**Overview**

This project implements a network-based drug repurposing approach to identify potential antiviral drugs. The analysis uses the human protein-protein interaction (PPI) network to quantify the proximity between virus-interacting human proteins (denoted as C) and drug target proteins (T\_d).

The methodology follows the workflow described in the manuscript: *Drug repurposing network-based proximity based on PPIs*. The goal is to prioritize drugs for repurposing against viral infections based on their network proximity to virus-interacting host proteins.

## Software and Libraries

The analysis is implemented in **Python 3.9.12** and utilizes the following packages:

* **Pandas 1.4.2** – for data handling
* **NetworkX 2.7.1** – for constructing and analyzing the PPI network
* **NumPy 1.22.4** – for numerical calculations

## Input Files

1. **Virus Proteins (VIRUS.xlsx)**
   * Contains the list of human proteins interacting with the virus of interest.
   * Column: gene\_id
2. **Protein-Protein Interaction Network (Interactome\_data.xlsx)**
   * Contains human PPI data, typically two columns: Protein1 and Protein2.
3. **Drug-Target Data (Drug\_Target\_Part\_1.xlsx)**
   * Contains drug-target interactions.
   * Columns: Drug and Target

**Note:** All input files should be located in the same folder as the script.

## Methodology

1. **Network Construction**
   * Build a PPI network from Interactome\_data.xlsx using NetworkX.
   * Filter virus-interacting proteins and drug targets to ensure they exist in the PPI network.
2. **Network Proximity Calculation**
   * For each virus-interacting protein set **C** and drug target set **T\_d**, compute the **network proximity** ​ as:

where *d(c, t)* is the shortest distance between proteins *c* and *t* within the human protein-protein interactome.

1. **Permutation Testing**

* Perform **1000 random permutations** to generate null distributions of network proximity​.
* Random protein sets are selected to preserve the degree distribution of the original proteins.
* Compute Z-score:

P-values are calculated as the proportion of random distances smaller than the observed *dCT* ​.

1. **Significance Criteria**
   * Z-score threshold: < −1.5
   * P-value threshold: P < 0.05
   * Refined candidates: P < 0.001

## Python Script Workflow

**File name:** Drug\_Proximity\_Analysis.py (or run in Jupyter Notebook)

Steps:

1. Load virus proteins, PPI network, and drug-target data.
2. Construct the PPI graph using NetworkX.
3. Filter out genes not present in the network.
4. Compute **network proximity (d\_CT)** for each drug.
5. Perform **degree-preserving permutation tests** to calculate Z-scores and P-values.
6. Combine and save results to New1.xlsx.
7. Optionally, the script uses ThreadPoolExecutor for parallel processing of drug-target files.