Simple k-RF Metrics for Labeled DAGs

- "sK-RFmeasures.py" contains the code to compute pairwise s-k-RF or m-k-RF distances of a set of (multi-labeled) labeled DAGs stored in an input JSON file. The command line to run the code is "python3 sK-RFmeasures.py inputfile k m n". The "inputfile" refers to a JSON file that consists of the DAGs whose pairwise distances must be computed. The parameter k is an integer that is greater than or equal to zero. The parameter m is a binary integer, where we set $m = 0$ if we want to compute s-k-RF distances and we set $m = 1$ if we want to compute m-k-RF distances. The parameter n is also a binary integer, where we set $n = 0$ if we want to compute unnormalized distances and we set $n = 1$ if we want to compute normalized distances.

  A multi-labeled DAG G in the input file needs to be represented as $[A, B]$, where A is the list of nodes in G with their labels, and B is the list of edges of G. Each node $v \in V(G)$ is represented as $[v, a_1, .., a_n]$ if $\ell(v) = \{a_1, ..., a_n\}$. Note that v must be in the first position of the array. Additionally, each edge from $v \in V(G)$ to $u \in V(G)$ is represented as $[v, u]$.

- "sKfunctions.py" contains the necessary functions required to run the code in "sK-RFmeasures.py".

- "255DAGs.json" is a sample input file containing all 255 DAGs used for our clustering experiment on labeled DAGs, as described in the paper.

- "250DAGsPC_sim.json" is a sample input file containing all 250 DAGs with the same number of edges, used for the correlation analysis of the s-k-RF measures with the m-k-RF measures, the generalized Kendall-tau (GKT) distance and the Katz dissimilarity, as described in the paper.

- "250DAGsPC_dif.json" is a sample input file containing all 250 DAGs with different number of edges, used for the correlation analysis of the s-k-RF measures with the m-k-RF measures, the generalized Kendall-tau (GKT) distance and the Katz dissimilarity, as described in the paper.

- "240mutation_trees.json" is a sample input file containing all 240 multi-labeled trees used for comparing clustering performances of the k-RF measures and the simple k-RF measures.

- "DAGs_resol.json" is a sample input file containing all DAGs used to compare the s-k-RF distances with the m-k-RF distances to show how the latter refine the former.
- "Resol_one_label_trees.json" is a sample input file containing all trees used to compare the s-k-RF distances with the k-RF distances to show how the former can improve the resolution of the latter for trees with different label sets.
- "combinedtree-PC.json" is a sample input file containing all 950 trees used for the correlation analysis performed to compare the s-k-RF measures with k-RF distances, DISC$\cap$, CASet$\cap$, and GRF distances.