# A NOTE ON A CONJECTURE OF HANSEN ET AL.

Ermelinda DeLaViña and Bill Waller
University of Houston-Downtown, Houston, TX, 77002

**Abstract**
Conjecture number 747 of Graffiti (circa 1992) states that the average distance of a simple, connected graph is not more than half the maximum order of an induced bipartite subgraph. Recently, P. Hansen et al. settled this conjecture by showing that the average distance is not more than half the maximum order of an induced forest. Moreover, they conjectured that the average distance is not more than half the maximum order of an induced linear forest. In this note, we give a partial resolution of this conjecture. Namely, we show that the average distance is less than half the maximum order of an induced linear forest, plus one-half.
**Keywords:** average distance, bipartite number, connected domination number, forest number, Graffiti, independence number, linear forest number.
**Mathematics Subject Classification:** 05C35.
**Abbreviated Title:** A Note on a Conjecture of Hansen

## Introduction and Key Definitions

Graffiti, a computer program that makes conjectures, was written by S. Fajtlowicz and dates from the mid-1980's. A numbered, annotated listing of several hundred of Graffiti's conjectures can be found in [4]. Graffiti has correctly conjectured a number of new bounds for several well studied graph invariants; bibliographical information on resulting papers can be found in [2].

We limit our discussion to graphs that are simple, connected and finite of order $n$. Although we often identify a graph $G$ with its set of vertices, in cases where we need to be explicit we write $V(G)$. We let $\alpha = \alpha(G)$ denote the *independence number* of $G$. If $u, v$ are vertices of $G$, then $\sigma_G(u, v)$ denotes the *distance* between $u$ and $v$ in $G$. This is the length of a shortest path in $G$ connecting $u$ and $v$. The *total distance from* $v$ in $G$, denoted by $w_G(v)$, is the sum of all distances from $v$ to the remaining vertices of $G$. The *average distance* of $G$, denoted by $\overline{D} = \overline{D}(G)$, is the average of all distances between pairs of distinct vertices of $G$. (In the degenerate case $n = 1$, we set $\overline{D}(G) = 0$.)

Unless stated otherwise, when we refer to a subgraph of a graph $G$, we mean an induced subgraph. We call the *bipartite number* of $G$ the maximum order of a bipartite subgraph. We denote this invariant by $b = b(G)$. One can make analogous definitions for the *forest number* and the *linear forest number* of $G$ (a linear forest is a forest where each connected component is a path). These invariants are denoted by $f = f(G)$ and $C = C(G)$, respectively. A few other more specialized definitions will be introduced in the next section. Standard graph theoretical terms not defined in this paper can be found in [6].

One of the earliest and best known of Graffiti's conjectures states that the average distance of a graph is not more than its independence number. This conjecture is listed as number 2 in [4].

Graffiti's conjecture number 2 was settled by F. Chung in [1], where the following theorem is proved.

**Theorem 1:** *Let $G$ be a graph. Then*

$$\overline{D} \leq \alpha,$$

*with equality holding if and only if $G$ is complete.*

In 1992, Graffiti formulated a generalization of its own conjecture number 2. This conjecture, listed as number 747 in [4], states that average distance of a graph $G$ is not more than half of $b(G)$. Recently, P. Hansen et al. in [5, Theorem 4.2] settled this conjecture by showing the following.

**Theorem 2:** *Let $G$ be a graph. Then*

$$\overline{D} \leq \frac{f}{2}.$$

Moreover, they close with the following stronger conjecture, numbered Conjecture 5.3 in their paper.

**Conjecture 1:** *Let $G$ be a graph. Then*

$$\overline{D} \leq \frac{C}{2}.$$

In this note, using results found in [3], we give a partial resolution of Conjecture 1. Namely, we show that:

**Theorem 3** *(Main Theorem)*: *Let $G$ be a graph. Then*

$$\overline{D} < \frac{C}{2} + \frac{1}{2}.$$

*Thus if $C$ is odd,*

$$\overline{D} < \left\lceil \frac{C}{2} \right\rceil.$$

## Other Definitions and Two Lemmas

A set of vertices $M$ of a graph $G$ is said to *dominate* $G$ provided each vertex of the graph is either in $M$ or adjacent to a vertex in $M$. The minimum order of a connected dominating set, called the *connected domination number* of $G$, is denoted by $\gamma_c = \gamma_c(G)$. A *trunk* for a graph $G$ is a sub-tree (not necessarily induced) that contains a dominating set of $G$. Hence, every spanning tree of $G$ is likewise a trunk for $G$, and every connected dominating set is the vertex set of some trunk. Therefore, if $G$ contains a trunk of order $t$, then $t \geq \gamma_c$. The following lemmas and their proofs are found in [3, Lemmas 5 and 7].

