Maximum Hitting Time
Longquan Qiao
February 13, 2013

Abstract

For a specific connected, finite graph, there is always a maximum hitting time between any two different vertices in the graph. We call it the hitting time of G to avoid confusion, though it is the maximum hitting time in the graph.

However, among various connected, finite graphs, we may be curious about the maximum value of hitting time for different graphs. This essay will give a thorough description and proof of the type of graph when the hitting time reaches its maximum value. The proof process will begin with a stronger induction and converge to the specific kind of graph eventually.

The graph we need to find is called lollipop, formed by a clique and a path, with the maximum hitting time $O[n^3]$, while the path only graph has a hitting time $O[n^2]$. Finally, we will calculate to get the special '21-lollipop' with the longest hitting time among all lollipops with different vertices distribution on clique and path.

1 Introduction

The path graph is shown below. When starting at point i and finishing at point j, the expected time to consume is $(j - i)^2$. This result can be obtained by adding $1, 3, 5 \ldots 2*(j – i) – 1$, which are the expected walking times for each edge from j to i. Therefore, the hitting time of path graph is $(n – 0)^2$, which means the time complexity is $O[n^2]$.

![Figure 1: path graph, length n+1](image)

The lollipop graph is coming from the name ‘lollipop’ intuitively. It contains two parts: a clique, which is a complete graph, and a path connected to the clique. The complete graph means there is an edge between any pair of vertices. For convenience, we will call a random point x in the clique, call the end of the path y and call the connected point z in both clique and path. In general, $L^m_n$ means a lollipop with n vertices and m of n vertices on the clique. In an extreme situation, all the vertices are in the clique, which is $L^m_n$, we can use $K_n$ instead. Finally, we will calculate to get the special '21-lollipop' with the longest hitting time among all lollipops with different vertices distribution on clique and path.

When walking on the lollipop graph, starting from clique and trying to reaching the end of the path, it will take a longer time than just a path which is easy to figure out based on the graph. However, the walk from x through z to y can be divided into
two parts: escaping from the clique and reaching Y before getting back to clique. The escaping probability is \(P(\text{reaching } z) \times P(\text{leaving clique})\). And the probability to reach \(y\) before back to \(z\) is \(1/(n - m)\). Therefore, the expected hitting time is \((n - 1) \times n \times (n - m)\), whose complexity can be denoted as \(O[n^3]\).

![Figure 2: various lollipop graphs](image)

2 Preliminaries

Firstly, to represent the hitting time, we have the following definition: for distinct vertices \(x\) and \(y\) of a graph \(G\), let’s define \(T_G(x, y)\) to be the expected hitting time of \(y\) for a random walk on \(G\) starting at \(x\).\(^3\)

However, to prove the lollipop graph is the one we need seems impossible because there are no existing rules or models which are suitable for all kinds of graphs to compare them in the same situation. So it is very necessary to build a model to uniform all graphs to compare and calculate the maximum value.

Based on this idea, and thanks to Professor Peter Winkler’s help, the following function can provide a stronger introduction by ignoring the structure of graphs and only paying attention to the number of edges, vertices and the neighbor of \(y\).

Here is a sound and altered version of \(T_G(x, y)\): for \(M \geq 0\), define \(T_G^M(x, y) = T_G(x, y) + M \times e(G)\), where \(e(G)\) is the number of edges of graph \(G\). Thus, if we find the lollipop graph can get the maximum value for \(T_G^M(x, y)\), then we will definitely get the result we want to prove (case \(M = 0\)).

Before to prove, I need to set the \(T^M(n)\) equal to the maximum of \(T_G^M(x, y)\) over all graphs \(G\) on \(n\) vertices. We may call the graph \(G\) ‘\((n, M)\)-extremal’ if \(T_G^M(x, y) = T^M(n)\) for future convenience.\(^3\)

As mentioned above, we need connect this function to our final result — lollipop graph. It is a useful notation to define a function \(f(n, M)\) as below.
\[ f(n, M) = T_{R_n}^{\text{III}}(x, y) \quad (M \geq n - 1) \]
\[ f(n, M) = T_{c_{n, m}}^{\text{M}}(x, y) \quad (M \leq n - 1) \]

3 Full Proof

For the full proof part, we need to consider the following two statements. If both of them hold, then we get the conclusion of lollipop above.

