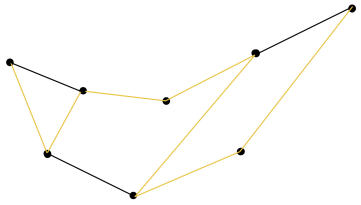
FINDING HAMILTONIAN PATHS IS NP-COMPLETE

In this note, we show that the problem whether a graph has a Hamiltonian path, is **NP**-complete. We start by recalling some definitions.

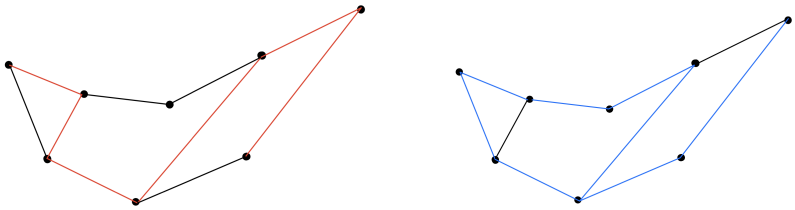
Definition 1. Let G be a graph. A path in G is called **simple** if it crosses every vertex at most once.

Definition 2. Let G be a graph. A path in G is called **Hamiltonian** if it crosses every vertex exactly once.

Note that every Hamiltonian path must be simple. For example, the yellow path in the graph below is Hamiltonian.



However, while the red path below is simple, it is not Hamiltonian, because it doesn't cross every vertex. The blue path below is neither Hamiltonian nor simple.



In the remainder of this note, we study the decision problem HamPath:

Given a graph G, does G have a Hamiltonian path?

Our goal is to show that HamPath is **NP**-complete, and for that, we must show the following

- HamPath $\in$ **NP**.
- HamPath is **NP**-hard.

Let us start by showing the first.

Proposition 3. HamPath $\in$ **NP**.

Proof. A certificate for this problem is a path p. To check whether p is Hamiltonian, we must count how often every vertex is crossed by this path. This can be done by going through the path and keeping track of how often we see every vertex. Since this can be done in polynomial time, we can conclude that HamPath $\in$ **NP**. □

To finish the proof of **NP**-completeness, it remains to show that HamPath is **NP**-hard. We prove this by constructing a reduction 3CNF $\leq_P$ HamPath. Specifically, this means that for every formula φ in 3-conjunctive normal form, we

must construct a graph G_φ such that G_φ has a Hamiltonian path if and only if φ is satisfiable.

Before we do so, let us start by thinking what it means for a formula φ in 3-conjunctive normal form to be satisfiable. By definition, φ is of the shape $\varphi = \bigwedge_{i=1}^{n} \varphi_i$ where each φ_i is a disjunction $\psi_{i,1} \vee \psi_{i,2} \vee \psi_{i,3}$ of literals $\psi_{i,j}$. If φ is satisfiable, then there is a model m for which we have $m(\varphi) = 1$. For such models, the following must hold.

Proposition 4. *For each conjunct φ_i, we either have $m(\psi_{i,1}) = 1$ or $m(\psi_{i,2}) = 1$ or $m(\psi_{i,3}) = 1$.*

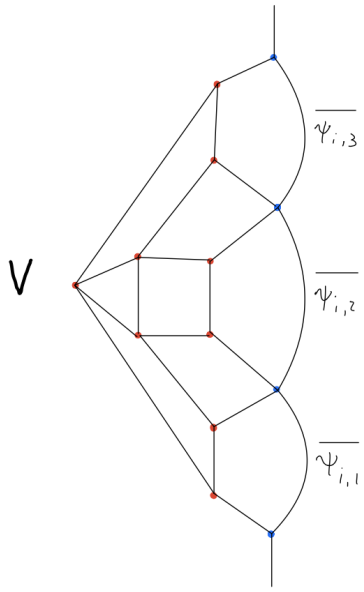
Proof. Note that we have $m(\bigwedge_{i=1}^{n} \varphi_i) = m(\varphi) = 1$. Since the left hand side is a conjuction, we must have $m(\varphi_i) = 1$ for every i. Next observe that we have $m(\psi_{i,1} \vee \psi_{i,2} \vee \psi_{i,3}) = m(\varphi_i) = 1$. A disjunction can only be evaluated to true if one of the disjuncts is evaluated to true. Hence, either $m(\psi_{i,1}) = 1$ or $m(\psi_{i,2}) = 1$ or $m(\psi_{i,3}) = 1$, which is precisely what we wanted to prove. □

