ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.11, 9541–9550
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.11.57 Spec. Iss. on ICRTAMS-2020

CONSTRUCTION OF CLADOGRAM USING BINARY MATRIX

RENJINI RAVEENDRAN P[1] AND S. BEENA

ABSTRACT. Mathematics and biology can be combined to get better solutions to problems. Graph theory has its application in all other fields. Phylogenetic analysis deals with the study of evolutionary history of organisms. Phylogenetic trees are well suited to represent evolutionary histories. A rooted network or a tree can be depicted either as a phylogram or as a cladogram. A cladogram is a drawing of a rooted phylogenetic tree or network aimed at clearly representing its topology, that is showing how clusters are nested within each other, and in which edge lengths do not have an interpretation. Morphological data of various organisms can be used for the construction of Cladograms. Tabular form of those data has to be made first. This work explains an algorithm with illustration for the construction of Cladogram after converting the data to a Binary Matrix.

1. INTRODUCTION

Mathematics has its application in almost all other branches of sciences. Graph theory is a branch of mathematics that has its influence in physics, chemistry, architecture, genetics, psychology, linguistics and so on. The concept of a rooted tree from Graph Theory is very simple and has proved to be extremely useful in many application domains.

Phylogenetic analysis deals with the study of evolutionary relationships between different species or taxa. It helps us to understand how genes, genomes,

---

[1] *corresponding author*

2010 *Mathematics Subject Classification.* 05C90.

Key words and phrases. cladogram, phylogenetic tree, rooted phylogenetic tree.

species, molecular sequences evolve and the evolution of life on Earth. The evolutionary history of a set of species is usually described by a rooted phylogenetic tree.

The word "cladogram", derived from Greek word "clados" meaning "branch" and "gramma" meaning "character" is a diagram to show relations among organisms. A Cladogram uses lines (edges) that branch off in different directions ending at a group of organisms with a last common ancestor. The lines (edges) represent characters. The length of the lines (edge length) does not have any interpretation in a Cladogram. There are many shapes of Cladograms but they all have lines that branch off from other lines. The lines can be traced back to where they branch off. These branching off points represent a hypothetical ancestor (not an actual entity) which can be inferred to exhibit the traits shared among the terminal taxa above it. This hypothetical ancestor might then provide clues about the order of evolution of various features, adaptation, and evolutionary history. Cladograms were generated largely on the basis of morphological characters. About graph theory and phylogenetic networks see [1–3].

This work deals with the construction of Cladogram as rooted tree. For this we need a morphological data of various organisms. This has to be first convert to a tabular form showing the presence or absence of certain characters. This will again be converted to a binary matrix, where 1 shows the presence of character and 0 otherwise. Then by step by step procedure we can construct the Cladogram.

2. PRELIMINARIES

(1.1) A graph G is a pair $G = (V, E)$ consisting of a finite set V and a set of 2-element subsets of V. The elements of V and E are called *vertices* and *edges* respectively.

(1.2) $X = \{x_1, \cdots, x_n\}$ is denoting a set of taxa, in which each taxon x_i represents some species, group or individual organism whose evolutionary history is of interest to us.

(1.3) Let X be a set of taxa. A cluster is any subset of X, excluding both the empty set $\emptyset$ and the full set X.

(1.4) A phylogeny describes the evolutionary history of a set of taxa.

(1.5) A tree is connected acyclic graph.

(1.6) Given a set of taxa X, a rooted phylogenetic tree consists of a rooted tree $T = (V, E, \rho)$ and a taxon labelling $\lambda : X \to V$ that assigns exactly one taxon to every leaf and none to any internal node. All nodes, except ρ, must have degree $\neq 2$.

(1.7) Binary matrix is matrix with entries $B = \{0, 1\}$. Such a matrix can be used to represent a binary relation.

3. ALGORITHM

Morphological data of various organisms can be used for the construction of Cladograms. Tabular form of those data has to be made first.
For example:

Organism/Character	Feather	Fur	Lungs	Gizzard	Jaws
Lamprey	✗	✗	✗	✗	✗
Antelope	✗	✓	✓	✗	✓
Sea Bass	✗	✗	✗	✗	✓
Bald Eagle	✓	✗	✓	✓	✓
Alligator	✗	✗	✓	✓	✓

TABLE 1

An Algorithm for the construction of Cladogram

The above data shows the relation between organisms and character (Presence or absence of certain characters).

Step I: Convert this tabular data to a Binary matrix. The rows represent various organisms and the columns represent the characters. The entries of the matrix $a_{ij} = 0$ implies the organism R_i doesn't have the character C_j, and $a_{ij} = 1$ implies the organism R_i possess the character C_j.

