Discrete Applied Mathematics 159 (2011) 945–952

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

On the distance distribution of trees

Peter Dankelmann

University of KwaZulu-Natal, Durban, South Africa

ARTICLE INFO

Article history:
Received 17 March 2009
Received in revised form 4 May 2010
Accepted 17 February 2011
Available online 31 March 2011

Keywords:
Distance
Distance distribution
Tree
Diameter
Radius

ABSTRACT

Let G be a tree and k a non-negative integer. We determine best possible upper and lower bounds on the number of pairs of vertices at distance exactly k in G in terms of order alone, and in terms of order and radius or diameter.

1. Introduction

Let G be a finite, connected graph of diameter d. For a non-negative integer $k \le d$, let $D_k = D_k(G)$ be the number of unordered pairs of vertices of G at distance exactly k. Following [1], we call the sequence $D_0, D_1, D_2, \ldots, D_d$ the *distance distribution* of G. It contains significantly more information than the classical distance parameters, diameter and average distance, which makes it a very useful tool in the analysis of transportation networks. Remarkably, very little is known about the distance distribution of graphs. In this paper, we study bounds on the individual terms of the distance distribution of trees.

Although the distance distribution of a graph G contains significant information, it does not necessarily determine G up to isomorphism. Slater [3] showed that even trees are not determined by their distance distributions. It is natural to ask which sequences are realisable as the distance distribution of some graph. This question appears difficult, even if restricted to trees. Wang and Amin [4] gave a characterisation of the pairs of positive integers (a, b) for which there exists a tree T with $D_2(T) = a$ and $D_3(T) = b$. Buckley and Superville [2] investigated graphs in which all terms of the distance distribution have the same value. In this paper we consider the individual terms D_k of the distance distribution of trees for different values of k. We give upper and lower bounds on D_k in terms of the order of a tree and in terms of its order and radius or diameter.

We use the following notation: the order (number of vertices) and size (number of edges) of a graph are denoted by n and q, respectively. If v is a vertex of a graph G then we denote its degree by $\deg_G(v)$. The distance between two vertices u and v, i.e., the minimum number of edges on a $u - v$ path, is denoted by $d(u, v)$. The distance $d(v, A)$ between a vertex v of T and a set A of vertices of T is defined as the minimum of all distances $d(v, a)$, $a \in A$. If k is a positive integer then a k-pair is an unordered pair of vertices at distance k. The diameter of a graph G is denoted by $\mathrm{diam}(G)$. For a vertex v of T and k a non-negative integer let $N_k(v, T)$ denote the set of vertices at distance exactly k from v in T, and let $n_k(v, T) = |N_k(v, T)|$. If there is no danger of confusion, we write $N_k(v)$ or even N_k, and similarly $n_k(v)$ or n_k. By $N_{\le k}(v)$ and $N_{\ge k}(v)$ we mean the set of vertices at distance at most k and at least k, respectively, from v. For $d \ge 2$ the broom $B(n, d)$ is the tree obtained from a path on d vertices by appending to one of its ends $n - d$ new end vertices. The double broom $DB(n, d)$ is the tree of order n

E-mail address: dankelma@ukzn.ac.za.

0166-218X/$ – see front matter
doi:10.1016/j.dam.2011.02.010

obtained from a path on $d-1$ vertices by appending $\lceil (n-d+1)/2\rceil$ and $\lfloor (n-d+1)/2\rfloor$ vertices, respectively, to its two ends.

For real valued functions $f(x), g(x)$ we write $f = O(g)$ if there exists a constant c such that, for large enough x, $|f(x)| \leq cg(x)$.

2. Bounds on $D_k(T)$ in terms of order

In this section we present bounds on the number of k-pairs of a tree in terms of order. In most cases, the bounds will depend on the parity of k. This is due to the fact that every tree has a bipartition of its vertex set into independent sets V_1, V_2. Hence for odd k, each k-pair has one vertex in V_1 and the other vertex in V_2, and so the number of k-pairs is bounded from above by $D_k(T) \leq \sum_{i \text{ odd}} D_i(T) = |V_1||V_2| \leq \frac{1}{4}n^2$. For even k, each k-pair has both vertices in the same V_i, and we have only the much weaker inequality $D_k(T) \leq \sum_{i \text{ even}} D_i(T) = \binom{|V_1|}{2} + \binom{|V_2|}{2} \leq \binom{n-1}{2}$. So while for odd k up to at most approximately half of all pairs of vertices can have distance k, it will turn out that for even k almost all pairs of vertices can be at distance k. We will be more specific in Theorem 1.

We begin with an upper bound on D_k for even k in terms of order alone. In the proof we use the following notation. Let T be a tree and let u, v be end vertices of T. Let $T' = T - uu' + uv'$, where u' and v' are the neighbours of u and v, respectively. We say that T' is obtained from T by making u a twin of v, and we will denote T' by $T(u \to v)$.

