



Also available at http://amc.imfm.si ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.) ARS MATHEMATICA CONTEMPORANEA 2 (2009) 77–92

Vulnerability bounds on the number of spanning tree leaves

Gábor Salamon *

Department of Computer Science and Information Theory Budapest University of Technology and Economics 1117 Budapest, Magyar tudósok körútja 2., Hungary

Received 28 September 2008, accepted 3 February 2008, published online 6 April 2009

Abstract

Hamiltonicity and vulnerability of graphs are in a strong connection. A basic necessary condition states that a graph containing a 2-leaf spanning tree (that is, a Hamiltonian path) cannot be split into more than k + 1 components by deleting k of its vertices. In this paper we consider a more general approach and investigate the connection between the number of spanning tree leaves and two vulnerability parameters, namely scattering number sc(G)[10] and cut-asymmetry ca(G) [16]. We prove that any spanning tree of a graph G has at least sc(G) + 1 leaves. We also show that if $X \subset V$ is a maximum cardinality independent set of G = (V, E), such that the elements of X are all leaves of a particular spanning tree then $|X| = \operatorname{ca}(G) + 1 = |V| - \operatorname{cvc}(G)$, where $\operatorname{cvc}(G)$ is the size of a minimum connected vertex cover of G. As a consequence we obtain a new proof for the following results: any spanning tree with independent leaves provides a 2-approximation for both the MAXIMUM INTERNAL SPANNING TREE [16] and the MINIMUM CONNECTED VERTEX COVER [17] problems. We also consider the opposite point of view by fixing the number of leaves to qand looking for a q-leaf subtree of G that spans a maximum number of vertices. Bermond [2] proved that a 2-connected graph on n vertices always contains a path (a 2-leaf subtree) of length min $\{n, \delta_2\}$, where δ_2 is the minimum degree-sum of a 2-element independent set. We generalize this result to obtain a sufficient condition for the existence of a large q-leaf subtree.

Keywords: Spanning tree leaves, vulnerability, Hamiltonian path.

Math. Subj. Class.: 05C38, 05C85

^{*}Research is supported by Grant No. 67651 of the Hungarian National Science Fund (OTKA). *E-mail address:* gsala@cs.bme.hu (Gábor Salamon)

1 Introduction

Hamiltonicity and vulnerability of graphs are closely related [5]. The most basic necessary condition of hamiltonicity claims that a graph containing a Hamiltonian path cannot be split into more than k + 1 components by deleting k of its vertices. This gives the intention for examining the existence of Hamiltonian cycles and paths as a function of different vulnerability parameters, such as e.g. toughness or scattering number. For a survey on this topic the reader is referred to [1]. A closely related approach is to generalize hamiltonicity to explore bounds on the length of the longest cycle or path of a graph. As an example we mention here a generalization of Ore's theorem [13] given by Bermond [2]. The length of the longest cycle in a 2-connected graph G on n vertices is at least min $\{n, \delta_2(G)\}$. where $\delta_2(G)$ is the minimum degree of a 2-element independent set of G. As an immediate consequence, a 2-connected graph G always has a path (a 2-leaf subtree) of length $\min\{n, \delta_2(G)\}$. In Section 6 we generalize the result of Bermond proving that if G is an arbitrary graph on n vertices then there exists a q-leaf subtree of G spanning at least $\min\{n, \rho_{q,2} + q - 1\}$ vertices, where $\rho_{q,2}$ is the minimum degree-sum of the two highest degree vertices of a q-element independent set. This is also a generalization of a result of Broersma and Tuinstra [4] that gives a sufficient degree condition on the existence of a q-leaf spanning tree.

The HAMILTONIAN PATH problem is among the first ones turned to be NP-hard [7]. However, its practical importance needs efficient approximation algorithms and heuristics for finding different generalizations of Hamiltonian paths.

Several problems of this kind were investigated. Karger et al. [11] dealt with the LONGEST PATH problem and proved that there is no constant factor approximation algorithm unless P=NP. Furer and Raghavachari [6] considered the MINIMUM DEGREE SPANNING TREE problem and gave an algorithm that approximates the optimal solution within an additive constant of 1. Gargano et al. [9] looked at the MINIMUM BRANCHING SPANNING TREE problem, a degree-based generalization of Hamiltonian paths. They proved that it is NP-complete to decide whether a spanning tree with at most k branchings (vertices of degree at least 3) exists (for any fix k). However, the approximability properties of this approach are still undiscovered.

In this paper we consider the following optimization problems:

Problem 1: MINIMUM LEAF SPANNING TREE

Input: A connected graph G.

Goal: Find a spanning tree of G with a minimum number of leaves (1-degree vertices).

Problem 2: MAXIMUM INTERNAL SPANNING TREE

Input: A connected graph G.

Goal: Find a spanning tree of G with a maximum number of internal vertices (non-leaves).

Clearly, both of these problems are generalizations of the HAMILTONIAN PATH problem, as a Hamiltonian path is, on one hand, a 2-leaf spanning tree, and on the other hand, a spanning tree with n - 2 internal vertices. From the optimization point of view, these problems are equivalent as only the cost function is complemented by counting internal nodes instead of leaves.

However, the two problems have different approximability behavior. Using the result of Karger et al. [11], Lu and Ravi [12] proved that there is no constant factor approximation for the MINIMUM LEAF SPANNING TREE problem unless P=NP.

Complementing the cost function yields a better situation as the MAXIMUM INTERNAL SPANNING TREE problem can be constant-factor approximated. A 2-approximation can be done in linear time [16] and a 7/4-approximation in $\mathcal{O}(n^4)$ time [15]. Both algorithms are based on finding a spanning tree whose leaf-set has a large independent subset. This pushes us to examine the spanning trees with many independent leaves. Our approach uses vulnerability parameters to bound the number of leaves and the number of independent leaves in a spanning tree.

