AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

22BIO211: Intelligence of Biological Systems - 2

STRING RECONSTRUCTION USING HAMILTONIAN PATH

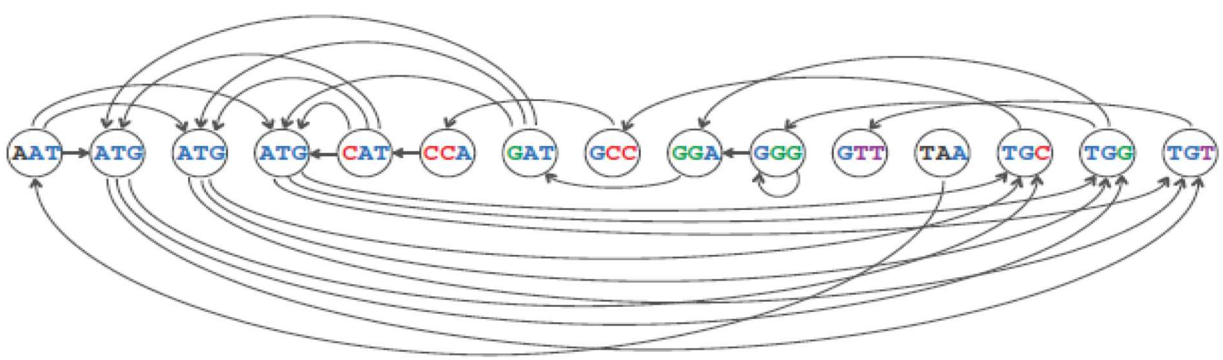
Dr. Manjusha Nair M
Amrita School of Computing, Amritapuri

Email : manjushanair@am.amrita.edu
Contact No: 9447745519

Overlap Graph Problem

Overlap Graph Problem:
Construct the overlap graph of a collection of k-mers.

Input: A collection *Patterns* of *k*-mers.
Output: The overlap graph OVERLAP(*Patterns*).



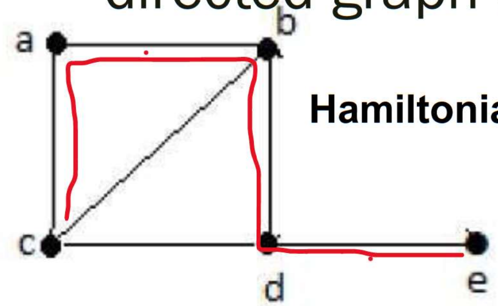

```
TAA
 AAT
  ATG
   TGC
    GCC
     CCA
      CAT
       ATG
        TGG
         GGG
          GGA
           GAT
            ATG
             TGT
              GTT
TAATGCCATGGGATGTT
```

String reconstruction using Overlap Graph and Hamiltonian Paths

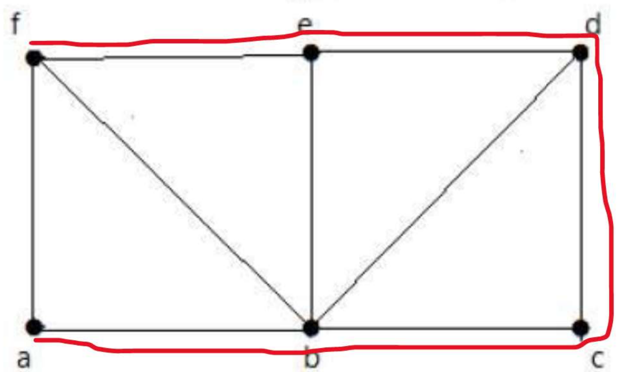
- To solve the String Reconstruction Problem, we are looking for a path in the overlap graph that visits every node exactly once.
  - *That is , a Hamiltonian path through an overlap graph*

Hamiltonian path

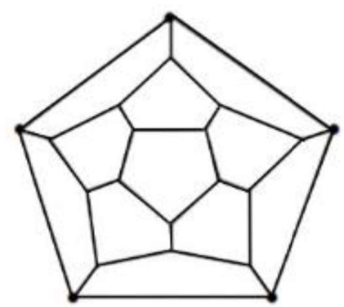
- **Hamiltonian path** is a path in an undirected or directed graph that visits each vertex exactly once.