Theorem 1. *Let T be a tree of order n. Then, for large n and fixed, even k with $2 \leq k \leq n-1$,*

$$D_k(T) \leq \begin{cases} \binom{n-1}{2} & \text{if } k = 2, \\ \frac{1}{2}n^2 - \sqrt{k-2}\,n^{3/2} + O(n) & \text{if } k \geq 4, \end{cases}$$

and this bound is best possible.

Proof. First let $k = 2$. Since T is a tree and thus bipartite, $V(T)$ has a bipartition into independent sets U and W. Since two vertices at distance two belong to the same partition set, we have

$$D_2(T) \leq \binom{|U|}{2} + \binom{|W|}{2} \leq \binom{n-1}{2}.$$

We note that equality holds if and only if one of the two partition sets has only one vertex, i.e., if T is a star.

Now let $k \geq 4$. For given n and k let T be a tree of order n for which $D_k(T)$ is maximum. We show that T has the property that for any two end vertices u and v of T with $d_T(u, v) \neq k$,

$$D_k(T(u \to v)) = D_k(T). \tag{1}$$

Clearly, making u a twin of v destroys $n_k(u, T)k$-pairs and creates $n_k(v, T)$ new ones, so $D_k(T(u \to v)) = D_k(T) - n_k(u, T) + n_k(v, T)$. By the maximality of $D_k(T)$ we have $n_k(u, T) \geq n_k(v, T)$. Similarly, by considering $T(v \to u)$, we obtain $n_k(u, T) \leq n_k(v, T)$. Hence $n_k(u, T) = n_k(v, T)$ and thus $D_k(T(u \to v)) = D_k(T)$, as desired.

We fix an end vertex v_1 of T. By successively making each end vertex not at distance k from v_1 a twin of v_1, we obtain a tree T_1 with the property that each end vertex of T_1 is either a twin of a vertex in $\{v_1\}$ or at distance exactly k from all vertices in $\{v_1\}$. We now fix an end vertex v_2 of T_1 which is not a twin of a vertex in $\{v_1\}$ (since T is not a star such a vertex exists), and we successively make every end vertex of T_1 which is not at distance k from a vertex $v_i \in \{v_1, v_2\}$ a twin of v_i, thus obtaining a tree T_2. Note that in T_2 every end vertex that is not a twin of a vertex in $\{v_1, v_2\}$ is at distance exactly k from all vertices in $\{v_1, v_2\}$. Now fix an end vertex v_3, if such a vertex exists, which is not a twin of any vertex in $\{v_1, v_2\}$, and successively make every end vertex not at distance k from some vertex $v_i \in \{v_1, v_2, v_3\}$ a twin of v_i. Continue this process until our tree contains a set $\{v_1, v_2, \ldots, v_r\}$ of end vertices which are pairwise at distance k, such that every end vertex of T_r either equals v_i or is a twin of v_i for some $i \in \{1, 2, \ldots, r\}$.

Note that, by (1), $D_k(T_r) = D_k(T)$, so $D_k(T_r)$ is maximum. Also, if u and v are two vertices at distance k in T_r, then u and v are twins of v_i and v_j, respectively, for some $i \neq j \in \{1, 2, \ldots, r\}$.

Let a_i be the number of twins of v_i, including v_i itself. Then

$$D_k(T_r) = \sum_{1 \leq i < j \leq r} a_i a_j. \tag{2}$$

Denote $\sum_{i=1}^{r} a_i$ by M. We now find an upper bound on M. Identifying all twins of v_i with v_i for $i = 1, 2, \ldots, r$, we obtain a tree T' containing r vertices which are pairwise at distance exactly k. Clearly,

$$\bigcup_{i=1}^{r} N_{\leq k/2-1}(v_i) \subset V(T').$$

Since each v_i has at least one vertex at distance exactly $k/2$, which is in none of the above sets, the inclusion is strict. Since in T', the sets $N_{\le k/2-1}(v_i)$ are disjoint, we have

$$|V(T')| \ge \sum_{i=1}^{r} |N_{\le k/2-1}(v_i)| + 1 \ge \frac{kr}{2} + 1.$$

Hence, by $n = |V(T')| + M - r$, we obtain

$$M = n - |V(T')| + r \le n - \frac{k-2}{2}r - 1. \tag{3}$$

For given n, r, k the value of $\sum_{i<j} a_i a_j$ is maximised, subject to (2) and (3), and $a_i \ge 1$ for $i \in \{1, 2, \dots, r\}$, if $\sum_{i=1}^{r} a_i = n - \frac{k-2}{2}r - 1$ and, moreover, all a_i have the same value, i.e., $a_i = \frac{n-1}{r} - \frac{k-2}{2}$ for $i = 1, 2, \dots, r$. So

$$D_k(T) = D_k(T_r) \le \binom{r}{2}\left(\frac{n-1}{r} - \frac{k-2}{2}\right)^2 = \frac{1}{2}\left(1 - \frac{1}{r}\right)\left(n - 1 - \frac{k-2}{2}r\right)^2.$$

Straightforward differentiation, with respect to r, of the right hand side above shows that it attains its maximum for $r = \frac{1}{4} + \sqrt{\frac{1}{16} + \frac{n-1}{k-2}}$. For constant k and large n this equals $\sqrt{\frac{n}{k-2}} + \frac{1}{4} + O(n^{-1/2})$. Substituting this value for r yields, after simplification, the desired result.

