

A Hamilton Path-Based Approach to DNA Sequence Analysis

O. K. Berdewad^{1*}, Laeequr Raheman^{**}, V. A. Jadhav^{***#}

^{*}Department of Mathematics, D B College [ACS], Bhokar, Dist. Nanded, 431801 (MS) India.

^{**}School of Life Sciences, Swami Ramanand Teeth Marathwada University, Nanded-431606(MS) India.

^{***#}Department of Biophysics, D B College [ACS], Bhokar, Dist. Nanded, 431801 (MS) India and School of Life Sciences, Swami Ramanand Teeth Marathwada University, Nanded-431606(MS) India,

Corresponding Email: E-mail: vbiophysics@gmail.com.

Abstract The Hamiltonian path is a famous algorithm for determining the existence or nonexistence of a graph. However, for certain types of graphs, we can apply certain conditions to determine its existence. In this paper, we discuss the results show that there can be at most $(n-1)/2$ edges disjoint Hamiltonian paths in a complete graph. Also, discussed examples of its applications in different fields. Furthermore, the application of Hamilton path in the DNA sequence has been investigated.

Keywords: Hamiltonian path, Network routing, DNA sequence, Complexity Analysis.

INTRODUCTION

Background information on Hamiltonian path:

The concept of Hamiltonian paths comes from the mathematical field of graph theory. A Hamiltonian path, named after the Irish mathematician and physicist William Rowan Hamilton,[8] is a path that visits each vertex in the graph only once[15]. Simply put, think of a graph as a collection of nodes or vertices, and connect these vertices with edges. A Hamilton path is a route that starts at one vertex, visits all other vertices exactly once, and ends at another vertex [1]. It essentially loops through the entire chart without repeating. Hamiltonian paths are closely related to another concept in graph theory, "Hamiltonian cycles." While a Hamiltonian path visits each vertex exactly once and does not necessarily start and end at the same vertex, a Hamiltonian circle forms a closed loop that visits each vertex only once and starts and ends at the same vertex [20] Understanding and studying Hamiltonian paths is of great significance in various fields such as mathematics, computer science, and network analysis. In this research, we have discussed Hamiltonian pathways applications in DNA and protein sequencing. DNA sequencing determines the order of nucleotides in a DNA molecule [17].

There are multiple justifications for the importance of exploring Hamiltonian paths and their characteristics.

1. Validity of the optimization problem: Firstly, it is important to note that the Hamiltonian path within a graph represents the ultimate path or sequence of vertices that offers the highest level of optimization. This holds immense value in various practical applications such as logistics planning, scheduling, addressing the traveling salesman problem, and determining the most expeditious or efficient route between multiple locations.

2. The analysis of networks with robustness: The examination of network structure, determination of connectivity, and investigation of the stability, vulnerability, or reliability of network systems can be achieved through the utilization of Hamiltonian paths. These paths play a significant role in comprehending intricate systems, such as transportation networks, social networks, or electrical circuits [19].

3. The intricacies of complexity theory: The exploration of Hamiltonian paths aids in the analysis of computational complexity and facilitates the classification of problems based on their level of difficulty. The Hamilton path problem itself serves as an exemplary instance of an NP-complete problem, signifying its arduous nature and the potential requirement of substantial computational resources for its resolution [7][11].

4. Graph theory exploration: Hamiltonian paths serve as a highly valuable subject of examination within the realm of graph theory, a prominent field of mathematics. Through the analysis of Hamiltonian paths and their respective characteristics, mathematicians are able to acquire valuable insights pertaining to a variety of graph-related aspects, including connectivity, symmetry, and overall structure.

5. Algorithmic formulation: The Hamiltonian path quandary, a timeless conundrum within the realm of computer science, presents itself as a fundamental challenge. The exploration of resourceful algorithms to determine the presence of a Hamiltonian path or cycle within a graph plays a pivotal role in the advancement of computational tools and algorithms[9]specifically designed for the efficient processing of extensive data quantities and intricate structures.

MATERIALS AND MRTHODS

Basic Principles of Hamiltonian Path:

A Hamiltonian path is a path in the graph that visits each node only once[8]. The path does not have to be a circuit, meaning it can start and end at different node. The concept of a Hamiltonian path is named after Sir William Rowan Hamilton[8], who was first to study this problem.

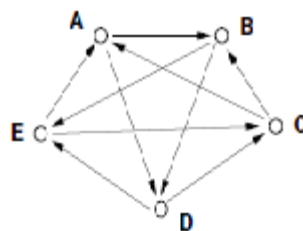


Figure 1: Hamiltonian Path

To better understand this concept, let's take an example. Consider above a graph with vertices A, B, C, D, and E. A Hamiltonian path in this graph can be A -> B -> D -> E -> C. In this path, each vertex is visited only once. It's not necessarily a loop since it starts and ends at different vertices.