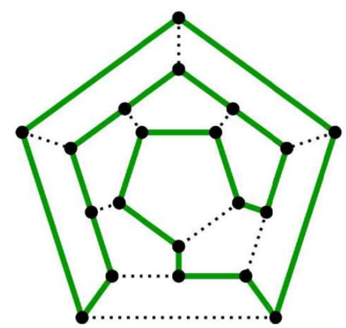


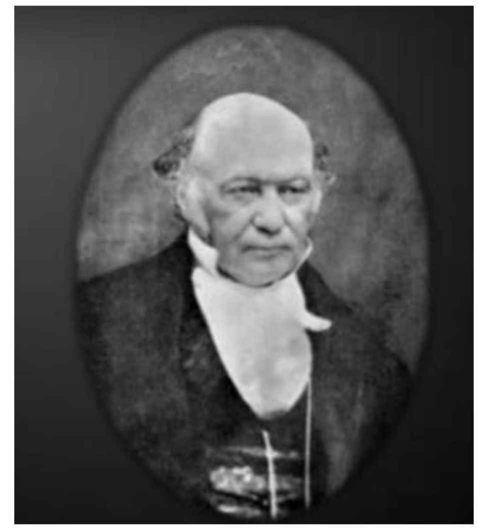
Hamiltonian Path: e-d-b-a-c



Hamiltonian Path : a-b-c-d-e-f





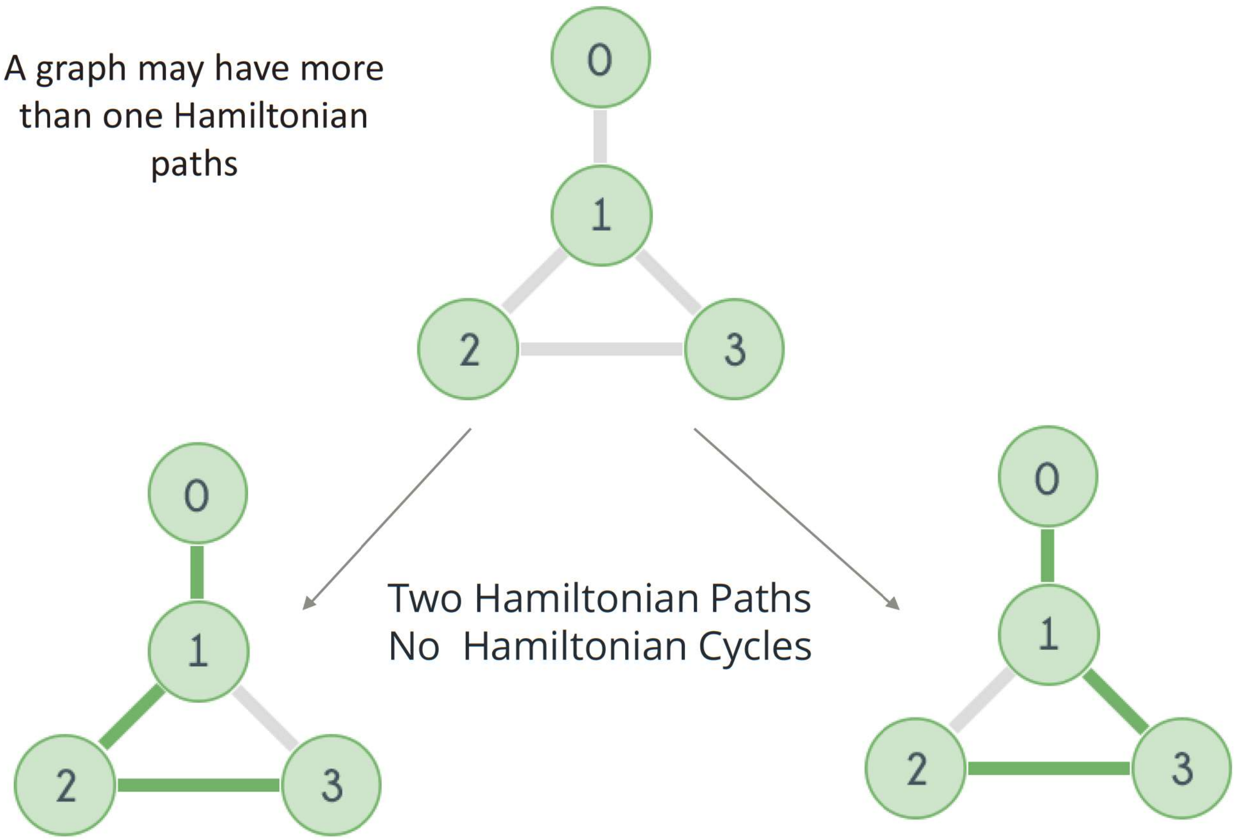


William Rowan Hamilton,1857

Icosian game

Hamiltonian : Examples

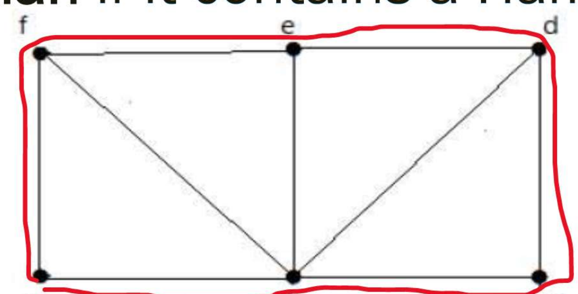
A graph may have more than one Hamiltonian paths



Two Hamiltonian Paths
No Hamiltonian Cycles

Hamiltonian cycle

- **Hamiltonian cycle** (or **Hamiltonian circuit**) is a Hamiltonian path that is a cycle.
  - *Visits each vertex exactly once*
  - *And returns to the starting vertex*
- Any Hamiltonian cycle can be converted to a Hamiltonian path by removing one of its edges
- A graph **is Hamiltonian** if it contains a Hamiltonian cycle.

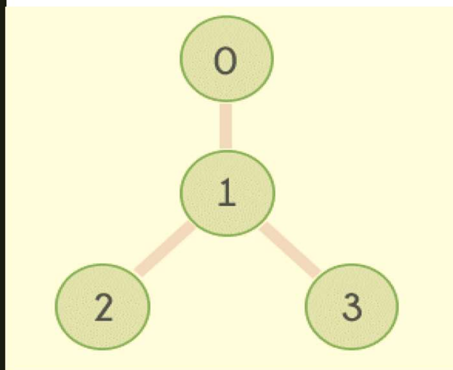


Hamiltonian Cycle : a-b-c-d-e-f-a

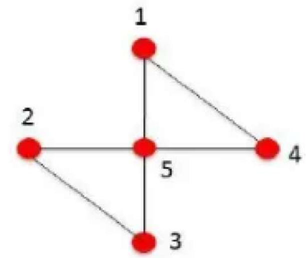
Hamiltonian cycle

- The number of different Hamiltonian cycles in a complete undirected graph on n vertices is $(n-1)!/2$