To see that the bound is best possible let k be constant and choose n such that $r = \frac{1}{4} + \sqrt{\frac{1}{16} + \frac{n-1}{k-2}}$ is an integer. Let T be the tree obtained from a star $K_{1,r}$ by subdividing each edge $\frac{1}{2}k - 2$ times and attaching $\lfloor \frac{n-1}{r} - \frac{k-2}{2} \rfloor$ or $\lceil \frac{n-1}{r} - \frac{k-2}{2} \rceil$ end vertices to each end vertex of the subdivided star, so that the resulting tree has order n. It is easy to verify that $D_k(T) = \frac{1}{2}n^2 - \sqrt{k-2}\,n^{3/2} + O(n)$. □

If k is odd, then the above-mentioned upper bound $\frac{1}{4}n^2$ on D_k can be improved as follows.

Theorem 2. *Let T be a tree of order n and $3 \le k \le n-1$, k odd. Then*

$$D_k(T) \le \left\lfloor \frac{n-k+1}{2} \right\rfloor \left\lceil \frac{n-k+1}{2} \right\rceil, \tag{4}$$

and this bound is sharp as shown by the double broom $DB(n, k)$.

We postpone the proof of Theorem 2 since it follows directly from a bound on D_k in terms of order and diameter (Theorem 4), which we present later. We note without proof that Theorem 2 can also be proved by induction, and that such a proof shows that equality holds if and only if T is a double broom $DB(n, k)$, or if $n = k + 1$ and T is a path.

We briefly discuss lower bounds on D_k in terms of order. A lower bound on D_k in terms of order alone would be meaningless for $k \ge 3$ since every tree of diameter less than k has $D_k = 0$. So we consider only D_2. It is reported in [1, Theorem 9.10] that the inequality

$$D_1(G) + D_2(G) \ge 2n - 3$$

which holds for all connected graphs G of order n, is due to Capobianco. It implies that $D_2(T) \ge n-2$ for all trees. This bound is sharp as $D_2(T) = n - 2$ if and only if T is a path.

3. Bounds on D_k in terms of order and radius or diameter

Next we present bounds on D_k in terms of order and either radius or diameter, whichever seems more natural for the specific bound.

We first present a lower bound on D_2 in terms of order and radius.

Theorem 3. *Let $r \in \mathbb{N}$. Then there exists a positive constant c_r such that for every tree T of order n and radius at most r,*

$$D_2(T) \ge c_r n^{2^r/(2^r-1)}.$$

For each fixed r there exist infinitely many values of n for which there exists a tree with n vertices and $D_2(T) = \left(\frac{r+1}{2} + o(1)\right)n^{2^r/(2^r-1)}$.

Proof. We show the theorem by induction on r. The theorem is clearly true for $r = 1$ since then T is a star and $D_2 = \binom{n-1}{2}$.

Let T be a tree of radius at most r, rooted at a central vertex v. For $i \ge 0$ let N_i be the set of vertices at distance exactly i from v and let $n_i = |N_i|$.

Now let $r \geq 2$. By our induction hypothesis there exists a real c_{r-1} such that for every tree T' of radius at most $r-1$ we have $D_2(T') \geq c_{r-1}|V(T')|^{2^{r-1}/(2^{r-1}-1)}$.

Let $s = n_1$ and let $w_1, w_2, \ldots, w_s$ be the neighbours of vertex v. For $i = 1, 2, \ldots, s$ let T_i be the branch of T at w_i. Then $\sum_{i=1}^{s} |V(T_i)| = n-1$. Each pair of vertices at distance 2 has either both vertices in N_1, one vertex in N_2 and the second vertex in N_0, or both vertices in the same branch T_i. Hence

$$D_2(T) = \binom{n_1}{2} + n_2 + \sum_{i=1}^{s} D_2(T_i) > \binom{n_1}{2} + \sum_{i=1}^{s} D_2(T_i).$$

By our induction hypothesis we have $D_2(T_i) \geq c_{r-1}|V(T_i)|^{2^{r-1}/(2^{r-1}-1)}$, hence

$$D_2(T) > \binom{n_1}{2} + c_{r-1}\sum_{i=1}^{s} |V(T_i)|^{2^{r-1}/(2^{r-1}-1)}.$$

