Mathematics Interdisciplinary Research **5** (2020) 103 – 111

The Zagreb Index of Bucket Recursive Trees

Ramin Kazemi⋆*, Ali Behtoei and Akram Kohansal*

Abstract

Bucket recursive trees are an interesting and natural generalization of recursive trees. In this model the nodes are buckets that can hold up to $b \geq 1$ labels. The (modified) Zagreb index of a graph is defined as the sum of the squares of the outdegrees of all vertices in the graph. We give the mean and variance of this index in random bucket recursive trees. Also, two limiting results on this index are given.

Keywords: Bucket recursive tree, the Zagreb index, limiting rule.

2010 Mathematics Subject Classification: Primary 05C05, Secondary 60F05.

How to cite this article
R. Kazemi, A. Behtoei and A. Kohansal, The Zagreb index of Bucket recursive trees, *Math. Interdisc. Res.* **5** (2020) 103 – 111.

1. Introduction

A graph is a collection of points and lines connecting a subset of them. The points and lines of a graph are also called vertices and edges of the graph, respectively. The vertex and edge sets of a graph G are denoted by $V(G)$ and $E(G)$, respectively. The order of a graph G is the cardinality of its vertex set, and the size of a graph is the cardinality of its edge set. The degree of a vertex v of a graph is the number of edges incident to the vertex v and is denoted by $d(v)$ (or d_v). The number of tail ends adjacent to a vertex v is called its outdegree and is denoted by $d^+(v)$. The first Zagreb index $Z(G)$ of G is defined as

$$Z(G) = \sum_{v \in V(G)} d(v)^2,$$

and the (modified) Zagreb index is defined as

$$\overline{Z}(G) = \sum_{v \in V(G)} d^+(v)^2.$$

⋆Corresponding author (E-mail: r.kazemi@sci.ikiu.ac.ir)
Academic Editor: Janez Žerovnik
Received 7 October 2019, Accepted 29 March 2020
DOI: 10.22052/mir.2020.204312.1166

Thus, the Zagreb index of a graph is defined as the sum of the squares of the outdegrees of all vertices in the graph [4]. This indices reflects the extent of branching of the molecular carbon-atom skeleton, and can thus be viewed as molecular structure-descriptors (see for examples, [1, 2, 3] and references therein).

Trees are defined as connected graphs without cycles, and their properties are basics of graph theory. A rooted tree is a tree with a countable number of nodes, in which a particular node is distinguished from the others and called the root node. A leaf of tree is a node with outdegree 0.

A recursive tree with n nodes is an unordered rooted tree, where the nodes are labelled by distinct integers from $\{1, 2, 3, ..., n\}$ in such a way that the sequence of labels lying on the unique path from the root node to any node in the tree are always forming an increasing sequence [7].

Mahmoud and Smythe [6] introduced bucket recursive trees as a generalization of recursive trees. Bucket recursive trees can model many possible recruiting situations. In their model the nodes of a bucket recursive tree are buckets that can hold up to $b \geq 1$ labels. A bucket recursive tree grows by the progressive attraction of increasing integer labels (usually 1 to n). At the $(n+1)$st stage, the n previous labels (say, $1, 2, \ldots, n$) compete to attract the $(n+1)$st label (i.e $n+1$) and all existing labels have the equal chance of recruiting the new label. Hence, a node with k labels has affinity $\frac{k}{n}$ at stage $n+1$. In other words, if the capacity of the node v is k, $c(v) = k$, then the probability of v of attracting the $(n+1)$st label is $p_v = \frac{k}{n}$. When the new label falls into an unsaturated bucket or node, it joins the labels in that bucket but, when the new label has been attracted by a saturated bucket or node, then it is placed in a new bucket which is attached as a child to the attracting node. This implies that, the first b labels $1, \ldots, b$ are assigned to the root node. The label $b+1$ is instead in a new bucket as a child to the root node. Then, the label $b+2$ can be either joined to the same bucket as label $b+1$ is joined to (with probability $\frac{1}{b+1}$) or start a new bucket (with probability $\frac{b}{b+1}$) and so on. For $b = 1$ the ordinary recursive trees are obtained. An example of a bucket recursive tree is given in Figure 1. Note that for a bucket recursive tree $T_{n,b}$ on n labels with maximal bucket size b, we have

