Correction in the reduction of the PPH problem to the graph realization problem.

The general idea of the reduction is correct, however there are errors in the details. If you are specifically interested in what the error is, read the explanation after the correction. However the easiest way to correct the error and also to simplify the reduction, is to present a new, complete discussion of the reduction. Then is done next.

1 Reducing the PPH problem to the Graph Realization Problem

The reduction has two steps.

Step 1: For a row i, which contains either a 1 or a 2, let $e_1, e_2, ...e_x$ be columns (possibly none) where row i contains 1 entries, ordered by their leaf-counts, largest first. As established earlier, if there is a PPH solution tree, then these columns must be on a directed path P from the root down to a leaf in tree T_1 . If there is one, let e be the column with the largest leaf count which has a 2 entry in row i .

We extend the ordered path P by adding a "common" glue edge g_0 to the start of it, and adding the edge e to the end of it. The resulting path is denoted P'. Note that if row i contains no 1 entries, then g_0 is added to the start of an empty path, and if row i contains no 2 entries, then no e is added to the end of P.

Recall that along a directed path from the root, in any solution of the PPH problem, the leaf counts of the column labels on the edges, must decrease (two adjacent edges with the same count on a directed path must have identical columns). Hence, path P' must be part of any PPH solution. Moreover, in any row i, if two columns with entries of 1 in row i have the same column counts, then the two columns must be identical, and the program should test for this. Similarly, if in any row i, a column with a 2 entry in row i must have a leaf count which is strictly less than the leaf count of any column that has a 1 in row i - otherwise, no tree is possible, and the program should check for this condidtion also. So after the leaf counts have been computed, and the columns sorted by leaf count, check these two conditions in each row.

Then, we create Π, by creating a path set for each adjacent pair of edges on P', and path set for each adjacent triple of edges on P'. It is easy to prove inductively that the only tree realization for these path-sets must be the ordered path P'.

By creating similar path-sets for each row (each time with the same g_0), we create a Π whose only realizing tree consists of the P' paths joined together with the glue edge g_0.

Step 2: Now consider a row i of S that contains more than one 2 value(s). Since at least one of column labels associated with a 2 value is already in the tree in the proper location, we only need to add to Π a path-set consisting of all the column labels in S where row i has a value of 2. A tree realization with this path set will create a path containing column labels associated with those 2 values, and hence that path will attach at the correct place in the tree, i.e., it will go through the last node on the path P for row i, as required by a PPH solution.

Creating such path sets for each row i creates a graph realization problem whose solution solves the PPH problem, after contracting edge g_0.

2 What was the problem with the original reduction

There is a small problem in section 3.5. The paper states

"However, if row i has no 2 value in a column of C_1, then in any solution to the PPH problem, the path between i and i' cannot go through an edge labeled with a column of C_1, and hence must go through a leaf of T_1. That leaf is easy to identify, as it is the only leaf v in T_1 whose entering edge has a label where row i has a value of 1 in S. So, for such a row i, we add to Π the path-set consisting of all the column labels in S where row i has a value of 2, along with the specific glue edge g_v. The use of g_v will ensure that the path realized is attached at the proper location in T^*."

One problem is that there may not be any 1's in row i. In that case, let the root of T play the role of v. But more problematic is that the general rule is not quite correct as shown in the following example:

	1	2	3
a	1	0	0
b	2	2	0

```
c  2  0  2
d  1  2  2
```

The reduction rules applied to the first three rows will ensure that the tree created has path 1,2 and 1,3 from the root, and will put a glue-edge g1 on the end of edge 1. Now for row d, the rule above creates the path requirement (2,3,g1). But edges 2 and 3 are already "in the tree", so there cannot be a single path through 2, 3 and g1. This is a problem, but it has a simple fix, as shown above.