Average Distance and Generalised Packing in Graphs

Peter Dankelmann, University of KwaZulu-Natal, Durban, South Africa

Abstract

Let $G$ be a connected finite graph. The average distance $\mu(G)$ of $G$ is the average of the distances between all pairs of vertices of $G$. For a positive integer $k$ a $k$-packing of $G$ is a subset $S$ of the vertex set of $G$ such that the distance between any two vertices in $S$ is greater than $k$. The $k$-packing number $\beta_k(G)$ of $G$ is the maximum cardinality of a $k$-packing of $G$. We prove upper bounds on the average distance in terms of $\beta_k(G)$ and show that for fixed $k$ the bounds are, up to an additive constant, best possible. As a corollary we obtain an upper bound on the average distance in terms of the $k$-domination number, the smallest cardinality of a set $S$ of vertices of $G$ such that every vertex of $G$ is within distance $k$ of some vertex of $S$.

1 Introduction

Let $G$ be a connected graph with vertex set $V(G)$ of order $n$. The average distance $\mu(G)$ is defined as $\mu(G) = \frac{1}{n^2} \sum_{\{u,v\} \subseteq V(G)} d_G(u,v)$, where $d_G(u,v)$ denotes the distance between the vertices $u$ and $v$ in $G$. The average distance, originally introduced by the chemist Wiener [16], provides a good tool to evaluate the performance of transportation networks as it is an indicator for the expected travel time between two randomly chosen points of the network. For a survey of results on the average distance before 1984 see [14]. More recent results can be found for example in [1, 6, 7] or in the survey on average distance in trees [10] by Entringer, Gutman and Dobrynin.

In 1986, the computer program GRAFFITI conjectured that for every connected graph $G$

$$\mu(G) \leq \beta(G),$$

where $\beta(G)$ is the independence number of $G$, i.e., the cardinality of a largest independent set of vertices of $G$. While Fajtlowicz and Waller [11] gave a proof of a slightly weaker inequality, $\mu(G) \leq \beta(G) + 1$, the conjecture was proved by F.R.K. Chung [3]. Her result was improved in [4], where the extremal graphs of given order and independence number and maximum average distance were determined. The aim of this paper is to generalise these results by presenting asymptotically sharp upper bounds on the average distance of graphs of given $k$-packing number. A $k$-packing of a graph $G$ is a set of vertices at distance greater than $k$ from each other, and the $k$-packing number $\beta_k(G)$ is the maximum
cardinality of a $k$-packing of $G$. Our upper bounds on $\mu(G)$ complement a result by Firby and Haviland [12], which states essentially that the \textit{minimum} average distance of a graph of given order and $k$-packing number $\beta_k$ is attained by a graph obtained from a large clique by appending $\beta_k$ paths of length $k/2$.

As a corollary we obtain similar results on the average distance of graphs of given order and $k$-domination number. A set $S \subseteq V(G)$ is $k$-\textit{dominating} in $G$ if each vertex of $G$ is within distance at most $k$ from some vertex in $S$. The minimum cardinality of a $k$-dominating set of $G$ is the \textit{$k$-domination number}, denoted by $\gamma_k(G)$. (Note that in the literature our $k$-domination number is sometimes referred to as the $k$-distance domination number, and that some authors use the term $k$-domination number for a different concept.)

In a recent paper, Tian and Xu [15] used different techniques to obtain sharp upper bounds on the average distance in terms of order and $k$-domination number. Our result for $k = 1$, i.e., for the ordinary domination number, is one of the two main results in [5]. We note that a related result, a bound on the average distance in terms of the total domination number, was given in [8].

A further motivation for our results is the fact that they relate a polynomial-time computable graph invariant, the average distance, and invariants that are NP-hard to compute, the $k$-packing number and the $k$-domination number. This can be used to obtain polynomial-time computable bounds on $k$-packing number and $k$-domination number in terms of the average distance.

We use standard notation. Specifically, if $G = (V, E)$ is a graph then $V = V(G)$ and $E = E(G)$ denote the vertex set and edge set of $G$, respectively. If $k$ is a positive integer then the $k$-th \textit{neighbourhood} of vertex $v$, denoted by $N^k_G(v)$, is the set of all vertices $u \neq v$ of $G$ with $d_G(u, v) \leq k$. For $N^1_G(v)$ we usually write $N_G(v)$. The degree of a vertex $v$, denoted by $\deg_G(v)$, is the number of its neighbours. An \textit{end vertex} of $G$ is a vertex of degree 1, and the set of all end vertices of $G$ is denoted by $\Gamma(G)$. If the graph $G$ is understood, we often drop the subscript or argument $G$. If $G$ and $H$ are graphs, then the union of $G$ and $H$, $G \cup H$, is the graph with vertex set $V(G) \cup V(H)$ and edge set $E(G) \cup E(H)$.

If $G$ is a tree and $u, v$ are vertices of $G$, then we denote the $(u, v)$-path of $G$ by $P_G(u, v)$. For three distinct vertices $u, v, w$ of $G$ there is a unique vertex $z$ that is on all three paths $P(u, v), P(u, w), P(v, w)$; this vertex is called the \textit{median vertex} of $u, v, w$. If $v$ is a vertex of $G$, not necessarily in $S$, then the \textit{distance between} $S$ and $v$, $d_G(v, S)$ is the minimum of the distances between $v$ and a vertex of $S$. If $c$ is a real valued weight function on the vertex set of $G$, and $S \subseteq V(G)$ then we follow the convention to write $c(S)$ for $\sum_{s \in S} c(s)$. The \textit{total distance} or \textit{distance} of $G$ is the sum of the distances between all unordered
pairs of vertices of $G$. It is denoted by $d(G)$. If $c$ is a non-negative weight function on $V$, then the distance of $G$ with respect to $c$ is defined by

$$d_c(G) = \sum_{(u,v) \subseteq V} c(u)c(v)d_G(u,v).$$

Hence, if $c(v) = 1$ for all $v \in V(G)$, then $d_c(G) = d(G)$. If $c(V(G)) = N > 1$, then we define the average distance of $G$ with respect to $c$ by

$$\mu_c(G) = \left(\frac{N}{2}\right)^{-1} d_c(G).$$

This definition is motivated as follows: If the vertices of a graph $G$ stand for sites of facilities, where vertex $v$ hosts $c(v)$ facilities, then $\mu_c(G)$ is the expected distance between two randomly selected facilities.

We define the distance of a vertex $v$ with respect to $c$, or $c$-distance, as

$$d_c(v) = d_c(v,G) = \sum_{w \in V - \{v\}} c(w)d_G(v,w).$$

## 2 A useful construction

The $c$-median of $G$ is the set of all vertices of minimum $c$-distance. Two properties of the ordinary median, the second one due to Barefoot, Entringer and Székely [2], also hold for the $c$-median if $c$ is strictly positive. The proofs for the ordinary median can be easily adapted to the $c$-median, so we omit it.

**Proposition 1** Let $T$ be a tree with a positive weight function $c$. Then the $c$-median of $T$ consists either of one vertex, or of two adjacent vertices.

**Lemma 1** [2] Let $T$ be a tree with a positive weight function $c$. If $v_1, v_2, \ldots, v_r$ is a path in $T$ and $v_1$ is a $c$-median vertex but $v_2$ is not, then

$$d_c(v_1) < d_c(v_2) < \ldots < d_c(v_r).$$

In particular, a vertex of maximum $c$-distance is an end vertex of $T$.

**Corollary 1** Let $T$ be a tree with a positive weight function $c$ and let $u, v, w \in V(T)$ such that $u$ is an internal vertex of $P(v,w)$. Then $d_c(v) > d_c(u)$ or $d_c(w) > d_c(u)$.

**Proof.** If $u$ is a $c$-median vertex of $T$, the statement follows from Proposition 1. So assume that $u$ is not a $c$-median vertex. Let $z$ be a $c$-median vertex of $T$ closest to the $(v,w)$-path. Then $u$ is an internal vertex of the $(z,v)$-path or of the $(z,w)$-path. Applying
Lemma 1 now yields the corollary. \( \square \)

In the following section we make extensive use of the following construction. Let \( T \) be a tree with a weight function \( c \) and let \( \emptyset \neq U \subset N(u) \) be a proper subset of the neighbourhood of a vertex \( u \). Let \( U = \{u_1, u_2, \ldots, u_p\} \) and let \( T_0, T_1, \ldots, T_p \) be the components of \( T - \{wu_1, wu_2, \ldots, wu_p\} \) containing \( u, u_1, \ldots, u_p \), respectively, and let \( A_0, A_1, \ldots, A_p \) be their respective vertex sets. If \( w \in A_0 \) then we refer to the tree \( T - \{ww_1, ww_2, \ldots, wu_p\} + \{wu_1, wu_2, \ldots, wu_p\} \) as the tree obtained from \( T \) by transferring \( U \) from \( u \) to \( w \). We note that Corollary 2 is similar to Lemma 1 in [7].