We mention here that Prieto and Sloper [14] investigated the MAXIMUM INTERNAL SPANNING TREE problem from a parameterized complexity point of view. They gave an $O(n^3)$ kernel showing FPT-membership of the problem.

Besides dealing with the MAXIMUM INTERNAL SPANNING TREE problem, we also consider the following problem:

Problem 3: MINIMUM CONNECTED VERTEX COVER

Input: A connected graph G.

Goal: Find a minimum size subset X of V(G) such that X spans a connected subgraph and meets each edge of G.

Savage gave a 2-approximation algorithm for this problem [17]. We prove that any spanning tree with independent leaves provides a 2-approximation for both the MAXIMUM INTERNAL SPANNING TREE and the MINIMUM CONNECTED VERTEX COVER problems. Although this does not improve directly the current approximation ratios, we believe that our new approach leads to better approximations in the future.

The rest of the paper is organized as follows. In Section 3 we show a connection between scattering number and the minimum number of leaves. In Section 4 we prove that cut-asymmetry exactly determines the size of a maximum independent subset of leaves. Several properties of cut-asymmetry are also presented, among others a sufficient condition for traceability. Namely, we prove that if $ca(G) \leq 1$ then G has a Hamiltonian path. We believe that this result has its own interest as cut-asymmetry and scattering number are very similar graph parameters and $sc(G) \leq 1$ is a well-known necessary condition of traceability. The algorithmic aspects of our results are presented in Section 5. We show how our approach can be used to directly obtain linear-time 2-approximation algorithms for both the MAXIMUM INTERNAL SPANNING TREE and the MINIMUM CONNECTED VERTEX COVER problems. In Section 6 we deal with large q-leaf subtrees of the input graph (given a fix parameter q). We provide a sufficient degree-based lower bound on the number of nodes that can be spanned by a q-leaf subtree.

2 Notation

By a graph G = (V, E) we mean an undirected simple connected graph on n = |V| vertices. We let $d_G(v)$ (or d(v)) denote the degree of vertex v in G, and for a set X of vertices $d_G(X) = \sum_{v \in X} d_G(v)$. We use $e_G(X, Y)$ to denote the number of edges between vertex disjoint sets X and Y of V(G). $N_G(v)$ is the set of neighbors of a vertex v. Let $V_i(G)$ ($V_{\geq i}(G)$) be the set of vertices having degree exactly i (at least i). Let G[X] denote the subgraph of G induced by X. A vertex-set X is called *independent* if G[X] has no edges. $\alpha(G)$ stands for the size of a maximum cardinality independent set of G. The number of components of a subgraph H is denoted by comp(H).

Let T be a spanning tree of G. The edges of T are called *tree-edges*, all other edges of G are *non-tree edges*. We denote by $P_T(x, y)$ the unique path in T between vertices x

and y. A vertex in $V_1(T)$ is called a *leaf*, a vertex in $V_2(T)$ is called a *forwarding vertex*, and a vertex in $V_{\geq 3}(T)$ is called a *branching*. Branchings and forwarding vertices are the *internal vertices* of T. For a leaf l of T the branching being closest to l is denoted by b(l). If (l, u) is a non-tree edge then $u^{\rightarrow l}$ is the neighbor of u along the path $P_T(u, l)$. The path $P_T(l, b(l))$ is called the *branch* of l, and $b^-(l)$ is the neighbor of b(l) in $P_T(l, b(l))$. The tree-edges that are not in branches are called *trunk-edges*. If the leaves of T form an independent set of G then T is called an *independence tree*.

G is called *traceable* if it has a Hamiltonian path. The minimum number of leaves of a spanning tree of G is denoted by ml(G). (Note that ml(G) = 2 if and only if G is traceable.)

For the sake of simplicity we use X + x, and X - x instead of $X \cup \{x\}$, and $X \setminus \{x\}$, respectively. Similarly, if X is a vertex-set of G = (V, E) then G - X denotes G[V - X].

3 The Scattering Number and the Minimum Number of Leaves

Scattering number is a vulnerability parameter which measures how much "structural damage" can be caused by removing some "important" vertices of the graph [10, 18]. It is defined as follows:

Definition 3.1. [10] The scattering number of a non-complete graph G = (V, E) is

$$\operatorname{sc}(G) = \max_{X \subset V, X \neq \emptyset} \left\{ \operatorname{comp}(G[V \setminus X]) - |X| : \operatorname{comp}(G[V \setminus X]) \ge 2 \right\}$$

By definition, $sc(K_n) = -\infty$.

This definition immediately implies that $sc(G) \leq 1$ if G is traceable. This can be interpreted as follows:

Proposition 3.2. If G has a 2-leaf spanning tree (i.e. a Hamiltonian path) then $ml(G) \ge sc(G) + 1$.

We show that every spanning tree of an arbitrary graph G has at least sc(G) + 1 leaves. Firstly, we prove this for trees.

Lemma 3.3. Let T be a tree with q leaves. Then $q \ge sc(T) + 1$.

Proof. To prove the upper bound let $X = \{x_1, x_2, \ldots, x_k\}$ be the vertex-set providing the maximum in Definition 3.1. If several such sets exist, let us choose one of minimum cardinality. Thus by definition $\operatorname{sc}(T) = \operatorname{comp}(T[V \setminus X]) - |X|$. Moreover, each vertex $x_j \in X$ is a branching of T, that is $d(x_j) \geq 3$ (for $j = 1, 2, \ldots, k$). Otherwise, if some $x_j \in X$ had degree less than 3 then for $X' = X - x_j$ we would have $\operatorname{comp}(T[V \setminus X']) \geq \operatorname{comp}(T[V \setminus X]) - 1$, and so we should have chosen X' instead of X.