This already gives an idea of what the desired graph G_φ should look like: a Hamiltonian path through G_φ must pick out a $\psi_{i,j}$ for every φ_i. To define G_φ like that, we start by adding vertices and edges for every conjunct φ_i:

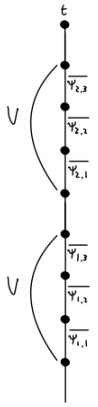
Going through the edge $\overline{\psi_{i,j}}$ means that in the model that will correspond to this graph, the literal $\psi_{i,j}$ is **false**. We do this for every conjunct and they are connected in serial. In addition, we add node t with an edge to it from the last node of the last conjunct. For a formula with 2 conjuncts, this looks as follows

However, this graph always has a Hamiltonian path, and we must do more to guarantee the graph has the desired property. First of all, we add vertices and

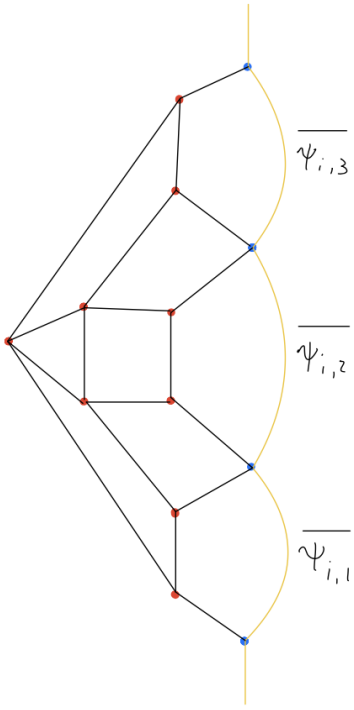
edges in such a way that a Hamiltonian path in G_φ is unable to cross all the edges $\overline{\psi_{i,1}}$, $\overline{\psi_{i,2}}$, and $\overline{\psi_{i,3}}$. For every conjunct φ_i we add the following



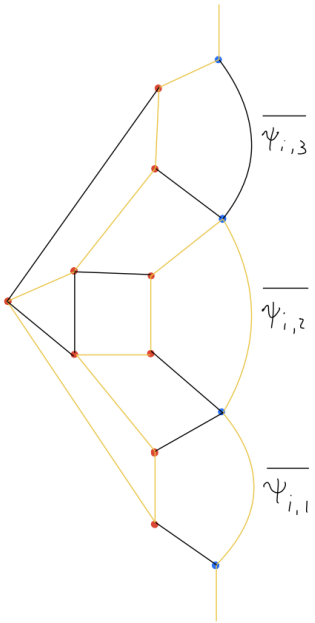
We call this part V, because we use it to guarantee that the formula is valid. Since we do this for every conjunct, the resulting graph would look as follow where the V refers to the part we discussed before.



Note that in the remainder of this construction, we are not going to add any edges with endpoints in one of the red vertices. If the resulting graph has a Hamiltonian path, then that path is unable to cross all the edges $\overline{\psi_{i,1}}$, $\overline{\psi_{i,2}}$, and $\overline{\psi_{i,3}}$. You can see this from the following pictures. First of all, we show why no Hamiltonian path is able to cross all these edges.