**Corollary 2** Let \( T \) be a tree with a strictly positive weight function and \( u, v, w \in V(T) \). Let \( u \) be an internal vertex of \( P(v, w) \) and let \( \emptyset \neq U \subseteq N(u) - V(P(v, w)) \). If \( T_v \) and \( T_w \) are the trees obtained from \( T \) by transferring \( U \) from \( u \) to \( v \) and \( w \), respectively, then

\[
d_c(T_v) > d_c(T) \quad \text{or} \quad d_c(T_w) > d_c(T).
\]

**Proof.** We compare the distance of \( T \) and the distance of \( T_w \). Let \( U, u_i, A_i \) and \( T_i \) be defined as above. Transferring \( U \) from \( u \) to \( w \) changes only the distances between vertices \( x, y \) with \( x \in A_0 \) and \( y \in A_i \) for some \( i \in \{1, 2, \ldots, p\} \). For such \( x, y \) we have

\[
d_{T_w}(x, y) - d_T(x, y) = \left( d_{T_w}(x, w) + 1 + d_{T_w}(u_i, y) \right) - \left( d_T(x, u) + 1 + d_T(u_i, y) \right) = d_T(x, w) - d_T(x, u).
\]

Hence

\[
d_c(T_w) - d_c(T) = \sum_{x \in A_0} \sum_{i=1}^{p} \sum_{y \in A_i} c(x)c(y) \left( d_T(x, w) - d_T(x, u) \right) = \sum_{i=1}^{p} c(A_i) \left( d_c(w, T_0) - d_c(u, T_0) \right).
\]

Similarly we have

\[
d_c(T_v) - d_c(T) = \sum_{i=1}^{p} c(A_i) \left( d_c(v, T_0) - d_c(u, T_0) \right).
\]

By Corollary 1 we have \( d_c(v, T_0) > d_c(u, T_0) \) or \( d_c(w, T_0) > d_c(u, T_0) \), and so \( d_c(T_v) - d_c(T) > 0 \) or \( d_c(T_v) - d_c(T) > 0 \), as desired. \( \square \)

**Definition 1** A \( k \)-star is a tree of radius at most \( k \).

**3 Results**

In what follows we consider a tree \( H \) that is obtained from the disjoint union of \( r \) trees \( T_1, T_2, \ldots, T_r \) by adding edges \( e_1, e_2, \ldots, e_{r-1} \), where each edge joins vertices in two different \( T_i \). We refer to the edges \( e_i \) as **link edges**. The set of link edges is denoted by \( E' \).
hence

\[ H = \bigcup_{i=1}^{r} T_i + E'. \]

A vertex incident with a link edge is a contact vertex. The set of contact vertices of \( T_i \) is denoted by \( W_i \). If \( T_i \) is joined to the remainder of the graph by only one link edge then we call \( T_i \) an end tree.

We also consider trees obtained from the disjoint union of trees \( T_1, T_2, \ldots, T_r \) by successively identifying \( r-1 \) pairs of vertices in different components. Hence a vertex of some \( T_i \) can be identified with more than one vertex. A vertex of a tree \( T_i \) that is identified with a vertex from some other tree \( T_j \) is also called a contact vertex.

A basic result on average distance states that among all connected graphs of given order \( n \) the path \( P_n \) has maximum average distance. Hence it is not surprising that our bounds (presented below) on the average distance of a graph \( G \) of given \( k \)-packing number depend on whether \( \beta_k(G) \leq \beta_k(P_n)(= 1 + \left\lfloor \frac{n-1}{k+1} \right\rfloor) \) or \( \beta_k(G) > \beta_k(P_n) \). For the former case we present a bound that is best possible apart from an additive term depending only on \( k \) and not on the order of \( G \). For the latter case we determine the unique extremal graph.

**Lemma 2** Let \( G \) be a tree and let \( c \) be a weight function on \( V(G) \) of total weight \( N \geq 3 \) with \( c(v) \geq 1 \) for all \( v \in V(G) \).

(a) If \( G \) is obtained from the vertex disjoint union of \( k \)-stars \( T_1, T_2, \ldots, T_r \) by adding \( r-1 \) edges, then

\[ \mu_c(G) \leq \frac{2k+1}{2} r + k - \frac{1}{2} + \frac{1}{\sqrt{N}}. \]

(b) If \( G \) is obtained from the vertex disjoint union of \( k \)-stars \( T_1, T_2, \ldots, T_r \) by identifying \( r-1 \) pairs of vertices of distinct \( T_i \), then

\[ \mu_c(G) \leq kr + k + \frac{1}{\sqrt{N}}. \]

**Proof.** We only prove part (a); the proof of (b) is almost identical.

For fixed \( k \) and \( N \) assume that \( G \) and \( c \) are chosen such that \( d_c(G) \) is maximum. If \( r = 1 \) then \( G \) has radius at most \( k \) and thus diameter at most \( 2k \), and so \( \mu_c(G) \leq 2k < 2k + \frac{1}{\sqrt{N}} \), hence the statement holds. So assume that \( r \geq 2 \). Then \( G \) has at least two end trees, say, \( T_1 \) and \( T_2 \). For each \( T_i, i = 1, 2, \ldots, r \), fix a vertex \( v_i \) (for example a centre vertex) so that every vertex of \( T_i \) is within distance \( k \) of \( v_i \), and let \( P := P(v_1, v_2) \).

**Claim 1:** All internal vertices of \( P \) have degree 2 in \( G \).

Suppose to the contrary that \( P \) contains an internal vertex \( z \) of degree greater than 2. Let \( z \) be contained in tree \( T_j \).
First assume that \( v_j \) is on \( P \). Define \( U = N_G(z) - V(P) \) and let \( G_1 \) and \( G_2 \) be the tree obtained from \( G \) by transferring \( U \) from \( z \) to \( v_1 \) and \( v_2 \), respectively. Clearly, \( G_1 \) and \( G_2 \) satisfy the hypothesis. (Those vertices \( x \) of \( T_j \) that are separated from \( v_j \) by \( z \) have to leave \( T_j \) and to be included in \( T_1 \) or \( T_2 \), but their distance to \( v_1 \) or \( v_2 \) in \( G_1 \) or \( G_2 \), respectively, is at most \( r - d_G(v_j, z) \). All other vertices remain in the same \( T_i \) and thus within distance \( k \) of \( v_i \).) But by Corollary 2 we have \( d_c(G_1) > d_c(G) \) or \( d_c(G_2) > d_c(G) \), a contradiction.

Now assume that \( v_j \) is not on \( P \). Define \( U' := N_G(z) - V(P(v_1, v_j)) \) and let \( G'_1 \) and \( G'_j \) be the tree obtained from \( G \) by transferring \( U' \) from \( z \) to \( v_1 \) and \( v_j \), respectively. As above, we obtain a contradiction to the maximality of \( d_c(G) \).

Claim 1 yields that \( T_1 \) and \( T_2 \) are the only end trees and that the other \( r - 2 \) \( k \)-stars \( T_j \) are paths. Each \( T_j \), \( j > 2 \), has radius at most \( k \) and thus at most \( 2k + 1 \) vertices. Hence \( P \) has at most \( (r - 2)(2k + 1) + 2(k + 1) = (r - 1)(2k + 1) + 1 \) vertices. Let \( T'_i := T_i - (V(P) - \{v_i\}) \) for \( i = 1, 2 \). Then \( T'_1 \) and \( T'_2 \) are \( k \)-stars that have only \( v_1 \) and \( v_2 \), respectively, in common with \( P \). This proves the following.

Claim 2: \( G = P \cup T'_1 \cup T'_2 \),

where \( P \) is a path of length at most \( (r - 1)(2k + 1) \) and \( T'_1 \) and \( T'_2 \) are vertex disjoint \( k \)-stars, and \( T'_i \) has only vertex \( v_i \) in common with \( P \) for \( i = 1, 2 \).