$$\frac{n}{b} \leq |V(T_{n,b})| \leq n - b + 1$$

and hence

$$\frac{n-b}{b} \leq |E(T_{n,b})| \leq n - b.$$

Mahmoud and Smythe [6] studied the multivariate structure of the tree and obtained a multivariate central limit theorem for the joint distribution of the number of nodes of different types for trees with bucket size $b \leq 26$. For trees with $b > 26$ a phase change in the distribution is detected and the central limit theorem does not hold. They studied two kinds of distance: the tree height and the depth of the nth label. Strong laws for the height are obtained via a technique based on introducing ghost nodes, then removing them. A weak law for the depth is introduced

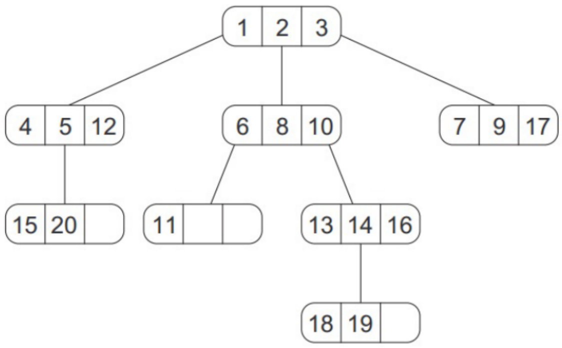

Figure 1: A bucket recursive tree on $n = 20$ labels with maximal bucket size $b = 3$.

by formulating and asymptotically manipulating the depth probability generating function. Let $X_n^{(i)}$ be the number of nodes of type i at stage n, where a node of type i is one containing i labels; the labels contained in a type i node will be called type i labels. They proved that as $n \to \infty$,

$$\mathbf{E}(X_n^{(i)}) = \frac{1}{i(i+1)H_b} n + O(n^{\alpha_b}), \qquad i = 1, 2, \ldots, b-1$$

and

$$\mathbf{E}(X_n^{(b)}) = \frac{1}{bH_b} n + O(n^{\alpha_b}), \tag{1}$$

where $\alpha_b < 1$ and H_n is the nth harmonic number. They proved that the asymptotic average number of buckets in a bucket recursive tree on n labels is n/H_b. One finds the asymptotic average number of internal nodes to be $n/(b+1)$. Summing up the average number of leaves of all types, one finds the asymptotic average total number of leaves to be $bn/((b+1)H_b)$; for large (but fixed) b, the asymptotic average number of buckets in the tree is about $n/\ln b$, and almost all the nodes of the tree are leaves.

2. Mean and Variance

Define the indicator function I on $V(T_{n,b})$, for each $v \in V(T_{n,b})$, as follows:

$$I(v) = \begin{cases} 1, & c(v) = b, \\ 0, & c(v) < b. \end{cases}$$

Lemma 2.1. *Let $M_1(T_{n,b})$ be the first Zagreb index and $\overline{M}_1(T_{n,b})$ be the modified first Zagreb index of a bucket recursive tree on n labels with maximal bucket size b. Then,*

i) $\sum_{v\in V(T_{n,b})} I(v)\,c(v) = b\,X_n^{(b)}$,

ii) $\sum_{v\in V(T_{n,b})} I(v)\,c(v)\,d^+(v) = b\,|E(T_{n,b})|$,

iii) $\sum_{v\in V(T_{n,b})} I(v)\,c(v)\,(d^+(v))^2 = b\left(M_1(T_{n,b}) - 3|E(T_{n,b})| + 2d^+(r_n)\right), \quad n > b.$

where $c(v)$ is the capacity of bucket v and r_n is the root vertex of $T_{n,b}$. Also, $|A|$ is the cardinality of set A.

