Spanning Trees with Many Leaves and Average Distance

Ermelinda DeLaViña and Bill Waller

Department of Computer and Mathematical Sciences
University of Houston-Downtown, Houston, TX, 77002
delavinae@uhd.edu   wallerw@uhd.edu

Submitted: Sep 18, 2006; Accepted: Jan 30, 2008; Published: Feb 11, 2008
Mathematics Subject Classification: 05C35

Abstract

In this paper we prove several new lower bounds on the maximum number of leaves of a spanning tree of a graph related to its order, independence number, local independence number, and the maximum order of a bipartite subgraph. These new lower bounds were conjectured by the program Graffiti.pc, a variant of the program Graffiti. We use two of these results to give two partial resolutions of conjecture 747 of Graffiti (circa 1992), which states that the average distance of a graph is not more than half the maximum order of an induced bipartite subgraph. If correct, this conjecture would generalize conjecture number 2 of Graffiti, which states that the average distance is not more than the independence number. Conjecture number 2 was first proved by F. Chung. In particular, we show that the average distance is less than half the maximum order of a bipartite subgraph, plus one-half; we also show that if the local independence number is at least five, then the average distance is less than half the maximum order of a bipartite subgraph. In conclusion, we give some open problems related to average distance or the maximum number of leaves of a spanning tree.

Introduction and Key Definitions

Graffiti, a computer program that makes conjectures, was written by S. Fajtlowicz and dates from the mid-1980's. Graffiti.pc, a program that makes graph theoretical conjectures utilizing conjecture making strategies similar to those found in Graffiti, was written by E. DeLaViña. The operation of Graffiti.pc and its similarities to Graffiti are described in [10] and [11]; its conjectures can be found in [13]. A numbered, annotated listing of several hundred of Graffiti's conjectures can be found in [19]. Both Graffiti and Graffiti.pc have correctly conjectured a number of new bounds for several well studied graph invariants; bibliographical information on resulting papers can be found in [12].

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 *Wiener index* or *total distance* of G, denoted by $W = W(G)$, is the sum of all distances between unordered pairs of distinct vertices of G [16]. Then the *average distance* of G, denoted by $\overline{D} = \overline{D}(G)$, is $2W(G)/[n(n-1)]$. Put another way, $\overline{D}(G)$ is the average distance between pairs of distinct vertices of G. (In the degenerate case $n = 1$, we set $W(G) = \overline{D}(G) = 0$.) Unless stated otherwise, when we refer to a subgraph of a graph G, we mean an induced subgraph.

Theorem 1 shown here is the first published result [20] concerning one of the earliest and best known of Graffiti's conjectures, which states that the average distance of a graph is not more than its independence number. This conjecture is listed as number 2 in [19].

Theorem 1 ([20]). *Let G be a graph. Then*

$$\overline{D} < \alpha + 1.$$

Graffiti's conjecture number 2 was then completely settled by F. Chung in [5], where the following theorem is proved.

Theorem 2 ([5]). *Let G be a graph. Then*

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

with equality holding if and only if G is complete.

In his Ph.D. dissertation [28], the second author generalized Theorem 2 somewhat by characterizing those graphs with order n and independence number α that have maximum average distance, for all possible values of n and α. A different, much shorter proof of this result was later discovered independently by P. Dankelmann [6]. In 1992, Graffiti formulated a new generalization of its own conjecture number 2. This conjecture, stated here as Conjecture 1, is listed as number 747 in [19]. For a graph G, we call the *bipartite number* of G the maximum order of an (induced) bipartite subgraph. We denote this invariant by $b = b(G)$. (There are many bounds for the maximum number of edges in a bipartite subgraph; for such results, see [1], [3] and [22]. Some results on the bipartite number as we define it can be found in [15].)

Conjecture 1 (Graffiti 747). *Let G be a graph. Then*

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

This conjecture has been one of the most circulated of Graffiti's open conjectures (see [29]). Fajtlowicz was interested in this conjecture in the hope that its proof might result in a more elegant proof of Theorem 2 (the current proofs are rather unwieldy). Note that the following Conjecture 2, which is slightly weaker than Conjecture 1, also generalizes conjecture number 2 of Graffiti. The main results of this paper are some partial resolutions of Conjecture 1 and a near resolution of Conjecture 2.

Conjecture 2. *Let G be a graph. Then*

$$\overline{D} \leq \left\lceil \frac{b}{2} \right\rceil.$$

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 $\gamma_c = \gamma_c(G)$. The maximum number of leaves contained on a spanning tree of G, called the *leaf number*, is denoted by $L = L(G)$. The problem of finding a spanning tree with a prescribed number of leaves has been shown to be NP-complete [24]; other computational aspects of L have also been considered (see [23]). Graffiti.pc recently conjectured various apparently new lower bounds for the leaf number L, two of which have implications for Conjecture 1. Lower bounds on L have received a lot of attention in the literature, partly because they imply upper bounds on the connected domination number γ_c, which has also received a lot of attention (note that $L = n - \gamma_c$). See [4], [26] for some references. The domination number and the k-distance domination number have moreover been related to the average distance of graphs in the recent papers [7], [8] and [9].