Step II: Identifying the Outgroup (outgroup is a taxon that is closely related to the main group of the taxa under consideration, but lies outside of it). Outgroups can be identified by their row entries. If all the entries in the row are 0, corresponding organism is an outgroup. Draw an edge and represent that organism at one end of the edge.

Step III: Delete that row from the Binary matrix.

Step IV: Consider columns having all entries 1 (if any). If not, select one column having more 1's as their entry. Represent the corresponding column name (character) along an edge and place it by forming a root with the outgroup.

Step V: Select the column containing less number of 1 in their entries than the columns selected in Step IV.

Step VI: Compare the entries of the columns selected in Step IV and V. Note the Position of 1 that changes to 0 row wise. When comparing C_s and C_t, row wise change occur in R_k, then part of cladogram becomes:

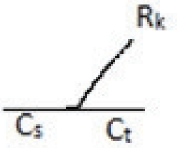

FIGURE 1

Now check all values of R_k, if there is any more 1 in the row then corresponding column C_d can be represented as:

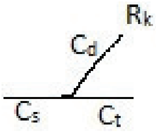

FIGURE 2

Step VII: Delete column C_s and row R_k.

Step VIII: Take C_t and a column having less number of 1 in their entries than that of C_t. Follow Step VI.

Step IX: Proceeding in this manner, last we obtain:

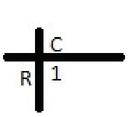

FIGURE 3

which means that the edge with symbol C ends in the organism R. Some relation may not end with single entry and that can be easily construct by the obtained simpler entry.

Step X: We obtain a rooted tree which shows the morphological relation among the given organisms.

4. ILLUSTRATION

Consider the Table 2:

TABLE 2

Organism/Character	Feather	Fur	Lungs	Gizzard	Jaws
Lamprey	✗	✗	✗	✗	✗
Antelope	✗	✓	✓	✗	✓
Sea Bass	✗	✗	✗	✗	✓
Bald Eagle	✓	✗	✓	✓	✓
Alligator	✗	✗	✓	✓	✓

Step I:

TABLE 3

Organism/Character	Feather	Fur	Lungs	Gizzard	Jaws
Lamprey	0	0	0	0	0
Antelope	0	1	1	0	1
Sea Bass	0	0	0	0	1
Bald Eagle	1	0	1	1	1
Alligator	0	0	1	1	1

Step II:

TABLE 4

Organism/Character	Feather	Fur	Lungs	Gizzard	Jaws
Lamprey	**0**	**0**	**0**	**0**	**0**
Antelope	0	1	1	0	1
Sea Bass	0	0	0	0	1
Bald Eagle	1	0	1	1	1
Alligator	0	0	1	1	1

Identifying **Outgroup - Lamprey**:

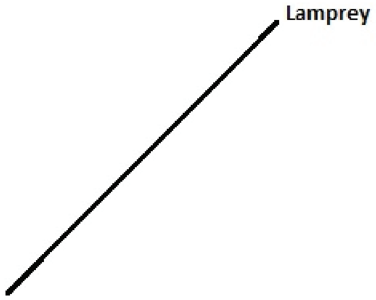

FIGURE 4

Step III:

TABLE 5

Organism/Character	Feather	Fur	Lungs	Gizzard	Jaws
Antelope	0	1	1	0	1
Sea Bass	0	0	0	0	1
Bald Eagle	1	0	1	1	1
Alligator	0	0	1	1	1

Step IV:

TABLE 6

Organism/Character	Feather	Fur	Lungs	Gizzard	**Jaws**
Antelope	0	1	1	0	**1**
Sea Bass	0	0	0	0	**1**
Bald Eagle	1	0	1	1	**1**
Alligator	0	0	1	1	**1**

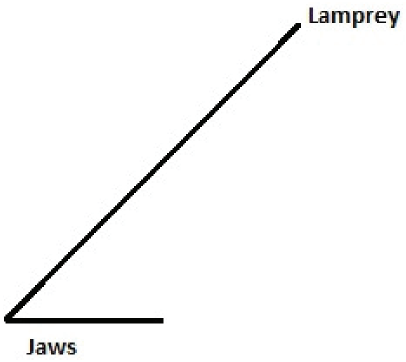

FIGURE 5

Step V and VI: Comparing the columns Jaws and Lungs:

TABLE 7

Organism/Character	Feather	Fur	Lungs	Gizzard	**Jaws**
Antelope	0	1	**1**	0	**1**
Sea Bass	0	0	**0**	0	**1**
Bald Eagle	1	0	**1**	1	**1**
Alligator	0	0	**1**	1	**1**

The 1 changes to 0 corresponding to the organism Sea Bass.