Proof. By definition of $X_n^{(b)}$, part (i) is obvious. Assume that r_n is the root of $T_{n,b}$ and hence $c(r_n) = b$. Each node u in $V(T_{n,b}) \setminus \{r_n\}$ is a child of another node, say v, and is counted in v's outdegree. Hence

$$\bigcup_{v\in V(T_{n,b})} \bigcup_{u\in N^+(v)} \{u\} = \bigcup_{v\in V(T_{n,b})} N^+(v) = V(T_{n,b}) \setminus \{r_n\}, \tag{2}$$

where $N^+(v)$ is the set of children of v. Thus

$$\sum_{v\in V(T_{n,b})} d^+(v) = |V(T_{n,b}) \setminus \{r_n\}| = |E(T_{n,b})|.$$

Since the outdegree of each leaf is 0 and the capacity of each non-leaf node is b, we have

$$\sum_{v\in V(T_{n,b})} d^+(v)\,c(v) = \sum_{v\in V(T_{n,b})} d^+(v)b = b \sum_{v\in V(T_{n,b})} d^+(v) = b|E(T_{n,b})|.$$

For each $v \in V(T_{n,b}) \setminus \{r_n\}$ we have $d^+(v) = d(v) - 1$, and $d^+(r_n) = d(r_n)$. Now by using the definition of the first Zagreb index we see that

$$\begin{aligned} M_1(T_{n,b}) &= \sum_{v\in V(T_{n,b})} (d(v))^2 \\ &= d^+(r_n)^2 + \sum_{v\in V(T_{n,b})\setminus\{r_n\}} (1 + d^+(v))^2 \\ &= d^+(r_n)^2 + \sum_{v\in V(T_{n,b})\setminus\{r_n\}} (1 + 2d^+(v) + d^+(v)^2) \\ &= (|V(T_{n,b})| - 1) + 2(|E(T_{n,b})| - d^+(r_n)) + \sum_{v\in V(T_{n,b})} (d^+(v))^2 \\ &= 3|E(T_{n,b})| - 2d^+(r_n) + \sum_{v\in V(T_{n,b})} (d^+(v))^2. \end{aligned}$$

Therefore,

$$\sum_{v\in V(T_{n,b})} (d^+(v))^2 = M_1(T_{n,b}) - 3|E(T_{n,b})| + 2d^+(r_n).$$

Similarly, since the outdegree of each leaf is 0 and the capacity of each non-leaf node is b, we have

$$\sum_{v\in V(T_{n,b})} (d^+(v))^2 c(v) = \sum_{v\in V(T_{n,b})} (d^+(v))^2 b = b(M_1(T_{n,b}) - 3|E(T_{n,b})| + 2d^+(r_n)).$$

Hence, proof is completed directly from the fact that $I(v) = \lfloor \frac{c(v)}{b} \rfloor$.

□

In Theorems 2.2 and 2.3 we compute the mean (expected value) and variance of $\overline{M}_1(T_{n,b})$, respectively.

Theorem 2.2. *Let $\overline{M}_1(T_{n,b})$ be the modified Zagreb index of a bucket recursive tree on n labels with maximal bucket size b. Then*

$$\mathbf{E}(\overline{M}_1(T_{n,b})) = b\sum_{i=b}^{n-1} \frac{\mathbf{E}(2|E(T_{i,b})|) + X_i^{(b)})}{i}, \qquad n > b.$$

