3TH INTERNATIONAL CONFERENCE ON COMBINATORICS, CRYPTOGRAPHY AND COMPUTATION
Iran University of Science &Technology

Proceedings of the 3nd International Conference on Combinatorics, Cryptography and Computation (I4C2018)

Exrermal trees with respect to the Sackin index

Mahsa Khatibi
Imam khomeini International University/Department of Mathematics
Imam khomeini International University, Qazvin, Iran
Mahsa72kh@gmail.com

Ali Behtoei
Imam khomeini International University/Department of Mathematics
Imam khomeini International University, Qazvin, Iran
a.behtoei@sci.ikiu.ac.ir

ABSTRACT

The Sackin index of a rooted tree T that summarizes the shape of T is defined as the sum of the depths of its leaves and is denoted by S(T). The Sackin index measures the degree of balance of rooted phylogenetic trees and acyclic molecular graphs. For a given number n of leaves (terminal taxa), extermal phylogenetic and binary trees with exactly n leaves and with respect to the Sackin index and Colless index are characteized in the literature. In this paper we show that $n-1 \leq S(T) \leq \left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil$ for each rooted n-vertex tree T and we characteize all rooted n-vertex trees with extermal Sackin index.Also, we provide exact lower and upper bounds for the Sackin index of rooted d–ary trees and we characteize extermal families of them.

KEYWORDS: Sackin index, phylogenetic tree, Molecular graph, Extrermality, d-ary tree.

1 INTRODUCTION

Trees are defined as connected simple graphs without cycles, and their properties are basics of graph theory. A chemical graph is a graph whose vertex set denotes the atoms and its edge set denotes the bonds between those atoms of the underlying chemical structure. The structures of many molecules such as alkanes, dendrimers and acyclic molecules are tree like. Structures of many chemical compounds may be synthesized and categorized by mathematical means. Chemists have a long tradition of using the atomic valences (i.e vertex degrees) and the dictances between atoms to find molecular structures graphically. The vertex set and the edge of a tree T are denoted by V(T) and E(T), respectively. Two vertices in T which are connected by an edge are called adjacent vertices. For each vertex $v \in V(T)$ the set of vertices in T which are adjacent to v is the neighbourhood of v and is denoted by $N_T(v)$.The degree of v is $deg_T(v) = |N_T(v)|$. Each vertex of degree one is called a pendant vertex, an end vertex or a leaf. A path in T is a sequence of adjacent edges which do not pass through the same vertex more than once. The length of each path is the number of its edges. The analysis of path lengths in different trees has received a lot of attention mostly because of its importancein the analysis of algorithms. A **rooted tree** is a tree in which a particular vertex is distinguished from the othersand it is called the root vertex. Rooted trees have wide applications in

chemical graph theory from which enumeration and coding problems of chmical structures can be indicated. The distance brtween the root and a vertex $v \in V(T)$ is the **depth** of v and is denoted by $D_T(v)$. Each vertex of degree at least three is called a **branch** vertex. We call a vertex u as a support vertex if it has at least one (non-root) pendant neighbour. In a rooted tree, the parent of a vertex is the vertex connected to it on the path to the root. Every vertex except the root has a unique parent. Achild of a vertex v is a vertex of which v is the parent. A descendant of any vertex v is any vertex which is either the child of v or is (recursively) the descendant of any of the children of v. For each vertex v of T, T_v is the (induced) subtree of T rooted at v, i.e, T_v includes v and all of v's descendants. There are many special families of graphs with interesting properties. A tree T is said to be **starlike** if it has exactlyine branch vertex. The **broom graph** $B_{n,d}$ is a graph consisting of a path P_d of order d, toghether with $(n-d)$ pendant vertices all adjacent to the same leaf of P_d, see [8]. Broom graphs are in fact one of the important types of chemical trees. For each fixed integer $d \geq 2$, a d-**ary tree** is a rooted tree in which each node has no more than d children. Almost all of rooted chemical trees are special kinds of d-ary trees for $d = 4$. Quintas and Szymanski in 1992 considered some connections of chemistry to random trees and they studied molecules with tree structures in which nodes of valence $d = 3$ are saturated and unsaturated nodes have affinity that is inversely proportional to their valence [10]. A **topological index** for a (chemical) graph G is a numerical quantity which is invariant under automorphisms of G and it does not depend on the labeling or pictorial representation of the graph. Topological indices and graph invariants based on the distances between pair of vertices are widely used for characterizing molecular graphs, establishing relationship between structure and properties of molecules, predicting biological activity of chemical compounds, and making their chemical applications. Many different indices have been proposed in the literature to measure the degree of balance of chemical trees or rooted phylogenetic trees, see[2]. These indices depend only on the topology (the shape) of tree, and it is invariant under the relabelings of leaves. The two most popular balance indices are Sackins and Colless. Asymptotics for the mean, variance and covariance of these two statistics as well as their limiting joint distribution for large phylogenies are obtained in [3]. Sackin index is one of the oldest measures that summarizes the shape of a tree, see [12] and [13]. The **Sackin index** of a rooted tree T is defined as the sum of the depths of its leaves and is denoted by $S(T)$. Clearly, if two tree structures (molecular graphs) with the same root have different and hence, they have some different physico-chemical properties. For a given number n leaves with respect to the Sackin index and Colless index are characteized, see [4], [9], [11] and [13]. The aim of this paper provide exact lower and upper bounds for the Sackin index of all rooted trees of order n (not only for phylogenetic or binary trees or trees with specified number of leaves) and to characteize all trees that attain this bounds.