Now let $X_0 = \emptyset$ and let $X_j = X_{j-1} + x_j$ (for j = 1, 2, ..., k). Furthermore let $T_j = T[V \setminus X_j]$ (for j = 0, 1, ..., k). That is, T_j is the forest obtained from T by removing the vertices of X_j . Then $\operatorname{comp}(T_j) \leq \operatorname{comp}(T_{j-1}) + d(x_j) - 1$, and hence

$$\operatorname{comp}(T_k) = \operatorname{comp}(T[V \setminus X]) \le 1 + \sum_{x_j \in X} d(x_j) - |X|.$$
(3.1)

On the other hand, let Y denote the set of branchings not in X, namely $Y = V_{\geq 3}(T) \setminus X$. Then counting the vertices of T according to their degree we obtain:

$$|V(T)| = q + |V_2(T)| + |X| + |Y|$$

= $q + |V_2(T)| + \sum_{i \ge 3} |V_i(T) \cap X| + \sum_{i \ge 3} |V_i(T) \cap Y|.$ (3.2)

Counting the total degree of vertices in T we have:

$$2|V(T)| - 2 = q + 2|V_2(T)| + \sum_{i \ge 3} i|V_i(T) \cap X| + \sum_{i \ge 3} i|V_i(T) \cap Y|, \qquad (3.3)$$

Equations (3.2) and (3.3) yield

$$\begin{split} \sum_{x_j \in X} d(x_j) &= \sum_{i \geq 3} i |V_i(T) \cap X| = \sum_{i \geq 3} (i-2) |V_i(T) \cap X| + 2|X| = \\ q - 2 - \sum_{i \geq 3} (i-2) |V_i(T) \cap Y| + 2|X| \leq q - 2 + 2|X|. \end{split}$$

Using (3.1) this implies

$$sc(T) = comp(T[V \setminus X]) - |X| \le q - 1$$

proving the upper bound on the scattering number.

The following theorem generalizes Lemma 3.3 for arbitrary graphs. Note that Theorem 3.4 is also a generalization of Proposition 3.2.

Theorem 3.4. If G is a connected graph then $ml(G) \ge sc(G) + 1$.

Proof. Let T be a minimum leaf spanning tree of G, that is $|V_1(T)| = ml(G)$, and let X be a set maximizing $comp(G[V \setminus X]) - |X|$. Then as $comp(G[V \setminus X]) \le comp(T[V \setminus X])$, by Lemma 3.3 we have:

$$\operatorname{sc}(G) = \operatorname{comp}(G[V \setminus X]) - |X| \le \operatorname{comp}(T[V \setminus X]) - |X| \le \operatorname{sc}(T) \le |V_1(T)| - 1 = \operatorname{ml}(G) - 1.$$

The scattering number of a tree T can also be used to form an upper bound on the number of leaves of T as the following theorem shows.

Theorem 3.5. Let T be a q-leaf tree on at least 3 vertices. Then $q \leq 2 \operatorname{sc}(T)$.

Proof. We construct a subset X of internal vertices of T such that each component of $T[V \setminus X]$ contains at most one leaf of T. Then we prove that $|X| \le q/2$. Combining these with the definition of scattering number directly yields the theorem.

Let us call a spanning tree that has at most one branching a *spider*. We build a sequence $\{T_1 = T, T_2, \ldots, T_k\}$ of trees such that T_k is a spider. If, for some *i*, T_i is not a spider we construct T_{i+1} as follows: let x_i be a branching of T_i with exactly one trunk-edge



Figure 1: Constructing T_{i+1} from T_i by removing the branching x_i

 (x_i, y_i) incident to it (such an x_i exists if T_i is not a spider). Let T'_i be the component of $T_i[V(T_i) - x_i]$ that contains y_i . If y_i is not a leaf of T'_i then $T_{i+1} = T'_i$. Otherwise, T_{i+1} is obtained by deleting the branch of y_i from T'_i (Figure 1). Observe that the leaves of T_{i+1} form a subset of leaves of T_i , and that a leaf l is in $V_1(T_i) \setminus V_1(T_{i+1})$ if and only if $b(l) = x_i$.

If T_i is a spider (thus i = k) then we define x_k to be the branching of the spider. (If T_k is a path then x_k is chosen to be any internal vertex of T_k .) Let $X = \{x_1, x_2, \ldots, x_k\}$. Observe that this construction ensures that each leaf of T is in a different component of $T[V \setminus X]$, thus $q \leq \operatorname{comp}(T[V \setminus X])$.

To see that $q \ge 2|X|$ let b_i denote the number of those branches of T_i that end in x_i . Then $b_i = d_{T_i}(x_i) - 1$ and $b_i \ge 2$ for $1 \le i \le k - 1$. This implies that

$$\sum_{i=1}^{k} b_i \ge 2k - 1.$$

As x_i cuts down exactly b_i leaves from T_i , and as T_k is a spider with $b_k + 1$ leaves we have $|V_1(T_{i+1})| = |V_1(T_i)| - b_i$, and $|V_1(T_k)| = b_k + 1$. Putting these together gives

$$q = \sum_{i=1}^{k} b_i + 1 \ge 2k = 2|X|.$$

Thus

$$\operatorname{sc}(T) \ge \operatorname{comp}(T[V \setminus X]) - |X| \ge q - \frac{q}{2} = \frac{q}{2}$$

which finishes the proof.

Notice that the bounds of Lemma 3.3 and Theorem 3.5 are sharp. On one hand, if T_1 is a *q*-leaf spider then $|V_1(T_1)| = q = sc(T_1) + 1$. On the other hand, let T_2 be a tree on vertices $V(T_2) = \{v_1, v_2, ..., v_k, u_1, u_2, ..., u_k, w_1, w_2, ..., w_k\}$ having edges $E(T_2) = \{(v_i, v_{i+1})\}_{i=1,2,...,k-1} \cup \{(v_i, u_i)\}_{i=1,2,...,k} \cup \{(v_i, w_i)\}_{i=1,2,...,k}$ (see Figure 2). Then $sc(T_2) = k = \frac{|V_1(T_2)|}{2}$.



