DISTRIBUTIONAL PROPERTIES OF MEASURES OF TREE TOPOLOGY

Jaap van Pelt, Harry B.M. Uylings and Ronald W.H. Verwer
Netherlands Institute for Brain Research
Meibergdreef 33, 1105 AZ Amsterdam, The Netherlands

ABSTRACT

An evaluation is made of measures of tree topology. These measures assign numbers to trees as a function of the topology. The use of such measures allows the researcher to study the variability in tree shapes on the basis of the distribution of numbers rather than of tree types. It is studied how many unique values the different measures can have in relation to the total number of different tree types. Such a property indicates to which extent the measures are able to distinguish the possible tree types. Additionally, the distribution of values for each measure is calculated for a uniform distribution of tree types. These data are important for selecting the most appropriate measure for the analysis of trees. It is shown that a measure for tree asymmetry on the basis of partition sets is most discriminative.

Keywords: branching patterns, distributions, measures, ordering of tree elements, rooted trees, topology.

INTRODUCTION

The topology of a tree is defined by the connectivity pattern of the segments in the tree. For a given number of segments several connectivity patterns or tree types are possible. In sets of natural trees, any tree type may occur, but with a probability that strongly depends on the developmental history. Populations of trees with different growth histories will therefore also differ in tree-type frequency distributions. The large number of different tree types (Table I, column 2), however, makes these distributions unmanageable. Therefore, the use of particular measures of tree topology may be preferred in the statistical analysis of tree populations. Then, by assigning numbers to the tree types, one can proceed with the statistical analysis of these number sets. In this paper several measures for binary trees that are frequently used in the literature, including one that was recently introduced, are discussed. It will be shown that any measure also implies a particular way of grouping of tree types.

MEASURES OF TREE TOPOLOGY

The elements that can be distinguished in a binary topological rooted
tree are points (root, bifurcation points and terminal tips) and segments connecting these points (intermediate and terminal segments), see Fig. la. The tree topology is defined by the number of these elements, e.g., segments (size), or terminal segments (degree) and the connectivity pattern. A measure of tree topology assigns a number to the tree as a function of the topology. Measures of topological trees can be evaluated on the basis of the distribution of their value set. The size of the value set (the number of unique values) becomes as large as the number of tree types if a measure assigns a unique number to each tree type. Then, the measure is maximally discriminative. The value set contains less numbers if the measure is not able to distinguish all tree types and assigns equal values to some of them. Apart from this discriminative aspect the numbers in the value set may have quite different distributional properties.

Recently, Uylings et al. (1989) have summarized measures that are currently used in the literature and they have noted that some of them are directly related to each other. A basic property of measures of tree topology is that they are based on the counting of tree elements. Consequently, all the discussed measures have or are based on rational or integer numbers as values. Three major categories of measures can be distinguished based on different ordering schemes of tree elements, viz. (a) the centrifugal ordering of tree elements, (b) the partitioning of tree elements at bifurcation points and (c) the Strahler ordering of segments.

![Diagram](image)

**Figure 1.** (a) elements of a topological tree; (b) numbering of tree elements by a centrifugal ordering scheme; (c) partitioning of terminal tips by bifurcation points (partitions) and (d) labeling of segments according to the Strahler ordering scheme.

