

Practical 1: Tree and Network Representation

Ximena Moure Eliya Tiram

12/09/2023 , submission deadline 18/09/2023

1. Extract The Phylogenetic Trees newick.tre File

```
1 import ete3
2 from IPython.display import display
3
4 # Function to read Newick trees from a file
5 def read_trees_from_file(filename):
6     with open(filename, "r") as file:
7         tree_strings = file.readlines()
8     return
9     [ete3.Tree(tree_string.strip(), format=8) for tree_string in tree_strings]
10
11 # Function to display trees in the IDE
12 def display_trees(trees):
13     for i, tree in enumerate(trees, start=1):
14         display(tree)
15
16
17 if __name__ == "__main__":
18     tre_file = "newick.tre"
19
20     trees = read_trees_from_file(tre_file)
21     display_trees(trees)
22
```

Listing 1: Python code ex1

2. How many phylogenetic trees are there?

There are 100 phylogenetic trees

3. Are these phylogenetic trees rooted or unrooted?

There are 74 unrooted trees: [1, 4, 5, 6, 7, 10, 11, 13, 18, 19, 20, 22, 23, 24, 25, 26, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 66, 67, 68, 69, 70, 72, 74, 75, 76, 77, 78, 79, 80, 81, 83, 84, 85, 86, 87, 90, 93, 94, 95, 96, 97, 98, 99]

There are 26 rooted trees: [2, 3, 8, 9, 12, 14, 15, 16, 17, 21, 27, 29, 40, 61, 62, 63, 64, 65, 71, 73, 82, 88, 89, 91, 92, 100]

4. Extract The Phylogenetic Trees enewick.tre File

```
1 import matplotlib.pyplot as plt
2 import networkx as nx
3
4 def read_phylogenetic_trees(path: str) -> list[str]:
5     with open(path, "r") as f:
6         tree_str = f.read()
7         return [x.replace("\n", "") + ";" for x in tree_str.split(";")][:-1]
8
9 # Constants
10 HYBRID_PREFIX = 'Node'
11 RETICULATION_NODE = '#'
12 SPECIAL_CHARS = ['(', ')', ';', RETICULATION_NODE]
13
14 def parse_enewick(s, pos=0, parent=None, G=None):
15     if G is None:
16         G = nx.Graph()
17
18     name, pos
19     = _extract_name_or_hybrid(s, pos, HYBRID_PREFIX + str(G.number_of_nodes()))
```

```

19     _add_node_and_edge(G, name, parent)
20
21     if pos < len(s) and s[pos] == '(':
22         pos = _parse_children(s, pos, name, G)
23
24     if pos < len(s) and s[pos] == RETICULATION_NODE:
25         hybrid, pos = _extract_name_or_hybrid(s, pos + 1, RETICULATION_NODE)
26         _add_node_and_edge(G, hybrid, name)
27
28     return G, pos
29
30 def _extract_name_or_hybrid(s, pos, default_name):
31     name = ''
32     while pos < len(s) and s[pos] not in SPECIAL_CHARS:
33         name += s[pos]
34         pos += 1
35     return name or default_name, pos
36
37 def _add_node_and_edge(G, node_name, parent):
38     G.add_node(node_name)
39     if parent:
40         G.add_edge(parent, node_name)
41
42 def _parse_children(s, pos, parent, G):
43     pos += 1 # Skip the opening bracket
44     while s[pos] != ')':
45         G, pos = parse_ewick(s, pos, parent, G)
46         if s[pos] == ',':
47             pos += 1
48     return pos + 1 # Skip the closing bracket
49
50 def draw_network(G):
51     pos = nx.spring_layout(G, seed=17)
52     nx.draw(G, pos, with_labels=True)
53     plt.show()
54
55 if __name__ == "__main__":
56     ewick_strings = read_phylogenetic_trees("ewick.tre")
57     for ewick_str in ewick_strings:
58         G, _ = parse_ewick(ewick_str)
59         draw_network(G)

```

Listing 2: Python code ex4

5. How many phylogenetic networks are there?

There are 100 phylogenetic networks.

6. Are these phylogenetic networks rooted or unrooted?

There are all rooted.