Now each vertex not on \( P \) is within distance \( k \) of \( v_1 \) or \( v_2 \). Let \( c' \) be the weight function obtained from \( c \) by moving the weights of the vertices in \( T'_1 \) and \( T'_2 \) to \( v_1 \) or \( v_2 \), respectively. More precisely, define for \( v \in V(G) \)

\[
c'(v) = \begin{cases} 
  c(v) & \text{if } v \in V(P) - \{v_1, v_2\}, \\
  c(V(T'_i)) & \text{if } v = v_i, \ i \in \{1, 2\}, \\
  0 & \text{otherwise}. 
\end{cases}
\]

Since no weight has been moved over a distance of more than \( k \), no distance between two weight units has changed by more than \( 2k \). So \( |\mu_c(G) - \mu_c(G)| \leq 2k \). The vertices not on \( P \) have weight 0, so they can be ignored. Hence, if \( c'' \) is the restriction of \( c' \) to \( V(P) \),

\[
\mu_c(G) \leq \mu_c(G) + 2k = \mu_{c''}(P) + 2k. \tag{1}
\]

Clearly the total weight \( c''(V(P)) \) equals \( N \) and every vertex has weight at least 1. Denote the length of \( P \) by \( \ell \).

Claim 3: If \( H \) is a path of length \( \ell \) and \( h \) a vertex weight function with \( h(V(H)) = N \) and \( h(v) \geq 1 \) for each vertex, then

\[
d_h(H) \leq \frac{1}{4} \ell N^2 - \frac{1}{12} (\ell - 1) \ell (\ell + 1).
\]

6
We may assume that $H$ is a path $a_0, a_1, \ldots, a_\ell$ with weight function $h$ such that $d_h(H)$ is maximum among all paths satisfying the above conditions. We first show that

$$h(a_i) = 1 \text{ for all } i \in \{1, 2, \ldots, \ell - 1\}. \quad (2)$$

Indeed, if $h(a_i) = 1 + \epsilon$, where $\epsilon > 0$, consider the weight function $\tilde{h}$ of total weight $N - \epsilon$ in which the weight of $a_i$ is reduced to 1, while the other weights remain unchanged. Clearly,

$$d_{\tilde{h}}(H) = d_h(H) + \epsilon d_h(a_i, H).$$

Now let $h'$ ($h''$) be the weight function obtained from $\tilde{h}$ by increasing the weight of $a_0$ ($a_\ell$) by $\epsilon$, so that $h'$ ($h''$) has total weight $N$. Then

$$d_{h'}(H) = d_{\tilde{h}}(H) + \epsilon d_{\tilde{h}}(a_0, H), \quad d_{h''}(H) = d_{\tilde{h}}(H) + \epsilon d_{\tilde{h}}(a_\ell, H).$$

By Corollary 1 we have $d_{\tilde{h}}(a_0) > d_{\tilde{h}}(a_i)$ or $d_{\tilde{h}}(a_\ell) > d_{\tilde{h}}(a_i)$, and so $d_{h'}(H) > d_h(H)$ or $d_{h''}(H) > d_h(H)$. This contradiction proves (2).

Hence only end vertices of $H$ have weight greater than 1. A simple maximisation now shows that the end vertices of $H$ have equal weight $\frac{1}{2} (N - \ell + 1)$ and that

$$d_h(H) = \frac{1}{4} \ell N^2 - \frac{1}{12} (\ell - 1)\ell(\ell + 1) \quad (3)$$

and so Claim 3 follows.

Now (1), after division by $\binom{N}{2}$, in conjunction with Claim 3 yield

$$\mu_c(G) \leq \frac{1}{2} + \frac{\ell}{2(N-1)} - \frac{(\ell - 1)\ell(\ell + 1)}{6N(N - 1)} + 2k.$$

A simple maximisation shows that $\frac{\ell}{2(N-1)} - \frac{(\ell - 1)\ell(\ell + 1)}{6N(N - 1)}$ is maximised, as a function of $\ell$, if $\ell = \sqrt{N + \frac{1}{3}}$, and can thus be bounded above by $(N + \frac{1}{3})^{3/2} / (3N(N - 1))$, which is less than $N^{-1/2}$ for $N \geq 3$. Since $\ell \leq (r - 1)(2k + 1)$ we obtain

$$\mu_c(G) \leq \frac{1}{2} (r - 1)(2k + 1) + 2k + \frac{1}{\sqrt{N}} = r \frac{2k + 1}{2} + k - \frac{1}{2} + \frac{1}{\sqrt{N}},$$

as desired. \qed

**Theorem 1** Let $G$ be a connected graph of order $n \geq 3$ and $k$-packing number $\beta_k \leq 1 + [(n - 1)/(k + 1)]$. Then

$$\mu(G) \leq \frac{k + 1}{2} \beta_k + \frac{3k + (-1)^{k+1}}{2} + \frac{1}{\sqrt{n}}.$$
Proof. (a) We first consider the case that \( k \) is even. We find a maximal \( k \)-packing \( A \) of \( G \) using the following procedure. Choose an arbitrary vertex \( v_1 \) of \( G \) and let \( A = \{ v_1 \} \). If there exists a vertex \( v_2 \) in \( G \) with \( d_G(v_2, A) = k + 1 \) add \( v_2 \) to \( A \). Add vertices \( v_i \) with \( d_G(v_i, A) = k + 1 \) to \( A \) until each of the vertices not in \( A \) is within distance \( k \) of \( A \). Let \( A = \{ v_1, v_2, \ldots, v_r \} \). Since \( A \) is a \( k \)-packing, we have

\[
r \leq \beta_k(G).
\]

For each vertex \( v_i \in A \) let \( T_i \) be a spanning tree of \( N_{G}^{k/2}(v_i) \cup \{ v_i \} \) that is distance preserving from \( v_i \), i.e., \( d_{T_i}(v_i, x) = d_G(v_i, x) \) for all \( x \in V(T_i) \). By our construction of \( A \), there exist \( r - 1 \) edges in \( G \), each of them joining an end vertex of some \( T_i \) to an end vertex of some \( T_j, i \neq j \), whose addition to \( \bigcup T_i \) yields a (not necessarily spanning) subtree \( H \) of \( G \).

The graph \( H \) is a union of \( k/2 \)-stars \( T_1, T_2, \ldots, T_r \) with additional edges, hence it satisfies the hypothesis of Lemma 2. We now define a weight function \( c \) on \( V(H) \), which is obtained from the constant unit weight function on \( V(G) \) by moving the weight of each vertex not in \( H \) to a nearest vertex in \( H \). More precisely, we define \( c \) as follows. For each vertex \( u \in V(G) \setminus V(H) \) choose a vertex \( u' \in V(H) \) which is closest to \( u \) and define a mapping \( f : V(G) \setminus V(H) \to V(H) \) by \( f(u) = u' \). Then \( d_G(u, f(u)) \leq k/2 \). For \( v \in V(G) \) let \( g(v) \) be the number of vertices \( u \) in \( V(G) \setminus V(H) \) with \( f(u) = v \). We now define the weight function \( c \) on \( V(G) \) by

\[
c(v) = \begin{cases} 
g(v) + 1 & \text{if } v \in V(H), \\ 0 & \text{if } v \in V(G) \setminus V(H). \end{cases}
\]

Since the weight of each vertex was moved over a distance not exceeding \( k/2 \), we have

\[
\mu(G) \leq \mu_c(G) + k. \tag{4}
\]

Since the vertices not in \( H \) have weight 0, they do not contribute to the total distance of \( G \) and \( c \). Hence \( \mu_c(G) \leq \mu_c(H) \) and thus

\[
\mu(G) \leq \mu_c(H) + k. \tag{5}
\]

Now \( H \) is obtained from the union of \( r \) disjoint \( k/2 \)-stars by adding edges. Each vertex has weight at least 1 and the total weight equals \( n \). Hence, by Lemma 2,

\[
\mu_c(H) \leq \frac{k + 1}{2} r + \frac{k - 1}{2} + \frac{1}{\sqrt{n}}. \tag{6}
\]

and thus, by \( r \leq \beta_k \),

\[
\mu(G) \leq \frac{k + 1}{2} r + \frac{k - 1}{2} + \frac{1}{\sqrt{n}} + k \leq \frac{k + 1}{2} \beta_k + \frac{3k - 1}{2} + \frac{1}{\sqrt{n}},
\]
as desired.

