2t - 3Ut^2 + t^3 - 3U^2 + 3U^2 - 6Ut + 3t^2 \\
&= 3(U^2t - Ut^2 - 2Ut) + t^2(t+3) \\
&\ge 3Ut(U-t-2).
\end{aligned}$$

□

Recalling the definition of $\Delta(n, M, r)$, we see that it is now sufficient to prove:

$$\frac{1}{9}(2n+M)(2n+M-2r+4/r)(r-1-2/r)+M/2$$
$$-(M+4/r)\left(\binom{r}{2}+(n-r)(r-2)\right)+r-1 \qquad \geq 0$$

for $3 \leq r \leq n-2$ and $0 \leq M \leq n-r-4/r$.

Now, the quantity above is easily seen to be at least

$$\frac{1}{9}(2n+M)(2n+M-2r+4/r)(r-1-2/r)-(M+4/r)(r-2)(n-r/2+1).$$

Dividing this through by $r-2$, we obtain:

$$\frac{1}{9}(2n+M)(2n+M-2r+4/r)(1+1/r)-(M+4/r)(n-r/2+1),$$

which we denote by Z.

Over our current range of M, Z is decreasing with M, since

$$\begin{aligned}\frac{\partial Z}{\partial M} &= \frac{1+1/r}{9}(4n+2M-2r+4/r)-(n-r/2+1)\\ &\leq \frac{1+1/r}{9}(6n-4r-4/r)-(n-r/2+1)\\ &\leq \frac{4/3}{9}\cdot 6(n-r/2+1)-(n-r/2+1) \qquad < 0.\end{aligned}$$

Hence Z is minimized for $M = n-r-4/r$, when

$$\begin{aligned}Z &= \frac{1}{9}(3n-r-4/r)(3n-3r)(1+1/r)-(n-r)(n-r/2+1)\\ &= \frac{n-r}{3}\Big[(3n-r-4/r)(1+1/r)-3(n-r/2+1)\Big].\end{aligned}$$

The expression inside the square brackets increases with n for fixed r, so is minimized if $n = r+2$. Hence,

$$\begin{aligned}Z &\geq \frac{n-r}{3}\Big[(2r+6-4/r)(1+1/r)-3(r/2+3)\Big]\\ &= \frac{n-r}{3}\Big[r/2-1+2/r-4/r^2\Big]\\ &= \frac{n-r}{3}(r/2-1)(1+4/r^2) \qquad > 0.\end{aligned}$$

This completes the proof of the theorem. □

Acknowledgments.

This work was done while the first author was visiting at Emory University. Special thanks are due to Herb Wilf for bringing the subject of extremal problems in random walks on graphs to the authors' attention, and to David Aldous for helping with background and sending a copy of his bibliography.

References.

[1] D.J.Aldous, An introduction to covering problems for random walks on graphs, *J. Theoretical Probab.* **2** (1989), 87-89.

[2] D.J. Aldous, Applications of random walks on graphs, preprint (1989).

[3] D.J. Aldous, Bibliography: random walks on graphs, preprint (1989).

[4] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovasz, and C. Rackoff, Random walks, universal traversal sequences, and the complexity of maze problems, *20th Annual Symposium on Foundations of Computer Science*, San Juan, Puerto Rico (October 1979), 218-223.

[5] K. Borre and P. Meissl, *Strength Analysis of Leveling-type Networks*, Geodaetisk Institut, Copenhagen (1974), vol. 50.

[6] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky and P. Tiwari, The electrical resistance of a graph captures its commute and cover times, *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, Seattle, Washington (May 1989), 574-586.

[7] P.G. Doyle and J.L. Snell, *Random Walks and Electric Networks*, Mathematical Assoc. of America (1974), Washington, DC.

[8] G.F. Lawler, Expected hitting times for a random walk on a connected graph, *Discrete Math.* **61** (1986), 5-15.