Figure 2: A graph proving that the bound of Theorem 3.5 is sharp.

These examples show that a difference of factor 2 can arise between the number of leaves and the scattering number of trees. In Section 4 we present cut-asymmetry as a vulnerability parameter which exactly determines the number of leaves when applied on a tree.

Finally, we give two basic properties of the minimum leaf spanning trees. They are used throughout the next sections.

Lemma 3.6. If T is a minimum leaf spanning tree of G then T is either a Hamiltonian path or an independence tree.

Proof. Suppose for a contradiction that T is not a Hamiltonian path and there exist two leaves l_1, l_2 of T such that $(l_1, l_2) \in E(G)$. The spanning tree T' with edge-set $E(T') = E(T) + (l_1, l_2) - (b(l_1), b^-(l_1))$ has less leaves than T contradicting the minimality of T (Figure 3(a)).

Recall that if T is a tree then $x \rightarrow l_1$ denotes the successor of x along the path $P_T(x, l_1)$.

Lemma 3.7. Let G be a non-traceable graph, and T be a minimum leaf spanning tree of G. Let l_1 be an arbitrary leaf of T, and $(l_1, x) \in E(G) \setminus E(T)$ be a non-tree edge. Then $d_T(x^{-l_1}) = 2$ and if $l_2 \neq l_1$ is a leaf of T then $(x^{-l_1}, l_2) \notin E(G)$.

Proof. Let us consider the spanning tree T' with edge-set $E(T') = E(T) + (l_1, x) - (x^{-l_1}, x)$. Observe that x^{-l_1} is an internal vertex of T. If $d_T(x^{-l_1}) > 2$ then $V_1(T') = V_1(T) - l_1$ and so T' has less leaves than T, a contradiction (Figure 3(b)). If $d_T(x^{-l_1}) = 2$ then $V_1(T') = V_1(T) - l_1 + x^{-l_1}$, that is T and T' has the same number of leaves. However, the edge (x^{-l_1}, l_2) connects two leaves of T' and thus (by Lemma 3.6) neither T' nor T is a minimum leaf spanning tree (Figure 3(c)).

We have the following trivial corollary of Lemma 3.6:

Corollary 3.8. If G is a non-complete connected graph then $ml(G) \le \alpha(G)$.

4 The Cut-asymmetry and the Maximum Number of Independent Leaves

In the previous section we have seen that scattering number provides both lower and upper bounds on the number of leaves of a tree. In this section we use another vulnerability measure, namely cut-asymmetry. Recall that scattering number shows how much structural



Figure 3: Decreasing the number of leaves by local improvements

damage can be caused by removing some individual vertices from the graph. We define cutasymmetry in a similar way but we count the connected subgraphs (instead of individual vertices) to be removed. We prove that this model exactly characterizes the number of leaves in a tree and can be used to determine the maximum number of independent leaves of a spanning tree of an arbitrary graph. Besides, we point out a connection between cutasymmetry and the minimum size of a connected vertex cover.

Let us recall the definition of cut-asymmetry.

Definition 4.1. [16] The cut-asymmetry of a graph G = (V, E) is

$$\operatorname{ca}(G) = \max_{X \subset V, X \neq \emptyset} \left\{ \operatorname{comp}(G[V \setminus X]) - \operatorname{comp}(G[X]) \right\}.$$

This definition immediately implies that $ca(G) \ge max \{sc(G), 0\}$, and that $ca(G) \le n - 2$ (equality holds if and only if G is a star). We mention here some further properties of cut-asymmetry. The following theorem is a consequence of Theorems 4.11 and 4.15, however a direct proof is also given below.

Theorem 4.2. ca(G) = 0 if and only if G is either a complete graph or a cycle.

Proof. If G is complete or a cycle then its cut-asymmetry is trivially 0. To see the other direction let G be a non-complete graph such that ca(G) = 0. Let $Z = \{z_1, z_2, \ldots, z_k\}$ be an arbitrary independent set. As ca(G) = 0 the graph $G[V \setminus Z]$ has k components

 C_1, C_2, \ldots, C_k . Let us contract each component C_i to a single vertex c_i . By this transformation we obtain the connected bipartite graph G' with classes Z and $C = \{c_1, c_2, \ldots, c_k\}$.

Proposition 4.3. G' is an even cycle.

Proof. For every $1 \leq j \leq k$ we have $d_{G'}(z_j) = 2$. Otherwise if $d_{G'}(z_j) = 1$ for some j then $\operatorname{comp}(G[Z-z_j]) = k-1$ and $\operatorname{comp}(G[V \setminus (Z-z_j)]) = k$ would imply $\operatorname{ca}(G) \geq 1$. If $d_{G'}(z_j) \geq 3$ for some j then $\operatorname{comp}(G[Z-z_j]) = k-1$ and $\operatorname{comp}(G[V \setminus (Z-z_j)]) \leq k-2$ would imply $\operatorname{ca}(G) \geq 1$. A similar reasoning shows that $d_{G'}(z_j) = 2$ for each $1 \leq j \leq k$. As a result G' is a 2-regular connected bipartite graph, that is, an even cycle.

Proposition 4.4. If vertex v is in the component C_i (for some i) such that $|V(C_i)| \ge 2$ then v has at most one neighbor in Z (in graph G).

Proof. Suppose that $z_{j_1} \in Z$ and $z_{j_2} \in Z$ are both neighbors of such a v. Then

$$\operatorname{comp}(G[Z+v]) = k-1$$
 and $\operatorname{comp}(G[V \setminus (Z+v)]) \ge k$

implies $ca(G) \ge 1$, a contradiction.

Proposition 4.5. There is at most one edge in G between vertex $z_j \in Z$ and component C_i for all i and j.