Theorem 3 (Graffiti.pc 177). *Let G be a graph. Then*

$$L \geq 2\alpha - b + 1.$$

Theorem 3 is actually weaker than the conjecture made by Graffiti.pc (number 177 in [13]), which replaces the constant 1 with the second smallest degree in the ordered degree sequence (this is sometimes the minimum degree and sometimes the second smallest degree). Conjecture 177 remains open. The proof of Theorem 3 is not difficult, but we defer all proofs to a later section. The next theorem is the basis for our main results.

Theorem 4 (Graffiti.pc 173). *Let G be a graph. Then*

$$L \geq n - b + 1.$$

Odd paths show that Theorem 3 is sharp, while odd cycles show that Theorem 4 is sharp. The proof of Theorem 4 is really just a by-product of the greedy algorithm for building a maximal connected bipartite subgraph, carried a little further. Theorem 4 can be combined with the Lemma 5 stated below to give a partial resolution of Conjecture 1 (Theorem 8). The *local independence of a vertex v*, denoted by $\mu(v)$, is the independence number of the subgraph induced by its neighborhood. The *local independence number* of a graph G, denoted $\mu = \mu(G)$, is the maximum of local independence taken over all vertices of G.

Conjecture 3 (Graffiti.pc 174). *Let G be a graph. Then*

$$L \geq n - b + \mu - 1.$$

Theorem 5. *Let G be a graph. Then*

$$L \geq n - b + \left\lceil \frac{\mu}{2} \right\rceil.$$

Conjecture 3, which generalizes Theorem 4 for graphs that are not complete, remains open; however, the similar but weaker statement proven in Theorem 5 combined with Lemma 5 gives a another partial resolution of Conjecture 1 (see Theorem 10).

The following lower bounds for L are shown in [14].

Theorem 6 ([14]). *Let G be a graph. Then*

$$L \geq n - 2\alpha + 1.$$

Theorem 7 ([14]). *Let G be a graph. Then*

$$L \geq n - 2\alpha + \mu - 1.$$

Since $2\alpha \geq b$, Theorem 4 provides an improvement to Theorem 6, which was motivated as a conjecture of J. R. Griggs [14]. (We recently discovered Griggs' conjecture is a result of P. Duchet and H. Meyniel [17].) If Graffiti.pc's conjecture 174 (listed here as Conjecture 3) is correct, it would provide an improvement to Theorem 7.

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 $t \geq \gamma_c$.

Lemma 5. *Let G be a graph with a trunk of order $t \geq 1$. Then*

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

Theorem 8. *Let G be a graph. Then*

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

Upon considering the proof of Theorem 1, we can use an additional lemma (Lemma 7) to give an improvement on Theorem 8 and a near resolution of Conjecture 2 (Theorem 9). Let G be a graph with v a vertex of G. Then 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.

Lemma 7. *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.*

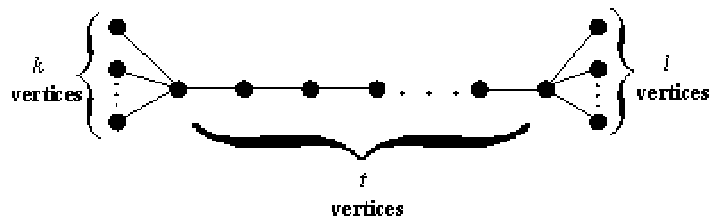

Figure 1: $R(k, t, l)$

Theorem 9 (Main Theorem). *Let G be a graph. Then*

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

Thus if b is odd,

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

Theorem 10. *Let G be a graph. If $\mu \geq 5$, then*

$$\overline{D} < \frac{b}{2}.$$

Let $R(k, t, l)$ denote the binary star on $k + t + l$ vertices, where the maximal interior path has order t and there are k leaves on one side of the binary star and l on the other. See Figure 1. Let $R(n, t)$ denote the binary star of order n where the maximal interior path has order t and the leaves are balanced as best possible on each side of the binary star.

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. Any other more specialized definitions will be introduced immediately prior to their first appearance. Standard graph theoretical terms not defined in this paper can be found in [30].

A Few Lemmas

Lemma 1 provides a useful method for comparing the total or average distance between two graphs with the same vertex sets.

