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:

- L.A. Szekely, Hua Wang, On subtrees of trees, Advances in Applied Mathematics, Volume 34, Issue 1, January 2005, Pages 138-155.
- 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.

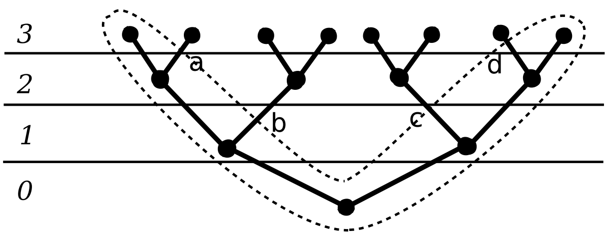

Figure 1: PBT with height 3. A longest path β^* of length 6 has been marked by dashes. It has branches $\mathcal{B}_{\beta^*} = \{a, b, c, d\}$, where the subtrees induced by b and c have longest paths of length $\sigma(b) = \sigma(c) = 2$. The path β^* induces $\rho(\beta^*) = 4! = 24$ backbone refinement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

1 Branches and Induced BBRCs

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

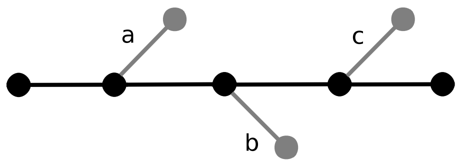

Figure 2: A backbone with branches (gray).

Let the branches be labelled $a, b, c, \ldots$. Consider the set $\mathcal{B}$ 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 "$\subset$", such as shown in Figure 3 for the set $\{a, b, c\}$. Note, that there is alway 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 $\rho(\beta) = \sigma(\beta)!$.

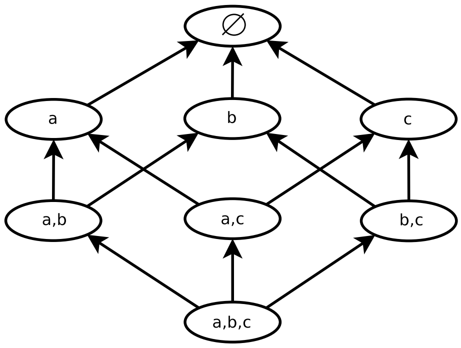

Figure 3: Subsets of branches.

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 $\mathcal{B}_1, \ldots, \mathcal{B}_n$ denote all the subsets of $\mathcal{B}$, including the full and empty sets. Associate $x \in \mathcal{B}_i$ 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.

□

2 Number of BBRCs

Lemma 2: The number of backbone refinement classes of a branch of height h can be recursively calculated as

$$b(h) = (h-1)! + (h-1)\sum_{i=1}^{h-1} b(i). \tag{1}$$

Proof:

From Lemma 1: The number of backbone refinement classes induced by a longest path β of this branch is $\rho(\beta) = (h-1)!$.
Then, for every branch b_i of the branches $b_1, \ldots, b_{h-1}$ of β, we recursively add its number of induced classes. Every branch can be combined uniquely with the $h-2$ other branches, or combined with none, thus appearing in a total of $h-1$ induced classes. □

We are now in the position to state the result.

Theorem 1: The number of backbone refinement classes of a complete binary tree of height h is

$$B(1) = 1 \tag{2}$$

$$B(h) = \sum_{i,j=2}^{h} 2^{i-1}2^{j-1}\ \left[\ (i+j-2)! + (i-2)\sum_{s=1}^{i-1} b(s) + (j-2)\sum_{t=1}^{j-1} b(t)\ \right], \qquad h \geq 2 \tag{3}$$

Proof:

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. $\square$

3 Number of Subtrees

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

$$F(h) = \lfloor q^{2^{h+1}} \rfloor \tag{4}$$

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.

h	B(h)	F(h)
1	1	5
2	8	26
3	632	677
4	60280	458331
5	11922296	2.101e+11
6	4.126e+09	4.413e+22
7	2.138e+12	1.947e+45
8	1.537e+15	3.792e+90

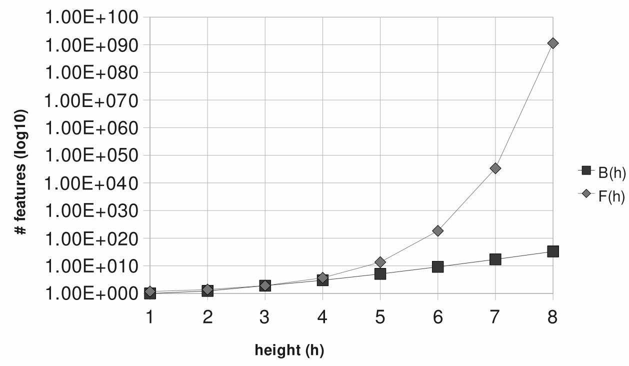

Figure 4: Comparison of BBRC and Tree Set Sizes