Now the function $f(x) = x^{\alpha}$ is concave up for every $\alpha > 1$, in particular for $\alpha = 2^{r-1}/(2^{r-1}-1)$. By Jensen's inequality, the sum $\sum_{i=1}^{s} x_i^{\alpha}$ is minimised, subject to $\sum_{i=1}^{s} x_i = n-1$, if all x_i equal $(n-1)/s$. Since $s = n_1$ we have

$$\begin{aligned} D_2(T) &\geq \frac{1}{2}(n_1^2 - n_1) + c_{r-1}n_1\left(\frac{n-1}{n_1}\right)^{2^{r-1}/(2^{r-1}-1)} \\ &> \frac{1}{2}n_1^2 - \frac{1}{2}n + c_{r-1}n_1^{-1/(2^{r-1}-1)}(n-1)^{2^{r-1}/(2^{r-1}-1)}. \end{aligned}$$

The derivative, with respect to n_1, of the right hand side of the last inequality vanishes for

$$n_1 = \left(\frac{c_{r-1}}{2^{r-1}-1}(n-1)\right)^{2^{r-1}/(2^{r}-1)}.$$

It is easy to verify that for this choice of n_1 the last expression is indeed minimised. Substituting n_1 now yields that there exists a positive constant $\tilde{c}_r$ such that for all trees of order n,

$$D_2(T) \geq \tilde{c}_r n^{2^r/(2^r-1)} - \frac{1}{2}n.$$

Since the exponent $2^r/(2^r-1)$ is greater than 1, and since the term $\tilde{c}_r n^{2^r/(2^r-1)} - \frac{1}{2}n$ is positive for all but finitely many values of n, there exists a positive constant c_r with $0 < c_r \leq \tilde{c}_r$ such that for all trees of order n

$$D_2(T) \geq c_r n^{2^r/(2^r-1)},$$

as desired.

To see that the order of the bound is best possible consider, for a given, even integer $r \geq 2$, the tree T of radius r constructed as follows. Let n be a large integer such that $n^{1/(2^r-1)}$ is an integer. Clearly there exist infinitely many such integers. Let v, the unique centre vertex of T, have degree $n^{2^{r-1}/(2^r-1)}$, let each neighbour of v have degree $n^{2^{r-2}/(2^r-1)}+1$ and let, generally, each vertex at distance i from v have degree $n^{2^{r-1-i}/(2^r-1)}+1$, for $i = 1, 2, \ldots, r-1$. A simple calculation shows that T has n end vertices and $O(n^{(2^r-2)/(2^r-1)})$ other vertices, and that $D_2(T) = \left(\frac{r}{2} + o(1)\right) n^{2^r/(2^r-1)} = \left(\frac{r}{2} + o(1)\right)|V(T)|^{2^r/(2^r-1)}$. $\square$

In Theorem 3 we determined the order of magnitude of the minimum number of 2-pairs in a tree of given order and radius. It seems to be difficult to determine also the largest possible value of the coefficient c_r.

Proposition 1. *Let T be a tree of order n and diameter d, and let $k \geq 3$. Then*

$$D_k \geq \begin{cases} n-k & \text{if } k \leq (d+3)/2, \\ d-k+1 & \text{if } (d+4)/2 \leq k \leq d. \end{cases} \tag{5}$$

Both bounds are sharp.

Proof. Fix a path $P = v_0, v_1, v_2, \ldots, v_d$ of length d. Then $d(v_i, v_{i+k}) = k$ for $i = 0, 1, \ldots, d-k$, so $D_k \geq d-k+1$, which proves the second bound.

Now let $k \leq (d+3)/2$. Denote $N_{\geq k}(v_0)$ and $N_{\leq k}(v_0)$ by $N_{\geq k}$ and $N_{\leq k}$, respectively. Then for each vertex $w \in N_{\geq k}$ there exists a vertex w' which is on the $v_0 - w$ path and at distance k from w. Now let $u \in N_{\leq k-1} - \{v_0, v_1, v_2, \ldots, v_{k-1}\}$. Since u is not on the $v_0 - v_d$ path, we have $d(v_0, u) + d(u, v_d) \geq d+2$ and thus

$$d(u, v_d) \geq d + 2 - d(v_0, u) \geq d + 2 - (k-1) \geq (2k-3) + 2 - (k-1) = k.$$

Hence there exists a vertex $\bar{u}$ which is on the $u - v_d$ path and at distance k from u.

