

ARGENET

To run ARGENET use the following command:

```
ARGENET_with_subsampling/ARGENET_with_subsampling.py -o observation_file -f output_folder -s sample_sizes -c correlations -i iterations_number -a alpha -t threads_number -h help
```

PARAMS:

- o *The path of the observation file containing sample data*
- f *The path of the folder where to save the computations*
- s *Sample sizes to use, given as list: e.g. [50,100,..]*
- c *The correlations to use, given as list: e.g. [0.5,0.6,...]*
- i *The number of iterations*
- a *The alpha value to use for the statistical test (p-value)*
- t *The number of threads to use*
- h *This tutorial*

OUTPUT:

The output of the program is saved in the output folder, which contains the additional timestamp in the folder name. In this way, no folder will be automatically deleted or overwritten. It contains a subfolder for each sample size, and each sample size contains each correlation threshold in a subfolder. At this level all files reporting the name of the samples used in each run (input of ARGENET) are reported in a file with the name perm-X-input, where X is a number that denotes an iteration. This approach was used to keep only the name of the samples to reduce the size of the output folder that otherwise, containing all input filtered matrices would carry an un-useful and repeated amount of heavy information. Each correlation threshold contains all the runs of ARGENET in subfolders and the output of the whole process with a frequency cut, from 50% to 100% with a 10% step. These files have a name like "over_threshold_setX", where X represents the number of times that the edge has to be found to be retained. The file named "over_threshold_numX" reports the length of the corresponding "over_threshold_setX" file. The folder contains also other useful information, like all the similarity among files in pairwise comparisons using the Jaccard formula, the average length of the files, the standard deviation of the length of the files, the size of the smaller file and the number of files that are empty. Moreover, the average similarity using the standard Jaccard, the subset Jaccard, the refined simple and subset and the weighted simple and subset. A file with the same name of the output folder with a "log" attached contains information about the run.

Requirements

1. Python
 - Numpy library - <http://www.numpy.org/>
2. R
 - VARBVS library
 - MASS library