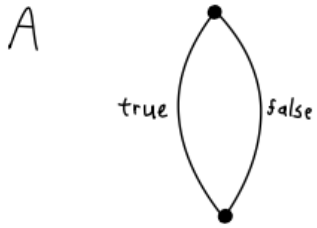
Since there are no additional edges going to the red nodes, it is impossible to visit any of those nodes, because doing so, would require the path to cross the blue vertices twice. However, it is possible to construct a Hamiltonian path through this part of the graph by avoiding at least one of the edges $\overline{\psi_{i,1}}$, $\overline{\psi_{i,2}}$, and $\overline{\psi_{i,3}}$. For example, if we do not cross $\overline{\psi_{i,3}}$, then we can consider the following path:



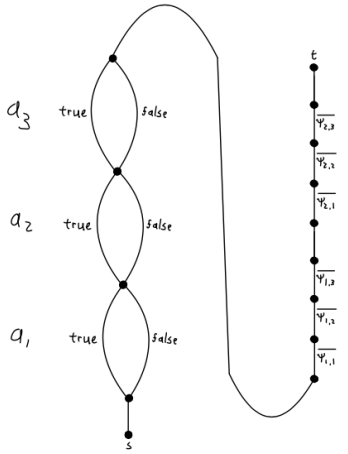
Exercise. Draw Hamiltonian paths for the case where you avoid only $\overline{\psi_{i,1}}$ and for the case where you avoid both $\overline{\psi_{i,1}}$ and $\overline{\psi_{i,3}}$.

Up to now, we have constructed a graph such that a Hamiltonian path through it picks out a $\psi_{i,j}$ from each φ_i. However, this does yet not give rise to a model: the choices must be coherent in some way. For example, if there is a conjunct $a \vee b \vee c$ and a conjunct $\neg a \vee \neg b \vee \neg c$, then it might pick out a from the first and $\neg a$ from the second. Since in models m we cannot have both $m(a) = m(\neg a) = 1$, we did not add enough yet to guarantee that this construction gives rise to a reduction.

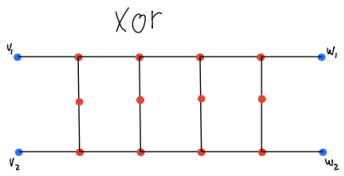
For that reason, we need to add more vertices and edges to the graph. We look at the following component



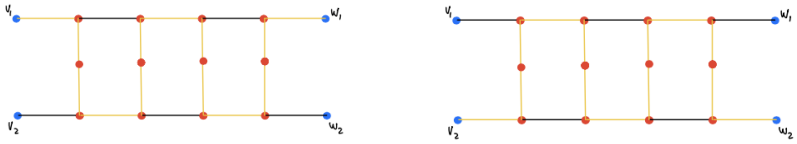
If this is a subgraph of some graph, then a Hamiltonian path is unable to cross both **True** and **False**. We add these parts in serial, and we add a node s, which we connect to the first node of the added component. We also connect all of this to the graph discussed before arising from the conjuncts. For a formula with two conjuncts and three atoms, we obtain the following graph (where we leave out the V-components).



The next component we need to consider, functions like the xor-operation from logic. More specifically, this component connects two edges, and in the resulting graph, a Hamiltonian path can only cross one of the two edges. Suppose, we have two edges $e_1 = \{v_1, w_1\}$ and $e_2 = \{v_2, w_2\}$. Consider the following graph:



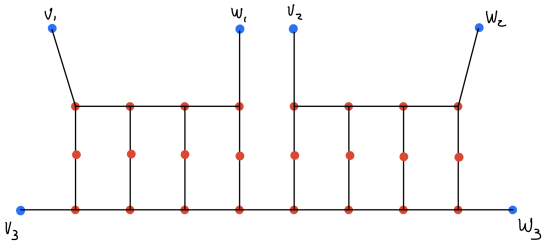
We call this graph an X-component. If this graph is a subgraph such that the red vertices are not connected to any other vertices, then there are at most two possibilities for a Hamiltonian path. These are given below



If we already have a graph G with two edges $e_1 = \{v_1, w_1\}$ and $e_2 = \{v_2, w_2\}$, then we can "add" this component in the following way

- Remove the edges e_1 and e_2
- Add the red vertices
- Add the edges from the graph using the v_1, w_1, v_2, and w_2.

