**Orientation of the human protein-protein interaction network elucidates drug targets and cancer driver genes**

Dana Silverbush1 and Roded Sharan1\*

1The Balavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel

\*To whom correspondence should be addressed: roded@post.tau.ac.il

**GitHub Repository**

**The repository includes the following:**

1. Code
   1. A python script performing Diffuse2Direct from start to end and can be used for
      1. Small networks or pathways with a single experiment
      2. Large networks with multiple experiments.
2. Examples:
   1. Small scale single experiments from KEGG

Input:

* + 1. A network file kegg.net
    2. A sources file
    3. A terminals file

Output

* + 1. D2D scores
    2. D2D scores distribution
  1. Large scale networks with multiple experiments

Input

* + 1. A drug targets file
    2. A drug differentially expressed genes

Output

* + 1. Precision recall curves

**To use:**

Download the code and examples.

In the code you will need to decide and change:

1. Input and output paths -match it to your paths
2. For large networks the script will randomly choose 100 experiments so to minimize running time. To faithfully orient a network it is recommended to use all available experiments.
3. Choose a balancing parameter C for the classifier. In the paper we used L1 regularization where the value of the regularization balancing parameter was chosen via a nested 3-fold cross-validation in the range of to .

**Files format:**

Network file

Columns: Gene1, Gene2, confidence, is\_interaction\_directed, interaction\_type

(tab delimited)

The information in column is\_interaction\_directed is not used in the script, however the code can be extended to diffuse while considering directed interactions in the network, as described in the SI.

interaction\_type: start with TRUE for known directions, FALSE for their opposite direction and UNDIRECTED for interaction that do not possess a direction. Any other type refers to interaction with unknown directionality.

Sources file

Columns: Experiment, source

(tab delimited)

For example, an experiment can be a drug and a source will be a drug target.

An experiment can have multiple sources.

Terminals file

Columns: Experiment, terminal

(tab delimited)

For example, an experiment can be a drug and a terminal will be a gene which was observed to change its expression in response to the drug.

An experiment can have multiple terminals.