**Lemma 1:** *Let $G$ be a graph with a trunk of order $t \geq 1$. Then*

$$\overline{D}(G) < \frac{t+3}{2}.$$

**Lemma 2:** *Let $G$ be a graph with a trunk $M$ of order more than one, and let $m$ be a vertex with maximum total distance in $G$. Then if $m \in M$, there exists a graph $F$ with $V(F) = V(G)$ and a vertex $x \in M$, such that $\overline{D}(F) \geq \overline{D}(G)$, and moreover such that $M - \{x\}$ is a trunk for $F$.*

One more piece of terminology is needed. Let $S$ be any subset of vertices of a graph $G$. Then the *open neighborhood* of $S$, denoted by $N(S)$, is the set of neighbors of all vertices in $S$, less $S$ itself.

## Main Results

Note that the following result is a modification of Theorem 4 found in [3] .

**Theorem 4:** *Let $G$ be a graph. Then*

$$\gamma_c \leq C - 1.$$

*Proof.* Choose an arbitrary vertex $x_0$ of $G$ and call it path $L_0$. If $G$ is not trivial, then we can choose a vertex $y$ in the open neighborhood $N(x_0)$ and append it to $L_0$. Next we choose a vertex $z$ in the open neighborhood of $L_0$ that is adjacent to exactly one endpoint of $L_0$, and no interior vertices of $L_0$. We then append $z$ to $L_0$, and we repeat this process until we can no longer choose such a vertex $z$.

Next choose a vertex $x_1$ outside of $L_0$ and its open neighborhood. Since $G$ is connected, we can assume there exists a vertex $c_0$ in $N(L_0)$ such that $c_0$ is adjacent to $x_1$. If no such vertex $x_1$ exists, then we quit. Otherwise, add $x_1$ to a path $L_1$ and continue as before. That is, choose a vertex $z$ in the open neighborhood of $L_1$ that is adjacent to exactly one endpoint of $L_1$, no interior vertices of $L_1$, and no vertices of $L_0$. We append $z$ to $L_1$, and we repeat this process until we can no longer choose such a vertex $z$.

When we reach stage $j$, we choose a vertex $x_j$ outside of $L_0 \cup L_1 \cup \ldots \cup L_{j-1}$ and its open neighborhood. Since $G$ is connected, we can assume there exists a vertex $c_{j-1}$ in $N(L_0 \cup L_1 \cup \ldots \cup L_{j-1})$ such that $c_{j-1}$ is adjacent to $x_j$. If no such vertex $x_j$ exists, then we quit. Otherwise, add $x_j$ to a path $L_j$ and continue as before. That is, choose a vertex $z$ in the open neighborhood of $L_j$ that is adjacent to exactly one endpoint of $L_j$, no interior vertices of $L_j$, and no vertices of $L_0 \cup L_1 \cup \ldots \cup L_{j-1}$. We append $z$ to $L_j$, and we repeat this process until we can no longer choose such a vertex $z$.

Once the algorithm terminates (assume after stage $k$), note that $L = L_0 \cup L_1 \cup \ldots \cup L_k$ induces a linear forest. Let $r_j$ be an endpoint of $L_j$ other than $x_j$. If $x_j$ is the only vertex of $L_j$, then put $r_j = x_j$. See Figure 1. Suppose $v$ is a vertex outside of the induced linear forest. Let $f(v)$ be the minimum integer such that $v$ is adjacent to some vertex of $L_{f(v)}$. Next we prove the following claim.

*Claim:* Let $v$ be a vertex outside of $L$. Then $v$ is adjacent to either both endpoints of $L_{f(v)}$ or an interior vertex in $L_{f(v)}$.

*Proof of claim.* If $f(v)$ is undefined, this implies the algorithm terminated prematurely. Hence we can assume $f(v)$ exists and $v$ is adjacent to a vertex in $L_{f(v)}$. By way of contradiction, suppose $v$ is adjacent to exactly one endpoint of $L_{f(v)}$ and no interior vertices of $L_{f(v)}$. But since $v$ is not adjacent to any vertex of $L_0 \cup L_1 \cup \ldots \cup L_{f(v)-1}$, then the algorithm would have selected $v$ for inclusion in $L_{f(v)}$, meaning that $v$ is contained in $L$, a contradiction. ∎