(b) We now consider the case that \( k \) is odd. As in the proof of (a), we find a maximal \( k \)-packing \( A = \{v_1, v_2, \ldots, v_r\} \). For each \( i \), \( 2 \leq i \leq r \), there exists a vertex \( v_j \), \( j < i \), such that \( d_G(v_j, v_i) = k + 1 \). Let \( w_{i,j} \) be a vertex with \( d_G(v_i, w_{i,j}) = d_G(v_j, w_{i,j}) = (k + 1)/2 \). For each \( i, 1 \leq i \leq r \) choose a set \( V_i \) such that \( N^{(k-1)/2}(v_i) \cup \{v_i\} \subseteq V_i \) and, in addition, each vertex \( v \) which is within distance \( (k + 1)/2 \) of \( A \) is contained in exactly one set \( V_i \), unless \( v = w_{i,j} \) for some \( i, j \), in which case \( v \) is contained in both, \( V_i \) and \( V_j \). Let \( T_i \) be a spanning tree of \( V_i \) which is distance preserving from \( v_i \). Hence the \( T_i \) are \( (k + 1)/2 \)-stars and \( G \) is obtained from \( \bigcup_{i=1}^r T_i \) by identifying \( r - 1 \) vertices, so we can apply Lemma 2(b). The remainder of the proof is analogous to the proof of part (a). \( \square \)

We now show that for fixed \( k \) the bound in Theorem 1 is, apart from an additive constant, best possible.

For \( k = 1 \) consider the graph \( G_{p,1,r} \) obtained from two disjoint cliques \( H_1 \) and \( H_2 \) of order \( p \) and a path of order \( 2r - 2 \) by identifying one of its end vertices with a vertex of \( H_1 \) and its other end vertex with a vertex of \( H_2 \). As shown in [4], graph \( G_{p,1,r} \) has independence number \( r \) and \( \mu(G_{p,1,r}) \) approaches \( r \) as \( p \) approaches infinity.

Given positive integers \( k \geq 2 \) and \( p \), define the graph \( I_{k,p} \) as follows. For \( k \geq 4 \) let \( I_{k,p} \) be the graph obtained from the star \( K_{1,p} \) by subdividing each edge \( \lfloor k/2 \rfloor - 2 \) times and then appending \( p \) end vertices to each end vertex of \( K_{1,p} \). Furthermore define \( I_{2,p} \) and \( I_{3,p} \) to be the star \( K_{1,p^2} \). Clearly \( I_{k,p} \) has order \( p^2 + p(\lfloor k/2 \rfloor - 1) + 1 \) and diameter \( 2\lfloor k/2 \rfloor \).

Let \( G_{p,k,r} \) be the graph obtained from two disjoint copies of \( I_{k,p} \) by identifying the centre vertex of one \( I_{k,p} \) with an end vertex of a path of order \( r(k + 1) - 2[\lfloor k/2 \rfloor] \), and the centre vertex of the other \( I_{k,p} \) with the other end vertex of this path. It is easy to verify that \( G_{p,k,r} \) has \( 2p^2 + (2[\lfloor k/2 \rfloor] - 2)p + rk + r - 2[\lfloor k/2 \rfloor] \) vertices, \( k \)-packing number \( r \), and diameter \( r(k + 1) - 1 \).

Let \( k \) and \( r \) be fixed and let \( p \) tend to infinity. A long but straightforward calculation shows that \( \mu(G_{p,k,r}) \) approaches \( r \frac{k+1}{2} + \frac{\lfloor k \rfloor}{2} - \frac{1}{2} \).

**Corollary 3** Let \( G \) be a connected graph on at least 3 vertices with \( k \)-packing number \( \beta_k \). Then

\[
\mu(G) \leq \frac{k+1}{2} \beta_k + \frac{3k + (-1)^{k+1}}{2} + \frac{1}{\sqrt{3}}.
\]

**Proof.** If \( \beta_k \leq 1 + [(n - 1)/(k + 1)] \) then the corollary follows by Theorem 1. For \( \beta_k > 1 + [(n - 1)/(k + 1)] \) we have \( \frac{k+1}{2} \beta_k + \frac{3k + (-1)^{k+1}}{2} + \frac{1}{\sqrt{3}} > \frac{n-1}{2} + 2 > \frac{n+1}{3} \), and so the corollary follows from the well known inequality \( \mu(G) \leq \mu(P_n) = \frac{n+1}{3} \). \( \square \)
For $k = 1$, Theorem 1 yields the bound $\mu(G) \leq \beta(G) + 2 + n^{-1/2}$, which differs from Fajtlowicz’ bound $\mu(G) \leq \beta(G) + 1$ (see [11]) and Chung’s bound $\mu(G) \leq \beta(G)$ (see [3]) only by an additive constant. For $\beta_2$, the packing number of $G$, we obtain the bound $\mu(G) \leq \frac{3}{2} \beta_2(G) + \frac{5}{2} + \frac{1}{\sqrt{n}}$.

We now consider graphs with $k$-packing number $\beta_k(G) > \beta_k(P_n) = 1 + \lfloor \frac{n-1}{k+1} \rfloor$.

**Definition 2** Given integers $k > 0$ and $\ell$ with $0 \leq \ell \leq 2k$. By $Y_{k,\ell}(a, b, c)$ we mean a tree $T$ containing vertices $a, b, c$ such that $d(a, b) = d(a, c) = k$ and $d(b, c) = \ell$, and either
(i) $\Gamma(T) = \{a, b\}$ and $\ell = 0$,
(ii) $\Gamma(T) = \{b, c\}$ with $\ell = 2k$, or
(iii) $\Gamma(T) = \{a, b, c\}$ with $0 < \ell < 2k$.

Note that $Y_{k,\ell}(a, b, c)$ is uniquely determined, up to isomorphism, and has $k + \frac{1}{2} \ell + 1$ vertices.

**Definition 3** (a) For given integers $n, r, k > 0$ with $r(k + 1) \leq n \leq (r - 1)(2k + 1)$ let $n - r(k + 1) = pk + q$, where $0 \leq q < k$, and let $\ell_1, \ell_2, \ldots, \ell_r$ be integers with
\[
\ell_i = 0 \quad \text{for } i \in \{1, 2, \ldots, \lfloor \frac{r-p-1}{2} \rfloor \},
\ell_i = 2q \quad \text{for } i = \lfloor \frac{r-p+1}{2} \rfloor,
\ell_i = 2k \quad \text{for } i \in \{ \lfloor \frac{r-p+1}{2} \rfloor + 1, \lfloor \frac{r-p+1}{2} \rfloor + 2, \ldots, r - \lfloor \frac{r-p-1}{2} \rfloor \},
\ell_i = 0 \quad \text{for } i \in \{ r - \lfloor \frac{r-p-1}{2} \rfloor + 1, r - \lfloor \frac{r-p-1}{2} \rfloor + 2, \ldots, r \}.
\]

We define $G_0(n, k, r)$ to be the tree obtained from the union of vertex disjoint trees $T_i$, $i = 1, 2, \ldots, r$, with $T_i = Y_{k,\ell_i}(v_i, w_{i,1}, w_{i,2})$, by adding link edges $w_{i,2}w_{i+1,1}$ for $i = 1, 2, \ldots, r - 1$. (We note that $G_0(n, k, r)$ has order $n$.)

(b) For given integers $n, r, k > 0$ with $rk + 1 \leq n \leq (r - 1)2k$ let $n - rk - 1 = pk + q$, where $0 \leq q < k$, and let $\ell_1, \ell_2, \ldots, \ell_r$ be defined by the four terms in (a).
We define $G_1(n, k, r)$ to be the tree obtained from the union of vertex disjoint trees $T_i$, $i = 1, 2, \ldots, r$, with $T_i = Y_{k,\ell_i}(v_i, w_{i,1}, w_{i,2})$, by identifying $w_{i,2}$ with $w_{i+1,1}$ for $i = 1, 2, \ldots, r - 1$. (We note that $G_1(n, k, r)$ has order $n$.)