Proof. Suppose that there exist vertices $u, v \in C_i$ and a vertex $z_j \in Z$ (for some *i* and *j*), such that both (u, z_j) and (v, z_j) are edges of *G*. Then by Proposition 4.4, $\operatorname{comp}(G[Z - z_j + u]) = k$ as *u* has no other neighbors in *Z*. We also have $\operatorname{comp}(G[V \setminus (Z - z_j + u])) = k - 1$ as z_j is neighboring to some other component $C_{i'}$. This implies $ca(G) \ge 1$, a contradiction.

As a result of Propositions 4.3 and 4.5 we obtain that exactly two edges of G leave each component C_i . For a fixed i let these edges be (z_{j_1}, u_i) and (z_{j_2}, v_i) $(u_i, v_i \in V(C_i))$. (Note that, by Proposition 4.4, $u_i = v_i$ if and only if C_i has a single vertex.) Suppose that C_i has a vertex x_i such that $C_i - x_i$ has a $u_i - v_i$ path. In this case $\operatorname{comp}(G[Z - z_{j_1} + x_i]) = k$ using Propositions 4.3 and 4.4, and $\operatorname{comp}(G[V \setminus (Z - z_{j_1} + x_i)]) = k - 1$ implying $\operatorname{ca}(G) \geq 1$, a contradiction. We conclude that component C_i is either a single vertex or a simple path connecting u_i and v_i . This, together with Proposition 4.3, proves the theorem.

Let us recall that $sc(G) \le 1$ is a necessary condition for the existence of a Hamiltonian path. The following theorem shows a sufficient condition for traceability by means of cut-asymmetry. Its proof is postponed here, as the theorem is a direct corollary of Theorems 4.11 and 4.15.

Theorem 4.6. If $ca(G) \le 1$ then G has a Hamiltonian path.

Unfortunately the traceability of G does not imply a low value of ca(G) as shown by the following theorem.

Theorem 4.7. If G = (V, E) is a traceable graph on n vertices then $ca(G) \leq \lfloor \frac{n-1}{2} \rfloor$. Moreover, if k is an arbitrary integer such that $0 \leq k \leq \lfloor \frac{n-1}{2} \rfloor$ then there exists a traceable graph G on n vertices for which ca(G) = k.

Proof. Property 3.2 implies that $sc(G) \leq 1$, that is, for any proper subset $X \subset V$ we have $comp(G[V \setminus X]) \leq |X|+1$. As $comp(G[V \setminus X]) \leq n-|X|$ we obtain $comp(G[V \setminus X]) \leq \lfloor \frac{n+1}{2} \rfloor$. Thus $ca(G) \leq \lfloor \frac{n-1}{2} \rfloor$.

To see the second part let us fix $k \leq \lfloor \frac{n-1}{2} \rfloor$. A graph on *n* vertices whose cutasymmetry is *k* is constructed as follows. If k = 0 then according to Theorem 4.2 the graph $G = K_n$ is a good example. So we suppose that $k \geq 1$. Let $P = x_1, x_2, \ldots, x_n$ be a path of length *n*. Now we obtain *G* by adding the edges of the complete graph on x_2, x_4, \ldots, x_{2k} to *P*. It is easy to check that $\operatorname{ca}(G) = k$.

The following result of Salamon and Wiener determines the cut-asymmetry of a tree.

Lemma 4.8. [16] If T is a tree then $|V_1(T)| = ca(T) + 1$.

Thus for any tree T we obtain

 $sc(T) + 1 \le |V_1(T)| = ca(T) + 1 \le 2sc(T).$

For a connected graph G we have proved so far

 $sc(G) + 1 \le \min \{ ml(G), ca(G) + 1 \}.$

In what follows we prove that $ml(G) \le ca(G) + 1$ holds for all graphs but the complete graph K_n and the cycle C_n . For this purpose we introduce a new parameter called leaf-independence.

Definition 4.9. Let G be any connected graph and T be a spanning tree of G. The *leaf-independence* of T in G (denoted by $li_G(T)$) is the cardinality of a maximum independent subset of $V_1(T)$. The *leaf independence* li(G) of the graph G is the maximum of $li_G(T)$ over all spanning trees of G.

It is important to point out that a spanning tree T can have more than $\lim_G(T)$ leaves. Indeed, $li_G(T) = |V_1(T)|$ if and only if T is an independence tree. Note that $\lim_G(T)$ is defined for all connected graphs, however, not all connected graphs have an independence tree. Böhme et al. [3] characterized the graphs that have no independence trees. (These graphs are the complete graph, the cycle and the complete bipartite graph $K_{n,n}$.) If an independence tree with α_t leaves exists then obviously $\lim_G O \ge \alpha_t$. The following conjecture would imply that leaf independence is a generalization of the concept of independence trees:

Conjecture 4.10. If a graph G has an independence tree then G has an independence tree with li(G) leaves.

The definition of leaf-independence directly yields $li(G) \leq \alpha(G)$ whenever G has at least 3 vertices. Also, the leaf-independence of a tree T (on at least 3 vertices) is equal to the number of its leaves. Thus, by Lemma 4.8, ca(T) = li(T) - 1. The following theorem states that this equality holds for any connected graph.

Theorem 4.11. Let G be a connected graph. Then ca(G) = li(G) - 1.

