

TO-DAG short manual

Timed Oncogenetic Directed Acyclic Graph (TO-DAG) infers the graph of the causal dependencies and the waiting times among mutational events from cross-sectional data of genetic alterations in independent human tumour samples. TO-DAG is provided as an R script (TO-DAG.R).

Requirements

1. R $\geq 3.x$ (available at <https://www.r-project.org/>)
2. R packages: `utils`, `igraph` (`igraph` version ≥ 1)

How to run

TO-DAG can be launched from Linux/Ubuntu terminal or from Windows command prompt as follows:

```
Rscript TO-DAG.R input_file weight_threshold
```

- `input.txt` is the input file. It is a text file format of a binary matrix, whose columns report the genetic events (e.g. E1, E2, ..., E10) and the rows report the samples (e.g. S1, S2, ... S49)
- `weight_threshold` is a real number in $[0,1]$ and denotes the cut-off on the edges weight

Example:

```
Rscript TO-DAG.R ./input.txt 0.03
```

TO-DAG saves all the edges whose weight is greater than 0.03.

Outputs

TO-DAG.R execution creates three folders: "plots", "networks", "degree":

- "networks": the inferred graph in SIF (i.e. source node, target node, and edge probability) and GraphML format are stored and named "inferred_graph.txt" and "inferred_graph.graphml", respectively; GraphML format files can be edited with Cytoscape (<http://www.cytoscape.org/>); simple paths in the graphs with probability of each event expressed as its conditional probability given its predecessors ("paths_prob.RData");
- "plots": the plots of the distribution of waiting times and node degrees of the inferred graphs in PDF format, and named "waiting_times.pdf", and "degree.pdf", respectively;
- "degree": the text file format of node degree and summaries of degree distributions are stored for the whole graph and for the sub-graphs of fast, slow and medium-speed transitions (corresponding to green, red, and orange coloured arcs, respectively).

How to cite:

Lecca P, Casiraghi N and Demichelis (2015) Defining order and timing of mutations during cancer progression: the TO-DAG probabilistic graphical model. *Front. Genet.* 6:309. doi: 10.3389/fgene.2015.00309