**UPGMA and Neighbor Joining Examples**

**Part I: Phylogenetic trees via UPGMA and Neighbor Joining algorithms. For problems 8 and 9, you are to consider the following distance matrix D:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** |
| **A** | **0** | **3** | **10** | **10** |
| **B** | **3** | **0** | **2** | **12** |
| **C** | **10** | **2** | **0** | **4** |
| **D** | **10** | **12** | **4** | **0** |

1. Run the UPGMA Algorithm on distance matrix D, using centroid based (average) linkage when computing inter-cluster distances, showing intermediate steps. Parts (a) and (c) are worth 5 points each. Parts (b), (d), and (e) are worth 10 points each.
   1. Which pair of nodes is Merged first? **B and C**
   2. What is the resulting inter-cluster distance matrix?

|  |  |  |  |
| --- | --- | --- | --- |
|  | **BC** | **A** | **D** |
| **BC** | **0** | **6.5** | **8** |
| **A** | **6.5** | **0** | **10** |
| **D** | **8** | **10** | **0** |

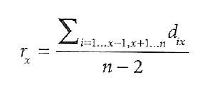
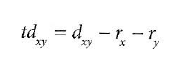
* 1. Which pair of nodes or clusters is Merged next? **(BC) and A**
  2. What is the resulting inter-cluster distance matrix?

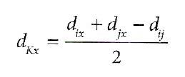
|  |  |  |
| --- | --- | --- |
|  | **ABC** | **D** |
| **ABC** | **0** | **26/3 = 8.66667** |
| **D** | **26/3 = 8.66667** | **0** |

* 1. What is the final phylogenetic tree based on the above UPGMA algorithm run?

**(((B C) A) D)**

1. Run the Neighbor-Join Algorithm on distance matrix D, showing the requested intermediate steps such as r-value computations in addition to intermediate transition distance matrices, recalculated distance matrices, and choices of the merged clusters. **You do not need to specify branch lengths of siblings X and Y to their new parent node (XY).** I will give the first set of r-values and transition distances. Relevant formulae are as follows:

**** ****

**where K is the newly merged cluster (i,j)**

Hence, the first set of r-values are: r(A) = 11.5, r(B) = 8.5, r(C) = 8, r(D) = 13

The first transition distance matrix TD is, accordingly:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | D |
| A | 0 | 3-11.5-8.5=-17 | 10-11.5-8=-9.5 | 10-11.5-13=-14.5 |
| B | -17 | 0 | 2-8.5-8=-14.5 | 12-8.5-13=-9.5 |
| C | -9.5 | -14.5 | 0 | 4-8-13=-17 |
| D | -14.5 | -9.5 | -17 | 0 |

Hence, there is a tie between (A,B) and (C,D) for the first pair of nodes to merge. **Break the tie in favor of the smaller original distance**, merging (A,B) first.

1. Compute the remaining entry of this updated distance matrix **(3 points)**:

|  |  |  |  |
| --- | --- | --- | --- |
|  | AB | C | D |
| AB | 0 | (10+2-3)/2=4.5 | (10+12-3)/2 = 9.5 |
| C | 4.5 | 0 | 4 |
| D | 9.5 | 4 | 0 |

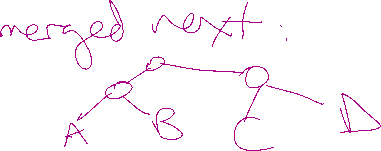
1. Now compute the new average distance values r(AB) and r(C) **(6 points)**:

r(AB) = 14, r(C) = 8.5, r(D) = 13.5

1. Now compute the new transition distance matrix TD **(12 points)**:

|  |  |  |  |
| --- | --- | --- | --- |
|  | AB | C | D |
| AB | 0 | 4.5-14-8.5=-18 | 9.5-14-13.5=-18 |
| C | -18 | 0 | 4-8.5-13.5=-18 |
| D | -18 | -18 | 0 |

1. Based on the transition distances computed in part (c), break any ties in favor of the smaller of the updated distances in part (a) to pick the nodes or clusters that will be merged next. Finally, output the resulting phylogenetic tree ignoring branch lengths **(4 points)**:



1. **EXTRA CREDIT:** We had to state tie-breakers for multiple steps in order to give a unique Neighbor-Join phylogenetic tree for the original distance matrix. However, there are actually 5 topologically distinct Neighbor-Join phylogenetic trees based on the original distances. What are all five such trees **(10 points extra credit)**?