Lemma 1. *Let G be a graph and $A \subset V(G)$. Let $B = V(G) - A$. Suppose G' is a graph such that $V(G') = V(G)$, and also such that:*

1) $\sum_{u \in A} \sum_{v \in A} \sigma_{G'}(u, v) \leq \sum_{u \in A} \sum_{v \in A} \sigma_G(u, v)$
2) $\sum_{u \in B} \sum_{v \in B} \sigma_{G'}(u, v) \geq \sum_{u \in B} \sum_{v \in B} \sigma_G(u, v)$
3) $\sum_{u \in A} w_{G'}(u) \geq \sum_{u \in A} w_G(u)$

Then $W(G') \geq W(G)$. Moreover, if any of these inequalities is strict, then $W(G') > W(G)$.

The proof of Lemma 2 involves only elementary algebra, counting, and limit arguments; we therefore omit it.

Lemma 2. *For integers $k \geq 0$ and $t \geq 1$,*

$$W(R(k,t,k)) = (t+3)k^2 + (t+2)(t-1)k + \frac{t(t+1)(t-1)}{6}, \text{ and}$$

$$W(R(k,t,k+1)) = (t+3)k^2 + k(t+1)^2 + \frac{t(t+1)(t+2)}{6}.$$

Moreover,

$$W(R(k,t,k)) \leq W(R(k,t,k+1)) \leq W(R(k+1,t,k+1)), \text{ and}$$

$$\lim_{k\to\infty} \overline{D}(R(k,t,k)) = \frac{t+3}{2}.$$

The following Lemma 3 can be immediately deduced from Lemma 2.

Lemma 3. *For integers $t \geq 1$ and $n \geq t$,*

$$\overline{D}(R(n,t)) < \frac{t+3}{2}.$$

The next lemma is essentially Theorem 2 from [20], although the proof given here is somewhat different. This lemma implies one of the most basic results about distance in graphs: Among all graphs of order n, the path on n vertices has the maximum total distance (and thus maximum average distance) [18].

Lemma 4. *Let G be a graph with a trunk of order $t \geq 1$. Then*

$$W(G) \leq W(R(n,t)),$$

with equality holding if and only if $G = R(n,t)$.

To deduce the corollary that among all graphs of order n, the path on n vertices has the maximum total distance, let T be a spanning tree of G. Then $|T| = |G| = n$. So by Lemma 4, $W(G) \leq W(R(n,n))$, with equality holding if and only if $G = R(n,n)$. But $R(n,n)$ is the path on n vertices.

Combining Lemmas 3 and 4 gives the following Lemma 5, which provides an upper bound on the average distance in graphs with a trunk of order t.

Lemma 5. *Let G be a graph with a trunk of order $t \geq 1$. Then*

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

The following lemma, which we state without proof, is an immediate consequence of results found in [18], Theorem 3.3].

Lemma 6. *Let T be a tree and let P be a path contained in T. Then if v is a vertex of P, there exists a leaf x of P such that $w_T(x) \geq w_T(v)$.*

Lemma 7. *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.*

Proofs

Lemma 1. *Let G be a graph and $A \subset V(G)$. Let $B = V(G) - A$. Suppose G' is a graph such that $V(G') = V(G)$, and also such that:*

1) $\sum_{u\in A} \sum_{v\in A} \sigma_{G'}(u,v) \leq \sum_{u\in A} \sum_{v\in A} \sigma_G(u,v)$
2) $\sum_{u\in B} \sum_{v\in B} \sigma_{G'}(u,v) \geq \sum_{u\in B} \sum_{v\in B} \sigma_G(u,v)$
3) $\sum_{u\in A} w_{G'}(u) \geq \sum_{u\in A} w_G(u)$

Then $W(G') \geq W(G)$. Moreover, if any of these inequalities is strict, then $W(G') > W(G)$.

Proof. It is enough to prove $2W(G') \geq 2W(G)$, from which the conclusion follows. Now,

$$\begin{aligned}
&2W(G') - 2W(G)\\
&= \sum_{u\in V} \sum_{v\in V} \sigma_{G'}(u,v) - \sum_{u\in V} \sum_{v\in V} \sigma_G(u,v)\\
&= \sum_{u\in V} \sum_{v\in V} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,]\\
&= 2 \sum_{u\in A} \sum_{v\in B} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,]\\
&\quad + \sum_{u\in A} \sum_{v\in A} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,] + \sum_{u\in B} \sum_{v\in B} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,].
\end{aligned}$$

By 2) the last term is non-negative, hence

$$\begin{aligned}
&2W(G') - 2W(G)\\
&\quad \geq 2 \sum_{u\in A} \sum_{v\in B} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,] + \sum_{u\in A} \sum_{v\in A} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,].
\end{aligned}$$

