

# Application of Graph Theory in Phylogenetics: The Primate Approach

Dhiraj Kumar Deka

*Abstract.* The theory of evolution has been the crowning achievement of mankind that has altered our perceptions of the world in a revolutionary way. This naïve looking theory whose only tool of operation is a simple mechanism called “natural selection” has the ability to challenge all the ideas relating to the origin of life that different civilisations nurtured for thousand of years. The theory of evolution has ceased to become a study of the domain of “biology” alone and different inter-disciplinary sciences have joined hands to create the perfect representation for the saga of life on this planet. This kind of cooperation is both unique and astonishing. Mathematics has also been a part of this endeavour and in this paper I explore how the concepts of “graph theory” can be exploited in studying “phylogeny”, which is the heart of the evolution analysis.

## 1. Introduction

We have inherited a unique planet that hosts an incredible bio-chemical feature called life. There are about 5 to 100 millions species of living organisms on earth today varying in shapes and sizes. Yet they exhibit morphological, biochemical and behavioural similarities that point towards some form of genealogical relationship. The study of such evolutionary relationship among varied groups of living organism is known as phylogeny. This relationship among wide range of species is pictorially represented in form of a graph called “phylogenetic tree”. A phylogenetic tree is a branched diagram showing evolutionary interrelations between a group of species which have evolved from a common ancestral form. Two closely related species are placed closely and by studying the degree of separation we can gauge how these species will behave. In this paper, we study various properties of the phylogenetic tree through the lens of graph theory using a concept called “rooted tree”.

## 2. What is Life?

There are lots of biological, chemical and philosophical definitions of life based on the three prominent

characteristics aspects of life: living beings grows, interchange energy and reproduce of the same kind. But the most ingenious definition of life from the point of view of physics was given by Erwin Schrödinger in 1944. He defined life as an organised structure that defies “the second law of thermodynamics”: the holy grail of classical physics. The celebrated second law of thermodynamics says that entropy of a closed system always increases [1]. In simple words, entropy is a measurable quantity which roughly tells about the disorderliness of the atoms in an object. Thus the first living object on earth, if such a thing really existed (?), had a chemical combination which could act against this disorderliness. It could keep its atoms in the desired spatial and physical state so that they could stay stable, which according to Richard Dawkins is the first rule of life [2]. And this organism much had a way to pass on this characteristic to its next lineage and so on. In each passage of genetic material from one generation to the other, there are some subtle variations called mutation which bring minute changes in the offspring. The nature selects the “favourable changes”. “Natural selection not only enhances the reproductive success of favourable variants but also diminishes the reproductive success of unfavourable ones” [3]. After deep time, these accumulated changes in the favourable variants bring out a new species.

The above concept of “natural selection” holding the key to the existence of human to whale on earth seems to be too naïve to be true. But this imaginative idea due to Charles Darwin is well supported by evidence and has been an accepted fact since its first publication in the book titled “On the origin of species by means of natural selection, or the preservation of the favoured races in the struggle for life” in 1859.

The most striking point of Darwin’s theory of evolution (Darwinism) is that a new species can originate only from a species which has been in existence already. If that is so, then all living organism are

inter-linked and there must be a way to obtain the complete picture of all the life forms through these connections.

This is done by using a concept called “Tree”.

“Tree” is an idea that originated in Kirchhoff’s work on electrical connections. Applying these ideas to phylogeny, we create the “phylogenetic tree”, which follows the rules of a “tree” as defined in graph theory. Hence we first look how a graph theorist defines a “tree”.

### 3. Definition of a Tree

A tree is a connected undirected graph with no simple circuits. A rooted tree is a tree in which one vertex has been designated as the root and every edge is directed away from it. A vertex  $u$  is called the parent of the vertex  $v$  if there exists a directed edge from  $u$  to  $v$ . When  $u$  is the parent of  $v$ , then we say that  $v$  is a child of  $u$ . Vertices with the same parents are called siblings. The ancestors of a vertex  $v$  are the vertices in the path from the root to the vertex  $v$ . The descendants of a vertex  $d$  are those vertices that have  $d$  as ancestors. A vertex of a tree is called a leaf if it has no children [4].

