

# Exact Subgraph Search: Subgraph Isomorphism Problem

---

## Introduction (from Wikipedia)

---

The **Subgraph Isomorphism** problem is a computational task in which two graphs  $G$  and  $H$  are given as input, and one must determine whether  $G$  contains a subgraph that is *isomorphic* to  $H$ . It can be shown that the decision problem is NP-Complete ( $\leq^p$  MAX-CLIQUE/Hamiltonian Cycle Problem) Here what we exactly need to do during queries is probably **Subgraph Matching**, which is a variation of the decision problem with more focus on *finding such a subgraph*.

## Formal Definition

(<http://theory.stanford.edu/~virgi/cs267/lecture1.pdf>)

---

Let  $H = (V_H, E_H)$  and  $G = (V, E)$  be graphs. A subgraph isomorphism from  $H$  to  $G$  is a function  $f: V_H \mapsto V$  such that if  $(u, v) \in E_H$ , then  $(f(u), f(v)) \in E$ .  $f$  is an **induced subgraph isomorphism** if in addition if  $(u, v) \notin E_H$ , then  $(f(u), f(v)) \notin E$ .

The point of subgraph isomorphism is essentially a vertex mapping that guarantees there's a corresponding edge mapping for every edge in the subgraph structure input  $H$ .

## Algorithms (Wikipedia & StackOverflow)

---

- Ullmann's Algorithm: it's basically a DFS traversal of the graph with a pruning subprocedure done by making sure for vertex  $i_H \in H$ , if it is matched by a vertex  $i$  of graph  $G$ , then we have that for every vertex  $v_H$  that is adjacent to vertex  $i_H$  in the subgraph  $H$ , there must exist a vertex  $v$  that is adjacent to node  $i$  of graph  $G$ , which eliminates the number of trials compared to the primitive DFS brute-force search. (<https://stackoverflow.com/questions/13537716/how-to-partially-compare-two-graphs/13537776#13537776>)
- Cordella's VF2 Algorithm (<https://ieeexplore.ieee.org/document/1323804>), a working example offered from StackOverflow (<https://stackoverflow.com/questions/8176298/vf2-algorithm-steps-with-example>) and a promising GitHub Repository for a simple implementation of the VF2 Algorithm with similar data representation. (<https://github.com/maaars/VF2-java>)
- **Bonnici**: proposed a subgraph isomorphism algorithm for biochemical data ([https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-S7-](https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-S7-2019-02-20)

S13), which incorporates a novel search strategy (*variable ordering* of the pattern graph vertices in the branches of the search tree, for instance, a variable ordering may begin with a pattern vertex having the highest degree or having the most uncommon label in the target graph, etc. The search strategy can thus narrow down the search space for the graph search).