By 1) the second term is non-positive, hence

$$\begin{aligned}
&2W(G') - 2W(G)\\
&\geq 2 \sum_{u\in A} \sum_{v\in B} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,] + 2 \sum_{u\in A} \sum_{v\in A} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,]\\
&= 2 \sum_{u\in A} \sum_{v\in V} [\, \sigma_{G'}(u,v) - \sigma_G(u,v)\,]\\
&= 2 \sum_{u\in A} [\, w_{G'}(u) - w_G(u)\,].
\end{aligned}$$

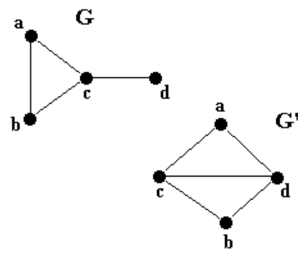

Figure 2: G and G'

By 3) the last term is non-negative, hence

$2W(G') - 2W(G) \geq 0.$

$\square$

Condition 1 may seem superfluous in light of Condition 3; nevertheless, it is sometimes necessary, as the two graphs G and G' in Figure 2 illustrate. Here we take $A = \{a, b\}$. It is easy to see $w_G(a) = w_G(b) = w_{G'}(a) = w_{G'}(b) = 4$, and $\sigma_G(c, d) = \sigma_{G'}(c, d) = 1$. But $W(G) = 8 > W(G') = 7$.

Lemma 4. *Let G be a graph with a trunk of order $t \geq 1$. Then*

$$W(G) \leq W(R(n, t)),$$

with equality holding if and only if $G = R(n, t)$.

Proof. Suppose G is chosen so that its total distance is maximum. It suffices to show $G = R(n, t)$. Let M be the given trunk for G of order t. We may assume G is a tree, since M can easily be extended to a spanning tree T of G with trunk M. We can dismiss the case $t = 1$ out of hand; for if $t = 1$, then G is a star, i.e. $G = R(n, 1)$.

Assume $t \geq 2$. Let L be a longest path in M, and suppose $|L| = \lambda$. Label the leaves of L, x and y. In addition, enumerate the non-trunk neighbors of x and y as $X = \{x_1, x_2, \ldots, x_p\}$ and $Y = \{y_1, y_2, \ldots, y_q\}$, respectively. Thus both $X, Y \subset G - M$. Let us assume $p \geq q$. Let z be the closest vertex to y on L other than y with degree greater than 2 in G.

Claim: Either no such vertex z exists, or $z = x$.

Proof of claim. By way of contradiction, suppose z exists and $z \neq x$. Moreover, suppose $\sigma_M(y, z) = \delta$. Since z is neither x nor y, $\delta \geq 1$ and $\lambda > \delta + 1$. Let $Z = \{z_1, z_2, \ldots, z_j\}$ denote the non-trunk neighbors of z, and let $F = \{f_1, f_2, \ldots, f_i\}$ denote the neighbors of z with respect to M not on L. Finally, let A denote the union of the components of $G - \{z\}$ which contain some vertex in $Z \cup F$. We derive a graph G' by first deleting from G all edges emanating from z which are sent to vertices in $Z \cup F$. In turn we add enough edges so that y is adjacent to each vertex in $Z \cup F$. This amounts to "transplanting" each of the components of A from z to y. See Figures 3 and 4.

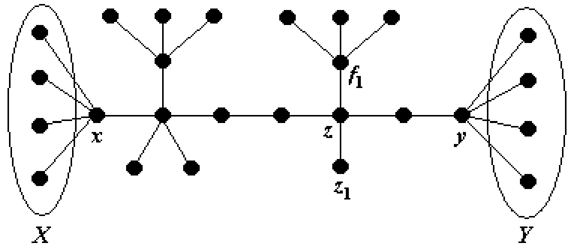

Figure 3: A hypothetical graph.

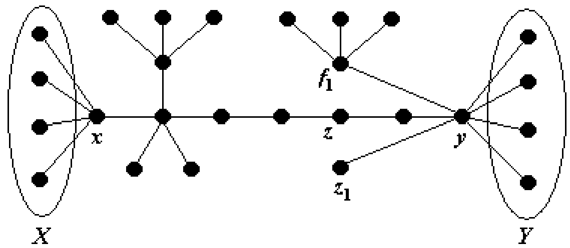

Figure 4: The graph G'.

We now apply Lemma 1 to G and G'. Clearly the first two conditions of the lemma are satisfied. By putting $C = G - \{X \cup Y \cup L\}$, it can be seen that for $a \in A$:

i) $\sum_{u \in X \cup Y} \sigma_{G'}(a, u) = \left(\sum_{u \in X \cup Y} \sigma_G(a, u)\right) + \delta(p - q)$,

ii) $\sum_{u \in L} \sigma_{G'}(a, u) = \left(\sum_{u \in L} \sigma_G(a, u)\right) + \delta[\lambda - (\delta - 1)]$, and

iii) for $u \in C$, $\sigma_{G'}(a, u) \geq \sigma_G(a, u)$.