**MEASURES BASED ON A CENTRIFUGAL ORDERING OF TREE ELEMENTS**

Centrifugal-order distribution. In the centrifugal ordering scheme of tree elements the root and the root segment are assigned centrifugal order (order) zero and each successive element in any path from the root to a terminal tip has an order incremented by one (Fig. 1b). The order of a point thus equals the length of the path (number of segments) between the root and this point. Evidently, the distribution, obtained by counting the number of elements per order depends on the tree type. This has been shown for the distribution of path lengths by Werner and Smart (1973) and for segment-orders by Van Pelt and Verwer (1987) and Van Pelt et al. (1989). The correspondence between path-length based and order based measures has been discussed in detail by Uylings et al. (1989). Different tree types may have equal order distributions, making the number of unique order distributions smaller than the number of different tree types. The algorithm of Werner and Smart (1973) has been used to calculate the number of unique order
Mean-order. A measure based on the order distribution is the sum of orders of all segments (order sum) or, alternatively, the mean order. The order sum is an even number because all non-zero orders have an even number of segments. For instance, the tree in Fig. 1b has an order sum of 16. There are two extreme order distributions. A 'thin' tree has one root segment and two segments at any other order. A 'compact' tree has \( 2^{n} \) segments at order \( n \), only the highest order may not be completely filled. The order sum for a thin tree of degree \( n \) (no. of terminal segments) is equal to \( n(n-1) \) and for a compact tree equal to \( 2(1-2^{-n}n_{m}) \) (Van Pelt and Verwer, 1987), with the maximal order \( n_{m} = \lfloor \log_{2} n \rfloor \). Here, \( \lfloor x \rfloor \) denotes the smallest integer, greater than or equal to \( x \). The number of unique values for the order sum (or for the mean order) is thus equal to \( n(n-1)/2 + 2^{n} - n_{m} \), being the number of even integers between and including the extreme values (Table 1, column 6). This number of unique values is smaller than the number of unique order distributions (Table I, column 5) because some order distributions have equal order sums. The frequencies of occurrence of the (equidistantly spaced) unique values for the tree types of degree \( n = 10 \) are shown in Fig. 2b. Ten trees appear to have a mean-order value of 3.05 (the order sum of 58, divided by the \( 2n-1 \) segments), indicating that they are indistinguishable for this measure. The frequency distribution of values is clearly not uniform.

Diameter. A second measure based on the order distribution is the diameter, being the largest path length in a tree, and equal to one plus the maximal segment order. This measure takes all integer values between its minimum for a 'compact' tree and its maximum for a 'thin' tree. It is a measure with only a few number of unique values as is illustrated in Table I, column 7. The non-uniform frequency distribution of the values for the tree types of degree 10 is shown in Fig. 2e.

MEASURES BASED ON THE PARTITIONING OF TREE ELEMENTS AT BIFURCATION POINTS

At each branching point in a binary tree the subsequent tree elements are partitioned over two subtrees. One can regard the topology of a tree as a particular scheme of successive partitions of terminal tips, as many as there are bifurcation points. Consequently, a tree can be represented by this scheme as is demonstrated by the use of a (unique) branching code, e.g. 11(5(2 3) 6(1 5(1 4(2 2)))).

Partition set. If the succession of partitions is ignored we obtain a set of individual partitions (partition set) \{(5,6),(1,5),(1,4),(2,3),(2,2)\}. Then, some trees become indistinguishable. For instance, the same partition set is obtained from the tree 11(5(1 4(2 2)) 6(1 5(2 3))). The number of unique partition sets derived from all trees of a given size is given in Table I, column 3. These numbers have been obtained by enumerating the sequence of all partition sets because an analytical expression is not available. The partition set appears to have highest discriminative power in comparison with any other representation of topological trees.

Tree asymmetry. A measure based on partitions is the partition asymmetry defined as \( |r-s|/(r+s-2) \) for the partition \( (r,s) \) (Van Pelt and Verwer, 1986). By averaging the asymmetries of all partitions \( (r,s) \) in a tree with \( r+s > 3 \), one obtains a measure for the tree asymmetry (Verwer and Van Pelt, 1986). The partition asymmetry is a rational number and the mean value for the partitions in a tree can consequently also be expressed as a rational
number. We have used this property in the selection of unique values for the tree asymmetry by comparing separately the (integer) values for the numerator and denominator of each rational number. The comparison on basis of 'floating point' representations of the asymmetry value would have introduced uncertainties due to 'rounding errors'. Some uniqueness in the partition sets is lost as is shown by the number of unique tree asymmetry values in Table I, column 4. However, this measure is still able to distinguish many tree types as is also shown in Fig. 2a. Only two values occur four times and most values occur only once in the set of trees of degree 10. In contrast to the measures based on the order distribution, the asymmetry measure does not have equidistant values but they concentrate in some areas of the value domain.

