Bounding the Sum of Distances of the Vertices of a Tree

JOSELITO A. UY

The distance $d(x,y)$ between the vertices x and y of a connected graph G is defined to be the length of a shortest path in G which joins x and y. It is clear that $d(x,y) = d(y,x)$, $d(x,y) \geq 0$ and $d(x,x) = 0$.

In this paper we will find a lower bound and an upper bound for the sum $\sum_{x,y\in V} d(x,y)$ of distances of vertices of a tree T = (V,E). To do this we shall investigate the family $\mathfrak{I}$ of trees of order n and find out which trees in $\mathfrak{I}$ will give extremum values for $\sum_{x,y\in V} d(x,y)$.

Lemma 1. $\sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1)$.

Proof We prove by induction on n. Let $S_n = \sum_{i=1}^{n} i^2$ and $f(n) = \frac{1}{6}n(n+1)(2n+1)$. Then $S_1 = 1 = f(1)$ and $S_2 = 5 = f(2)$. Suppose that $n > 2$ and $S_{n-1} = f(n-1)$. Now

$$S_n = S_{n-1} + n^2 = \frac{1}{6}(n-1)n(2n-1) + n^2$$

$$= \frac{1}{6}n(n+1)(2n+1).$$

This means that $S_n = f(n)$ for $n > 2$. □

Lemma 2. Let T = (V,E) be a tree of order n. (1) If $T = K_{1,n-1}$ (star graph), then

$$\sum_{x,y\in V} d(x,y) = 2(n-1)^2.$$

JOSELITO A. UY is a Professor of Mathematics at MSU-IIT. His paper was presented during the "22nd Annual Convention of the Mathematical Society of the Philippines" on 27-28 May 1995 at the Polytechnic University of the Philippines, Sta. Mesa, Manila.

(2) If $T = P_n$ (path), then $\sum_{x,y \in V} d(x,y) = \frac{1}{3}n(n^2 - 1)$.

Proof. (1) Let $T = K_{1,n-1}$. Then consider the following cases:

Case 1. $d(x,y) = 1$. Then there are $2(n-1)$ terms of $\sum_{x,y\in V} d(x,y)$ for this case.

Case 2. $d(x,y) = 2$. Then there are $2\binom{n-1}{2} = (n-1)(n-2)$ terms of $\sum_{x,y\in V} d(x,y)$ for this case.

Case 3. $d(x,y) > 2$. Then there are no terms of $\sum_{x,y\in V} d(x,y)$ for this case Hence

$$\sum_{x,y \in V} d(x,y) = 2(n-1) + 2(n-1)(n-2) = 2(n-1)^2.$$

(2) For the second part we take $T = P_n = [1, 2, ..., n]$. Then consider the following cases:

Case 1. $d(x,y) = 1$. Then there are $2(n-1)$ terms of $\sum_{x,y\in V} d(x,y)$ in this case.

Case 2. $d(x,y) = 2$. Then $[1, 2, 3]$, $[2, 3, 4]$, ..., $[n-2, n-1, n]$ are the subtrees of T whose lengths are each equal to 2. It follows that there are $2(n-2)$ terms of $\sum_{x,y\in V} d(x,y)$ in this case.

Case 3. $d(x,y) = 3$. Then $[1, 2, 3, 4]$, $[2, 3, 4, 5]$, ..., $[n-3, n-2, n-1, n]$ are the subtrees of T whose lengths are each equal to 3. Hence there are $2(n-3)$ terms of $\sum_{x,y\in V} d(x,y)$ in this case.

We continue in this fashion until we reach the last case where $d(x,y) = n-1$. There are only 2 terms of $\sum_{x,y\in V} d(x,y)$ in this case. So for any tree $T = (V, E)$ of order n,

$$\sum_{x,y \in V} d(x,y) = 2[(n-1) + 2(n-2) + 3(n-3) + \ldots + (n-1)]$$

$$= 2\sum_{i=1}^{n-1} i(n-i) = 2\left[n\sum_{i=1}^{n-1} i - \sum_{i=1}^{n-1} i^2\right]$$