For each vertex $v$ outside of $L$, let $a_v$ denote the neighbor of $v$ in $L_{f(v)}$ other than $r_{f(v)}$. The prior claim guarantees that $a_v$ exists. We are now in a position to complete the proof. We will construct a spanning tree $T'$ for a dominating set $M'$of $G$ with order at most $C - 1$. Thus $T'$ is the required trunk and we are finished. First, though, we construct a spanning tree $T$ for a somewhat larger dominating set $M$. The vertices of $M$ are $L \cup \{c_0, c_1, \ldots, c_{k-1}\}$. (Note: The $c_j$'s may not be unique.) The edges of $T$ are the edges of each path $L_j$ along with each edge $\{c_j, x_{j+1}\}$ and $\{c_j, a_{c_j}\}$. Since $f(c_j) \leq j$ and $c_j$ is adjacent to $x_{j+1}$ for each $j$, this implies there exists a path in $T$ from each vertex of $M$ to $x_0$. Thus $M$ spans a connected subgraph. Moreover, the claim implies that $M$ dominates $G$, so $T$ is a trunk. We now construct $M'$ and $T'$ by deleting each $r_j$ from $M$ and $T$ along with any incident edges in $T$. Recall $r_j \neq a_v$ for any vertex $v$ outside of $L$. Also, either $r_j$ is adjacent to some vertex of $L_j$ or $r_j$ is adjacent to $c_j$. Hence $M'$ continues to dominate $G$. We want to show $T'$ is a spanning tree for $M'$. Choose a vertex $v$ in $M'$. Because $r_j$ is an endpoint of $L_j$, then the path in $T$ from $v$ to $x_0$ remains intact in $T'$, unless $r_p = x_p$ for some integer $p$ and the path from $v$ to $x_0$ in $T$ contains the edges $\{c_{p-1}, x_p\}$ and $\{c_q, x_p\}$, for some integer $q > p$. Therefore, $f(c_q) = p$ and $a_{c_q} = x_p = r_p$, a contradiction to our choice of $a_{c_q}$.

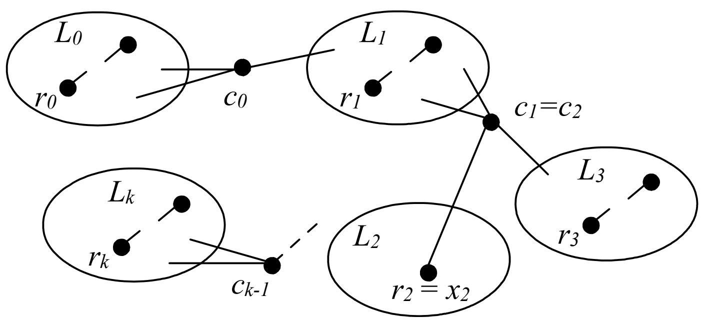


FIGURE 1: Algorithm diagram.

We now know that $T'$ is a trunk. But

$$\begin{aligned}|M'| &= |L \cup \{c_0, c_1, \ldots, c_{k-1}\} - \{r_0, r_1, \ldots, r_k\}| \\ &= |L_0 \cup L_1 \cup \ldots \cup L_k \cup \{c_0, c_1, \ldots, c_{k-1}\} - \{r_0, r_1, \ldots, r_k\}| \\ &= |L_0 \cup L_1 \cup \ldots \cup L_k| + |\{c_0, c_1, \ldots, c_{k-1}\}| - |\{r_0, r_1, \ldots, r_k\}| \\ &\leq C + k - (k+1) \\ &= C - 1.\end{aligned}$$

■

**Theorem 3** *(Main Theorem)*: *Let G be a graph. Then*

$$\overline{D} < \frac{C}{2} + \frac{1}{2}.$$

*Thus if C is odd,*

$$\overline{D} < \left\lceil \frac{C}{2} \right\rceil.$$

*Proof.* The algorithm described in the proof of Theorem 4 starts with an arbitrary vertex $x_0$, and if $G$ is not trivial, then $x_0$ is an element of the final trunk $T'$ of order at most $C-1$ constructed in the proof. Hence, we can run the algorithm choosing $x_0$ as a vertex of maximum total distance. Then by the Lemmas 1 and 2,

$$\overline{D}(G) \leq \overline{D}(F) < \frac{\gamma_c(F)+3}{2} \leq \frac{C-2+3}{2} = \frac{C}{2} + \frac{1}{2}. \qquad \blacksquare$$

## Bibliography


[1] F. Chung, *The average distance is not more than the independence number*, J. Graph Theory, 12(1988), p. 229-235.

[2] E. DeLaViña, Web site of bibliographical information on conjectures of Graffiti and Graffiti.pc, Web address: http://cms.uhd.edu/faculty/delavinae/wowref.html

[3] E. DeLaViña and B. Waller, *Spanning trees with many leaves and average distance*, Electronic J. of Combinatorics, 15(2008).

[4] S. Fajtlowicz, "Written on the Wall" (manuscript), Web address: http://math.uh.edu/~siemion

[5] P. Hansen, A. Hertz, R. Kilani, O. Marcotte and D. Schindl, *Average distance and maximum induced forest*, J. Graph Theory, 1(2009), p. 31-54.

[6] D. B. West, "Introduction to Graph Theory (2nd. ed.)," Prentice-Hall, NJ, 2001.