In total this yields $|N_{\ge k}| + |N_{\le k-1}| - k = n - k$ pairs of vertices at distance k. It remains to show that these pairs are distinct. Clearly, if $u_1, u_2 \in N_{\le k-1} - \{v_0, v_1, \ldots, v_{k-1}\}$ with $u_1 \neq u_2$ then $\{u_1, \bar{u_1}\} \neq \{u_2, \bar{u_2}\}$. Also, if $w_1, w_2 \in N_{\ge k-1}$ with $w_1 \neq w_2$, then $\{w_1, w_1'\} \neq \{w_2, w_2'\}$. So suppose that there exist $u \in N_{\le k-1} - \{v_0, v_1, \ldots, v_{k-1}\}$ and $w \in N_{\ge k}$ with $\{w, w'\} = \{u, \bar{u}\}$. Then $u = w'$ and $w = \bar{u}$. Now $u = w'$ implies that u is on a $w - v_0$ path P_w. Since P_w consists of a (possibly empty) path from w to the nearest vertex on P followed by a subpath of P, we have $d(u, V(P)) \le d(w, V(P))$. On the other hand $w = \bar{u}$ implies that w is on a $u - v_d$ path P_u. Since P_u consists of a (non-empty) path from u to the nearest vertex on P followed by a subpath of P, we have $d(u, V(P)) > d(w, V(P))$, a contradiction. Hence all $n - k$ pairs are distinct.

The broom $B(n, d)$ shows that the first bound is sharp. The tree obtained from a path on $d + 1$ vertices by appending $n - d - 1$ new end vertices to one of its centre vertices shows that the second bound is sharp. □

Theorem 4. *Let T be a tree of order n and diameter d. If k is odd, $k \ge 3$, then*

$$D_k(T) \le \begin{cases} \left\lceil \dfrac{n-d-1}{2} \right\rceil \left\lfloor \dfrac{n-d-1}{2} \right\rfloor + n - k & \text{if } k \le d \le 2k-3, \\ \left\lceil \dfrac{n-d-1}{2} \right\rceil \left\lfloor \dfrac{n-d+1}{2} \right\rfloor + n - k & \text{if } 2k-2 \le d \le 3k-5, \\ \left\lceil \dfrac{n-d-1}{2} \right\rceil \left\lfloor \dfrac{n-d-1}{2} \right\rfloor + 2n - d - k - 1 & \text{if } 3k-4 \le d, \end{cases}$$

and these bounds are sharp.

Proof. Since every tree is bipartite, $V(T)$ has a bipartition into independent sets V_1 and V_2. Let $P = v_0, v_1, \ldots, v_d$ be a path of length d. We denote the set of the vertices not on P by $\overline{V(P)}$, and we define $\overline{V_i(P)} = V_i - V(P)$ for $i = 1, 2$.

Define a value $\rho(w)$ for each vertex $w \in \overline{V(P)}$ as follows:

$$\rho(w) = 2|N_k(w) \cap V(P)| + |N_k(w) \cap \overline{V(P)}|.$$

We claim that

$$D_k(T) = d + 1 - k + \frac{1}{2} \sum_{w \in \overline{V(P)}} \rho(w). \tag{6}$$

Indeed, P contains exactly $d + 1 - k$ pairs of vertices at distance k. Let $\{x, y\} \not\subseteq V(P)$ be a pair of vertices at distance k. If exactly one of the two vertices, x say, is on P then the pair $\{x, y\}$ contributes 2 to $\rho(y)$ and thus to $\sum_w \rho(w)$. If neither x nor y is on P, then the pair $\{x, y\}$ contributes 1 to $\rho(x)$ and 1 to $\rho(y)$. In either case, the pair $\{x, y\}$ contributes 2 to $\sum_w \rho(w)$, and so the number of pairs of vertices at distance k which are not contained in P equals $\frac{1}{2} \sum_w \rho(w)$, and (6) follows.

To prove the bound in Theorem 4, it suffices by (6) to prove the following three upper bounds on $\sum \rho(w)$ in Claims 1–3.

Claim 1: If $d \ge 3k - 4$ then

$$\frac{1}{2} \sum_{w \in \overline{V(P)}} \rho(w) \le \left\lfloor \frac{n-d-1}{2} \right\rfloor \left\lceil \frac{n-d-1}{2} \right\rceil + 2(n-d-1).$$

Since T is a tree, each vertex $w \in \overline{V(P)}$ is at distance k from at most 2 vertices on P, so $|N_k(w) \cap \overline{V(P)}| \le 2$. Furthermore, if $w \in \overline{V_i(P)}$, then no vertex in the same partite set is at distance k from w. Hence $\rho(w) \le 2 + |\overline{V_{3-i}(P)}|$ for all $w \in \overline{V_i(P)}$, and so

$$\frac{1}{2} \sum_{w \in \overline{V(P)}} \rho(w) \le 2|\overline{V(P)}| + |\overline{V_1(P)}||\overline{V_2(P)}| \le 2(n-d-1) + \left\lfloor \frac{n-d-1}{2} \right\rfloor \left\lceil \frac{n-d-1}{2} \right\rceil,$$

since $|\overline{V_1(P)}| + |\overline{V_2(P)}| = n - d - 1$, and Claim 1 follows.