- The number of different Hamiltonian cycles in a complete directed graph on n vertices is $(n-1)!$

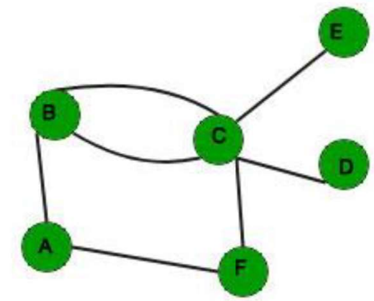
Non-Hamiltonian Graphs



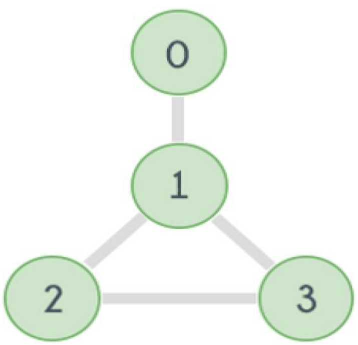
No Hamiltonian Path



No Hamiltonian cycle

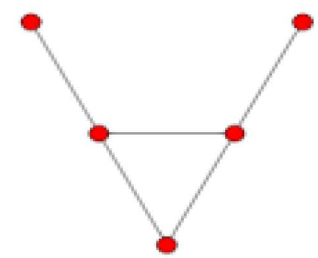


No Hamiltonian Path



Hamiltonian Path exist,
No Hamiltonian Cycle

Hamiltonian Path exist,
No Hamiltonian Cycle



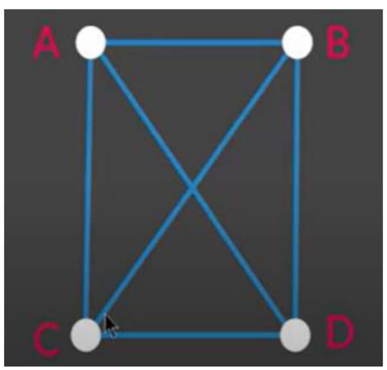
Hamiltonian Path exist,
No Hamiltonian Cycle

Check whether a Graph is Hamiltonian

- There is no simple necessary and sufficient criteria to determine if there are any Hamiltonian paths or circuits in a graph.
- But there are certain criteria which rule out the existence of a Hamiltonian circuit in a graph
  - *such as- if there is a vertex of degree one in a graph then it is impossible for it to have a Hamiltonian circuit.*
- There are certain theorems which give sufficient but not necessary conditions for the existence of Hamiltonian graphs.