Now we list the three most important properties of the tree which will be exploited in this study:

1. Tree is a connected graph. Hence there always exists a path in between any two pair of vertices. But what is unique about the paths in tree is that these paths are unique.
2. Tree is a simple graph. It does not contain any multiple edges.
3. Tree does not contain any cycle or self loop.

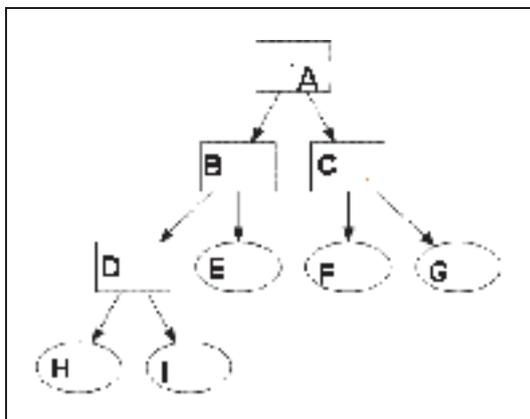


Fig. 1. A rooted tree.

In the above figure, the vertex A is the parent of both B and C. Thus B and C are the children of A. B

and C are siblings as they have a common parent but E and F are not siblings as their parents are different. The vertex D, B and A are the ancestors of H and I. Thus the descendent of B are D, E, H and I. In this figure E, F, G, H and I (denoted in circle) are leaves as they have no children. A is the root of the tree as all the other vertices are its descendent.

### 4. Phylogenetic Tree

When the vertices of a rooted tree represent species and the edges represent their inter connections with other species then we call that tree to be a phylogenetic tree or evolutionary tree. Thus a phylogenetic tree is a branching diagram that shows the evolutionary relationship between related species which have a common ancestor. For example if each letter A, B, C, D, E, H, I and J in Fig. 1 represents a distinct species, then that diagram will be an example of phylogenetic tree of the species.

This important tree catches the entire hullabaloo that compels the species to evolve and represents history spanning a million years with an edge no longer than five centimeters long! For example, if the paths from two different species originate from a single vertex then we conclude that they had the same ancestor. Similarly, a long path between two species will necessarily mean that they would have less resemblance.

In Fig. 2, we present the phylogenetic tree of the different species of gorilla. Here all the vertices represent the species (like Eastern lowland gorilla, mountain gorilla, etc). We see that all these different variants of gorilla originated from single species which is the ancestor of all other gorilla species. Likewise we see that Eastern gorilla is the parent of the Eastern lowland gorilla and thus the later is the children of the former’s. Also as both Eastern lowland gorilla and Mountain gorilla have a common parent, hence they are siblings. We notice that Eastern lowland gorilla, Mountain gorilla, Western lowland gorilla and Cross river gorilla are “leaves” as no species evolved from them.

### 5. The Phylogenetic Tree of the Primates

Now we come to the “meat” of the paper. We will be exploring the concept from the graph theory in studying the phylogenetic tree and will try to arrive on conclusion simply based on mathematical deduction. As a case study, we have selected the phylogenetic

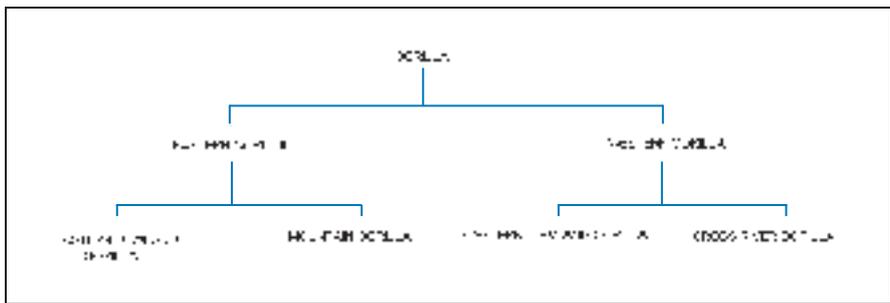


Fig. 2. Phylogenetic tree of gorilla.