Let U be the set of vertices not on P but adjacent to a vertex of P, and let $W = V(T) - (V(P) \cup U)$. As usual write U_i and W_i for $U \cap V_i$ and $W \cap V_i$, $i = 1, 2$.

Claim 2: If $k \le d \le 2k - 3$ then

$$\frac{1}{2} \sum_{w \in W} \rho(w) \le \left\lfloor \frac{n-d-1}{2} \right\rfloor \left\lceil \frac{n-d-1}{2} \right\rceil + n - d - 1.$$

Since $d \le 2k - 3$, a vertex $w \in U$ is at distance k from at most one vertex on P. A vertex $w \in W$ is at distance k from at most two vertices of P. Furthermore, if w is in $\overline{V_i(P)}$, then all vertices at distance k from w not on P are in V_{3-i} and not adjacent to w. Hence at most $|\overline{V_{3-i}}| - \deg_{T-V(P)}(w)$ vertices not on P are at distance k from w, and so

$$\rho(w) \leq \begin{cases} 2 + |\overline{V_{3-i}}| - \deg_{T-V(P)}(w) & \text{if } w \in U_i, \\ 4 + |\overline{V_{3-i}}| - \deg_{T-V(P)}(w) & \text{if } w \in W_i. \end{cases} \tag{7}$$

Summation over all $w \in \overline{V(P)}$ yields

$$\begin{aligned}\sum_{w \in \overline{V(P)}} \rho(w) &\leq 2|W| + \sum_{i=1}^{2} \left(\sum_{w \in \overline{V_i(P)}} 2 + |\overline{V_{3-i}(P)}| - \deg_{T-V(P)}(w) \right) \\ &= 2|W| + 2|\overline{V_1(P)}||\overline{V_2(P)}| + 2|\overline{V(P)}| - 2q(T - V(P)),\end{aligned}$$

where $q(T - V(P))$ is the number of edges in the forest $T - V(P)$. We have $q(T - V(P)) = |W|$ since there is a bijection that maps each $w \in W$ to the edge in $T - V(P)$ incident with w which is on the path from w to P in T. As in Claim 1 we have $|\overline{V_1(P)}||\overline{V_2(P)}| \leq \lfloor \frac{n-d-1}{2} \rfloor \lceil \frac{n-d-1}{2} \rceil$. Hence

$$\frac{1}{2} \sum_{w \in W} \rho(w) \leq \left\lfloor \frac{n-d-1}{2} \right\rfloor \left\lceil \frac{n-d-1}{2} \right\rceil + n - d - 1,$$

which is Claim 2.

Claim 3: If $2k - 2 \leq d \leq 3k - 5$ then

$$\frac{1}{2} \sum_{w \in W} \rho(w) \leq \left\lceil \frac{n-d-1}{2} \right\rceil \left\lfloor \frac{n-d+1}{2} \right\rfloor + n - d - 1.$$

We show that if $w \in \overline{V_1(P)}$ and $w' \in \overline{V_2(P)}$ then

$$\rho(w) + \rho(w') \leq \begin{cases} n - d + 5 - \deg_{T-V(P)}(w) - \deg_{T-V(P)}(w') & \text{if } \{w, w'\} \subseteq U, \\ n - d + 7 - \deg_{T-V(P)}(w) - \deg_{T-V(P)}(w') & \text{if } \{w, w'\} \not\subseteq U. \end{cases} \tag{8}$$

As above, w and w' have at most two vertices each on P at distance k. As in Claim 2, at most $|\overline{V_2}| - \deg_{T-V(P)}(w)$ vertices not on P are at distance k from w, and at most $|\overline{V_1}| - \deg_{T-V(P)}(w')$ vertices not on P are at distance k from w'. Hence

$$\begin{aligned}\rho(w) + \rho(w') &\leq 8 + |\overline{V_2(P)}| + |\overline{V_1(P)}| - \deg_{T-V(P)}(w) - \deg_{T-V(P)}(w') \\ &= n - d + 7 - \deg_{T-V(P)}(w) - \deg_{T-V(P)}(w').\end{aligned}$$

Hence (8) for the case $\{w, w'\} \not\subseteq U$ follows.

Assume that $\{w, w'\} \subseteq U$ and $d_T(w, w') \neq k$. As above, w and w' are at distance k from at most 2 vertices each on P. Vertex w is not at distance k from its neighbours and from w', hence w is at distance k from at most $|\overline{V_2(P)}| - 1 - \deg_{T-V(P)}(w)$ vertices in $\overline{V(P)}$. (Note that w and w' are not adjacent since both are in U.) Similarly, w' is not at distance k from at most $|\overline{V_1(P)}| - 1 - \deg_{T-V(P)}(w')$ vertices in $\overline{V(P)}$. Hence $\rho(w) + \rho(w') \leq |\overline{V(P)}| - 2 - \deg_{T-V(P)}(w) - \deg_{T-V(P)}(w') + 8 = n - d + 5 - \deg_{T-V(P)}(w) - \deg_{T-V(P)}(w')$, and so (8) follows for this case.