**First-order partition asymmetry.** Representing a tree by the single asymmetry of the first-order partition reduces the number of unique values considerably. Then, only \( \left\lceil \frac{n}{2} \right\rceil \) unique (rational) values are possible (Table I, last column). Here, \( \left\lceil x \right\rceil \) denotes the greatest integer smaller than or equal to \( x \). Almost half of the number of tree types are most asymmetric in the 1st-order partition (Fig. 2d).

**Vertex Ratio.** The Vertex Ratio (VR), as a measure for tree topology introduced by Berry and Flinn (1984), is defined as the ratio \( N(1,l)/N(1,x) \) with \( x \times 1 \), where \( N(1,j) \) is the number of \( (1,j) \) partitions in the partition set. This ratio can also be written as \( N(1,1)/(n-2N(1,1)) \) for trees of degree \( n \) (cf. Verver and Van Pelt, 1985), and becomes infinite large for pure symmetrical trees (then, \( n=2N(1,1) \)). This measure has only \( \left\lceil \frac{n}{2} \right\rceil \) unique values because of the \( \left\lceil \frac{n}{2} \right\rceil \) possible values for \( N(1,1) \). Horsfield and Woldenberg (1986) prefer to use the inverse vertex ratio to express asymmetry in a tree.

<table>
<thead>
<tr>
<th>degree n</th>
<th>tree types</th>
<th>part. sets</th>
<th>asymmetry</th>
<th>order distr.</th>
<th>mean order meter</th>
<th>Re-br. [n/2] ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>23</td>
<td>18</td>
<td>16</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>46</td>
<td>45</td>
<td>40</td>
<td>28</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>98</td>
<td>95</td>
<td>78</td>
<td>50</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>207</td>
<td>195</td>
<td>154</td>
<td>89</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>451</td>
<td>417</td>
<td>298</td>
<td>159</td>
<td>34</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>983</td>
<td>875</td>
<td>690</td>
<td>285</td>
<td>42</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>2179</td>
<td>1887</td>
<td>1105</td>
<td>510</td>
<td>51</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>4850</td>
<td>4021</td>
<td>2820</td>
<td>914</td>
<td>61</td>
<td>11</td>
</tr>
<tr>
<td>16</td>
<td>10905</td>
<td>8727</td>
<td>5300</td>
<td>1639</td>
<td>72</td>
<td>12</td>
</tr>
<tr>
<td>17</td>
<td>24631</td>
<td>18755</td>
<td>9185</td>
<td>2938</td>
<td>83</td>
<td>12</td>
</tr>
<tr>
<td>18</td>
<td>56011</td>
<td>40850</td>
<td>19666</td>
<td>5269</td>
<td>95</td>
<td>13</td>
</tr>
<tr>
<td>19</td>
<td>127912</td>
<td>88366</td>
<td>47692</td>
<td>9451</td>
<td>108</td>
<td>14</td>
</tr>
<tr>
<td>20</td>
<td>293547</td>
<td>192991</td>
<td>78264</td>
<td>16952</td>
<td>122</td>
<td>15</td>
</tr>
</tbody>
</table>

**Table I. Value Set Size of Measures of Tree Topology**

**MEASURES BASED ON THE STRAHLER ORDERING OF SEGMENTS**

In the Strahler ordering scheme, originally used for the analysis of river patterns, all terminal segments have Strahler order 1. At junctions of
segments the Strahler order of the parent segment is incremented with one if and only if both daughter segments have the same Strahler order (Fig. 1d). A sequence of segments with equal Strahler order is called a Strahler branch. A tree can be characterized by its stream set, indicating the number of branches for each Strahler order.