Proof. Let $\mathcal{F}_n$ be the sigma-field generated by the first n stages of the bucket recursive tree [5]. The label n is attracted by a randomly chosen bucket U_{n-1} in (random) bucket recursive tree $T_{n-1,b}$, and bucket recursive tree $T_{n,b}$ is obtained. If $c(U_{n-1}) < b$, then U_{n-1} is an unsaturated bucket and the label n is added to this node. In this case, each node preserves its previous outdegree. If $c(U_{n-1}) = b$ (i.e U_{n-1} is a saturated bucket), the the label n is attached to this node as a new bucket containing only the element n. Then

$$\overline{M}_1(T_{n,b}) = \overline{M}_1(T_{n-1,b}) + I(U_{n-1})(2d^+(U_{n-1}) + 1)$$

From Lemma 2.1,

$$\begin{aligned}\mathbb{E}\left(\overline{M}_1(T_{n,b}|\mathcal{F}_{n-1}\right) &= \overline{M}_1(T_{n-1,b}) + \frac{1}{n-1}\sum_{v\in V(T_{n-1,b})} I(v)c(v)\,(1 + 2d^+(v)) \\ &= \overline{M}_1(T_{n-1,b}) + \frac{b}{n-1}\left(2|E(T_{n-1,b})| + X_{n-1}^{(b)}\right). \qquad (3)\end{aligned}$$

Taking expectation of the relation (3), proof is completed since $\overline{M}_1(T_{i,b}) = 0$ for $i \le b$. □

The covariance between two jointly distributed real-valued random variables X and Y with finite second moments is defined as the expected product of their deviations from their individual expected values:

$$\mathbf{C}ov(X,Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))).$$

Theorem 2.3. *Let $\overline{M}_1(T_{n,b})$ be the modified Zagreb index of a bucket recursive tree on n labels with maximal bucket size b. Then the following relation holds:*

$$\begin{aligned}\mathbf{V}ar(\overline{M}_1(T_{n,b})) &= b\sum_{i=b}^{n-1}\Big(\frac{1}{i}\mathbf{E}\Big(4(M_1(T_{i,b})-8|E(T_{i,b})|+X_i^{(b)}+8d^+(r_i)\Big)\\ &- \frac{b^2}{i^2}\mathbf{E}^2(2|E(T_{i,b})|+X_i^{(b)})+2A_{i,b}\Big),\end{aligned}$$

where

$$A_{i,b}=\frac{2b}{i}\mathbf{C}ov(\overline{M}_1(T_{i,b}),X_i^{(b)})+\frac{b}{i}\mathbf{C}ov(\overline{M}_1(T_{i,b}),|E(T_{i,b})|).$$

Proof. Let $A_{n,n-1,b}=\mathbf{E}(\overline{M}_1(T_{n,b})-\overline{M}_1(T_{n-1,b}))^2$. From Lemma 2.1,

$$\begin{aligned}A_{n,n-1,b} &= \frac{1}{n-1}\sum_{v\in V(T_{n-1,b})}\mathbf{E}(I(v)(2d^+(v)+1)^2)c(v) \quad (4)\\ &= \frac{1}{n-1}\mathbf{E}\Big(4b(M_1(T_{n-1,b})-8b|E(T_{n-1,b})|+bX_{n-1}^{(b)}+8bd^+(r_{n-1})\Big).\end{aligned}$$

Since

$$\begin{aligned}\mathbf{E}(\overline{M}_1(T_{n,b})-\mathbf{E}(\overline{M}_1(T_{n,b}))|\mathcal{F}_{n-1})) &= \overline{M}_1(T_{n-1,b})-\mathbf{E}(\overline{M}_1(T_{n-1,b}))\\ &+ \frac{2b}{n-1}(|E(T_{n-1,b})|-\mathbf{E}(|E(T_{n-1,b})|))\\ &+ \frac{b}{n-1}(X_{n-1}^{(b)}-\mathbf{E}(X_{n-1}^{(b)})),\end{aligned}$$

we have

$$\mathbf{C}ov(\overline{M}_1(T_{n,b}),\overline{M}_1(T_{n-1,b}))=\mathbf{V}ar(\overline{M}_1(T_{n-1,b}))+A_{n-1,b},$$

where

$$A_{n-1,b}=\frac{b}{n-1}\mathbf{C}ov(\overline{M}_1(T_{n-1,b}),X_{n-1}^{(b)})+\frac{2b}{n-1}\mathbf{C}ov(\overline{M}_1(T_{n-1,b}),|E(T_{n-1,b})|).$$

