**Network-Based Separation and Mean Shortest Distance Calculation**

This repository provides Python scripts to calculate **network-based distance** and **network separation** between two sets of genes (or proteins) on a given interaction network.

It includes automated scripts to run calculations across multiple pairs of gene sets.

## Repository Structure

* separation.py – Core script to calculate mean shortest distances and network separation between two gene sets.
* automated\_separation.py – Script to run separation.py for multiple gene set pairs automatically.
* interactome.tsv–Protein–protein interaction (PPI) network file in tab-separated format (gene1, gene2).
* \*.txt – Text files with virus/gene sets (one gene per line) for input to the scripts.
* output/ – Folder where results for all pairwise calculations are stored.

## Requirements

* Python 2.7 or 3.x
* Libraries:
  + networkx
  + numpy

Install required libraries via pip if needed:

pip install networkx numpy

Usage

1. Calculate separation for a single pair of gene sets

python separation.py --g1 VirusA.txt --g2 VirusB.txt -o output.txt

**Parameters:**

* --g1 : File containing gene set A (one gene per line)
* --g2 : File containing gene set B (one gene per line)
* -n : Optional, network file in tab-separated edge list format (default: interactome.tsv)
* -o : Output file to save results (default: separation\_results.txt)

**Output includes:**

* Mean shortest distance within each gene set (d\_A, d\_B)
* Mean shortest distance between sets (d\_AB)
* Network separation (s\_AB)

### 2. Automate separation calculations for multiple pairs

Use automated\_separation.py to run separation analysis for multiple virus pairs automatically:

python automated\_separation.py

Features:

 This script scans the current folder for .txt gene set files.

 It calculates separation for all possible pairs and stores results in the output/ folder.