Data and Computer Programs Used in the Research Project

titled

"The Bourque Distance for Mutation Trees of Cancers"

by K. Jahn,, N. Beerenwinkel and LX Zhang

1. 30node_randomtrees.txt

   This file contains 20,000 rooted trees with 30 uniquely label nodes in prufer format, which were generated by applying the nearest neighbour interchanges (NNI) in 400 steps. In each step, apply an NNI to a fix random tree produced in the previous step to generate a random tree for 5 times. The program used to generate this data set is "Random_tree_generator.c"

2. 30node_uneuqal_data1.txt

   The file contains 20,000 rooted trees given by edges, which were obtained by randomly removing 1, 2, or 3 nodes with prob. 1/200, 1./200, 1/400, respectively, and then merging randomly chosen 2 or 3 nodes with probability 1/100 and 1/200, respectively. As such, the trees have different label sets and have multiple-labelled nodes. The computer program used to generate this dataset is "Multi_Labeled_Tree_Generation.c".

3. 30node_unequal_data2_5percent.txt

   The file contains 20,000 rooted trees given by edges, which were obtained by randomly removing 1, 2, or 3 nodes from each tree with prob. 4/100, 4./100, 2/100, respectively, and then merging randomly chosen 2 or 3 nodes with probability 6/100 and 2/100, respectively. As such, the trees have different label sets and have multiple-labelled nodes. The computer program used to generate this dataset is "Multi_Labeled_Tree_Generation.c".

4. Random_tree_generator.c

   It was used to generate the random tree dataset in Item 1. The compile and run commands can be found on the top in the file.

5. Multi_Labeled_Tree_Generation.c

   It was used to generate the random tree datasets in Item 2 and Item 3. The compile and run commands can be found on the top in the file.

6. Pairwise_Bourque_DISTS_v3.c

   It computes the pairwise Bourque distance (BD), 1-Bourque distance (1-BD) and 2-Bourque distance (2-BD) between the trees in the input tree file, which are introduced in this work. The compile and run command can be found on the top of the file.

7. Pairwise_CASet_v2.c, Pairwise_DISC_v2.c, Pairwise_AD_v2.c

   They compute the pairwise CASet and DISC distances between trees in the input tree file, given in the reference [18] in our paper.

8. Pairwise_Triplets_v2.c

   It computes the pairwise the pairwise a triplet-based distance between the trees in the input tree file, which are introduced in reference [5] in our paper. The compile and run command can be found on the top of the file.