But

$$\begin{aligned}A_{n,n-1,b} &= \mathbf{V}ar(\overline{M}_1(T_{n,b}))+\mathbf{V}ar(\overline{M}_1(T_{n-1,b}))\\ &- 2\mathbf{C}ov(\overline{M}_1(T_{n,b}),\overline{M}_1(T_{n-1,b}))\\ &- \mathbf{E}(\mathbf{E}(\overline{M}_1(T_{n,b}))-\mathbf{E}(\overline{M}_1(T_{n-1,b})))^2\\ &= \mathbf{V}ar(\overline{M}_1(T_{n,b}))-\mathbf{V}ar(\overline{M}_1(T_{n-1,b}))-2A_{n-1,b}\\ &+ \frac{b^2}{(n-1)^2}\mathbf{E}^2(2|E(T_{n-1,b})|+X_{n-1}^{(b)}). \quad (5)\end{aligned}$$

From (4) and then (5),

$$\mathbf{V}ar(\overline{M}_1(T_{n,b})) = \mathbf{V}ar(\overline{M}_1(T_{n-1,b})) + B_{n-1},$$

where

$$B_i = \frac{1}{i}\mathbf{E}\Big(4b(M_1(T_{i,b}) - 8b|E(T_{i,b})| + bX_i^{(b)} + 8bd^+(r_i)\Big) - \frac{b^2}{i^2}\mathbf{E}^2(2|E(T_{i,b})| + X_i^{(b)}) + 2A_{i,b}.$$

By iteration, proof is completed. □

3. Two Limiting Results

A sequence of random variables, $X_1, X_2, ...,$ converges in probability to a random variable X if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$$

or equivalently,

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1.$$

We use the notation $\xrightarrow{P}$ to denote convergence in probability.

Theorem 3.1. *As* $n \to \infty$,

$$\frac{\overline{M}_1(T_{n,b})}{n} \xrightarrow{P} \frac{2b+1}{H_b}.$$

Proof. It is an immediate consequence of Chebyshev's inequality [5], since $\mathbf{E}(\overline{M}_1(T_{n,b})) = \frac{2b+1}{H_b}n + o(n)$. □

Theorem 3.2. *Suppose*

$$W_{n,b} = \overline{M}_1(T_{n,b}) - \overline{M}_1(T_{n-1,b}) - a_{n-1,b}, \quad n > b$$

where

$$a_{n-1,b} = \frac{b}{n-1}\Big(2|E(T_{n-1,b})| + X_{n-1}^{(b)}\Big).$$

Then

$$\frac{1}{n}\sum_{j=b}^{n}\frac{1}{j}\sum_{i=b}^{j} W_{i,b} \xrightarrow{P} 0.$$

Proof. We have

$$\mathbf{E}((\overline{M}_1(T_{n,b}) - \overline{M}_1(T_{n-1,b}))^2|\mathcal{F}_{n-1}) = \mathbf{E}(W_{n,b}^2|\mathcal{F}_{n-1}) + a_{n-1,b}^2,$$

since $\mathbf{E}(W_{n,b}|\mathcal{F}_{n-1}) = 0$ and $a_{n-1,b}$ is $\mathcal{F}_{n-1}$-measurable [5]. Then

$$A_{n,n-1,b} = \mathbf{E}(W_{n,b}^2) + \mathbf{E}(a_{n-1,b}^2)$$

From Theorem 2.3,

$$\begin{aligned}\mathbf{E}(W_{n,b}^2) &= A_{n,n-1,b} - \mathbf{E}(a_{n-1,b}^2)\\ &= (2b+6) + O\Big(\frac{\log n}{n}\Big),\end{aligned}$$