Examples for the graphs defined above are given below. Both graphs are obtained from four trees $Y_{2,0}$, one tree $Y_{2,2}$, and one tree $Y_{2,4}$. In the first figure, the link edges of $G_0(21, 2, 6)$ are indicated by thin lines, while the edges of the $T_i$ are thick lines.
Lemma 3 Let $G$ be a tree of order $n$ and let $r$ and $k$ be positive integers such that $r > 1 + \lfloor \frac{n-1}{2k+1} \rfloor$. 

(a) If $G$ is obtained from vertex disjoint trees $T_1, T_2, \ldots, T_r$, rooted at $v_1, v_2, \ldots, v_r$, respectively, by adding $r-1$ link edges such that $d_G(v_i, w_i) \geq k$ for each contact vertex $w_i$ of $T_i$, $i = 1, 2, \ldots, r$, then

$$d(G) \leq d(G_0(n, k, r)).$$

(b) If $G$ is obtained from vertex disjoint trees $T_1, T_2, \ldots, T_r$, rooted at $v_1, v_2, \ldots, v_r$ by identifying $r-1$ pairs of contact vertices of different trees $T_i$ such that $d_G(v_i, w_i) \geq k$ for each contact vertex $w_i$ of $T_i$, $i = 1, 2, \ldots, r$, then

$$d(G) \leq d(G_1(n, k, r)).$$

**Proof.** (a) We first observe that $G$ is not a path since it is easy to see that for each decomposition of a path of order $n$ into $r$ trees as above we have $r \leq 1 + \lfloor \frac{n-1}{2k+1} \rfloor$. We also note that $r \geq 3$. This follows from the fact that $n \geq r(k+1)$ since each $T_i$ has at least $k+1$ vertices, which implies $r > 1 + \lfloor \frac{n-1}{2k+1} \rfloor \geq 1 + \lfloor \frac{r(k+1)-1}{2k+1} \rfloor$. The latter inequality does not hold for $r = 1$ or $r = 2$, so $r \geq 3$.

We may assume that $G$ has, among all trees satisfying the hypothesis, maximum total distance. We also assume that the decomposition of $G$ into $r$ rooted trees $T_i$ is such that
the number of roots that are end vertices of \( G \) is as large as possible. We denote the set of contact vertices of \( T_i \) by \( W_i \).

Our strategy is as follows. The first claims will show that each \( T_i \) has at most two contact vertices, and is isomorphic to some graph \( Y_{k,\ell_i}(a, b, c) \). Then we will show that the \( \ell_i \) satisfy certain conditions, which eventually leads to identifying \( G_0 \) and \( G_1 \) as the extremal graphs.

CLAIM 1: \( \Gamma(T_i) \subseteq W_i \cup \{v_i\} \) for all \( i \).

Suppose to the contrary that \( T_i \) has an end vertex \( u \) not in \( W_i \cup \{v_i\} \). Let \( P: u, u_1, \ldots, u_s \) be the path to the closest vertex \( u_s \) of degree at least 3 in \( G \).

Then \( P - u_s \) contains neither a contact vertex nor the root \( v_i \). To see this suppose that there exists a \( j \leq s - 1 \) with \( u_j \in W_i \cup \{v_i\} \), and that \( j \) is minimum. If \( u_j \) is a contact vertex then \( T_i \) contains only \( u, u_1, \ldots, u_j \) but no root; but if \( u_j \) is the root \( v_i \) then we could root \( T_i \) at \( u \) instead of \( u_j \), thus increasing the number of roots that are end vertices. Now \( u_s \) has at least two neighbours \( u', u'' \neq u_{s-1} \). Let \( G_1 \) and \( G_2 \) be the tree obtained by transferring \( \{u_{s-1}\} \) from \( u_s \) to \( u' \) and \( u'' \), respectively. Although \( u, u_1, \ldots, u_{s-1} \) might become part of another tree \( T_j \), this does not change the set of contact vertices. Hence the graphs \( G_1 \) and \( G_2 \) satisfy the hypothesis of the lemma. Applying Corollary 2 to the path \( u', u_s, u'' \) now yields that \( G_1 \) or \( G_2 \) has larger distance than \( G \). This contradiction to the choice of \( G \) proves Claim 1.

CLAIM 2: Each \( T_i \) is incident with at most 2 link edges.

Suppose to the contrary that \( T_i \) is incident with three link edges \( w_1w'_1, w_2w'_2 \) and \( w_3w'_3 \), where \( w_1, w_2, w_3 \) are in \( T_i \). Note that some of these vertices might coincide. Let \( z \in V(T_i) \) be the median vertex of \( w_1, w_2, w_3 \). We may assume that \( v_i \) is not in the component of \( T_i - z \) containing \( w_3 \). Then \( z \) has a neighbour \( a \) closest to \( w'_3 \) (possibly \( a = w'_3 \)). Let \( G_1 \) and \( G_2 \) be the tree obtained from \( G \) by transferring \( \{a\} \) from \( z \) to \( w'_1 \) and to \( w'_2 \). It is easy to verify that \( G_1 \) and \( G_2 \) satisfy the hypothesis of the Lemma. Since \( w'_1, z, \) and \( w'_2 \) lie on a path, Corollary 2 yields a contradiction to the maximality of \( d(G) \), and so Claim 2 follows.

Now each \( T_i \) is incident with at most two link edges. So \( T_1, T_2, \ldots, T_r \) are connected, by link edges, in a path-like fashion. We can assume that \( T_i \) is joined, via a link edge, to \( T_{i-1} \) and to \( T_{i+1} \) for \( i = 2, 3, \ldots, r - 1 \). Denote the contact vertex of \( T_i \) that is adjacent to a contact vertex of \( T_{i-1} \) (or \( T_{i+1} \)) by \( w_{i,1} \) (or \( w_{i,2} \)). Because it will turn out to be convenient we also define \( w_{1,1} := w_{1,2} \) and \( w_{r,2} := w_{r,1} \).

CLAIM 3: \( T_1 \) (or \( T_r \)) is a path of length at least \( k \) with end vertices \( v_1 \) and \( w_{1,2} \) (or \( v_r \) and \( w_{r,1} \)).
$T_1$ has only one contact vertex, $w_{1,2}$. It follows from Claim 1 that $\Gamma(T_1) \subseteq \{v_1, w_{1,2}\}$. Hence $T_1$ is a path with end vertices $v_1$ and $w_{1,2}$. Since $d(v_1, w_{1,2}) \geq k$, $T_1$ has length at least $k$, as desired. The statement for $T_r$ is proved analogously.

Claim 4: Let $v_i \notin \Gamma(T_i)$. Then $T_i$ is a path of length at least $2k$ with end vertices $w_{i,1}$ and $w_{i,2}$.

By Claim 1 and $v_i \notin \Gamma(T_i)$ we have $\Gamma(T_i) \subseteq W_i$. Now $\Gamma(T_i)$ contains at least two vertices, and, by Claim 2, $W_i$ contains at most two vertices. Hence $\Gamma(T_i) = W_i$ and $T_i$ is a path with end vertices $w_{i,1}$ and $w_{i,2}$. Since $d(v_i, w_{i,1}) \geq k$ and $d(v_i, w_{i,2}) \geq k$, $T_i$ has length $d(w_{i,1}, v_i) + d(v_i, w_{i,2}) \geq 2k$, as desired.

Claim 5: Let $v_i \in \Gamma(T_i)$ and $1 < i < r$. Then $T_i = Y_{k,\ell}(v_i, w_{i,1}, w_{i,2})$ and $0 \leq \ell < 2k$.

First consider the case that $T_i$ is not a path. Then $T_i$ has at least three end vertices, and so $|W_i| \geq 2$ by Claim 1. Hence $w_{i,1} \neq w_{i,2}$ and in conjunction with Claim 2 we obtain that $\Gamma(T_i) = \{w_{i,1}, w_{i,2}, v_i\}$. Furthermore $T_i$ has exactly one vertex $z$ of degree 3, and all vertices in $V(T_i) - \{v_i, w_{i,1}, w_{i,2}, z\}$ have degree 2. Let $a, b$, and $c$ be the neighbours of $z$ closest to $w_{i,1}, v_i$, and $w_{i,2}$, respectively. Without loss of generality we assume $d(v_i, w_{i,1}) \geq d(v_i, w_{i,2})$. Let $\ell = d_{T_i}(v_i, w_{i,2})$. To prove that $T_i$ is a $Y_{k,\ell}(v_i, w_{i,1}, w_{i,2})$ it suffices to show that $d(v_i, w_{i,1}) \leq k$. Suppose to the contrary that $d_{T_i}(v_i, w_{i,1}) \geq k + 1$. Let $G_1$ and $G_2$ be the graphs obtained from $G$ by transferring $\{a\}$ from $z$ to $b$ and $c$, respectively. The distance between $v_i$ and $w_{i,1}$ in $G_1$ (in $G_2$) is one less (one more) than in $G$, hence both graphs satisfy the hypothesis of the lemma. By Corollary 2, at least one of $G_1$ and $G_2$ has greater distance than $G$. This contradiction shows that $d_{T_i}(v_i, w_{i,1}) = k$ and so $d_{T_i}(v_i, w_{i,2}) = k$. It follows that $T_i$ is a $Y_{k,\ell}(v_i, w_{i,1}, w_{i,2})$. Since $v_i$ is not on $P(w_{i,1}, w_{i,2})$ we have $\ell = d(w_{i,1}, w_{i,2}) < d(w_{i,1}, v_i) + d(v_i, w_{i,2}) = 2k$, as desired.