Now assume that $\{w, w'\} \subseteq U$ and $d_T(w, w') = k$. We show that w and w' have, in total, at most three vertices at distance k on P. Suppose to the contrary that each of w, w' has two vertices at distance k on P. If v_i and v_j are the neighbours of w and w', respectively, on P, then the vertices at distance k from $w(w')$ are v_{i-k+1} and v_{i+k-1} (v_{j-k+1} and v_{j+k-1}). We can assume without loss of generality that $i < j$. Then $j = i + k - 2$, and so v_{i-k+1} and $v_{j+k-1} = v_{i+2k-3}$ are vertices of P. But $d(v_{i-k+1}, v_{i+2k-3}) = 3k - 4 > \operatorname{diam}(T)$, a contradiction. Hence w and w' have, in total, at most three vertices at distance k on P. In conjunction with $|N_k(w) \cap \overline{V(P)}| \leq |\overline{V_2(P)}| - \deg_{T-V(P)}(w)$ and $|N_k(w') \cap \overline{V(P)}| \leq |\overline{V_1(P)}| - \deg_{T-V(P)}(w')$ we obtain (8).

Let $\overline{V_1(P)} = \{w_1, w_2, \ldots, w_m\}$ and $\overline{V_2(P)} = \{w'_1, w'_2, \ldots, w'_\ell\}$. Without loss of generality we assume that $m \geq \ell$ and that there exists an integer s such that $\{w_i, w'_i\} \subseteq U$ for $i = 1, 2 \ldots, s$ and $\{w_i, w'_i\} \not\subseteq U$ for $i = s + 1, s + 2, \ldots, \ell$. We distinguish two cases.

Case 1: $m = \ell$.

Then $m = \frac{1}{2}(n - d - 1)$. By (8) we have

$$\begin{aligned}\sum_{w \in \overline{V(P)}} \rho(w) &= \sum_{i=1}^{m} \left(\rho(w_i) + \rho(w'_i)\right) \\ &\leq \sum_{i=1}^{s} (n - d + 5 - \deg_{T-V(P)}(w_i) - \deg_{T-V(P)}(w'_i))\end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=s+1}^{m}(n-d+7-\deg_{T-V(P)}(w_i)-\deg_{T-V(P)}(w_i'))\\
&= m(n-d+5)+2(m-s)-\sum_{w\in\overline{V(P)}}\deg_{T-V(P)}(w)\\
&\le \frac{n-d-1}{2}(n-d+5)+2|W|-2q(T-V(P)).
\end{aligned}
$$

The last inequality follows from the fact that each of the $m-s$ pairs $\{w_i, w_i'\}$, $i=s+1, s+2, \ldots, m$, contains at least one vertex of W, so $|W|\ge m-s$. As in Claim 2, we have $|W|=q(T-V(P))$. Hence the last inequality in conjunction with the fact that $\frac{1}{2}\sum_{w\in\overline{V(P)}}\rho(w)$ is an integer, yields Claim 3 for this case.

Case 2: $m>\ell$.

We bound $\rho(w_i)$ for $i=\ell+1, \ell+2, \ldots, m$. Since w_i has at most two vertices at distance k on P, and at most $|\overline{V_2(P)}|-\deg_{T-V(P)}(w_i)$ vertices at distance k in $\overline{V(P)}$, we have

$$\rho(w_i)\le 4+|\overline{V_2(P)}|-\deg_{T-V(P)}(w_i).$$

If now $m\ge \ell+2$ then $|\overline{V_2(P)}|=\ell\le\frac{n-d-3}{2}$, and if $m=\ell+1$ then $|\overline{V_2(P)}|=\ell=\frac{n-d-2}{2}$. Hence $\rho(w_i)\le\frac{n-d+5}{2}$ if $m\ge\ell+2$, and $\rho(w_i)\le\frac{n-d+6}{2}$ if $m=\ell+1$. In both cases we get