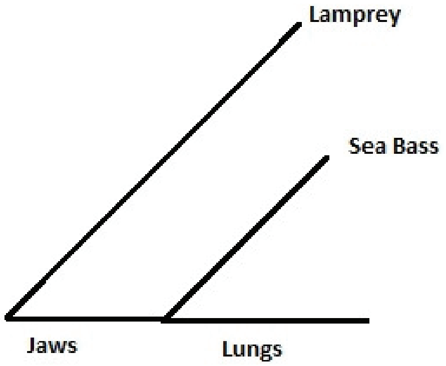

FIGURE 6

Step VII: Delete column Jaws and Row Sea Bass (As all other entries in that row are zero).

TABLE 8

Organism/Character	Feather	Fur	Lungs	Gizzard
Antelope	0	1	**1**	**0**
Bald Eagle	1	0	**1**	**1**
Alligator	0	0	**1**	**1**

Step VIII, IX and X: Compare the columns Lungs and Gizzard. The 1 changes to 0 corresponding to Row Antilope. But here the Row Antilop contains one more non zero entry - corresponds to column Fur.

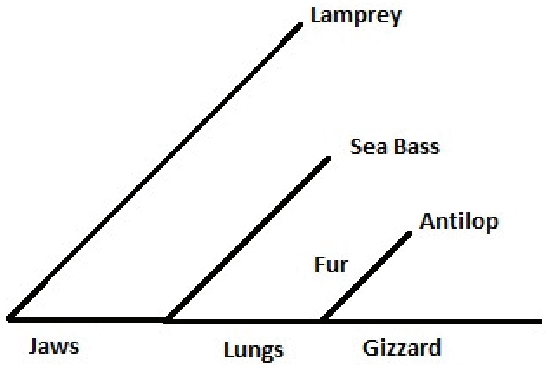

FIGURE 7

Delete the Column Lungs, Fur (as there is no more 1 entry) and the Row Antilop.

TABLE 9

Organism/Character	Feather	**Gizzard**
Bald Eagle	1	**1**
Alligator	0	**1**

Now compare the columns Gizzard and Feather. The 1 changes to 0 corresponding to the Row Alligator.

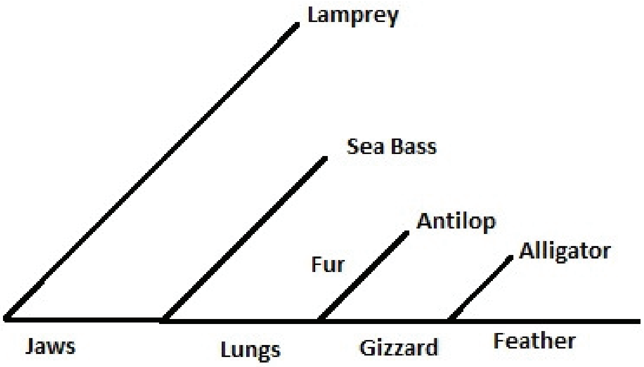

FIGURE 8

Delete the column Gizzard and the row Alligator.

TABLE 10

Organism/Character	Feather
Bald Eagle	1

This means that the edge Feather ends with Bald Eagle.

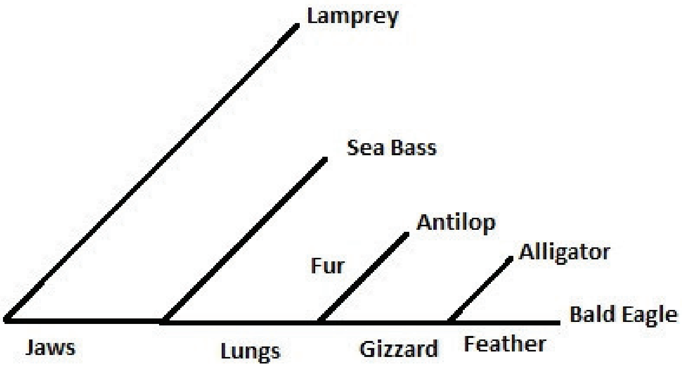

FIGURE 9. Cladogram

REFERENCES

[1] G. S. SINGH: *Graph Theory*, PHI Private Limited, New Delhi, 2010.

[2] N. DEO: *Graph Theory with Applications to Engineering and Computer Science*, PHI Learning, 2016.

[3] D. H. HUSON, R. RUPP, C. SCORNAVACCA: *Phylogenetic Networks: Concepts, Algorithms and Applications,* Cambridge University Press, 2011.

DEPARTMENT OF MATHEMATICS
ALL SAINTS' COLLEGE
THIRUVANANTHAPURAM-695007
KERALA, INDIA
Email address: `renjiniraveendran7@gmail.com`

MMNSS COLLEGE
KOTTIYAM