1.1 The main results

In this section we provide a lower bound and an upper bound for the Sachin index of general rooted trees in terms of the order of tree and we show that these bounds are exact and tight. Then, we characterize all trees that attain these bounds. Also, we obtain similar results for rooted d-ary trees.

1.2 Theorem

For the sackin index of each rooted n-vertex tree T we have $S(T) \geq n-1$. Moreover, in this lower bound the equality holds if and only if T is a path or a starlike tree (whose root is its unique branch vertex).

Proof. If T is a path of order n, then $S(T) = S(P_n) = n-1$. Note that each tree with the maximum degree 2 is a path. Let T be a starlike tree with the root r and with the maximum degree $\Delta(T) = deg_T(r) = k \geq 3$. Assume that $L(T) = \{x_1, x_2, \dots, x_k\}$ is the set of all leaves in T. Hence, T consists of the root vertex r and k paths. The order of each path is equal to its length plus one. The number of vertices (except the root) in each one of these paths is equal to the depth of the corresponding leaf in it. Thus, we have $n = 1 + \sum_{i=1}^{k} D_T(x_i)$ which implies that

$$S(T) = \sum_{i=1}^{k} D_T(x_i) = n - 1$$

Therefore the Sackin index of each rooted *n*-vertex path or starlike is $n - 1$. Now assume that *T* is a rooted *n*-vertex tree (with the root vertex *r*) which is not a path and is not a starlike tree. Thus, *T* has at least two branch vertices. Let $u \neq r$ be a branch vertex in *T*. Let u_1, u_2 be two children of *u*. Let T' be a rooted *n*-vertex tree obtained from *T* by removing the edge uu_1 and adding the edge ru_1 i.e $T' = T - uu_1 + ru_1$. Thus, the set of leaves in T' is equal to the set of leaves in *T*, i.e. $L(T') = L(T)$. Also, for each $y \in L(T) \cap L(T_{u_1})$ we have

$$D_{T'}(y) = D_T(y) - D_T(u) < D_T(y)$$

Hence, $S(T') < S(T)$. Therefore, *T* is not a rooted *n*-vertex tree with the minimum Sackin index among all rooted trees of order *n.* By the repetition of this process, finally we will reach a tree whose unique branch vertex is its root, i.e a starlike tree rooted at its branch vertex. The proof is completed.

1.3 Corollary

For the Sackin index of each rooted *n*-vertex *d*-ary tree *T* we have $S(T) \geq n - 1$. Also, the equality holds if and only if *T* is a path or a starlike tree whose root is its central vertex with degree at most *d.*

Proof. Using the proof of Theorem 1.2 and by the difinition of *d*-ary trees, the proof is straight forward because the root vertex may have at most *d* children.

1.4 Theorem

For each rooted tree *T* of order $n \geq 2$ we have $S(T) \leq \left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil$. Moreover, in this upper bound the equality holds if and only if *T* is a broom graph and $T \in \left\{ B_{n,\left\lfloor \frac{n}{2} \right\rfloor}, B_{n,\left\lceil \frac{n}{2} \right\rceil} \right\}$.

Proof. Let T_0 be a rooted *n*-vertex tree (with the root vertex *r*) whose Sackin index is maxzimum among all rooted *n*-vertex trees. At first we want to show that T_0 is a broom graph. Let $L(T_0)$ be the set of leaves in T_0 and let x_0 be a leaf with the largest depth amoung all leaves in T_0 an let y_0 be the neighbour of x_0. For each leaf x in To we have $D_{T_0}(x) \leq D_{T_0}(x_0)$.
If there exists a leaf x' with the support vertex y' such that $D_{T_0}(x') < D_{T_0}(x_0)$, then $y' \neq y_0$ and $D_{T_0}(y') < D_{T_0}(y_0)$ This implies that $S(T_0 - x'y' + x'y_0) > S(T_0)$ which contradicts the maximality of T_0. Hence, in T_0 all leaves are at the same depth. Now assume that there exists another support vertex $y \neq y_0$. Assume that $L(T_0) \cap N_{T_0(y)} = \{x_1, x_2, \dots, x_t\}$ for some $t \geq 1$. Since $D_{T_0}(u) = D_{T_0}(x_0)$ for each $u \in L(T_0)$, we have $deg_{T_0}(y) = 1 + t$. Note that $D_{T_0}(y) = D_{T_0}(y_0)$ and $y \neq y_0$imply that $y \neq r$ and hence $D_{T_0}(y) \geq 1$. Let

$$\tilde{T}_0 = T_0 - x_1y - x_2y - \cdots - x_ty + x_1y_0 + x_2y_0 + \cdots + x_ty_0.$$