Now consider the case that $T_i$ is a path. The two end vertices of $T_i$ are $v_i$ and $z$, respectively. Now transferring $\{b\}$ from $w_{i,1}$ to $a$ yields a tree $G_1$ satisfying the hypothesis of the lemma since $d(a, v_i) = d(v_i, w_2) - 1 \geq k$, so that each vertex transferred to $a$ is at distance at least $k$ from $v_i$. Transferring $\{b\}$ from $w_{i,1}$ to $w_{i-1,2}$ (thus removing $b$ and its descendants in $T_i$, and attaching them to the tree $T_{i-1}$) also yields a tree, $G_2$, which satisfies the hypothesis of the lemma. Again by Corollary 2 we obtain $d(G_1) > d(G)$ or $d(G_2) > d(G)$, a contradiction to the maximality of $d(G)$. Hence $z$ is the only contact vertex of $T_i$, and so $w_{i,1} = w_{i,2} = z$.

It remains to show that $T_i$ is a path of length $k$. Suppose to the contrary that $T_i$ is a path of length greater than $k$. Let $a$ be the neighbour of $z$ in $T_i$. Let $G_1$ and $G_2$ be the
trees obtained from $G$ by transferring $\{w_{i+1,1}\}$ from $w_{i,2}$ to $w_{i-1,2}$ and to $a$, respectively. It is easy to see that both trees satisfy the hypothesis of the Lemma. Again, Corollary 2 yields a contradiction to the maximality of $d(G)$. Hence $T_i$ is a path of length $k$ and so $T_i = Y_{k,0}(v_i, w_{i,1}, w_{i,2})$.

**Claim 6:** For all $i \in \{1, 2, \ldots, r\}$ there is some $\ell \in \{0, 1, \ldots, 2k\}$ with $T_i = Y_{k,\ell}(v_i, w_{i,1}, w_{i,2})$.

Suppose to the contrary that there exists an $i$ such that $T_i$ is not of the form $Y_{k,\ell}(v_i, w_{i,1}, w_{i,2})$.

It follows from Claims 3 to 5 that either $i \in \{1, r\}$ and $T_i$ is a path of length greater than $k$, or that $1 < i < r$ and $T_i$ is a path of length greater than $2k$. Since $G$ is not a path, there exists $j \in \{1, 2, \ldots, r\}$ such that $T_j = Y_{k,\ell}(v_j, w_{j,1}, w_{j,2})$ and $0 \leq \ell < 2k$. Assume without loss of generality that $i < j$. We can further assume that $i$ and $j$ are chosen such that $j - i$ is minimum. Then the trees $T_{i+1}, T_{i+2}, \ldots, T_{j-1}$ are paths of length $2k$ and the their vertices induce the path $P(w_{i,2}, w_{j,1})$. Now modify the trees $T_{i+1}, \ldots, T_{j-1}$ by “shifting” them towards $T_i$, removing vertex $w_{i,2}$ from $T_i$, and adding vertex $w_{j-1,2}$ to $T_j$. More precisely define $T'_i = T_i - w_{i,2}$, $T_j = G[V(T_j) \cup \{w_{j-1,2}\}]$ and $T'_s = G[V(T_s) - \{w_{s,2}\} \cup \{w_{s-1,2}\}]$ for $s = i + 1, i + 2, \ldots, j - 1$. It is easy to see that, after choosing suitable roots for $T'_i, \ldots, T'_j$, the decomposition of $V(G)$ into trees $T_1, \ldots, T_{i-1}, T'_i, \ldots, T'_j, T_{j+1}, \ldots, T_r$ satisfies the hypothesis of the lemma. However, the tree $T'_j$ is neither a path nor of the form $Y_{k,\ell}(v_j, w_{j,1}, w_{j,2})$, therefore $d(G)$ is not maximal. This contradiction proves Claim 6.

In what follows denote the set $V(T_i) \cup V(T_{i+1}) \cup \ldots \cup V(T_j)$ by $V_{i,j}$ and let $T_{i,j}$ be its induced subtree.

**Claim 7:** Let $1 \leq i < j \leq r$.

(i) If $|V_{i,j-1}| < |V_{j+1,r}|$ then $d(w_{i,1}, T_{i,j}) \leq d(w_{j,2}, T_{i,j})$.

(ii) If $|V_{i,j-1}| > |V_{j+1,r}|$ then $d(w_{i,1}, T_{i,j}) \geq d(w_{j,2}, T_{i,j})$.

We only show (i), the proof of (ii) being similar. We also assume that $i > 1$ since for $i = 1$ the proof is along the same lines. Suppose to the contrary that $d(w_{i,1}, T_{i,j}) > d(w_{j,2}, T_{i,j})$. We now “reverse” the order of the trees $T_i, T_{i+1}, \ldots, T_j$. Define a new graph $G_1 = G - \{w_{i-1,2}w_{i,1}, w_{j,2}w_{j+1,1}\} + \{w_{i-2}w_{j,2}, w_{i,1}w_{j+1,1}\}$. It is easy to see that $G_1$ satisfies the hypothesis of the lemma and that the distance between two vertices $x$ and $y$ remains unchanged unless $x \in V_{i,j-1} \cup V_{j+1,r}$ and $y \in V_{i,j}$, or vice versa. Let $A = V_{i,j-1}$, $B = V_{i,j}$ and $C = V_{j+1,r}$. If $x \in A$ and $y \in B$ we have

\[
\begin{align*}
    d_{G_1}(x, y) - d_G(x, y) &= d_{G_1}(x, w_{i-1,2}) + 1 + d_{G_1}(w_{j,2}, y) \\
    &\quad -(d_G(x, w_{i-1,2}) + 1 + d_G(w_{i,1}, y)) \\
    &= d_G(w_{j,2}, y) - d_G(w_{i,1}, y).
\end{align*}
\]
Similarly we have \( d_{G_1}(x, y) - d_G(x, y) = d_{G}(w_{i,1}, y) - d_{G}(w_{j,2}, y) \) for \( x \in C \) and \( y \in B \).

Hence we have in total
\[
d(G_1) - d(G) = \sum_{x \in A} \sum_{y \in B} \left( d_{G_1}(x, y) - d_{G}(x, y) \right) + \sum_{x \in C} \sum_{y \in B} \left( d_{G_1}(x, y) - d_{G}(x, y) \right)
\]
\[
= \sum_{x \in A} \sum_{y \in B} \left( d_{G}(w_{j,2}, y) - d_{G}(w_{i,1}, y) \right)
\]
\[
+ \sum_{x \in C} \sum_{y \in B} \left( d_{G}(w_{i,1}, y) - d_{G}(w_{j,2}, y) \right)
\]
\[
= (|A| - |C|)(d(w_{j,2}, T_{i,j}) - d(w_{i,1}, T_{i,j}))
\]
\[
> 0,
\]
by our assumption \( |V_{k,i-1}| < |V_{j+1,r}| \) and \( d(w_{i,1}, T_{i,j}) > d(w_{j,2}, T_{i,j}) \). This contradiction proves Claim 7.

It follows from Claim 6 that we can associate with \( G \) a sequence \( \ell_1, \ell_2, \ldots, \ell_r \) such that \( \ell_i = a \) if \( T_i = Y_{k,a}(v_i, w_{i,1}, w_{i,2}) \) for \( i = 1, 2, \ldots, r \). This sequence determines \( G \) up to isomorphism. It remains to show that the \( \ell_i \) satisfy the conditions in Definition 3.