$$= 2\left[\tfrac{1}{2}n^2(n-1) - \sum_{i=1}^{n-1} i^2\right].$$

Using Lemma 1, we can simplify the last expression above to obtain

$$\sum_{x,y \in V} d(x,y) = n^2(n-1) - 2\sum_{i=1}^{n-1} i^2$$

$$= n^2(n-1) - \tfrac{1}{3}n(n-1)(2n-1)$$

$$= \tfrac{1}{3}n(n^2-1). \quad \square$$

Theorem 3. Let $\mathfrak{J}$ be the family of trees T = (V,E) of order n. Then

$$\min\left\{\sum_{x,y \in V} d(x,y) : T \in \mathfrak{J}\right\} = 2(n-1)^2.$$

Proof. A tree of order n has $n - 1$ edges. Then in the sum $\sum_{x,y \in V} d(x,y)$ there are exactly $2(n-1)$ terms each equal to 1 while the other nonzero terms are greater than or equal to 2. If $T = K_{1,n-1}$, then all the terms different from 0 and 1 are each equal to 2. If $T \neq K_{1,n-1}$, then there is at least one term which is equal to 3. Hence, the minimum value of $\sum_{x,y \in V} d(x,y)$ is attained only if $T = K_{1,n-1}$. Lemma 2 completes the proof of this theorem. $\square$

Lemma 4. Let $\mathfrak{J}$ be the family of trees T = (V,E) of order n. The sum

$$s(x) = \sum_{y \in V} d(x,y)$$

is a maximum in $\mathfrak{J}$ if T is a path and x is an end-vertex.

Proof. Let T be of order n. For $s(x)$ to be maximum, the $n - 1$ edges of T must be constructed in such a way that $d(x,y)$ is as large as possible for each $y \in V$. To do this, let $x_1, x_2, \ldots, x_n$ be the vertices of T where the indices correspond to the order in which T is contructed. Without loss of generality, let $x = x_1$. We begin with the edge $[x,x_2]$. For $d(x,x_3)$ to be maximum, we require that $[x_2,x_3] \in E$. Also, $[x_3,x_4]$ must be an edge of T for $d(x,x_4)$ to be maximum. Continuing in this manner, we can eventually require that $[x_{n-1}, x_n] \in E$. Hence it is necessary that P $= [x_1, x_2, x_3, \ldots, x_n]$ is a path in T. But P is of order n and it has $n - 1$ edges; hence T = P and x is an end-vertex. □

Theorem 5. Let $\mathfrak{I}$ be the family of trees T = (V, E) of order n. Then

$$\max \left\{ \sum_{x,y \in V} d(x,y) : T \in \mathfrak{I} \right\} = \tfrac{1}{3} n(n^2 - 1).$$

Proof. First, we claim that $\sum_{x,y \in V} d(x,y)$ is a maximum in $\mathfrak{I}$ if T is a path. We prove this claim by induction on n. For $1 \le n \le 3$, the property is immediate since every tree in this case is a path. Suppose now that the property is true for all trees with at most $n - 1$ vertices. Let z be an end-vertex of T. Then

$$\sum_{x,y \in V} d(x,y) = 2s(z) + \sum_{x,y \in W} d(x,y),$$

where W = V − $\{z\}$. Consider the right side of the last equation above. By Lemma 4, $s(z)$ in the first term is maximum if T is a path and z is one of its end-vertices. The second term is also maximum by applying induction hypothesis to the subtree whose $n - 1$ vertices make up the set W. Thus $\sum_{x,y \in V} d(x,y)$ is a maximum. This completes the proof of the claim. By Lemma 2, the proof of the theorem is also completed. □

Corollary 6. For any tree T = (V,E) of order n,

$$2(n-1)^2 \le \sum_{x,y \in V} d(x,y) \le \tfrac{1}{3} n(n^2 - 1).$$

Proof. This corollary results from the combination of Theorem 3 and Theorem 5. □

References

[1] J.A. Bondy & U.S.R. Murty, *Graph Theory With Applications*, Macmillan Press, New York, 1976.

[2] Frank Harary, *Graph Theory*, Addison-Wesley, Massachussetts, 1971.

[3] Ioan Tomescu, *Problems in Combinatorics and Graph Theory*, John Wiley & Sons, New York, 1985.

[4] H. P. Yap, *Some Topics in Graph Theory*, London Math Society Lecture Notes, Series 108, Cambridge University Press, 1986.