since for a random variable X, $\mathbf{E}(X^2) \geq \mathbf{E}^2(X)$. Since

$$\sum_{j=b}^{n} \frac{1}{j} \sum_{i=b}^{j} W_{i,b} \leq (H_n - 1) \sum_{j=b}^{n} W_{j,b},$$

proof is completed by Chebyshev's inequality. □

4. Two Special Cases

Case 1.
Suppose $b = 1$. Then a bucket recursive tree reduce to an ordinary recursive tree. Thus $|E(T_{n,1})| = n - 1$. Since that in a random recursive tree of order n, $\mathbf{E}(r_n^+) = H_{n-1}$ [8], we have

$$\mathbf{E}(\overline{M}_1(T_{n,1})) = 1 \sum_{i=1}^{n-1} \frac{\mathbf{E}(2(i-1)+i)}{i} = 3n - 2H_{n-1} - 3 = 3n + O(\log n).$$

We have,

$$\overline{M}_1(T_{n,1}) = M_1(T_{n,1}) - 3(n-1) + 2d^+(r_n).$$

Then

$$\mathbf{E}(M_1(T_{n,1})) = 6n - 4H_{n-1} - 6 = 6n + O(\log n).$$

Also,

$$\mathbf{V}ar(\overline{M}_1(T_{n,1})) = 8n + \mathcal{O}(\log^2 n)$$

and as $n \to \infty$,

$$\frac{\overline{M}_1(T_{n,1})}{n} \xrightarrow{P} 3.$$

Case 2.
Suppose that the capacity of all leaves is b. I.e., all buckets are saturated buckets. Thus $|E(T_{n,b})| = \frac{n}{b} - 1$ and

$$\mathbf{E}(\overline{M}_1(T_{n,b})) = 3(n-b) - 2b(H_{n-1} - H_{b-1}).$$

Also,

$$\mathbf{V}ar(\overline{M}_1(T_{n,b})) = \frac{8}{b}n + \mathcal{O}(\log^2 n).$$

Confilcts of Interest. The authors declare that they have no conflicts of interest.

References

[1] A. T. Balaban, I. Motoc, D. Bondchev and O. Mekenyan, Topological indices for structure-activity correlations, *Topics Curr. Chem.* **114** (1983) 21–55.

[2] K. Ch. Das, I. Gutman and B. Zhou, New upper bounds on Zagreb indices, *J. Math. Chem.* **46**2 (2009) 514–521.

[3] K. Ch. Das and I. Gutman, Some properties of the second Zagreb index, *MATCH Commun. Math. Comput. Chem.* **52** (2004) 103–112.

[4] I. Gutman and N. Trinajstić, Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons, *Chem. Phys. Lett.* **17** (1972) 535–538.

[5] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*, Academic Press, New York, 1980.

[6] H. M. Mahmoud and R. T. Smythe, Probabilistic analysis of bucket recursive trees, *Theor. Comput. Sci.* **144** (1995) 221–249.

[7] A. Meir and J. W. Moon, Recursive trees with no nodes of out-degree one, *Congressus Numerantium.* **66** (1988) 49–62.

[8] R. T. Smythe and H. Mahmoud, A survey of recursive trees, *Theor. Probab. Math. Statist.* **51** (1995) 1–27.

Ramin Kazemi
Department of Statistics,
Imam Khomeini International University,
Qazvin, I. R. Iran
e-mail: r.kazemi@sci.ikiu.ac.ir

Ali Behtoei
Department of Pure Mathematics,
Imam Khomeini International University,
Qazvin, I. R. Iran
e-mail: a.behtoei@sci.ikiu.ac.ir

Akram Kohansal
Department of Statistics,
Imam Khomeini International University,
Qazvin, I. R. Iran
e-mail: kohansal@sci.ikiu.ac.ir