$$\sum_{i=\ell+1}^{m}\rho(w_i)\le(m-\ell)\frac{n-d+5}{2}+\frac{1}{2}.$$

By (8) we have

$$
\begin{aligned}
\sum_{w\in\overline{V(P)}}\rho(w) &= \sum_{i=1}^{s}(\rho(w_i)+\rho(w_i'))+\sum_{i=s+1}^{\ell}(\rho(w_i)+\rho(w_i'))+\sum_{i=\ell+1}^{m}\rho(w_i)\\
&\le \sum_{i=1}^{s}(n-d+5-\deg_{T-V(P)}(w_i)-\deg_{T-V(P)}(w_i'))\\
&\quad+\sum_{i=s+1}^{\ell}(n-d+7-\deg_{T-V(P)}(w_i)-\deg_{T-V(P)}(w_i'))+(m-\ell)\frac{n-d+5}{2}+\frac{1}{2}\\
&= \ell(n-d+5)+2(\ell-s)+(m-\ell)\frac{n-d+5}{2}+\frac{1}{2}-\sum_{w\in\overline{V(P)}}\deg_{T-V(P)}(w)\\
&< \frac{(n-d-1)(n-d+5)}{2}+2|W|-2q(T-V(P)),
\end{aligned}
$$

as desired. The last inequality follows from the fact that $\frac{1}{2}+2(\ell-s)<2(m-s)$ and that $|W|\ge m-s$ as above. As in Case 1, the last inequality in conjunction with $|W|=q(T-V(P))$ and the fact that $\frac{1}{2}\sum_{w\in\overline{V(P)}}\rho(w)$ is an integer, yields Claim 3. This completes the proof of the upper bound.

The following trees show that the bounds in Theorem 4 are sharp. If $k\le d\le 2k-3$ then let T be the tree obtained from a path $P=v_0, v_1, \ldots, v_d$ by attaching $\lfloor(n-d-1)/2\rfloor$ new end vertices to v_1 and $\lceil(n-d-1)/2\rceil$ new end vertices to v_{k-1}. If $2k-2\le d\le 3k-5$ then let T be the same tree. If $d\ge 3k-4$ then let T be the tree obtained from a path $P=v_0, v_1, \ldots, v_d$ by attaching $\lfloor(n-d-1)/2\rfloor$ new end vertices to v_{k-1} and $\lceil(n-d-1)/2\rceil$ new end vertices to v_{2k-3}. It is easy to verify that in each case T attains the bound proved in this theorem. □

Theorem 4 implies Theorem 2 since for fixed k and n the bound in Theorem 4 is maximised for $d=k$, and setting $k=d$ in Theorem 4 yields Theorem 2.

Proposition 2. *Let T be a tree of order n and diameter d. If k is even, $k\ge 2$, then*

$$D_k(T)\le\binom{n-d-1}{2}+2n-d-k-1.$$

Proof. Let $P=v_0, v_1, \ldots, v_d$ be a path of length d. Clearly, there are exactly $d+1-k$ k-pairs on P, viz., the pairs $\{v_i, v_{i+k}\}$ for $i=0, 1, \ldots, d-k$. There are at most $2(n-d-1)k$-pairs $\{u, v_i\}$ with $u\in V-V(P)$ and $v_i\in V(P)$ since each such u is at distance k from at most 2 vertices on P. Finally, there are at most $\binom{n-d-1}{2}$ k-pairs that are contained in $V-V(P)$ since $|V-V(P)|=n-d-1$. Adding these three terms yields the bound. □

We remark that the above bound is sharp only for $k = 2$. For $k \geq 4$ the following construction shows that the bound in Proposition 2 is best possible in the sense that for fixed d and k, and for $n \to \infty$, the coefficient 1 of the dominating term $\binom{n-d-1}{2}$ cannot be improved. Choose n such that $\frac{1}{4} + \sqrt{\frac{1}{16} + \frac{n'-1}{k-2}}$ is an integer, where $n' = n - d$. Let T' be the tree of order n' constructed at the end of the proof of Theorem 1. Then $D_k(T') = \frac{1}{2}n'^2 - \sqrt{k-2}n'^{3/2} + O(n')$. Let T be the tree obtained from T' by identifying the centre vertex of T' with a centre vertex of a path of length d. Then

$$\begin{aligned} D_k(T) &= D_k(T') + d + 1 - k + 2(n - d - 1) \\ &= \binom{n-d-1}{2} - \sqrt{k-2}(n-d)^{3/2} + O(n). \end{aligned}$$

Acknowledgement

The author would like to thank Earl Glen Whitehead (University of Pittsburgh) for bringing the topic of this paper to his attention, and for several invaluable discussions.

References

[1] F. Buckley, F. Harary, Distances in Graphs, Addison Wesley, New York, 1990.
[2] F. Buckley, L. Superville, Distance distributions and mean distance problems, in: Proceedings of the Third Caribean Conference on Combinatorics and Computing, Bridgetown, 1981, University of the West Indies, Barbados, 1981, pp. 67–76.
[3] P. Slater, Counterexample to Randic's conjecture on distance degree sequences for trees, J. Graph Theory 6 (1982) 89–92.
[4] H. Wang, A.T. Amin, Realizability of a tree with a given sequence as its partial distance distribution, Congr. Numer. 110 (1995) 193–199.