On the Number of Backbone Refinement Classes

Andreas Maunz
Freiburg Center for Data Analysis and Modelling (FDM), Hermann-Herder-Str. 3a, 79104 Freiburg, Germany

This document contains a formula for calculating the number of BBRCs in a perfect binary tree. It is compared to the complete set of trees. See also:

- Andreas Maunz, Christoph Helma, Stefan Kramer, Large-Scale Graph Mining using Backbone Refinement Classes, in KDD '09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (forthcoming).

An example perfect binary tree (PBT) of height 3 is shown in Figure 1.

Figure 1: PBT with height 3. A longest path $\beta^*$ of length 6 has been marked by dashes. It has branches $B_{\beta^*} = \{a, b, c, d\}$, where the subtrees induced by $b$ and $c$ have longest paths of length $\sigma(b) = \sigma(c) = 2$. The path $\beta^*$ induces $\rho(\beta^*) = 4! = 24$ backbone refinement classes.

1 Branches and Induced BBRCs

Consider a path $\beta$ such as the one in Fig. 2. A branch (gray) is either present or not present (the picture shows all branches present). The number of branches is $\sigma(\beta) = \text{length}(\beta) - 1$, where $\text{length}(\beta)$ is the number of edges in $\beta$.

Let the branches be labelled $a, b, c, \ldots$. Consider the set $B$ of branches $\{a, b, c, \ldots\}$ and all its subsets including the empty set. The subsets can be partially ordered in a (directed) set graph according to “$\subseteq$”, such as shown in Figure 3 for the set $\{a, b, c\}$. Note, that there is always a node corresponding to the full set of branches (with indeg 0 and outdeg $\sigma(\beta)$) and node corresponding to the empty set (with indeg $\sigma(\beta)$ and outdeg 0).

**Lemma 1**: The number of paths from the full set node to the empty set node in the set graph of a backbone equals the number of backbone refinement classes induced by this backbone. It is $\rho(\beta) = \sigma(\beta)!$.
Figure 3: Subsets of branches.

Proof: Starting at the full set node, after having crossed \(i\) edges, there are \(\sigma(\beta) - i\) ways to remove exactly one item from the remaining set, which in combination yields \(\sigma(\beta)!\) different paths.

Let \(B_1, \ldots, B_n\) denote all the subsets of \(B\), including the full and empty sets. Associate \(x \in B_i\) with the following meaning: \(x\) is not present on the backbone.

It follows that two sibling nodes in the set tree induce two different backbone refinement classes, since all elements in one class contain a branch that no element from the other class contains (no two trees are refinements of each other).

On the other hand, the subgraph relation between a child and a parent node does not induce a new backbone refinement class, since all branches on the parent are still allowed on the child.

\[b(1) = 1\]

\[B(h) = \sum_{i,j=2}^{h} 2^{i-1}2^{j-1} \left( (i + j - 2)! + (i - 2) \sum_{s=1}^{i-1} b(s) + (j - 2) \sum_{t=1}^{j-1} b(t) \right), \quad h \geq 2\]
There are $\sum_{i,j=2}^{h} 2^{i-1}2^{j-1}$ paths containing the root in a PBT of height $h$. For each pair $(i, j)$, the corresponding path has $i + j - 2$ subtree inducing branches (because of the missing branch at the root node) and $(i + j - 2)!$ induced backbone refinement classes. Then, for every branch, we add its number of induced classes. On each side of the root, every branch can be combined uniquely with the $i - 2$ and $j - 2$ other branches, respectively, or combined with none.

3 Number of Subtrees

Szekely and Wang showed that a PBT with height $h$ has exactly

$$F(h) = \lfloor q^{2h+1} \rfloor$$

non-empty subtrees containing the root, where $q \approx 1.502837$.

4 Comparison of BBRC and Tree Set Sizes

For different values of height $h$ we calculate the number of BBRCs $B(h)$ according to Theorem 1, as well as the complete tree set size $F(h)$ according to the result of Szekely and Wang. Figure 4 compares the set sizes for heights 1 to 7.

<table>
<thead>
<tr>
<th>h</th>
<th>B(h)</th>
<th>F(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>632</td>
<td>677</td>
</tr>
<tr>
<td>4</td>
<td>60280</td>
<td>458331</td>
</tr>
<tr>
<td>5</td>
<td>1192296</td>
<td>2.101e+11</td>
</tr>
<tr>
<td>6</td>
<td>4.126e+09</td>
<td>4.413e+22</td>
</tr>
<tr>
<td>7</td>
<td>2.138e+12</td>
<td>1.947e+45</td>
</tr>
<tr>
<td>8</td>
<td>1.537e+15</td>
<td>3.792e+90</td>
</tr>
</tbody>
</table>

Figure 4: Comparison of BBRC and Tree Set Sizes