Thus, $L(\tilde{T}_0) = L(T_0) \cup \{y\}$ and $S(\tilde{T}_0) = S(T_0) + D_{T_0}(y) > S(T_0)$, which is a contradiction. Therefore, y_0 is the unique support vertex for all (non-root) leaves in T_0. This means that $T_0$is isomorphic to the broom graph $B_{n,n-s}$ whose root is the remained pendant vertex of its main path P_{n-s}. Now assume that $N_{T_0}(y_0) \cap L(T_0) = \{x_0, x_1, \dots, x_{s-1}\}$ for some $s \geq 1$. Since $n = 1 + D_{T_0}(y_0) + s$, we have $D_{T_0}(y_0) = n - 1 - s$. Hence, $D_{T_0}(x_i) = n - s$ for each $i \in \{0,1, \dots, s-1\}$. Therefore,

$$S(T_0) = \sum_{i=0}^{s-1} D_T(x_i) = s(n-s).$$

It is easy to see that

$$\max\{s(n-s) : 1 \le s \le n, \qquad s \in N\} = \left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil.$$

Hence, among all rooted n-vertex brooms, $B_{n,\lfloor \frac{n}{2} \rfloor}$ and $B_{n,\lceil \frac{n}{2} \rceil}$ have the maximum Sackin index $\left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil$. This completes the proof.

1.5 Corollary

For the Sackin index of each rooted n-vertex d-ary tree T we have $S(T) \le f(n,d)$ where

$$f(n,d) = \begin{cases} \left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil & d \ge \left\lfloor \frac{n}{2} \right\rfloor \\ d(n-d) & d < \left\lfloor \frac{n}{2} \right\rfloor \end{cases}$$

Moreover, in this upper bound the equality holds if and only if $\left(d \ge \left\lfloor \frac{n}{2} \right\rfloor \ \ and\ T \in \left\{ B_{n,\lfloor \frac{n}{2} \rfloor}, B_{n,\lceil \frac{n}{2} \rceil} \right\}\right)$ or $\left(d < \left\lfloor \frac{n}{2} \right\rfloor \ \ and\ T = B_{n,n-d}\right)$.

Proof. Using the proof of Theorem 1.5 and by the definition of a *d*-ary tree, the proof is straightforward because y_0 may have at most *d* children.

1.6 Reference

S. Z. Aghamohammadi, on computing the general Narumi-Katayama index of some graphs, International Journal of Industrial Mathematics, 7 (1) (2015) 45-50.

M. G. Blum, O. Francois, on statistical tests of phylogenetic tree imbalance: the Sackin and other indices revisited. Mathematical biosciences 195 (2005) 141-153.

M. G. Blum, O. Francois, S. Janson, The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree balance, The Annals of Applied probability, 16 (4) (2006) 2195-2214.

S. B. Heard, Patterns in Tree Balance among Cladistic, Phenetic, and Randomly Generated phylogenetic Trees. Evolution, 46 (1992) 1818-1826.

S. M. Hosamani, Ashwini Index of a Graph, International Journal of Industral Mathematics, 8 (4) (2016) 377-384.

H. M. Mahmoud, Distances in random plane-oriented recursive trees, Jurnal of Computational and Applied Mathematic, 41 (1992) 237-245.

K. Moradian, R. Kazemi, M. H. Behzadi, The Sackin index of random recursive trees, U.P.B. Sci. Bull., Series A, 79 (2) (2017) 125-130.

M. J. Morgan, S. Mukwembi, H. C. Swart, on the eccentric connectivity index of a graph, Discete Math. 311 (2011) 1229-1234.

A. Mir, F. Rossello, L. Rotger, A new balance index for phylogenetic trees, Mathematical Biosciences, 241 (1) (2013) 125-136.

L. Quintas, J. Szymanski, Nonuniform random recursive trees with bounded degree, In Sets, Graphs, and numbers: Colloquia Mathematica Societas Janos Bolyai, 60 (1992) 611-620.

J. S. Rogers, Central moments and probability distributions of three measures of phylogenetic tree imbalance, Systematic Biology, 45 (1) (1996) 99-110.

M. J. Sackin, Good and bad phenograms, Systematic Zoology, 21 (1972) 225-226.

K. T. Shao, R. Sokal, Tree balance. Systematic Zoology, 39 (1990) 226-276.