1. In every \((n, M)\)-extremal graph \(G\), the set of all neighbors of \(y\), \(R(y)\), is a complete sub-graph of \(G\).
2. Any graph which is not a lollipop graph cannot be \((n, M)\)-extremal.

Now let’s begin to prove the first statement.

Consider the following three graphs: \(G\), \(A\) and \(B\).

![Graph G, A and B.](image)

Suppose \(a\) and \(b\) are non-adjacent neighbors of \(y\). In graph \(A\), I remove edge \(a-y\) and add edge \(a-b\). In graph \(B\), I remove edge \(b-y\) and add edge \(a-b\). We use this case to claim that the neighbors of vertex \(y\) must form a completer clique. Otherwise, either \(A\) or \(B\) will use a longer time to reach \(y\) than graph \(G\). And because \(G\), \(A\) and \(B\) have the same number of edges, \(G\) cannot be \((n, M)\)—extremal if \(y\) has two non-adjacent neighbors.

Let \(S\) denote the set \(\{a, b, y\}\) in any of the three graphs. We have:

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

\(p(x, a)\): probability of random walk, starting at \(x\), hits \(a\) before \(b\) and \(y\).

\(T(a, S)\): expected time for a random walk from \(a\), leaving \(S\), to next hit \(S\).

\(p(a, a/b)\): probability of random walk, starting from \(a\) and leaving \(S\), hits \(a/b\) before \(b\) and \(y/a\) and \(y\).

\(d(a)\): degree of \(a\) without considering \(y\) and \(b\).

If we can prove that in one of \(A\) and \(B\), represented as \(C\), \(T_C(a, y) > T_G(a, y)\) and \(T_C(b, y) > T_G(b, y)\) will sufficient for our statement one.

\[ T_G(a, y) = \frac{1}{d(a) + 1} + \frac{d(a)}{d(a) + 1} \left( T(a, S) + p(a, a)T_G(a, y) + p(a, b)T_G(b, y) \right) \]
\[ T_G(b, y) = \frac{1}{d(b) + 1} + \frac{d(b)}{d(b) + 1} \left( T(b, S) + p(b, a)T_G(a, y) + p(b, b)T_G(b, y) \right) \]
Let \( D(a) = 1 + d(a) - d(a)p(a, a) \) and \( U(a) = 1 + d(a)T(a, S) \) and \( D(b), U(b) \) as the same. Without loss of generality, we assume that \( U(b)d(a)p(a, y) \leq U(a)d(b)p(b, y) \).

\((*)\)

\[
T_G(a, y) = \frac{D(b)U(a) + d(a)p(a,b)U(b)}{D(a)D(b) - d(a)d(b)p(a,b)p(b,a)} \quad \text{note as} \quad \frac{X_{Ga}}{Y_{Ga}}
\]

\[
T_G(b, y) = \frac{D(a)U(b) + d(b)p(b,a)U(a)}{D(a)D(b) - d(a)d(b)p(a,b)p(b,a)} \quad \text{note as} \quad \frac{X_{Gb}}{Y_{Gb}}
\]

Now we repeat the calculation steps for graph A.

We have:

\[
T_A(a, y) = \frac{(D(b)+1)U(a)+(1+d(a)p(a,b))(U(b)+1)}{D(a)(D(b)+1) - (1+d(a)p(a,b))(1+d(b)p(b,a))} \quad \text{note as} \quad \frac{X_{Aa}}{Y_{Aa}}
\]

\[
T_A(b, y) = \frac{D(a)(U(b)+1)+(1+d(b)p(b,a))U(a)}{D(a)(D(b)+1) - (1+d(a)p(a,b))(1+d(b)p(b,a))} \quad \text{note as} \quad \frac{X_{Ab}}{Y_{Ab}}
\]

