

Practical 1: Tree and Network Representation

Sara Montese

18 September 2023

1 Phylogenetic Trees

1.1 Python script to extract the phylogenetic trees into a graph representation

```
import ete3
def parse_newick_trees_ete(filename):
    trees = []
    with open(filename, 'r') as file:
        for line in file:
            newick_str = line.strip()
            if newick_str:
                tree = ete3.PhyloTree(newick_str, format=1)
                trees.append(tree)
    return trees

input_file = "newick.tre"
trees = parse_newick_trees_ete(input_file)

print(f"there are {len(trees)} phylogenetic trees. ")
unrooted = 0
for i, tree in enumerate(trees, start=1):
    #print(tree)
    children = tree.get_children()
    if len(children) > 2:
        unrooted += 1
print(f"there are {unrooted} unrooted trees. ")
print(f"there are {len(trees) - unrooted} rooted trees. ")
```

Plot the trees (Fig. 1):

```
from Bio import Phylo
import networkx, pylab

trees = Phylo.parse(input_file, "newick")
for tree in trees:
    net = Phylo.to_networkx(tree)
    networkx.draw(net)
    pylab.show()
    print(Phylo.draw_ascii(tree))
```

1.2 How many phylogenetic trees are there?

There are 100 phylogenetic trees in the file.

1.3 Are these phylogenetic trees rooted or unrooted?

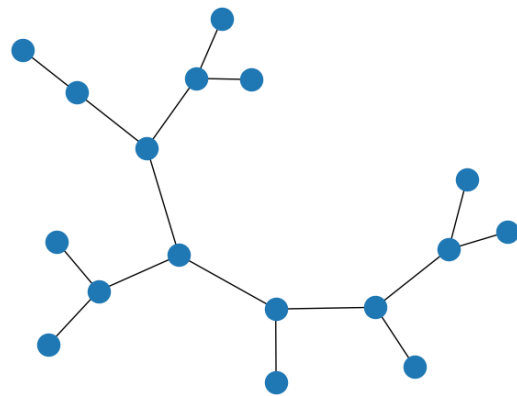
You can know if a tree is rooted or unrooted by counting the number of children of the root node. If this number is greater than 2, the tree is unrooted. In the file *newick.tre* there are 7 unrooted trees and 93 rooted trees.

2 Phylogenetic Networks

2.1 Python script to extract the phylogenetic networks into a graph representation



(a)



(b)

Figure 1: Graph Representation of a Phylogenetic Tree

```
import phylonetwork as pn
import networkx as nx
import matplotlib.pyplot as plt

unrooted = 0
networks = []
with open("../enewick.tre", 'r') as file:
    for line in file:
        enewick_line = line.strip()
        if enewick_line:
            net = pn.PhylogeneticNetwork(eNewick=enewick_line)
            phylo_net = nx.DiGraph(net)
            nodes_indegree_0 = [node for node in phylo_net.nodes() \
                                if phylo_net.in_degree(node) == 0]

            # to say if network is unrooted, check:
            #if there is more than one node of indegree 0 OR
            #if a single node of indegree 0 and if that node has more than
            #2 children:

            if len(nodes_indegree_0) >1 or \
               ( len(nodes_indegree_0) == 1 and \
                 len(list(phylo_net.successors(nodes_indegree_0[0]))) >2):
                unrooted+=1

            networks.append(phylo_net)

print(f"There are {unrooted} unrooted networks over {len(networks)}.")
Plot the networks (Fig. 2):
for i, graph in enumerate(networks, start=1):
    plt.figure(figsize=(6, 6))
    plt.title(f'Graph {i}')
    pos = nx.spring_layout(graph)
    nx.draw(graph, pos, with_labels=True, node_color='yellow')
    plt.axis('off')
    plt.show()
```

2.2 How many phylogenetic networks are there?

There are 100 phylogenetic networks.

2.3 Are these phylogenetic networks rooted or unrooted?

In the file *enewick.tre* there are 100 rooted phylogenetic networks and 0 unrooted phylogenetic networks.

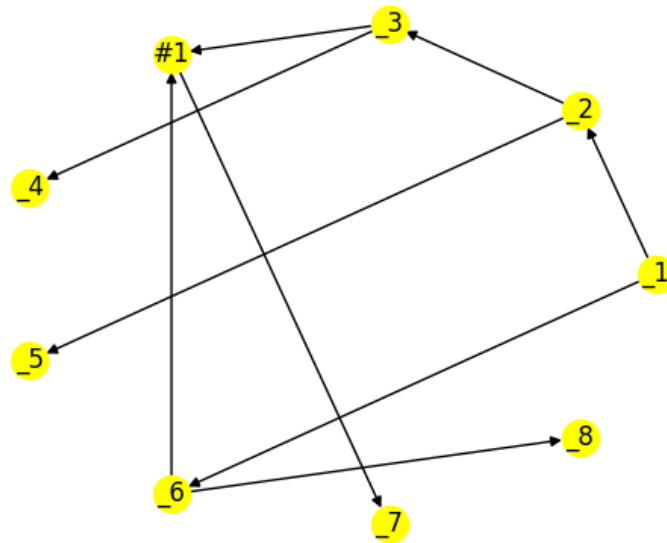


Figure 2: Graph Representation of a Phylogenetic Network