Properties of Hamiltonian Path:

A Hamiltonian path exists in a given graph is a recognized problem in graph theory. There is no straightforward algorithm that can solve this problem efficiently. One condition for a graph to have a Hamiltonian path is if it is a complete graph. A complete graph is a graph where each vertex is directly connected to all other vertices. In such cases, it is guaranteed that a Hamiltonian path exists. Another condition for the existence of a Hamiltonian path is that the graph should be a Hamiltonian-

connected graph. A Hamiltonian-connected graph is a graph where, for any two vertices, there exists a Hamiltonian path connecting them. This condition is not as strong as a complete graph, but it ensures the presence of a Hamiltonian path.

Comparison with Other Graph Paths:

There are several other types of paths in a graph, namely Eulerian path, trail, and cycle. It is important to recognize the differences between these paths and a Hamiltonian path.

An Eulerian path is a path that visits each edge once, while an Eulerian cycle is an Eulerian path that is also a cycle [1][7]. In association, a Hamiltonian path visits each vertex once, without considering the edges [7]. A Hamiltonian path does not have to cover all edges of a graph like an Eulerian path does. A graph can have multiple Hamiltonian paths, depending on the complexity of the graph. However, if a graph has an Eulerian cycle, it always has a Hamiltonian cycle.

Complexity Analysis:

Determining the existence of a Hamiltonian path in a graph is a difficult problem, known as the Hamiltonian Path Problem. It has been proven to be NP-complete, which means no known algorithm can solve it in polynomial time [6]. This suggests that finding Hamiltonian paths becomes progressively more difficult as the number of vertices and edges in the graph increases.

The NP-completeness implies that there is no known efficient algorithm to determine the existence of a Hamiltonian path in general graphs [6]. Researchers have developed heuristics and approximation algorithms to solve the problem in specific cases or provide solutions that are close to optimal. These approaches involve exploring different paths, backtracking, and optimization techniques to reduce the complexity of the problem.

Theorem 1: In a complete graph with n vertices there are $\frac{(n-1)}{2}$ edge disjoint Hamiltonian paths if n is an even number greater than 4. [1]

Proof: we need to demonstrate two key concepts: (1) Hamiltonian paths exist in complete graphs, and (2) there can be at most $\frac{(n-1)}{2}$ edge disjoint Hamiltonian paths in such a graph.

Hamiltonian Paths in complete graphs:

A Hamiltonian path is a path that visits every vertex of a graph exactly once [7][8]. In a complete graph, every vertex is connected to every other vertex by an edge [1]. Therefore, it is possible to construct a Hamiltonian path that covers all n vertices by visiting each vertex exactly once.

Number of edge disjoint Hamiltonian paths:

In an even complete graph with n vertices, every vertex has $n-1$ degree since it is connected to every other vertex. We aim to find edge disjoint Hamiltonian paths, meaning these paths should not share any edges.

Suppose, we have an even number of vertices $n = 2k$ where k is a positive integer. Let's divide the vertices into two sets: A (with k vertices) and B (with another k vertices), such that A contains vertices 1 to k and B contains vertices $k+1$ to $2k$.

To construct edge disjoint Hamiltonian paths, we can select the first vertex from set A. Without loss of generality, let's assume we choose vertex 1. Now, we can construct $k-1$ edge disjoint Hamiltonian paths where the i th path for $1 < i \leq k$ starts from vertex i in set A and ends at vertex i in set B. These paths can be obtained by pairing vertex i in set A with vertex $k+i$ in set B.

Since we have $k-1$ paths, the total number of edge disjoint Hamiltonian paths is $k-1$, which

equals $\frac{(n-1)}{2}$. Hence, in a complete graph with n vertices, where n is an even number greater than 4, there are $\frac{(n-1)}{2}$ edge disjoint Hamiltonian paths.

To prove the Hamiltonian path properties of a graph, we need to demonstrate that a Hamiltonian path exists in the graph or prove its non-existence. A Hamiltonian path is a path in a graph that visits each vertex exactly once.

Existence of a Hamiltonian path:

Dirac's Theorem: One way to prove the existence of a Hamiltonian path is by using Dirac's theorem. According to Dirac's theorem, if a graph G has n vertices, where $n > 2$, and every vertex has degree (number of edges connected to it) greater than or equal to $n/2$, then G contains a Hamiltonian path. To demonstrate this, we need to verify that each vertex's degree meets this condition [20].