**Claim 8:** Let \( 1 \leq i < j \leq r \) such that \( \ell_{i+1} = \ell_{i+2} = \ldots = \ell_{j-1} = 2k \).

(i) If \( |V_{i,i-1}| < |V_{j+1,r}| \) then \( \ell_i \leq \ell_j \), and \( \ell_i = 0 \) or \( \ell_j = 2k \).

(ii) If \( |V_{i,i-1}| > |V_{j+1,r}| \) then \( \ell_i \geq \ell_j \), and \( \ell_i = 2k \) or \( \ell_j = 0 \).

(iii) If \( |V_{i,i-1}| = |V_{j+1,r}| \) then \( \ell_i \in \{0, 2k\} \) or \( \ell_j \in \{0, 2k\} \).

To prove (i) suppose to the contrary that \( \ell_i > \ell_j \). Then \( T_{i,j} \) is associated with the sequence \( \ell_i, 2k, 2k, \ldots, 2k, \ell_j \). Now simple but straightforward calculations yield that \( d(w_{i,1}, T_{i,j}) > d(w_{j,2}, T_{i,j}) \), which contradicts Claim 7(i). Hence \( \ell_i \leq \ell_j \).

Now suppose that \( \ell_i > 0 \) and \( \ell_j < 2k \). Consider the graph \( G_1 \) obtained from \( G \) by replacing \( T_i \) by \( Y_{k,\ell_i}(v_i, w_{i,1}, w_{i,2}) \) and \( T_j \) by \( Y_{k,\ell_j}(v_j, w_{j,1}, w_{j,2}) \). So the sequence associated with \( G_1 \) is \( \ell_1, \ell_2, \ldots, \ell_{i-1}, \ell_i - 1, \ell_{i+1}, \ldots, \ell_{j-1}, \ell_j, \ell_{j+1}, \ldots, \ell_r \). It is easy to see that \( G_1 \) has order \( n \). We now compare \( d(G) \) and \( d(G_1) \). As above we let \( A = V_{i,i-1}, B = V_{i,j}, C = V_{j+1,r} \) and \( B' := B - (V_i \cup V_j) \cup (V(T_i') \cup V(T_j')) \). Let \( H \) be the subgraph of \( G \) induced by \( B \), and let \( H_1 \) be the subgraph of \( G_1 \) induced by \( B' \). Since \( d_{H_1}(w_{i,1}, w_{j,2}) = d_H(w_{i,1}, w_{j,2}) \), the distance between any two vertices \( x, y \in A \cup C \) in \( G \) equals their distance in \( G_1 \). Hence
\[
d(G_1) - d(G) = \sum_{x \in A} \sum_{y \in B'} d_{G_1}(x, y) - \sum_{x \in A} \sum_{y \in B} d_{G}(x, y) + \sum_{x \in C} \sum_{y \in B'} d_{G_1}(x, y)
\]
\[
- \sum_{x \in C} \sum_{y \in B} d_{G}(x, y) + d(H_1) - d(H).
\] (7)

The four double sums on the right hand side above can be expressed in terms of \(|A|\),
\[|B| = |B'|, |C|, \text{ and the distance of the contact vertices } w_{i,1} \text{ and } w_{j,2} \text{ in both, } H \text{ and } H_1.\]

\[
\sum_{x \in A} \sum_{y \in B'} d_{G_1}(x, y) = \sum_{x \in A} \sum_{y \in B'} \left(d_{G_1}(x, w_{i-1,2}) + 1 + d_{G_1}(w_{i,1}, y)\right)
= |B'|d(w_{i-1,2}, T_{1,i-1}) + |A||B'| + |A|d(w_{i,1}, H_1).
\]

Similarly we have

\[
\sum_{x \in A} \sum_{y \in B} d_G(x, y) = |B|d(w_{i-1,2}, T_{1,i-1}) + |A||B| + |A|d(w_{i,1}, H),
\]

\[
\sum_{x \in C} \sum_{y \in B'} d_{G_1}(x, y) = |B'|d(w_{j+1,1}, T_{j+1,r}) + |C||B'| + |C|d(w_{j,2}, H_1),
\]

\[
\sum_{x \in C} \sum_{y \in B} d_G(x, y) = |B|d(w_{j+1,1}, T_{j+1,r}) + |C||B| + |C|d(w_{j,2}, H).
\]

Now \(H\) is associated with the sequence \(\ell_i, \ell_{i+1}, \ldots, \ell_j\), and \(H_1\) with \(\ell_i - 1, \ell_{i+1}, \ldots, \ell_{j-1}, \ell_j + 1\). Long but straightforward calculations show that

\[
d(w_{j,2}, H_1) - d(w_{j,2}, H) = \frac{1}{2}(\ell_j - \ell_i) + 1 + (j - i)(2k + 1),
\]

\[
d(w_{i,1}, H_1) - d(w_{i,1}, H) = \frac{1}{2}(\ell_j - \ell_i) + 1 - (j - i)(2k + 1),
\]

\[
d(H_1) - d(H) = \left(\frac{\ell_i + \ell_j}{2} + k + 1 + (j - i - 1)(2k + 1)\right)(\ell_j - \ell_i + 2) > 0.
\]

Substituting these terms into (7), and using the fact that \(|B| = |B'|\), and that distances between vertices in \(A \cup C\) in \(G_1\) are the same as in \(G\), we obtain, after simplifications,

\[
d(G_1) - d(G) = |A|\left(\frac{d(w_{i,1}, H_1) - d(w_{i,1}, H)}{2}\right)
+ |C|\left(\frac{d(w_{j,2}, H_1) - d(w_{j,2}, H)}{2}\right) + d(H_1) - d(H)
= (j - i)(2k + 1)(|C| - |A|) + \frac{\ell_j - \ell_i + 2}{2}(|A| + |C|) + d(H_1) - d(H) > 0
\]

since \(\ell_j \geq \ell_i\) and \(d(H_1) - d(H) > 0\). This contradicts the maximality of \(d(G)\), hence (i) follows.

The proof of (ii) is analogous and thus omitted. To prove (iii) we may assume that, without loss of generality, \(\ell_i \leq \ell_j\). Then the proof is identical to the proof of (i). Hence Claim 8 follows.

If \(\ell_1 = \ldots = \ell_r = 0\), then each \(T_i\) has \(k + 1\) vertices and so \(n = r(k + 1)\). Hence \(p = q = 0\) (where \(p, q\) are as in Definition 3), and the sequence \(\ell_1, \ldots, \ell_r\) satisfies Definition 3 and so \(G = G_0(n, k, r)\). Hence assume that not all \(\ell_i\) equal 0.
CLM 9: There exist $a, b \in \{1, 2, \ldots, r\}$, $a \leq b$, such that

(i) $\ell_i = 0$ for $i \in \{1, 2, \ldots, a - 1\} \cup \{b, b + 1, \ldots, r\}$,

(ii) $\ell_i = 2k$ for all $i \in \{a + 1, a + 2, \ldots, b - 1\}$,

(iii) $\ell_a = 2q$, where $q$ is as defined in the statement of the lemma.

(iv) $r - b + 1 \leq a \leq r - b + 2$.

Clearly the term $|V_{i,i-1}| - |V_{i+2,r}|$ is strictly increasing in $i$. Define $s \in \{1, 2, \ldots, r\}$ to be the smallest integer such that $|V_{i,s-1}| - |V_{s+2,r}| \geq 0$. If $i < s$ then Claim 8(i), with $j = i + 1$, yields that $\ell_i \leq \ell_{i+1}$ and $\ell_i = 0$ or $\ell_{i+1} = 2k$. Applying this to $i = 1, 2, \ldots, s - 1$ yields that there exists $a \in \{1, 2, \ldots, s\}$ such that $\ell_i = 0$ for $1 \leq i < a$, and $\ell_i = 2k$ for $a < i \leq s$. If $i > s$ then $|V_{i,i-1}| > |V_{i+2,r}|$ and as above we obtain that there exists $b \in \{s + 1, s + 2, \ldots, r\}$ such that $\ell_i = 2k$ for $s + 1 \leq i < b$, and $\ell_i = 0$ for $b < i \leq r$. Hence $\ell_i = 0$ for $i = \{1, 2, \ldots, a - 1\} \cup \{b + 1, \ldots, r\}$, and $\ell_i = 2k$ for $i \in \{a + 1, a + 2, \ldots, b - 1\}$. Applying Claim 8 with $i = a$ and $j = b$ yields that $\ell_a \in \{0, 2k\}$ or $\ell_b \in \{0, 2k\}$. We can assume that $\ell_b \in \{0, 2k\}$ (if not we reverse the order of the $\ell_i$). Hence (i) and (ii) follow.

