How to generate/replicate Suppl. Figure 1:

Suppl. Figure 1. Relative deviation of topological measures for PPI networks.

1. Running the code

To run the code, execute the function *run_suppl_fig1* with one of these options:

- Option 1: 1 to generate item with existing results. Usage: *run_suppl_fig1(1).*
- Option 2: 2 to recreate item from original data, involving all required computations. Usage: *run_suppl_fig1(2).*

Here is an overview of the execution of each option. The execution times reported below are measured executing the code in Windows 10 Pro with 256 GB RAM, and AMD Ryzen Threadripper PRO 3995WX 64-Cores CPU with 2.70 GHz. The software environment is MATLAB 2019a.

Option 1: Total execution time negligible.
Runs *plot_suppl_fig1* to generate item with existing results located in data folder. In data folder, there is:

- original_data: contains the raw data downloaded from their source.
- statistics: contains Excel sheets of the topological measures values of the original network and topological measures values averaged over 10 reduced networks (either 10% or 50% links removal) of each organism. The subfolder “results” contains the topological measures computed for 10 reduced networks and the original network of each organism saved in Matlab.
- script: contains *plot_suppl_fig1*.

Option 2: Total execution time **~40 h**

All the results of the following scripts are stored in the directory data_replicated. Below are the different steps to implement the computation:

- In order to download the data, follow the instructions reported in the file *instructions to download the data* located inside the folder “data”.
- *create_[NAME PPI NET]* : create the adjacency matrix for each 15 PPI networks. Outputs in “matrix”. Total execution time: ~**7 min**.
- *run_link_removal_10* and *run_link_removal_50*: remove at random 10% and 50% of the total links in the network. Outputs in “sparsified_matrices”. Total execution time: ~**1 min 30**.
- *Run_compute_statistics_original, Run_compute_statistics_linkrem10* and *Run_compute_statistics_linkrem50*: computes several topological measures based on original network and perturbed networks (10% and 50% of links deleted) for various organisms. Outputs in “statistics/results”
- *Create_table_statistics_original, Create_table_statistics_linkrem10, Create_table_statistics_linkrem50*: create tables of topological measures values based on original network and perturbed networks (10% and 50% of links deleted) for various organisms. Outputs in “statistics”. Total execution time: ~**1 min**
- *plot_suppl_fig1*: Plot Suppl. Figure 1. Total execution time: **negligible time.**

Note 1: running option 2 for multiple independent times might generate results slightly different because of the random link removal process performed by the function *run_link_removal_10*.

Note 2: The section of the main code that computes the topological measures for the original networks, as well as those networks perturbed by 10% and 50% link removal, is time consuming. To minimize the computation time, it is recommended to create three copies of the main function *run_suppl_fig1*. In each copy, one of the three scripts ("Run_compute_statistics_original," "Run_compute_statistics_linkrem10," and "Run_compute_statistics_linkrem50") is designated for execution while the other two are commented out. This enables the concurrent execution of the scripts, significantly reducing the total execution time by 3 roughly. Following the completion of these three scripts, the rest of the main script can be run.