Now it’s time to see the result:

\[
T_A(a, y) - T_G(a, y) \Rightarrow X_{Aa}*Y_{Ga} - X_{Ga}*Y_{Aa} \\
\geq d(a)p(a, b)[U(a)d(b)(1 - p(b, a) - p(b, b))] - U(b)d(a)p(a,b)[1 - p(a, a) - p(a, b)]
\]
\[
\geq 0;
\]

It holds by using our assumption \((*)\) above.

\[
T_A(b, y) - T_G(b, y) \Rightarrow X_{Ab}*Y_{Gb} - X_{Gb}*Y_{Ab} \\
\geq D(a)[U(a)d(b)(1 - p(b, a) - p(b, b))] - U(b)d(a)(1 - p(a, a) - p(a, b))
\]
\[
\geq 0. \quad [3]
\]

Therefore, we successfully proved the statement one. Because there is always at least one graph having a longer hitting time to \( y \) than the graph \( G \) (a and b are two non-adjacent neighbors of \( y \)) with a specific \( n \) and \( M \).

We can put this into another way:

Fix \( n \) and \( M \), if \((n, M)\)—extremal value is obtained from graph \( G \), the neighborhood \( R(y) \) of \( y \) is a clique.

It means \( R(y) \) may consist only one vertex \( y’ \) or consist the whole graph except \( y \). That’s exactly what we want as a lollipop graph.

Before proving the statement two, here is a useful lemma:

Lemma 1: If a vertex \( y \) has a unique neighbor \( y’ \), vertex \( x \) is not as same as \( Y \), we have:

\[
(i) \quad T_G^M(x, y) = T_{G-y}^M(x, y') + M + 1.
\]

\[
(ii) \quad \text{And for } M \leq n - 1, f(n, M) = f(n - 1, M + 2) + M + 1.
\]

It is not difficult to prove this lemma \((i)\): because we have \( T_G(x, y) = T_{G-y}(x, y') + T_G(y', y) \), and \( T_G(y', y) = T_G(y, y) - 1 = 2*e(G) - 1 = 2*e(G - y) + 1 \).

When we replace the graph \( G \) with the lollipop graph \( L_n^m \), it is obvious that the lemma \((ii)\) holds. [3]
Having this lemma, we can prove the second statement now:
Assume we have a graph $G$ is $(n, M)$–extremal but not a lollipop graph.

Case 1: $|R(y)| = 2$, by lemma (i), we know $G - y'$ is $(n - 1, M + 2)$–extremal if $G$ is $(n, M)$–extremal. Since $G - y'$ is a lollipop graph, $G$ is also a lollipop, contradicting the assumption.

Case 2: $|R(y)| = n$, then $G$ is complete graph $K_n$, contradicting the assumption.

Case 3: $|R(y)| = n - 1$, in this case, $y$ is connected to all vertices except $x$. We have:

$$T^M_G(x, y) = M \binom{n}{2} - M(n - s - 1) + n + 1 + \frac{2(n-s-2)}{n-1}. \quad \text{For } M \geq 2, \text{ it is less than } (n - 1 + M \binom{n}{2}),$$

which is $T^M_K_n(x, y)$. For $M \leq 2$, it is less than $(n - 3 + (M + 2) \binom{n-1}{2})$,

Case 4: $|R(y)| = r \in [3, n - 2]$, this is the last part we need to prove. We need another lemma to help work it out.

Lemma 2: $T_G(x, y) \leq T_{G/R}(x, R) + 4*e(G)/r - r + 1$, where $G/R$ means contracting all the neighbors of $y$ to one point.

