USERS MANUAL FOR DADA

WHAT IS DADA?

DADA is a network-based script implemented in MATLAB. The goal is to prioritize candidate set of genes based on their association level with a disease of interest. Network proximity calculations of the candidates with respect to the seed set are based on Random Walk with Restarts method. DADA provides several statistical adjustment models to remove the bias with respect to degree of target genes in existing algorithms based on Random Walk with Restarts simulation. Please refer to the manuscript for detailed explanations of algorithms used.

RUNNING DADA

REQUIRED ARGUMENTS:

DADA_rank_genes script requires 3 arguments:

- **network_file** : this is an ASCII file in which each line consists of two gene names and the reliability score [0,1] as the third column. If you have an unweighted network, set this score to 1 for all interactions.
    - gene_name1 gene_name2 interaction_score
- **candidate_file**: this ASCII file lists each candidate gene to be ranked, one gene per line.
    - gene_name
- **seed_file** : this ASCII file lists seed genes to be used as prior information, as well as their association score in [0,1] with the disease of interest.
    - gene_name association_score

OPTIONAL ARGUMENTS

By default, DADA uses statistical adjustment based on **centrality** and uniform hybridization based on **seed degrees.** User can set DADA settings using following optional arguments:

- ‘SIG’ argument followed by one of the options {‘CENT’, ‘CAND’ , ‘SEED’} to set the statistical adjustment model to use.
    - ‘CENT’ : statistical adjustment based on centrality scores (default)
    - ‘CAND’: statistical adjustment based on candidate degrees
    - ‘SEED’: statistical adjustment based on seed degrees

- ‘HYB argument followed by one of the {‘SEED, ‘OPT’ , ‘CAND’} options to set the uniform prioritization model to use.
    - ‘SEED’ : uniform prioritization based on average seed scores (default)
    - ‘OPT’ : uniform prioritization by using the best ranking available for each gene (optimistic)
    - ‘CAND’: uniform prioritization based on average candidate scores

Please refer to the manuscript for detailed explanations and mathematical formulations of each method listed above.

OUTPUTS

DADA script returns two sorted lists of candidate genes:

- rwr: sorted list of candidate genes using random walk with restart method
- dada: sorted list of candidate genes using DADA (with given settings)

SAMPLE RUNS

```
>> [rwr dada]=DADA_rank_genes('ppi.txt','candidates.txt', 'seedSet.txt');

loading ppi data... this operation may take a while...
5 % of the interactions read
10 % of the interactions read
15 % of the interactions read
20 % of the interactions read
25 % of the interactions read
30 % of the interactions read
35 % of the interactions read
40 % of the interactions read
45 % of the interactions read
50 % of the interactions read
55 % of the interactions read
60 % of the interactions read
65 % of the interactions read
70 % of the interactions read
75 % of the interactions read
80 % of the interactions read
85 % of the interactions read
90 % of the interactions read
95 % of the interactions read
100 % of the interactions read
ppi data loaded
candidate set loaded
```

```
seed set loaded

computing ranks...
statistical adjustment method: based on network centrality
uniform prioritization method: based on average seed degree
rankings are computed


>> [rwr dada]=DADA_rank_genes('ppi.txt','candidates.txt', 'seedSet.txt',
'SIG', ‘SEED’, 'HYB', 'CAND');

loading ppi data... this operation may take a while...
5 % of the interactions read
10 % of the interactions read
…
95 % of the interactions read
100 % of the interactions read
ppi data loaded
candidate set loaded
seed set loaded

computing ranks...
statistical adjustment method: based on seed degrees
generating random seed sets...
uniform prioritization method: based on average candidate degree
rankings are computed


>> [rwr dada] = DADA_rank_genes('ppi.txt','candidates.txt', 'seedSet.txt',
'HYB', 'CAND');

loading ppi data... this operation may take a while...
5 % of the interactions read
10 % of the interactions read
…
95 % of the interactions read
100 % of the interactions read
ppi data loaded
candidate set loaded
seed set loaded

computing ranks...
statistical adjustment method: based on network centrality
uniform prioritization method: based on average candidate degree
rankings are computed
```