To prove (iii) we first note that we can assume that $\ell_a < 2k$ (if not we decrease $a$ by 1). We first determine the order of $T_a$. Since $T_i$ has order $k + 1$ (for $i \in \{1, 2, \ldots, a - 1\} \cup \{b, \ldots, r\}$ or $2k + 1$ for $i \in \{a + 1, a + 2, \ldots, b - 1\}$), we have

$$\sum_{i \neq a} |V(T_i)| = (r + a - b)(k + 1) + (b - 1 - a)(2k + 1) = r(k + 1) - (b - a)k - 1,$$

and thus

$$|V(T_a)| = n - \sum_{i \neq a} |V(T_i)| = n - r(k + 1) + (b - a)k + 1 = pk + q + (b - a)k + 1.$$

Hence $|V(T_a)| \equiv q + 1 \pmod{q}$. On the other hand $T_a = Y_{k,\ell_a}(v_a, w_{i,1}, w_{a,2})$ and so $T_a$ has order $k + 1 + \frac{1}{2}\ell_a$. Hence $\frac{1}{2}\ell_a \equiv q \pmod{k}$ and so, by $0 \leq \ell_a < 2k$, we have $\ell_a = 2q$.

We now show the first inequality of (iv). If $q = 0$ then we can assume, without loss of generality, that the initial sequence of $a$ zeroes is at least as long as the terminal sequence of $r - b + 1$ zeroes (otherwise we reverse the order of the $\ell_i$), hence we have $a \geq r - b + 1$ in this case. Hence we may assume that $0 < q < k$. We apply Claim 8 with $i = a$ and $j = b$. Let $H$ be the graph induced by $V_{i,j}$. Then $H$ has the associated sequence $\ell_a, 2k, 2k, \ldots, 2k, 0$ and so $d(w_{i,1}, H) > d(w_{j,2}, H)$. Now Claim 8 yields that $|V_{i,i-1}| \geq |V_{j+1,r}|$, i.e., $(a - 1)(k + 1) \geq (r - b)(k + 1)$, which implies $a \geq r - b + 1$, as desired.

We derive the second inequality by applying Claim 8 with $i = a$ and $j = b - 1$. Again let $H$ be the graph induced by $V_{i,j}$. Then $H$ has the associated sequence $\ell_a, 2k, 2k, \ldots, 2k$ and so $d(w_{i,1}, H) < d(w_{j,2}, H)$ by $\ell_a < 2k$. Claim 8 now yields that $|V_{i,i-1}| \leq |V_{j+1,r}|$, i.e.,
\[(a - 1)(k + 1) \leq (r - b + 1)(k + 1),\] which implies \(a \leq r - b + 2,\) as desired.

It is now easy to verify that Claim 9 implies that \(G = G_0(n, k, r),\) and so part (a) of the lemma follows.

(b) The proof of part (b) of the lemma is very similar to the proof of part (a), only some of the claims need to be modified. Claim 2 should be rephrased as follows “Each \(T_i\) has at most 2 contact vertices, and if a vertex \(w\) is a contact vertex of three or more trees, then at most one of the trees \(T_i\) containing \(w\) has more than one contact vertex.” This implies that the trees \(T_i\) can be numbered such that there exist \(s, t \in \{1, 2, \ldots, r\}\) such that \(T_1, \ldots, T_{s-1}\) are end trees with a common contact vertex which they share with \(T_s.\) Also \(T_{r}, T_{r-1}, \ldots, T_{t+1}\) are end trees with a common contact vertex which they share with \(T_t,\) and \(T_s, T_{s+1}, \ldots, T_t\) have two contact vertices each, so that \(T_i\) and \(T_{i+1}\) share a contact vertex. The remainder of the proof goes through with only minor modifications. \(\square\)

**Theorem 2** Let \(G\) be a graph of order \(n\) and \(k\)-packing number \(\beta_k > 1 + \frac{n - 1}{k + 1}.\)

(a) If \(k\) is even then
\[\mu(G) \leq \mu(G_0(n, k/2, r)),\]
with equality if and only if \(G = G_0(n, k/2, r).\)

(b) If \(k\) is odd then
\[\mu(G) \leq \mu(G_1(n, (k + 1)/2, r)),\]
with equality if and only if \(G = G_1(n, (k + 1)/2, r).\)

**Proof.** (a) Let \(G\) be a connected graph of order \(n\) and \(k\)-packing number \(\beta_k(G) =: r.\)

Then \(G\) contains a \(k\)-packing \(S = \{v_1, v_2, \ldots, v_r\} \subseteq V(G).\) Hence the sets \(N^{k/2}(v_i)\) are disjoint. For \(i = 1, 2, \ldots, r\) let \(T_i'\) be a spanning tree of the graph induced by \(\{v_i\} \cup N^{k/2}(v_i).\)

The forest \(\bigcup_{i=1}^r T_i'\) can be extended to a spanning forest \(\bigcup_{i=1}^r T_i =: F\) of \(G\) for which \(T_i' \leq T_i\) for \(i = 1, 2, \ldots, r.\) From \(F\) we obtain a spanning tree \(H\) of \(G\) by adding \(r - 1\) edges of the form \(w_iw_j,\) where \(w_i\) and \(w_j\) belong to different trees \(T_i\) and \(T_j.\) Since \(T_i\) and \(T_j\) contain \(N^{k/2}(v_i)\) and \(N^{k/2}(v_j),\) we have \(d_{T_i}(v_i, w_i) \geq k/2\) and \(d_{T_j}(v_j, w_j) \geq k/2,\) respectively. Hence \(H\) satisfies the hypothesis of Lemma 3 (with \(k/2\) instead of \(k\).) Rewriting the statement of Lemma 3 in terms of average distance we have
\[\mu(G) \leq \mu(H) \leq \mu(G_0(n, k/2, r)).\]

Assume equality holds. Then we have \(G = H\) since deletion of an edge strictly decreases the average distance of a graph. By Lemma 3 we have \(H = G_0(n, k/2, r).\) Hence \(G = G_0(n, k/2, r).\)
4 Average distance and $k$-domination number

As a corollary to the above results, we obtain upper bounds on the average distance of a graph of given order and $k$-domination number. We need the following result by Meir and Moon [13].

**Theorem 3** [13] Let $T$ be a tree and $k$ a positive integer. Then

$$\gamma_k(T) = \beta_{2k}(T).$$

Let $G$ be a graph of given order and $k$-domination number and maximum average distance. Then $G$ is a tree since every connected graph has a spanning tree with the same $k$-domination number. Hence $\gamma_k(G) = \beta_{2k}(G)$ and thus, by Theorems 1 and 2 we have the following result. We remark that the case $k = 1$ of part (b) is one of the two main results (Theorem 2) in [5].

**Corollary 4** Let $G$ be a graph of order $n \geq 3$ and $k$-domination number $\gamma_k$.

(a) If $\gamma_k \leq 1 + \left\lfloor \frac{n-1}{2k+1} \right\rfloor$ then

$$\mu(G) \leq \frac{2k+1}{2} \gamma_k + \frac{6k-1}{2} + \frac{1}{\sqrt{n}},$$

and this bound is, apart from an additive constant, best possible.

(b) If $\gamma_k > 1 + \left\lfloor \frac{n-1}{2k+1} \right\rfloor$ then

$$\mu(G) \leq \mu(G_0(n, 2k, r)),$$

and this bound is sharp.

Using different methods, Tian and Xu [15] recently proved a sharp upper bound on the average distance of graphs of given order and $k$-domination number. For $\gamma_k > 1 + \left\lfloor \frac{n-1}{2k+1} \right\rfloor$, their result is equivalent to Corollary 4 above, and for $\gamma_k \leq 1 + \left\lfloor \frac{n-1}{2k+1} \right\rfloor$ their result is slightly stronger. Using Theorem 3, it can be seen that, if one neglects additive constants in the bounds, that their results are equivalent to our bounds on the average distance of graphs of given order and $k$-packing number for even $k$. 


References