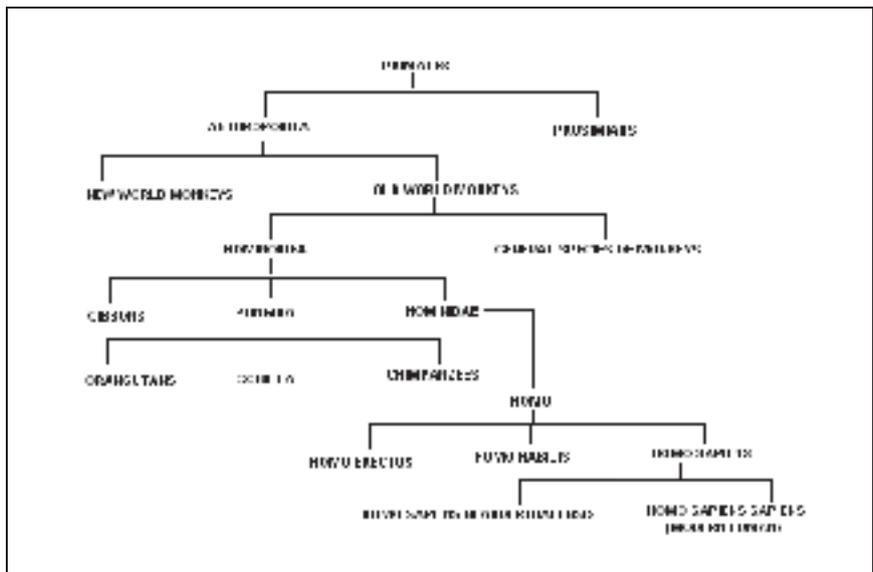


Fig. 3. Phylogenetic tree of the primates.

tree of the primate. There are two main reasons for this:

1. The complete phylogenetic tree of all the species on earth is still a distant dream as thousands of new species are being discovered every year.
2. As we also belong to the primate order, it will be a huge privilege to study the branch of the tree of life that hosts us.

The above phylogenetic tree of the primate represented in Fig. 3 will be our focal point of reference for the rest of the discourse in this paper.

### 6. Ancestor’s Tale<sup>a</sup>: Finding the Root of the Tree

It is an exciting idea to learn that all living forms on earth originated from some earliest primitive form of life-like complex biochemical compounds that was

formed some 3.5 billion years ago in the sea. Thus at some point human, a dog, crow, lizards, dinosaurs, fishes, worms and cockroaches shared the common ancestor. Suppose we want to find the common ancestor of modern human (that is us) and orangutans, which is one of the great apes. From the phylogenetic tree of the primate we find the ancestors of both the species.

From Table 1, we see that at some point of time both Human and Orangutan had a common ancestor called Hominoidea. Thus at distant past, there lived a mammal which was neither human nor orangutan but it branched in two different directions to give rise to two completely contrasting species.

Now let us find the relative evolutionary distance of human and orangutan from their common ancestor. We say that two species are closely related if they are found in same branch of the phylogenetic tree. We see that there are three transitional species in between human and Hominoidea where there is only one transitional species between orangutan and Hominoidea. Hence we can safely conclude that the extinct species Hominoidea would resemble more to

<sup>a</sup>“The Ancestor’s Tale: A Pilgrimage to the Dawn of Life” is the name of a book by Richard Dawkins.

Table 1. Common ancestors of modern human and orangutan

SPECIES	ANCESTOR1	ANCESTOR2	ANCESTOR3	ANCESTOR4
MODERN HUMAN	Homo Sapiens	Homo	Hominidae	<u>Hominoidea</u>
ORANGUTAN	Pongidae	<u>Hominoidea</u>	Catarrhini	Anthropoidea

Table 2. Human ancestors

ANCESTOR1	ANCESTOR2	ANCESTOR3	ANCESTOR4	ANCESTOR5	ANCESTOR6	ANCESTOR7
Homo Sapiens	Homo	Hominidae	Hominoidea	Catarrhini	Anthropoide	Earliest Primates

orangutan than a human being. Thus using the simple concept of “ancestor” in a phylogenetic tree we can arrive at some startling conclusion.