Hence $w_{G'}(a) \geq w_G(a)$, which implies the third condition holds as well. Therefore $W(G') \geq W(G)$. It is easy to see that we can form a trunk of order t for G' by first deleting the edges $\{z, f_1\}, \{z, f_2\}, \ldots, \{z, f_i\}$ from M, and in turn adding the edges $\{y, f_1\}, \{y, f_2\}, \ldots, \{y, f_i\}$. But this contradicts our choice of G. □

Now the claim implies $G = R(p, t, q)$. If $p \leq q + 1$, then $G = R(n, t)$. So suppose $p > q + 1$. We derive a graph G' by fist deleting the edge $\{x_1, x\}$, and in turn adding the edge $\{x_1, y\}$. Applying Lemma 1 to G and G' with $A = \{x_1\}$, we have $W(G') > W(G)$. This follows from the supposition $p > q + 1$. But again we have a contradiction. Hence $G = R(n, t)$. □

Lemma 7. *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.*

Proof. Since M is a trunk for G, M can easily be extended to a spanning tree T for G with trunk M. Clearly $V(T) = V(G)$ and $\overline{D}(T) \geq \overline{D}(G)$, and also $w_T(m) \geq w_G(m)$. Let

L be the longest path in M containing m. Then by Lemma 6, there exists a leaf x of L such that $w_T(x) \geq w_T(m)$. If x is a leaf of T, then $M - \{x\}$ is a trunk for T, hence by putting $F = T$ we are done. Otherwise, let Z denote the set of neighbors of x with respect to T that are leaves of T. Let y denote the unique neighbor of x with respect to M. We derive a graph F by adding enough edges to T so that $Z \cup \{x, y\}$ induces a clique. We now apply Lemma 1 to G and F with $A = Z$ and $G' = F$. The first two conditions of the lemma clearly hold. Because for $z \in Z$,

$w_F(z) = w_F(x) = w_T(x) \geq w_T(m) \geq w_G(m) \geq w_G(z)$, the third condition holds also. Thus $\overline{D}(F) \geq \overline{D}(G)$. But $M - \{x\}$ is a trunk for F, so we are finished. □

Theorem 3 (Graffiti.pc). *Let G be a graph. Then*

$$L \geq 2\alpha - b + 1.$$

Proof. Let A be a maximum independent set, and let B be the complement of A. Suppose F is the subgraph induced by B. Moreover, suppose $C_1, C_2, \ldots, C_m$ are the connected components of F. If we color each of the vertices of A red, and color one vertex out of each of the components C_j green, it is easy to see that the colored vertices induce a bipartite subgraph. Thus $b \geq \alpha + m$. Since G is connected, then each vertex of A must be adjacent to a vertex of B. Thus B is a dominating set, but may not induce a connected subgraph, in particular when $m > 1$. However, again since G is connected, there exist vertices $a_1, a_2, \ldots, a_k \in A$ where $k < m$ such that $M = B \cup \{a_1, a_2, \ldots, a_k\}$ induces a connected subgraph. So M is contained in a trunk T' for G. We can now use T' to create a spanning tree T for G, where each of the vertices in $A - \{a_1, a_2, \ldots, a_k\}$ is a leaf of T. Therefore,

$$\begin{aligned} L &\geq |A - \{a_1, a_2, \ldots, a_k\}| \\ &= |A| - |\{a_1, a_2, \ldots, a_k\}| \\ &= \alpha - k \\ &\geq \alpha - (m-1) \\ &= 2\alpha - (\alpha + m) + 1 \\ &\geq 2\alpha - b + 1. \end{aligned}$$

□

Theorem 4 (Graffiti.pc 173). *Let G be a graph. Then*

$$L \geq n - b + 1.$$

Proof. We will show $\gamma_c \leq b - 1$, from which the result follows. Choose an arbitrary vertex x_0 of G and color it, say red. If G is not trivial, then we can choose a vertex y in the open neighborhood $N(x_0)$ and color it another color, say green. Next we choose a vertex z in the open neighborhood of the colored vertices that is not adjacent to both colors red and green (adjacent to either x_0 or y but not both, in the first instance). We color z the opposite color from its colored neighbors, and we repeat this process until we can no longer choose such a vertex z. Notice that the set of colored vertices induce a connected subgraph. We will refer to these colored vertices as the set B_0, and suppose T_0 is a spanning tree of the subgraph induced by B_0.