Proof. Let T be a spanning tree of G, such that $li_G(T) = li(G)$. Let Z be a maximum independent set of $V_1(T)$, and let $X = V \setminus Z$. Then on one hand

$$\operatorname{comp}(G[V \setminus X]) = |V \setminus X| = \operatorname{li}(G),$$

and on the other hand

$$1 \le \operatorname{comp}(G[X]) \le \operatorname{comp}(T[X]) = 1$$

Therefore

$$\operatorname{ca}(G) \ge \operatorname{comp}(G[V \setminus X]) - \operatorname{comp}(G[X]) = \operatorname{li}(G) - 1.$$

To prove the reverse direction let us choose X^* to be the set giving the maximum in Definition 4.1. If several such sets exist we take one of maximum cardinality. First we show that in this case $V \setminus X^*$ is an independent set of G. Assume that this is not the case, and there is an edge (u, v) in $G[V \setminus X^*]$. Then considering $X' = X^* + u$ we have $\operatorname{comp}(G[X']) \leq \operatorname{comp}(G[X^*])$, and $\operatorname{comp}(G[V \setminus X']) \geq \operatorname{comp}(G[V \setminus X^*])$. This forms a contradiction as $\operatorname{comp}(G[V \setminus X']) - \operatorname{comp}(G[X']) \geq \operatorname{comp}(G[V \setminus X^*]) - \operatorname{comp}(G[X^*])$, and $|X'| > |X^*|$.

In the follows we show that X^* spans a connected subgraph of G. Suppose for a contradiction that $G[X^*]$ has more than one components. Then, since $V \setminus X^*$ is an independent set, there must exist two components C_1 and C_2 of $G[X^*]$, and a vertex $u \in V \setminus X^*$ such that $G[V(C_1) \cup V(C_2) + u]$ is connected. Hence for the independent set $V \setminus X' = (V \setminus X^*) - u$ we have:

$$\operatorname{comp}(G[V \setminus X']) = |V \setminus X'| = |V \setminus X^*| - 1 = \operatorname{comp}(G[V \setminus X^*]) - 1,$$

and

$$\operatorname{comp}(G[X']) \le \operatorname{comp}(G[X^*]) - 1.$$

Thus $\operatorname{comp}(G[V \setminus X']) - \operatorname{comp}(G[X']) \ge \operatorname{comp}(G[V \setminus X^*]) - \operatorname{comp}(G[X^*])$, that is a contradiction as $|X'| > |X^*|$.

To finish the proof let T^* be a spanning tree of $G[X^*]$. We then connect each vertex of $V \setminus X^*$ to T^* . Thus we obtain a spanning tree of G in which all vertices of $V \setminus X^*$ are leaves. Thus

$$\operatorname{li}(G) \ge |V \setminus X^*| = \operatorname{comp}(G[V \setminus X^*]) - \operatorname{comp}(G[X^*]) + 1 = \operatorname{ca}(G) + 1,$$

using that $V \setminus X^*$ is independent and that X^* spans a connected subgraph of G.

In proof of Theorem 4.11 we have seen that there exists an independent set $Z^* = V \setminus X^*$ such that $G[V \setminus Z^*]$ is connected and that $\operatorname{comp}(G[Z^*]) - \operatorname{comp}(G[V \setminus Z^*]) = ca(G)$. This yields us an alternative definition of cut-asymmetry, that is

Corollary 4.12. Let Z be a maximum size independent set of V for which $G[V \setminus Z]$ is connected. Then ca(G) = |Z| - 1.

Recall that a connected vertex cover of G is a vertex-set that spans a connected subgraph and that meets all edges of G. Let us denote by cvc(G) the size of a minimum cardinality connected vertex cover of G. The set $Z^* = V \setminus X^*$ of the above proof is a maximum-size vertex-set that is independent and whose complement X^* spans a connected subgraph. This means that X^* is a minimum-size connected vertex cover of G, and proves the following:

Corollary 4.13. For any connected graph G on n vertices li(G) + cvc(G) = n.

Note that actually the above arguments also show that the complement of a minimum connected vertex cover is always a maximum size independent subset of leaves of a spanning tree.

A minimum connected vertex cover is NP-hard to compute [8]. Thus Corollary 4.13 and Theorem 4.11 imply that it is also NP-hard to compute both cut-asymmetry and leaf-independence.

In what follows we show that the leaf-independence provides an upper bound on the minimum number of leaves in almost all graphs. By Lemma 3.6 we obtain that if G has no Hamiltonian path then the leaves of any minimum leaf spanning tree of G are independent. In this case $ml(G) \leq li(G)$ is straightforward. If G is traceable then ml(G) = 2 and so $ml(G) \leq li(G)$ is true if and only if G has a spanning tree with at least two independent leaves. We use a result of Böhme et al. to show that this condition is satisfied whenever G is neither the complete graph K_n nor the cycle C_n .

Theorem 4.14. [3] If G is traceable and each Hamiltonian path of G is the part of a Hamiltonian cycle then G is either the complete graph K_n or the cycle C_n or the complete bipartite graph $K_{n,n}$.

This theorem immediately implies that $li(G) \ge 2$ if G is not isomorphic to K_n , C_n or $K_{n,n}$. It is easy to see that for these special graphs we have $li(K_n) = li(C_n) = 1$ and $li(K_{n,n}) = n - 1$. Thus we obtain the following theorem.

Theorem 4.15. Let G be any connected graph but the complete graph K_n and the cycle C_n . Then $ml(G) \leq li(G)$.

Summarizing the above results we have

Corollary 4.16. For any connected graph G but K_n and C_n :

$$\operatorname{sc}(G) + 1 \le \operatorname{ml}(G) \le \operatorname{li}(G) = \operatorname{ca}(G) + 1 \le \alpha(G)$$

In the next section we use these inequalities to prove approximation ratios on two NPhard optimization problems.