## 7. Unique Way of Origin

When we look at our surroundings, we are astonished to see the variety of life forms that have flourished on this planet earth. And on top of this extravaganza sits the human intelligence which is challenging every domain of the nature to reveal its card more generously. It is astonishing to believe that such intelligence originated from one cell organism only through the process of natural selection. Now a question definitely arises: is the passage of life from the single cell root to its prime achievement through zillion of transitional species is unique?

Fortunately, graph theory has a very clear and confident answer for this mysterious question and the answer is YES. This can be explained from the fact that the phylogenetic tree, being a tree by definition, does not contain multiple paths. For example, in the phylogenetic tree of the primates, the passage of life from the primates to the modern human is unique. To illustrate this point, we observe the path of transition of earliest primates that first walked this planet during Jurassic age that evolved into modern human in 65 million years (Table 2).

If we study the phylogenetic tree, we clearly see that this is the only way of coming down to Modern Human from the Earliest Primates. Thus Human evolved on earth through a unique lineage and graph theory satisfactorily asserts that.

## 8. Arguments Against Intelligent Design

The phylogenetic tree can give an excellent argument

against all those academics who believe in pseudo-science like “intelligent design”. As we move up the ladder in the phylogenetic tree, we observe less and less complex life form. As the phylogenetic tree is rooted, there must be a simplest form of life from which all other life form originated. This nullifies the Intelligent Design’s hypothesis of the existence of a designer’s hand in the process of origin of life.

## 9. Who Will Inherit this Earth from Us?

We must note that no species in the phylogenetic tree has existed forever. With passage of time, each one has given space to newer and advanced species. Thus all the species we list in the phylogenetic tree are transitional species: they are heading somewhere. Now where is the human heading for? Being a “leaf” in the evolutionary tree, currently our species do not have any child. This means that there is no species on earth that has descended from us taking along all our genetic materials. Now what is the probability that our species go through “negative evolution” and arrive at some ancestral point? In simple words, can human go through some weird metamorphosis and become an ape again? The graph theory says that it is not possible because the phylogenetic tree do not have any cycle. Hence no path exists from a descendent vertex to an ancestor vertex in an opposite direction. Thus there is no evolutionary track through which human can go back and become ape again.

Though it is true that we cannot go backward in an evolutionary tree, there is always a way forward. We are evolving to a newer species every moment as evolution is a continuous process. Definitely human will have to vacate this earth for some advanced species originating from itself in distant future, if the earth stays habitable up to that point.

## 10. Conclusion

The phylogenetic tree is the quick reference guide of the process of natural selection that is happening on this earth for the last 3.5 billion years. The use of graph theory in phylogeny not only makes the study simpler but it also gives some astounding answers to some very important questions. This amalgamation of science from two different disciplines in studying the most vital feature of earth is a concrete example of a shared world we live in. Only through such comprehensive study, we can explore newer horizons in understandings the mechanism of the “*Greatest Show on Earth*”<sup>b</sup>: *LIFE*”.

---

<sup>b</sup>“The Greatest Show on Earth: The Evidence for Evolution” is the name of a book by Richard Dawkins.

## References

- [1] E. Schrödinger, *What is Life: The Physical Aspect of the Living Cell* (The University Press, 1944).
- [2] R. Dawkins, *The Selfish Gene* (Oxford University Press, 1976).
- [3] Monroe W. Strickberger, *Evolution*, 2nd Edition (Jones and Bartlett Publishers, Inc., 1995).
- [4] K. H. Rosen, *Discreet Mathematics and Its Application* (Tata McGraw-Hill, 2012).
- [5] Wikipedia.org and other internet sites.



### R Dhiraj Kumar Deka

dhirajkdeka@gmail.com

Dhiraj Kumar Deka grew up in Assam, India in a teacher’s family. A gold-medalist and a rank holder from his state, he went to win many other awards including the Commonwealth Youth Games Essay Contest. His short stories have been included in three anthologies published by Unisun and broadcasted by All India Radio. His primary interests are cosmology and evolutionary biology.