****
Introduction
****

MNJSTXL is a python based tool for computing species tree from a set of incongruent gene trees:

1. For individual trees within the specified input treelist, following couplet based information is to be computed:
    1. Accumulated couplet internode count
    2. Accumulated couplet based extra lineage
2. The distance matrix is formed by computing either of the following statistic between individual couplets:
    1. Average (or mode based average) couplet internode count (mode based average is computed by mean of those internode count values whose frequency is at least 50% of the mode frequency).
    2. Average couplet extra lineage count.

This couplet distance matrix is then used in NJ method, to generate the species tree.

Input source trees can be either in NEWICK format or in NEXUS format.
However, all the source trees should have identical input formats. They should be placed in a standard tree list file, according to the syntax of NEXUS or NEWICK formats. Such a tree list text file is to be provided as an input of this executable.

Output weighted supertree is generated in the NEWICK format.

****
Dependencies
****
This package is developed in Linux Systems (Ubuntu 14.04), using Python 2.7. It is tested and meant for systems having linux OS (Fedora / Ubuntu).

Development was done using Python 2.7.
Note: We plan to support Python 3 environment in some future release.

We have used the phylogenetic library Dendropy 3.12.0 (available on the link: https://pythonhosted.org/DendroPy/) for implementation.

Note: there is a new release of Dendropy 4.0 but we have used 3.12.0 for the implementation. We did not upgrade the code for Dendropy 4.0 support, and plan it as a future work.

Note: We do not support development version corresponding to Windows XP and MacOS, although that will be done in some future release.

********** User do not need to install anything.
Please follow the instructions below to execute and use the implementation **********

****
Execution
****

Upon extracting the archieve, user needs to go inside the directory 'MNJSTXL', where a standalone

executable **MNJSTXL** is provided. It is the main executable of python script.

At first, change the permissions of the executables by first going into the directory containing this executable and then writing following commands:
chmod +x MNJSTXL

****
EXAMPLE OF COMMANDS
****

(assuming the current directory is within the extracted archieve)

./MNJSTXL -I '*source_tree_input_filename*' -p '*inp_file_format*' -m '*method*'

Command descriptions:

1. Using -I command we specify the input filename (denoted by '*source_tree_input_filename*'). User need to specify the absolute or relative path of the file containing the input gene tree dataset (maintained in a text file of standard tree list, in either **nexus** or **newick** format).
2. -p option is for specifying the input tree format (as denoted by '*inp_file_format*'). If input file contains the trees in NEWICK format, then specify the option as (-p 1) (1 stands for newick). If input file contains the trees in NEXUS format, then specify the option as (-p 2) (2 stands for nexus). By default, p = 1 is set.
3. -m option is required to specify the '*method'* for species tree construction.

Its value is either 1 or 2, or 3 (default).

- m = 1 stands for using classical NJ_st method (Liu et. al. 2011) for species tree construction,
- m = 2 is for using the method MNJst, which uses mode based filtered average internode count (for individual couplets) for species tree construction. The mode based averaging uses only those internode count values whose occurrence frequencies are at least 50% of the modal frequency.
- m = 3 corresponds to the method using above mentioned mode based average internode count (for individual couplets) and also using the couplet based average extra lineage information for species tree generation.

In addition, the package contains another option:
-O '*output_file_name*'

Here, user can specify the output file name containing the derived species tree file.

If no such option is provided, our method performs the following operations:

- If m = 1, a directory “NJ_ST” is created within the same directory containing the input treelist file. Within this new created directory, one file *'outtree_newick.tre'* is created, which contains the derived species tree. Another text file named *'Complete_Desription.txt'* is created, which contains execution and timing information for the method.
- For m = 2, and m = 3, directory “M_NJ_ST” or 'M_NJ_ST_XL', respectively, is created within the same directory containing the input treelist file. Above mentioned files within the new directory are generated as per the execution.

****
For any queries, please contact
****

Sourya Bhattacharyya
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
<sourya.bhatta@gmail.com>