**Branching ratio.** A measure, based on this ordering scheme is the Rb-branching ratio, being the averaged ratio between the number of branches of two successive Strahler orders (e.g. Smart, 1972). The value set size of this measure is equal to the number of different stream sets (Table 1, column 8). This number can be obtained by counting the number of different sequences of positive integers, with the degree as first integer and each following integer maximally equal to half its predecessor.

Horsfield et al. (1987) use the B-branching ratio, being the ratio of the number of branches of Strahler-order 1 and 2 and prove that it directly relates to the Vertex Ratio (B=2+1/VR), thus belonging also to the class with \( |n/2| \) unique values (Table 1, column 9).

Both measures Rb and B show a highly non-uniform distribution of values between the minimum equal to 2 and a maximum (for a 'thin' tree) that is equal to the degree of the tree. This is illustrated in Figs. 2(c) and (f) for the set of tree types of degree 10.

![Graphs of various measures](image)

**Figure 2.** Value distributions of six measures of tree topology, obtained by calculating the values for the 98 tree types (nt) of degree (deg) 10. The (nval) unique values for each measure are exactly marked by vertical lines, indicating their frequencies. Note the single occurrence of the value of 10 in Figs. 2(e), (c) and (f) for the diameter, and the branching ratios Rb and B, respectively.

**DISCUSSION**

Several measures of tree topology have been evaluated on the basis of the number of unique values and the value distribution for a set of tree
types of degree 10. None of the measures is able to distinguish all the tree
types and each measure thus introduces a particular way of grouping. The
measure for tree asymmetry appears to have most discriminative power, e.g.,
it can distinguish 78 from the 98 trees of degree 10. Some other measures
belong to the category with \( n/2 \) unique values and distinguish only 5 of the
98 trees. Whether a combined use of these measures may result in a unique
vectorial representation of tree types, even for large trees, is an
interesting but still open question.

The unique values of the measures may uniformly fill the value domain
(e.g. mean order, 1st-order partition and diameter) or may show some
clustering (e.g. tree asymmetry and branching ratios). None of the measures
has a uniform distribution of values for the set of all tree types.

Apart from the value distribution of the measure, the tree types
themselves may also occur with non-uniform probabilities. In fact, any set of
natural trees will have a non-uniform probability distribution of the tree
types (as far as we know). The value distribution for an observed set of
branching patterns therefore depends both on the used measure and on the
tree-type frequencies. The difference in probabilities of occurrence between
tree types may be correlated with particular properties of these tree types.
Then, it makes sense to use a measure that is optimally sensitive for this
property such that the variance in the tree-type distribution is optimally
expressed by this measure. By using growth models we will be able to combine
both variance sources and this is the subject of our current studies.

ACKNOWLEDGEMENT

We are grateful to G. van der Meulen for help with figures.

REFERENCES

Berry M., Flinn R. Vertex analysis of Purkinje cell dendritic trees in the
Horsfield K, Woldenberg MJ. Branching ratio and growth of tree-like
Horsfield K, Woldenberg MJ, Bowes CL. Sequential and synchronous growth
1987; 49: 413-429.
Uylings HBM, Van Pelt J, Verwer RW. Topological analysis of individual
Van Pelt J, Verwer RW. Topological properties of binary trees grown with
197-211.
Van Pelt J, Verwer RW. Mean centrifugal order as a measure for branching
Van Pelt J, Verwer RW, Uylings HBM. Centrifugal-order distributions in
Verwer RW, Van Pelt J. Topological analysis of binary tree structures when
Verwer RW, Van Pelt J. Descriptive and comparative analysis of geometrical
properties of neuronal tree structures. J. Neurosci. Meth. 1986; 18:
179-206.
Werner C, Smart JS. Some new methods of toplologic classification of channel