Check whether a Graph is Hamiltonian

- **Dirac's Theorem (1952)-**
  - A Graph G with *n* vertices (*n* >= 3) is Hamiltonian if every vertex has degree *n*/2 or greater.
- **Ore's Theorem (1952)-**
  - *If for any pair of non-adjacent vertices u and v, deg(u) + deg(v) >= n, then G contains a Hamilton circuit*.



No of Vertices , n = 4
n/2 = 2

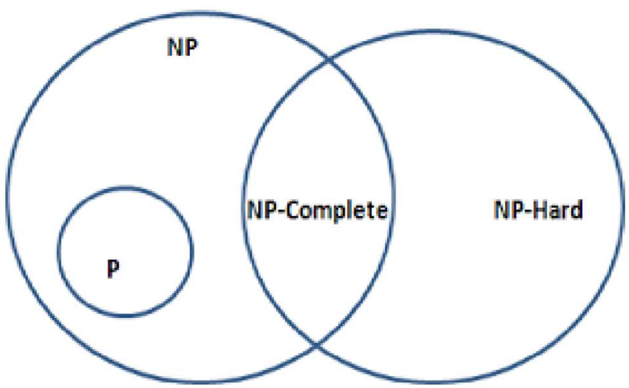
Vertices	Degree
A	3
B	3
C	3
D	3

The Graph is Hamiltonial as per Dirac's Theorem

Check whether s Graph is Hamiltonian

- There are known necessary conditions needed for a graph to be Hamiltonian. But,no necessary and sufficient (if and only if) is known.
- There are certain graphs which have a Hamiltonian circuit but do not follow the conditions in the above-mentioned theorem
- The problem on deciding whether a graph is Hamiltonian or not is an **NP-complete** problem (no algorithm exists that runs in polynomial time).

Complexity Classes : P, NP,NP Hard , NP Complete,

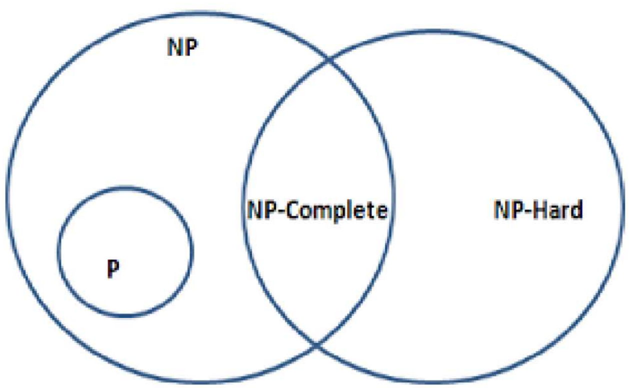


P -Polynomial time - These problems can be solved in polynomial time, i.e.; O(n), O(n^2), or O(n^k), , where k is a constant.
There exist a deterministic algorithm.

NP - Non deterministic polynomial time - “yes/no” instances of these problems can be solved in polynomial time.
There exist a non deterministic algorithm.

$P \subseteq NP$

Complexity Classes : P, NP,NP Hard , NP Complete,



NP Hard - These problems need not have any bound on their running time and need not be in NP, meaning the solutions need not be verifiable in polynomial time.

NP Complete - A problem is NP-complete if it is both NP and NP-hard
If an NP Hard problem is non-deterministic polynomial time solvable, it is a NPC problem.

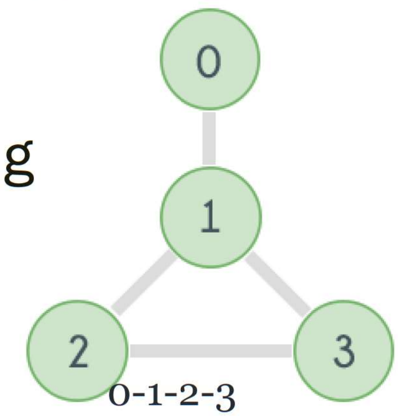
Hamiltonian cycle : Application

- Travelling Sales-man Problem (TSP)
  - *"Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city?“*
- The Hamiltonian cycle problem is to find if there exist a tour that visits every city exactly once.
- Here we know that Hamiltonian Tour exists and in fact many such tours exist, the problem is to find a minimum weight Hamiltonian Cycle.