It is obvious that $T_G(x, R) \leq T_{G/R}(x, R)$. So it is sufficient to prove

$$\max_{w \in R} T_G(w, y) \leq 4*e(G)/r - r + 1. \text{ Get the maximum one of } T_G(w, y) \text{ in } R \text{ and suppose the } d(w) = r + k - 1. \text{ The vertex } w \text{ sends out } k \text{ edges out of } R \text{ and } G' \text{ is the graph formed by removing } \binom{r}{2} \text{ edges inside } R \text{ from } G. \text{ Because } T_{G'}(w, R) + \max_{u \in R} T_G(u, y) \leq T_G(w, w) + T_G(w, y), \text{ we have:}

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

Because: $\sum_{u \in R, u \neq z, y} (T_G(u, y) + 1) = 2e(G) - T_G(w, y) - 1$, we can get the following:

$$(r + k - 1)T_G(w, y) \leq 1 + 2e(G) - T_G(w, y) - 1 + 2e(G) - r(r - 1) + kT_G(w, y) \leq (4*e(G)/r - r + 1)(r + k - 1),$$

which proves the lemma 2.

By Lemma 2, we have:

$$T^M_G(x, y) \leq T_{G/R}(x, R) + (M + 4/r)e(G) - r + 1 \leq T_{G/R}(x, R) + (M + 4/r)(e(G/R) + \binom{r}{2} + (n - r)(r - 2)) - r + 1.$$ 

Therefore,

$$T^M_G(x, y) \leq T^{M+4/r}(n - r + 1) + (M + 4/r)[\binom{r}{2} + (n - r)(r - 2)] - r + 1 \leq f(n - r + 1, M + 4/r) + (M + 4/r)[\binom{r}{2} + (n - r)(r - 2)] - r + 1.$$

As coming to so far, we shall show that

$$\Delta = f(n, M) - f(n - r + 1, M + 4/r) - (M + 4/r) \left[ \binom{r}{2} + (n - r)(r - 2) \right] + r - 1 > 0$$

when $M \geq 0$ and $r \in [3, n - 2]$. 
When $M \geq n - 1$, $\Delta = 2(r - 1) + (M + 4/r)(n - r) - 2n(n - 1)/r$
\[ \geq 2(r - 1) + (n - 1 + 4/r)(n - r) - 2n(n - 1)/r \]
\[ > 0. \] While $n > r \geq 3$.

When $n - 1 \geq M \geq n - r - 4/r$, $K_{n-r+1}$ is $(n - r + 1, M + 4/r)$-extremal among lollipop graphs, it means we can prove as $M \geq n - 1$.

When $M \leq n - r - 4/r$, we can divide this into several smaller ranges to prove it one by one and get the final result is always greater than zero. [3]

4 Final Conclusion

After finding out the lollipop graph, our task is now to calculate how to divide the $n$ vertices to get the maximum hitting time among various lollipop graphs.

Since $T_G(x, y) = T_G(x, z) + T_G(z, y) = m - 1 + T_G(z, y)$.

Now I label the path as $z(v_0), v_1, v_2 \ldots v_{t-1}, y(v_t)$. $T_G(z, y) = \sum_{i=1}^{t} T_G(v_{i-1}, v_{i})$
\[ = \sum_{i=1}^{t} (2e(G - \{v_{i+1}, \ldots, v_{t}\}) - 1) = 2t \left( \binom{M}{2} + \sum_{i=1}^{t} (2i - 1) \right) = t^2m^*(m - 1) + t^2. \]

Consider $T^{M}_{ln}(x,y) - T^{M}_{ln+1}(x,y) = (m - 1)(3m - 2n - M + 2) \geq 0$, we can get:
\[ m \geq \frac{1}{3}(2n + M - 2) \]

When $M = 0$, $m = \left[ \frac{2n-2}{3} \right]$, the lollipop can get the maximum hitting time from $x$ to $y$. We may call this distribution ’21-lollipop’ with $\frac{2}{3}$ vertices on the clique and $\frac{1}{3}$ on the path approximately.

In a word, the lollipop is the graph having the maximum hitting time among all kinds of graphs. The ’21-lollipop’ graph is the lollipop having the maximum hitting time among all lollipops with a fixed total vertex number $n$.

Finally, for all lollipops, the complexity of hitting time is $O[n^3]$ and for the special ’21-lollipop’, the hitting time is $\frac{4}{27}n^3$ approximately. [1][2][3]

References