5 Algorithmic Aspects

In this section we consider two NP-hard optimization problems. The MAXIMUM INTER-NAL SPANNING TREE problem [16] is about finding a spanning tree of the input graph that maximizes the number of internal vertices. The MINIMUM CONNECTED VERTEX COVER problem [17] is about finding a minimum size subset of vertices that spans a connected subgraph and meets all edges of the graph. Both cited papers give a linear-time 2approximation algorithm for their problem under investigation. The common point is that firstly an independence tree is found by a slight modification of depth first search. Then the authors prove that such a tree provides a 2-approximation for their problem. It turned out that this property of independence trees is a direct consequence of Corollary 4.16. Thus we can hope that the approximation ratio of 2 can be improved in the future for some special graph classes by further investigating the graph parameters present in Corollary 4.16. Currently, the best known approximation factors are 2 for the MINIMUM CONNECTED VERTEX COVER problem [17], and 7/4 for the MAXIMUM INTERNAL SPANNING TREE problem [15]. We need the following proposition to show that independence trees provide 2-approximation for both problems.

Proposition 5.1. If G is a graph on n vertices then $2(n - \alpha(G)) \ge n - \operatorname{sc}(G)$

Proof. Let X be an independent set of size $\alpha(G)$. Then

$$\operatorname{sc}(G) \ge \operatorname{comp}(G[X]) - |V \setminus X| = \alpha(G) - (n - \alpha(G)).$$

Subtracting both sides from n yields the proposition.

Therefore by Corollary 4.16 we have

Corollary 5.2. For any graph G on n nodes but K_n and C_n :

 $2(n - \alpha(G)) \ge n - (\operatorname{sc}(G) + 1) \ge n - \operatorname{ml}(G) \ge n - \operatorname{li}(G) \ge n - \alpha(G)$

Recall that by definition n - ml(G) is the maximum number of internal vertices in a spanning tree, and that by Corollary 4.13 n - li(G) is the cardinality of a minimum size connected vertex cover of G.

Now let T be any independence tree of G. Then $ml(G) \leq |V_1(T)| \leq li(G)$, and Corollary 5.2 shows that $n - |V_1(T)| \geq \frac{1}{2}(n - ml(G))$ and that $n - |V_1(T)| \leq 2(n - li(G))$. This means that T is a 2-approximation for both MAXIMUM INTERNAL SPANNING TREE and MINIMUM CONNECTED VERTEX COVER problems.

6 Spanning Many Vertices with a *q*-Leaf Tree

Up to now we were focusing on spanning trees with few leaves. In this section we do the contrary, we fix the number of leaves to q and examine how many vertices can be spanned by a q-leaf subtree. This approach is a generalization of finding a long path in a graph, as a path is just a 2-leaf subtree. Let $\delta_q(G)$ denote the minimum degree-sum of a q-element independent subset of V(G), that is $\delta_q(G) = \min_X \{\sum_{v \in X} d(v) : |X| = q, X \text{ is independent}\}.$

Let us recall Ore's theorem:

Theorem 6.1. [13] Let G be a graph on n vertices. If $\delta_2(G) \ge n$ then G has a Hamiltonian path.

For 2-connected graphs, Bermond proved the following generalization:

Theorem 6.2. [2] Let G be a 2-connected graph on n vertices. Then there exists a path of length min $\{n, \delta_2(G)\}$ in G.

Broersma and Tuinstra used a different point of view to generalize Ore's result. They examined the existence of q-leaf spanning trees and obtained the following:

Theorem 6.3. [4] Let G be a graph on n vertices. If $\delta_2 \ge n - q + 1$ for some integer $2 \le q \le n - 1$ then G has a q-leaf spanning tree.

In what follows we give a common generalization of the above results. We give a sufficient condition on the existence of a *q*-leaf subtree that spans *many* vertices.

To formulate our statement on subtrees we use $\rho_{q,k}$ to denote the minimum degreesum of the k highest degree vertices of a q-element independent set $(2 \le k \le q)$. Formally, $\rho_{q,k} = \min_X \left\{ \sum_{i=1}^k d(x_i) : X = \{x_1, x_2, \dots, x_q\}, d(x_1) \ge d(x_2) \ge \dots \ge d(x_q), X \text{ is independent} \right\}$. Clearly $\delta_q(G) = \rho_{q,q}$ and for $k \le q$: $\frac{\delta_q}{q} \le \frac{\rho_{q,k}}{k}$.

Theorem 6.4. Let G = (V, E) be a connected graph and let $2 \le q < \alpha(G)$ be an integer. Then G has a subtree with at most q leaves that spans at least min $\{\rho_{q,2} + q - 1, |V|\}$ vertices of G.

Proof. Let T be a maximum cardinality subtree of G with at most q leaves. If T spans G then we are done. Otherwise let $R = V(G) \setminus V(T) \neq \emptyset$. As T is maximal, none of its leaves is adjacent to R and thus for each leaf l we have $d_{G[V(T)]}(l) = d_G(l)$. Moreover T must be a minimum leaf spanning tree of G[V(T)]. Suppose that this is not the case and there exists a tree T' spanning V(T) with less than q leaves. Let e be an edge between V(T) and R. Then T' + e is a tree with at most q leaves that spans more vertices than T.

Thus T must have all properties of a minimum leaf spanning tree. According to Lemma 3.6, the set L of leaves of T is independent in G. Let l_1 and l_2 be the first and the second highest degree vertices in L. For the leaf l_2 we define the set $S = \{v : \exists u \in V(T) \text{ s.t. } (l_2, u) \in E(G) \setminus E(T), v = u^{\rightarrow l_2}\}$. By Lemma 3.7 we obtain that the vertices of S are forwarding vertices of T and so by the definition of S we have $|S| = d_G(l_2) - 1$. Moreover, also by Lemma 3.7, the vertices of S are not adjacent to l_1 , that is S and $N_G(l_1)$ are disjoint. Therefore

$$|V(T)| \ge |V_1(T)| + |S| + |N_G(l_1)| = q + d_G(l_2) - 1 + d_G(l_1) \ge \rho_{q,2}(G) + q - 1,$$

using the fact that the leaves of T form a q-size independent set.

The above bound is sharp as shown by the complete bipartite graph $G = (A \cup B, E) = K_{\delta,n-\delta}$ (for any $\delta = \frac{\delta_q}{q} < n/2$). To see this, let T be any non-spanning subtree of G having exactly q leaves and t = |V(T)| vertices.