Hamiltonian path/cycle problem

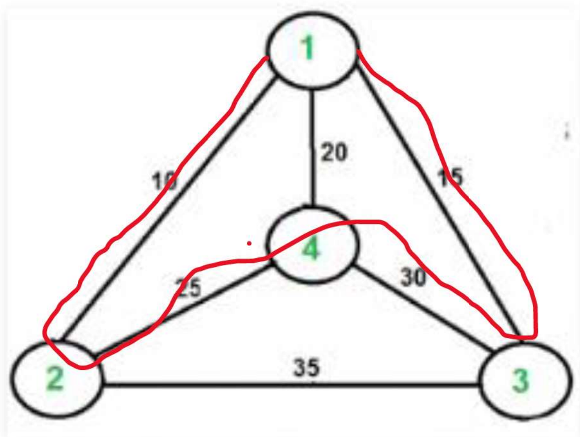
- Determining if Hamiltonian path/cycle exist in a graph.
  - *Hamiltonian cycle can be converted into a Hamiltonian path by removing one edge*
- NP-complete
- A Hamiltonian Path in a graph having N vertices is nothing but a permutation of the vertices of the graph [v_1, v_2, v_3, ……v_{N-1}, v_N] , such that there is an edge between v_i and v_{i+1} where $1 \leq i \leq N-1$.
  - *Can checked for all permutations of the vertices whether any of them represents a Hamiltonian Path or not*

4 vertices - 24 possible permutations



0-1-2-3
3-2-1-0
0-1-3-2
2-3-1-0

Travelling Sales man Problem, Revisited…



A TSP tour in the graph is **1-2-4-3-1**.
The cost of the tour is
10+25+30+15 which is **80.**

Naive Solution (Brute Force Approach):

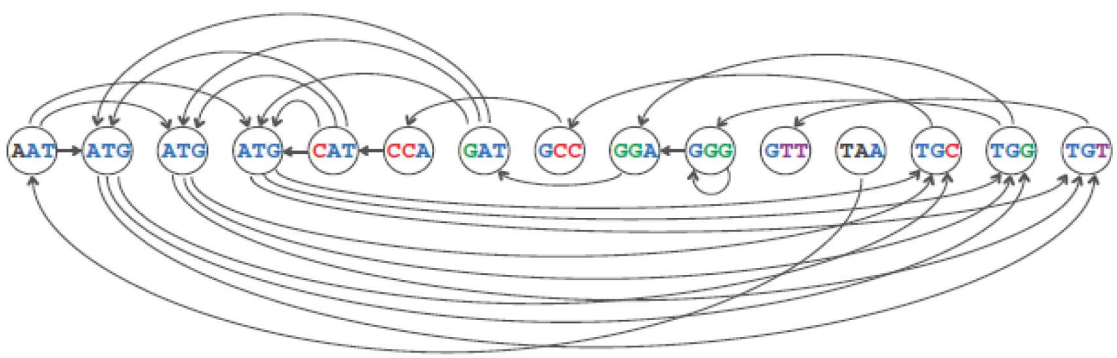
1) Consider city 1 as the starting and ending point.
2) Generate all (n-1)! Permutations of cities.
3) Calculate cost of every permutation and keep track of minimum cost permutation.
4) Return the permutation with minimum cost.