Note that any number of X-components can share an edge. For example, if we have edges $e_1 = \{v_1, w_1\}, e_2 = \{v_2, w_2\}$, and $e_3 = \{v_3, w_3\}$, then we use the following graph.



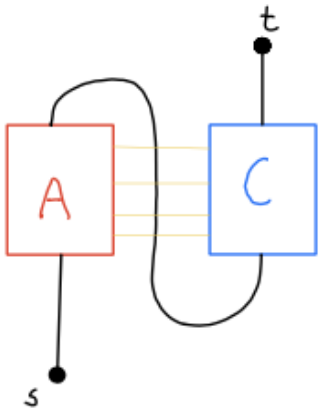
Now we add the following every literal $\psi_{i,j}$

- If $\psi_{i,j}$ is an atom a_k, then we add an X-component between $\overline{\psi_{i,j}}$ and the **True** edge of a_k.
- If $\psi_{i,j}$ is a negation $\neg a_k$, then we add an X-component between $\overline{\psi_{i,j}}$ and the **False** edge of a_k.

We call the resulting graph G_φ. Let us briefly recap what it looks like

- We have a node s.
- From that node s, we have one edge going to a part A, where we added a component A for each atom a_i.
- We have a part C, where we added a component V for each conjunct φ_i
- For each literal $\psi_{i,j}$, we added a component X between $\psi_{i,j}$ and an edge in A depending on whether this literal is a negation or not.
- We have a node t and an edge from the component V.

Schematically, G_φ looks as follows



where the red part represents the atoms and the blue part represents the conjuncts. The yellow part connects the blue and red part in such a way that the edges of representing literals are connected to the correct edge in the part of the atoms.

Now we show that this gives rise to a reduction.

Lemma 5. *If G_φ has a Hamiltonian path, then φ is satisfiable.*

Proof. Suppose, G_φ has a Hamiltonian path p. This path must cross the vertex s, and go through the edge from s to component A. In A, this path goes through either **True** or **False** for every atom a, and this gives rise to a model m. To show that $m(\varphi) = 1$, we must show that $m(\varphi_i) = 1$ for every i. More specifically, p should not cross the edge $\overline{\psi_{i,j}}$ for every i and j. We observed before that this must be the case for every Hamiltonian path. □

Lemma 6. *If φ is satisfiable, then G_φ has a Hamiltonian path.*

Proof. Let m be a model such that $m(\varphi) = 1$. We construct the following path p

- We start at s and we go through A
- For every atom a, we cross **True** if $m(a) = 1$ and we cross **False** if $m(a) = 0$. This way, we cross all nodes in A.
- Next we go to C. For each literal $\psi_{i,j}$, we cross $\overline{\psi_{i,j}}$ only if $m(\psi_{i,j}) = 0$.

Note that there also are X-components through which the path must go. Each of these X-components is added between $\overline{\psi_{i,j}}$ and e, where e is either **True** or **False**. By construction, p crosses $\overline{\psi_{i,j}}$ only if $m(\psi_{i,j}) = 0$, and $m(e)$ only if $m(\psi_{i,j}) = 1$. This allows us to modify p in such a way that all vertices in the X-components are crossed. Since all vertices are crossed exactly once, we can conclude that G_φ has a Hamiltonian path. □

Theorem 7. HamPath *is* **NP***-complete.*

Proof. To prove this theorem, we need to show that HamPath $\in$ **NP** and that HamPath is **NP**-hard. In Proposition 3, we showed the former. For the latter, note that we constructed a reduction from 3CNF to HamPath, and we proved the correctness of this in Lemmas 5 and 6. In addition, the amount of vertices and edges in G_φ depends polynomially on the size of φ, so this construction can be done in polynomial time. Hence, HamPath is indeed **NP**-complete. □