Next choose a vertex x_1 outside of B_0 and its open neighborhood. Since G is connected, we can assume there exists a vertex c_0 in $N(B_0)$ such that c_0 is adjacent to x_1. If no such vertex x_1 exists, then we quit. Otherwise, color x_1 red and continue as before. That is, choose a vertex z in the open neighborhood of either x_1 or the vertices colored after x_1 that is not adjacent to both colors red and green. We color z the opposite color from its colored neighbors, and we repeat this process until we can no longer choose such a vertex z. Notice again that x_1 and the set of vertices colored after x_1 induce a connected subgraph. We will refer to these colored vertices as the set B_1, and suppose T_1 is a spanning tree of the subgraph induced by B_1.

When we reach stage j, we choose a vertex x_j outside of $B_0 \cup B_1 \cup \ldots \cup B_{j-1}$ and its open neighborhood. Since G is connected, we can assume there exists a vertex c_{j-1} in $N(B_0 \cup B_1 \cup \ldots \cup B_{j-1})$ such that c_{j-1} is adjacent to x_j. If no such vertex x_j exists, then we quit. Otherwise, color x_j red and continue as before. That is, choose a vertex z in the open neighborhood of either x_j or the vertices colored after x_j that is not adjacent to both colors red and green. We color z the opposite color from its colored neighbors, and we repeat this process until we can no longer choose such a vertex z. Notice that x_j and the set of vertices colored after x_j induce a connected subgraph. We will refer to these colored vertices as the set B_j, and suppose T_j is a spanning tree of the subgraph induced by B_j.

Once the algorithm terminates (assume after stage k), note that $B_0 \cup B_1 \cup \ldots \cup B_k$ induces a bipartite subgraph. Let r_j be a leaf of T_j other than x_j. If x_j is the only vertex of T_j, then put $r_j = x_j$. See Figure 4. Suppose v is an uncolored vertex. Let $f(v)$ be the minimum integer such that v is adjacent to some vertex of $B_{f(v)}$. Next we prove the following claim.

Claim: Let v be an uncolored vertex. Then v is adjacent to both a red vertex and a green vertex in $B_{f(v)}$.

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 $B_{f(v)}$ colored red. Next suppose there does not exist a green vertex in $B_{f(v)}$ to which v is adjacent. But since v is not adjacent to any vertex in $B_0 \cup B_1 \cup \ldots \cup B_{f(v)-1}$, then the algorithm would have selected v for inclusion in $B_{f(v)}$, meaning that v is a colored vertex, a contradiction. $\square$

For each uncolored vertex v, let a_v denote the neighbor of v in $B_{f(v)}$ 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 $b-1$. 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

$B_0 \cup B_1 \cup \ldots \cup B_k \cup \{c_0, c_1, \ldots, c_{k-1}\}$. (Note: The c_j's may not be unique.) The edges of T are the edges of each tree T_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 uncolored vertex v. Also, either r_j is adjacent to some vertex of B_j or r_j is adjacent to c_j. Hence

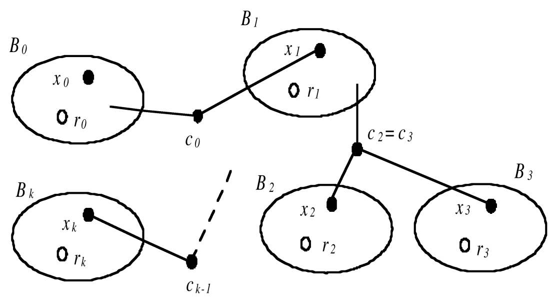

Figure 5:

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 a leaf of T_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}.

We now know that T' is a trunk. But

$$\begin{aligned}|M'| &= |B_0 \cup B_1 \cup \ldots \cup B_k \cup \{c_0, c_1, \ldots, c_{k-1}\} - \{r_0, r_1, \ldots, r_k\}| \\ &= |B_0 \cup B_1 \cup \ldots \cup B_k| + |\{c_0, c_1, \ldots, c_{k-1}\}| - |\{r_0, r_1, \ldots, r_k\}| \\ &\leq b + k - (k+1) \\ &= b - 1.\end{aligned}$$

□

Theorem 5. *Let G be a graph. Then*

$$L \geq n - b + \left\lceil \frac{\mu}{2} \right\rceil.$$

Proof. We will show $\gamma_c \leq b - \left\lceil \frac{\mu}{2} \right\rceil$, from which the result follows. 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' constructed in the proof. Hence, we can run the algorithm choosing x_0 as a vertex of maximum local independence. By our choice of x_0, the spanning tree T_0 of the first bipartite component B_0 has at least μ leaves, of which $\left\lceil \frac{\mu}{2} \right\rceil$ are in a common color class of B_0; let R_0 be a set of $\left\lceil \frac{\mu}{2} \right\rceil$ monochromatic leaves of B_0. Since B_0 is a maximal bipartite subgraph, every vertex not in B_0 but adjacent to a vertex of B_0 must be adjacent to vertices of both color classes. Thus,