Time Complexity: **$\Theta(n!)$**

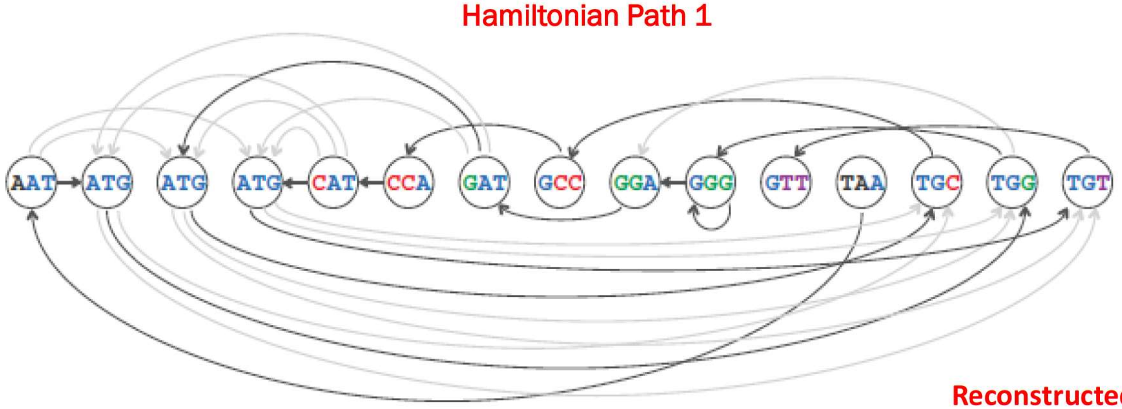
NP Hard Problem

Example : Hamiltonian Path 1

Overlap Graph



Hamiltonian Path 1



Genome path

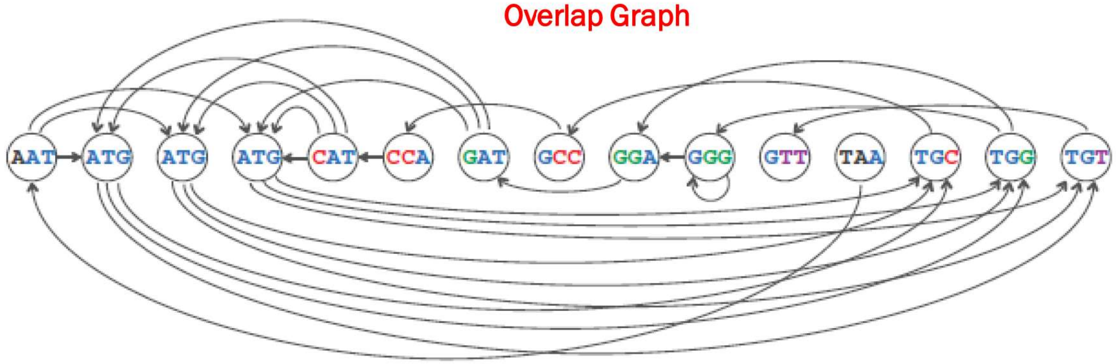
TAA
AAT
ATG
TGG
GGG
GGA
GAT
ATG
TGC
GCC
CCA
CAT
ATG
TGT
GTT

Reconstructed Genome

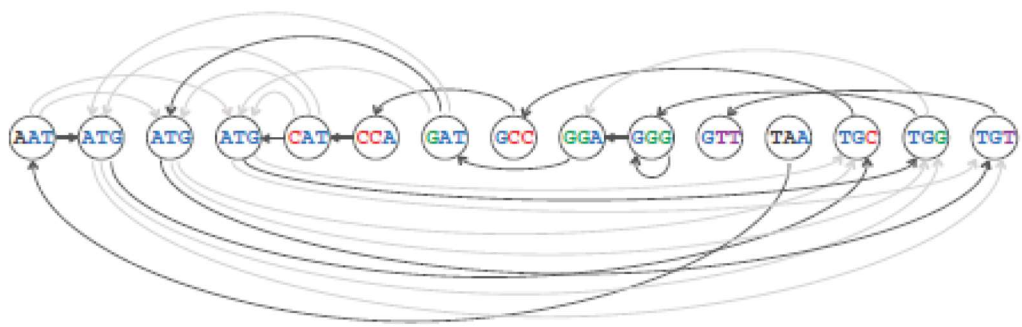
TAATGGGATGCCATGTT

Example : Hamiltonian Path 2

Overlap Graph



Hamiltonian Path 2



Genome path

TAA
AAT
ATG
TGC
GCC
CCA
CAT
ATG
TGG
GGG
GGA
GAT
ATG
TGT
GTT

TAATGCCATGGGATGTT

Reconstructed Genome

Hamiltonian Path Problem

Hamiltonian Path Problem:
Construct a Hamiltonian path in a graph.

> **Input:** A directed graph.
> **Output:** A path visiting every node in the graph exactly once (if such a path exists).

- String Reconstruction using Hamiltonian Path : Algorithm
  - *Construct Overlap(Patterns)*
  - *Find a Hamiltonian path in Overlap(Patterns)*
  - *Output the string formed by the k-mers on the path*
  - *If there is no Hamiltonian path, there is no solution to the String Reconstruction Problem*

Hamiltonian Path Problem

- NP Hard
- Hamiltonian Path problem can be solved only for small or easy graphs
  - *What would be an easy graph?*
- **How we could design an efficient algorithm for it ??**

Summary

- Overlap graph Problem
- String Reconstruction using Hamiltonian path
- Complexity classes
- Hamiltonian Path Problem