Time Complexity Analysis for the Symptom-Disease Prime Mapping Algorithm

Overview

The algorithm consists of three main components:

1. **Prime Generation:** Generating the first S prime numbers to map to unique symptoms.
2. **Mapping and Matching:** Computing the square-free (SQF) integer for each disease and matching user-input symptom primes against these SQFs.
3. **Graph Visualization:** Plotting a bipartite graph with symptom primes on the left and disease SQFs on the right.

1. Prime Generation

For S unique symptoms, we need to generate S prime numbers. The helper function `is_prime` checks if a number is prime in $\mathcal{O}(\sqrt{P})$, where P is the candidate number. Therefore, the overall complexity for generating S primes is:

$$\mathcal{O}\Big(S \times \sqrt{P}\Big)$$

Since S is typically small (e.g., 10–20), this step is efficient in practice.

2. Mapping and Matching (Graph Construction)

Mapping

Each disease is associated with a set of symptoms. The disease's SQF integer is computed as the product of the primes corresponding to symptoms where the indicator is 1. This operation requires iterating over the symptoms for each disease, which is done in one pass.

Matching

Let:

- M be the number of user-selected symptom primes.
- D be the number of diseases.

For each disease, we check if each of the M primes divides the disease's SQF. The matching phase therefore has a time complexity of:

$$\mathcal{O}(M \times D)$$

Given that M and D are very small in typical cases (e.g., $M \approx 2$–5, $D \approx 5$), this is very efficient.

3. Graph Visualization

The visualization process also iterates over the $M \times D$ pairs to draw edges (i.e., check divisibility for plotting). Its time complexity is:

$$\mathcal{O}(M \times D)$$

Again, this is negligible for small values of M and D.

Summary of Overall Complexity

The overall worst-case time complexity is the sum of the prime generation and the matching phases:

$$\textbf{Total Time Complexity} = \mathcal{O}\left(S \times \sqrt{P}\right) + \mathcal{O}(M \times D)$$

Worst-case: If the number of symptoms S or diseases D were to scale up significantly, the $\mathcal{O}(M \times D)$ term could dominate. However, in our intended use case, both S and D are very small.

Regular use case: With $S \approx 10$, $M \approx 2$–5, and $D \approx 5$, all operations are effectively constant time relative to the problem size, and the overall computation is extremely efficient.