Step-by-step Time-Complexity Derivation for DEviRank Path Aggregation (L_{max}=3, Shortest Paths)

This note provides a step-by-step derivation of the per-drug time complexity for the DEviRank scoring term that aggregates contributions over shortest paths of bounded length (maximum length L = 3) in a weighted protein–protein interaction (PPI) network.

1 Notation and Assumptions

Let G = (V, E) be the PPI graph with |V| nodes and |E| edges.
Let S be the set of drug target proteins and T the set of disease-associated proteins.
Let Δ denote the maximum node degree in G, and deg(v) the degree of node v.
We enumerate only sample paths (no repeated nodes) with length at most L = 3.
We assume O(1) lookup for membership in T.

We analyze the dominant path-aggregation term used to score a drug d:

$$w_d(d) = \sum_{s \in S} DGI_t * \left(\sum_{\substack{t \epsilon T, \\ p \epsilon P[s \to t]}} \left(\prod_{e \in p} PPI_e * \left(1 + \sum_{n \in N[t]} PPI_n \right) \right) \right)$$

where $N[t]$ denotes the set of edges connecting protein t to its neighboring proteins, $P[s \to \mathrm{t}]$ denotes the set of paths from s to t, e represents an edge along path p, and PPI_e is the confidence score of interaction e.

The per-path work includes multiplying up to 3 edge weights ($PPI_{e)}$ and a constant number of scalar multiplications/additions. Thus, the cost per enumerated path is O(1) when L is a small constant.

2 Step A: Compute Disease Weights $w_t(\mathrm{t})$

The disease weight for each t ∈ T is:

$$w_t(\mathrm{t}) = \sum_{n \in N[t]} PPI_n$$

Computing $w_t(\mathrm{t})$ for a single t visits all incident edges once, costing O(deg(t)). Therefore, computing all disease weights costs:

$$T_{total-disease-targets} = O(\sum_{t \in T} \deg(t))$$

3 Step B: Bound the Number of Shortest Paths from a Fixed s (L_{max}=3)

We bound the number of candidate simple paths starting at a fixed drug target node s, for lengths 0, 1, 2, and 3.

Length 0 (zero edge): disease target node t is the drug target s.

$$\#paths_{len0(s)} = 1$$

Length 1 (one edge): the number of neighbors is at most deg(s).

$$\#paths_{len1}(s) \leq deg(s)$$

Length 2 (two edges): after choosing the first neighbor, the second step can go to at most (Δ − 1) new nodes (we cannot return to s if paths are simple).

$$\#paths_{len2}(s) \leq deg(s) \cdot (\Delta - 1)$$

Length 3 (three edges): similarly, after two steps, the third step has at most (Δ − 1) choices again to avoid revisiting the previous node(s), yielding:

$$\#paths_{len3}(s) \leq deg(s) \cdot (\Delta - 1)^2$$

Thus, the total number of simple paths from s of length at most 3 is bounded by:

$$\#paths_{len\leq 3}(s) \leq 1 + deg(s) + deg(s)(\Delta - 1) + deg(s)(\Delta - 1)^2$$
$$\Rightarrow T_{paths}(s) = O(\deg(s)\,\Delta^2)$$

Where Δ denotes the maximum degree of the PPI network.

4 Step C: Total Path Enumeration Cost Over All Drug Targets

Because the work per enumerated path is O(1) for L_{max} = 3, the time to enumerate and score all bounded-length paths originating from s is:

$$T_{paths}(s) = O(\deg(s)\,\Delta^2)$$

Summing over all s ∈ S gives the per-drug path aggregation cost:

$$T_{total\ path\ from\ S\ to\ T} = O(\sum_{s\in S} \deg(s)\Delta^2)$$

A simple worst-case bound follows from deg(s) ≤ Δ for all s:

$$T_{total\ path\ from\ S\ to\ T} = O(|S|\Delta^3)\ \ (worst\ case)$$

5 Final Per-Drug Time Complexity

Computations from steps 2,3 and 4 yields:

$$T_{total} = O(\sum_{t\in T} \deg(t) + \sum_{s\in S} \deg(s)\Delta^2)$$

and in the worst case:

$$T_{total} = O\left(\sum_{t\in T} \deg(t) + |S|\Delta^3\right)$$

6 Notes on Practical Implementation

1) The bound assumes path enumeration is performed from each s and endpoints are checked against T in O(1).

2) Restricting to shortest paths avoids repeated-node walks, which would otherwise increase path counts.

3) Since drugs are independent, total runtime scales linearly with the number of drugs and is embarrassingly parallel across drugs.