If the leaves (being independent in G) are all in B then $|E(T)| = t - 1 = e_T(A, B) = e_T(A, B \cap V_{\geq 2}(T)) + q$, and each internal vertex of B has at least 2 neighbors in A, so we have $2|B \cap V_{\geq 2}(T)| \leq e_T(A, B \cap V_{\geq 2}(T)) = t - q - 1$. This, combined to $t \leq q + |B \cap V_{\geq 2}(T)| + \delta$ results $t \leq 2\delta + q - 1 = 2\frac{\delta_q}{q} + q - 1$. If the leaves are all in A then we take $G' = G - V_1(T)$ and a subtree $T' = T - V_1(T)$. It is easy to see that T' has all of its $q' \leq q$ leaves in B and following to the above argument we have $|V(T')| \leq 2(\delta - q) + q' - 1 \leq 2\delta - q - 1$ and so $t = |V(T')| + q \leq 2\delta - 1$.

As a result, at most $2\frac{\delta_q}{q} + q - 1 \le \rho_{q,2}(G) + q - 1$ vertices of G can be spanned by a subtree of at most q leaves.

Putting Corollary 3.8 and Theorem 6.4 together yields the following:

Corollary 6.5. Given a connected graph G = (V, E) and an integer $2 \le q$. If $q \ge \alpha(G)$ or $\rho_{q,2}(G) \ge n - q + 1$ then G has a spanning tree with at most q leaves.

This is a generalization of the result of Broersma and Tuinstra [4] as $\rho_{q,2}(G)$ (if defined) is an upper bound on $\delta_2(G)$.

7 Concluding Remarks

In this paper we discussed the connection between vulnerability of a graph G and the number of leaves of its spanning trees. We gave a lower bound on the number of leaves by means of scattering number sc(G) and an upper bound on the number of independent spanning tree leaves by means of cut-asymmetry ca(G). These bounds are used to prove that any independence tree provides a 2-approximation for both MAXIMUM INTERNAL SPANNING TREE and MINIMUM CONNECTED VERTEX COVER problems. We also gave some new cut-asymmetry related results. We proved that the leaf-set of an arbitrary spanning tree can contain at most ca(G) + 1 independence tree has at most li(G) leaves. However, it is an open question whether there always exists an independence tree with exactly li(G) leaves. If the answer is positive then the notion of leaf-independence is a generalization of the concept of independence trees.

At last, we gave a lower bound—by means of vertex degrees of q-element independent sets—on the number of vertices spanned by a q-leaf subtree.

Acknowledgments

Author thanks to Gábor Wiener and to the anonymous referees for their valuable comments.

References

- D. Bauer, H. Broersma and E. Schmeichel, Toughness in Graphs A Survey, *Graphs and Combinatorics* 22 (2006), 1–35.
- [2] J-C. Bermond, On Hamiltonian Walks, in: Proc. of the Fifth British Combinatorial Conference, 1975, 41–51.
- [3] T. Böhme, H. J. Broersma, F. Göbel, A. V. Kostochka and M. Stiebitz, Spanning trees with pairwise nonadjacent end-vertices, *Discrete Mathematics* 170 (1997), 219–222.
- [4] H. Broersma and H. Tuinstra, Independence Trees and Hamilton Cycles, Journal of Graph Theory 29 (1998), 227–237.
- [5] V. Chvátal, Tough graphs and on Hamiltonian circuits, *Discrete Mathematics* 5 (1973), 215–228.
- [6] M. Furer and B. Raghavachari, Approximating the Minimum Degree Spanning Tree to within One from the Optimal Degree, in: *Proc. of the 3rd Annual ACM-SIAM Symp. on Discrete Algorithms*, 1992, 317–324.
- [7] M. R. Garey and D. S. Johnson, *Computers and intractability: a guide to NP-completeness*, Freeman, 1979.
- [8] M. R. Garey and D. S. Johnson, The rectilinear Steiner tree problem is NP-complete, SIAM Journal of Applied Mathematics 32 (1977), 826–834.
- [9] L. Gargano, P. Hell, L. Stacho and U. Vaccaro, Spanning Trees with Bounded Number of Branch Vertices, in: *Proc. of ICALP'02*, Springer LNCS Vol. 2380, 2002, 355–365.
- [10] H. A. Jung, On a class of posets and the corresponding comparability graphs, *Journal on Combinatorial Theory, Series B* 24 (1978), 125–133.
- [11] D. Karger, R. Motwani and G. D. S. Ramkumar, On Approximating the Longest Path in a Graph, in: Proc. of WADS'93, 1993, 421–432.

- [12] H.-I. Lu and R. Ravi, The Power of Local Optimization: Approximation Algorithms for Maximum-leaf Spanning Tree (DRAFT), *Technical report number CS-96-05*, Department of Computer Science, Brown University, Providence, Rhode Island, 1996.
- [13] O. Ore, Note on Hamiltonian Circuits, American Mathematical Monthly 67 (1960), 55.
- [14] E. Prieto and C. Sloper, Either/or: Using vertex cover structure in designing FPT algorithms the case of *k*-internal spanning tree, in: *Proc. of WADS 2003, LNCS 2748, 2003, 465–483.*
- [15] G. Salamon, Approximation Algorithms for the Maximum Internal Spanning Tree Problem, in: *Proc. of MFCS 2007, LNCS 4708*, 2007, 90–102.
- [16] G. Salamon and G. Wiener, On Finding Spanning Trees with Few Leaves, *Information Processing Letters* 105 (2008), 164–169.
- [17] C. Savage, Depth-first search and the vertex cover problem, *Information Processing Letters* **14** (1982), 233–235.
- [18] S. Zhang and Z. Wang, Scattering number in graphs, Networks 37 (2001), 102–106.