$B_0 \cup B_1 \cup \ldots \cup B_k \cup \{c_0, c_1, \ldots, c_{k-1}\} - (R_0 \cup \{r_1, \ldots, r_k\})$ is a connected dominating set, and

$$\begin{aligned}|M'| &= |B_0 \cup B_1 \cup \ldots \cup B_k \cup \{c_0, c_1, \ldots, c_{k-1}\} - (R_0 \cup \{r_1, \ldots, r_k\})| \\ &= |B_0 \cup B_1 \cup \ldots \cup B_k| + |\{c_0, c_1, \ldots, c_{k-1}\}| - |R_0 \cup \{r_1, \ldots, r_k\}| \\ &\leq b + k - (k + \left\lceil \tfrac{\mu}{2} \right\rceil) \\ &= b - \left\lceil \tfrac{\mu}{2} \right\rceil.\end{aligned}$$

□

Theorem 8. *Let G be a graph. Then*

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

Proof. Combining $L \geq n - b + 1$ and $L = n - \gamma_c$ gives $\gamma_c \leq b - 1$. Then by Lemma 5,

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

□

Theorem 9. *Let G be a graph. Then*

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

Thus if b is odd,

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

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 $b - 1$ constructed in the proof. Hence, we can run the algorithm choosing x_0 as a vertex of maximum total distance. Then by the Lemmas 5 and 7,

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

□

Theorem 10. *Let G be a graph. If $\mu \geq 5$, then*

$$\overline{D} < \frac{b}{2}.$$

Proof. Combining $L \geq n - b + \left\lceil \frac{\mu}{2} \right\rceil$ and $L = n - \gamma_c$ gives $\gamma_c \leq b - \left\lceil \frac{\mu}{2} \right\rceil$. Then by Lemma 5, and since $\mu \geq 5$,

$$\overline{D}(G) < \frac{\gamma_c + 3}{2} \leq \frac{b - \left\lceil \frac{\mu}{2} \right\rceil + 3}{2} \leq \frac{b}{2}.$$

□

Some Open Problems

In addition to Conjecture 1 (Graffiti 747), Graffiti.pc 177, and Conjecture 2 (Graffiti.pc 174), all of which remain open, we are aware of numerous other open problems related to average distance or the leaf number that are worth mentioning. Many, but not all of these conjectures are due to either Graffiti or Graffiti.pc. The following conjecture (circa 1996) is another long standing open conjecture of Graffiti.

Conjecture 4 (Graffiti [13], conjecture 2). *Let G be a graph. Then*

$$L \geq 2(\textit{average of } \mu(v) - 1).$$

Recently, Graffiti.pc made the following conjecture related to L, which is reminiscent of Dirac's famous sufficient conditions for a graph to have a Hamiltonian cycle or path. Let $\delta = \delta(G)$ be the minimum degree of a graph G.

Conjecture 5 (Graffiti.pc 190). *Let G be a graph. If $\delta \geq \dfrac{L+1}{2}$, then G contains a Hamiltonian path.*

For a graph G, let B be the set of vertices of maximum eccentricity. Then the average distance from the boundary vertices of G, denoted by $\overline{D}(B, V)$, is the average of the nonzero distances between vertices of B and vertices of $V(G)$. The following conjecture is sometimes an improvement of Conjecture 1.

Conjecture 6 (Graffiti.pc 21). *Let G be a graph. Then*

$$b \geq 2 \cdot \overline{D}(B, V).$$

Two of the best sources for problems and conjectures related to distance in graphs are Graffiti (see, for instance, [2]) and the classical 1984 paper by J. Plesnik [27]. For example, a problem that Plesnik addresses but remains unresolved is the following:

Problem 1. *What is the maximum total or average distance among all graphs of order n with diameter d?*

This problem seems to be quite challenging. See [21] or [31] for recent related results. Conjecture 7 is an attempt to make the problem more tractable, in a special case. Let $C(n)$ denote the cycle of order n.

Conjecture 7. *Let G be a graph with diameter $d > 2$ and order $2d+1$. Then*

$$W(G) \leq W(C(2d+1)).$$

Analogous to the notion of a trunk, we call a *hoop* for a graph G a cycle subgraph (not necessarily induced) that contains a dominating set of G. We know by Lemma 5 an upper bound on the average distance of graphs with a trunk of order t, in terms of t. This lemma provides an upper bound on the average distance of graphs with a hoop of order t as well.

Problem 2. *What is a better upper bound on the average distance of graphs with a hoop of order t, other than that given by Lemma 5?*

To close, let's return to Theorems 3 and 4. We already observed that odd paths and cycles imply the lower bounds for L contained in these theorems are sharp. It would be an interesting (if perhaps formidable) undertaking to characterize the case of equality for these two theorems.

