

TP : Random walk with restart on networks using NetworkX

```
In [2]: import networkx as nx
import csv
import os

def rwr(network_file, weights_file):
    g = nx.Graph()

    with open(network_file, "r") as f:
        csv_reader = csv.reader(f, delimiter="\t")
        for row in csv_reader:
            g.add_edge(row[0], row[1])

    weight_dict = {}

    with open(weights_file, "r") as f:
        csv_reader = csv.reader(f, delimiter="\t")
        for name, weight in csv_reader:
            weight_dict[name] = float(weight)

    pagerank_results = nx.pagerank(g, personalization=weight_dict, alpha=0.9)

    return pagerank_results

if __name__ == "__main__":
    rwr = "."
    network_file = os.path.join(rwr, "data\\PPI_HiUnion_LitBM_APIID_gene_names_1
    weights_file = os.path.join(rwr, "data\\tAML_P3_SNV_polyphen_15390_nodes.ts
    # I don't use the output file as it is not recommended in jupyter notebooks.
    res = rwr(network_file, weights_file)
```

Use cases of the RWR or Personalised PageRank: Outside of the scientific area, these algorithms are commonly used in social networks to recommend new friends to users, or in e-commerce and streaming services in order to recommend products based on a user's previous purchases and browsing history.

Uses of the techniques in computational biology: They are used in the scientific and clinical areas to predict drug-drug combinations as well as drug-target interactions. They can also be applied to gene regulatory network inference. In the specific context of Protein-Protein interaction networks, they are used to predict protein complexes.

SNV data: = single nucleotide variance data, referring to a DNA sequence alteration of one single nucleotide.

Polyphen score: It represents the probability that an aminoacid substitution is damaging to the organism. Some aminoacids have similar chemical properties or are located in a part of the protein where the substitution goes "unnoticed". If a substitution (due to DNA

sequence alteration) happens between aminoacids that have very different chemical properties or are crucial to the functioning of the protein, it can create a lot of damage. Although in many cases, there are mechanisms to detect wrong proteins and delete them before they can actually harm the organism. The score ranges from 0 to 1, where 0 means the substitution can be tolerated and 1 means it is harmful.

Damping factor alpha: It defines the probability of the random walk restarting from the/a core node at each step of the algorithm. It is, again, a value between 0 and 1. A high alpha value can be useful to focus on one or multiple specific nodes of the graph, as it increases the probability for the walk to restart back from there. on the other hand, a low alpha value allows to explore the network very broadly and not get stuck in one and the same area.

Ten most interesting proteins ranked in order:

```
In [9]: for i in sorted(res.items(), key=lambda item: item[1], reverse=True)[:10]:  
        print(i)
```

```
('NTN1', 0.004471455725879966)  
('SDCBP', 0.0034897788409275635)  
('APP', 0.003487723047085233)  
('OPRL1', 0.0034851871389207706)  
('RHOD', 0.0033874984940111237)  
('MEOX2', 0.003337762250098252)  
('TTN', 0.0032002033356381384)  
('EXOC1', 0.0031301135552706783)  
('PPP3CA', 0.0030979654554612267)  
('NSD2', 0.002991751661900752)
```