Nota bene: P. Hansen et al. have now shown in [25] that the average distance of a simple, connected graph is at most half the maximum order of an induced forest, using strategies different than our own. This settles Conjecture 1 (Graffiti 747).

References

[1] N. Alon, *Bipartite subgraphs*, Combinatorica, 16(1996), p. 301-311.

[2] R. Beezer, J. Riegsecker and B. Smith, *Using minimum degree to bound average distance*, Discrete Mathematics, 226(2001), p. 365-371.

[3] J. A. Bondy and S. C. Locke, *Largest bipartite subgraphs in triangle-free graphs with maximum degree three*, J. Graph Theory, 10(1986), p. 477-504.

[4] Y. Caro, D. West and R. Yuster, *Connected domination and spanning trees with many leaves*, SIAM J. Disc. Math., 13(2000), p. 202-211.

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

[6] P. Dankelmann, *Average distance and the independence number*, Discrete Applied Mathematics, 51(1994), p. 73-83.

[7] P. Dankelmann, *Average distance and the domination number*, Discrete Applied Mathematics, 80(1997), p. 21-35.

[8] P. Dankelmann, *Average distance and generalized packing in graphs*, preprint.

[9] P. Dankelmann and R. Entringer, *Average distance, minimum degree, and spanning trees*, J. Graph Theory, 33(2000), p. 1-13.

[10] E. DeLaViña, *Graffiti.pc*, Graph Theory Notes of New York, 42(2002), p. 26-30.

[11] E. DeLaViña, *Some history of the development of Graffiti*, DIMACS volume "Graphs and Discovery: Proceedings of the 2001 Working Group on Computer-Generated Conjectures from Graph Theoretic and Chemical Databases," 69(2005), p. 81-118.

[12] E. DeLaViña, Web site of bibliographical information on conjectures of Graffiti and Graffiti.pc,
Web address: http://cms.dt.uh.edu/faculty/delavinae/research/wowref.htm.

[13] E. DeLaViña, "Written on the Wall II,"
Web address: http://cms.dt.uh.edu/faculty/delavinae/research/wowII.

[14] E. DeLaViña, S. Fajtlowicz, and B. Waller, *On some conjectures of Griggs and Graffiti*, DIMACS volume "Graphs and Discovery: Proceedings of the 2001 Working Group on Computer-Generated Conjectures from Graph Theoretic and Chemical Databases ," 69(2005), p. 119-125.

[15] E. DeLaViña and B. Waller, *Some conjectures of Graffiti.pc on the maximum order of induced subgraphs*, Congressus Numerantium, 166 (2004), p. 11-32.

[16] A. Dobrynin, R. Entringer and I. Gutman, *Wiener index of trees: Theory and applications*, Acta Applicandae Mathematicae, 66(2001), p. 211-249.

[17] P. Duchet and H. Meyniel, *On Hadwiger's number and the stability number*, Ann. Discrete Math., 13(1982), p. 71-74.

[18] R. C. Entringer, D. E. Jackson and D. A. Snyder, *Distance in graphs*, Czech. Math. J., 26(1976), p. 283-296.

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

[20] S. Fajtlowicz and W. Waller, *On two conjectures of Graffiti*, Congressus Numerantium, 55(1986), p. 51-56.

[21] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Szekely and L. Volkmann, *Wiener index versus maximum degree in trees*, Discrete Applied Mathematics, 122(2002), p. 127-137.

[22] K. Fraughnaugh Jones, *Maximum bipartite subgraphs and independence*, Congressus Numerantium, 48(1985), p. 219-224.

[23] T. Fujie, *An exact algorithm for the maximum leaf spanning tree problem*, Computers & Operation Research, 30(2003), p. 1931-1944.

[24] M. R. Garey and D. S. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness," W. H. Freeman and Company, San Francisco, CA, 1979.

[25] P. Hansen, A. Hertz, R. Kilani, O. Marcotte and D. Schindl, *Average distance and maximum induced forest*, Les Cahiers du GERAD, 2007, preprint.

[26] T. W Haynes, S. T. Hedetniemi and P. J. Slater, "Fundamentals of Domination in Graphs," Marcel Dekker, Inc., NY, 1998.

[27] J. Plesnik, *On the sum of all distances in a graph or digraph*, J. Graph Theory, 8(1984), p. 1-24.

[28] W. Waller, "Average Distance in Graphs with Prescribed Order and Independence Number," Ph.D. Dissertation, University of Houston, Houston, TX (1989).

[29] D. B. West, Open problems column #23, SIAM Activity Group Newsletter in Discrete Mathematics, (1996).

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

[31] T. Zhou, J. Xu and J. Liu, *Extremal problem on diameter and average distance of graphs*, J. Univ. Sci. Technol. China, 34(2004), p. 410-413.