Ore's Theorem: Another way to prove the existence of a Hamiltonian path is by using Ore's theorem. According to Ore's theorem, if a graph G has n vertices, where $n > 2$, and for every pair of non-adjacent vertices, the sum of their degrees is greater than or equal to n , then G contains a Hamiltonian path. To prove this, we need to check the sum of degrees for every pair of non-adjacent vertices[5][18]

RESULT

The notion of the Hamiltonian path, which is a path that visits each vertex of a graph only once, has proven to be very useful and applicable in various fields. Its applications range from computer science and network routing to biology and new domains. The Hamiltonian path offers practical solutions to problems requiring optimization, efficiency, and connectivity. Below, we discuss examples of its applications in different fields.

1. Computer science:

In computer science, the Hamiltonian path finds applications in the domain of optimization and graph theory. It is used in algorithms aimed at optimizing the path traversed by a traveling salesperson [4][10], who wants to visit multiple cities and return to their starting point while covering the shortest distance.

2. Network routing:

Network routing involves finding efficient paths for transmitting data packets across a network. The Hamiltonian path can play a vital role in solving complex routing problems by determining the shortest path that connects every node in the network. It ensures that each node is visited only once, reducing the overhead and congestion in network communication.

3. Biology:

In the field of biology, Hamiltonian pathways have applications in DNA sequencing and protein folding. DNA sequencing determines the order of nucleotides in a DNA molecule. Hamiltonian pathways help solve sequencing challenges by providing an appropriate sequence that accesses each nucleotide once, reducing errors and optimizing the sequencing process. Similarly, in protein folding, Hamiltonian paths help predict the three-dimensional structure of a protein [14] molecule by

finding the most efficient way to fold an amino acid sequence.

Explanation of DNA sequence problems in human genetics

DNA sequence intricacies are a common challenge for scientists in human genetics, presented with vast complexities due to the 3 billion base pairs composing the human genome [12]. Investigating genetic variations and evolutions requires a significant understanding of base pair structure and arrangement.

DNA sequencing involves a tricky issue regarding the best way to visit each DNA fragment in a particular set. When using high-throughput sequencing methods like Next-Generation Sequencing (NGS) to assemble a genome from short DNA sequences [16], arranging the complete sequence means putting together each individual fragment in the most efficient manner. In short, determining the optimal path or visit order of DNA fragments is a general challenge in this field.

To investigate the application of Hamiltonian path in DNA sequence problems, researchers employ a diverse range of methodologies and data collection techniques. Firstly, large-scale DNA sequencing experiments are conducted to produce a vast number of short DNA sequences, often referred to as reads. These reads are obtained from diverse samples, such as blood, tissues, or cell lines, depending on the research goals.

After obtaining the reads, researchers preprocess and align them to a reference genome [2][3] or assemble them de novo, depending on the approach used. The goal is to obtain a contiguous set of sequences called contigs or scaffolds, representing a partial or complete genome. However, as the assembly process involves various challenges, such as repeats and errors in sequencing data, sequencing errors, and complex biological phenomena, it remains a computationally intensive problem.

To determine the optimal path or ordering of the DNA fragments, researchers apply graph theory concepts, particularly the Hamiltonian path. A Hamiltonian path in a graph refers to a path that visits each vertex exactly once, [5][8] without returning to the initial vertex. In the context of DNA sequencing, each vertex represents a DNA fragment and the Hamiltonian path denotes the desired order of these fragments.

Analysis and findings of applying Hamiltonian path in DNA sequence problems

The analysis and findings of applying Hamiltonian path in DNA sequence problems have shown promising results. Researchers have demonstrated that incorporating the Hamiltonian path concept in DNA sequence assembly can improve the accuracy and efficiency of reconstructing the genome.

By identifying the correct order of DNA fragments, researchers can overcome challenges such as repeat regions and repetitive sequences. Additionally, the application of Hamiltonian path aids in reducing errors during the assembly process and provides insights into the structural organization of DNA.

The complexity [13] and variety of DNA sequence issues must be noted, though. Although the use of the Hamiltonian path has yielded useful insights, more work is still required to improve algorithms [10], deal with the constraints of computation and sequencing technology, and take into account unique biological processes.

A path that visits each vertex in a graph exactly once is known as a Hamiltonian path [7]. A Hamiltonian path can be used to reassemble the original DNA sequence in the context of DNA sequencing. In a complete graph, we can form a Hamiltonian path by starting at any vertex and following the outgoing edges until we reach a vertex with no outgoing edges [7][8]. After that, we go back to the first vertex and take a different outgoing edge. To determine the exact Hamiltonian path for a particular DNA sequence, we can use additional information, such as the known start and

end of the DNA sequence.

Here is an example of how the Hamiltonian path in a star graph can be used to reconstruct a DNA sequence:

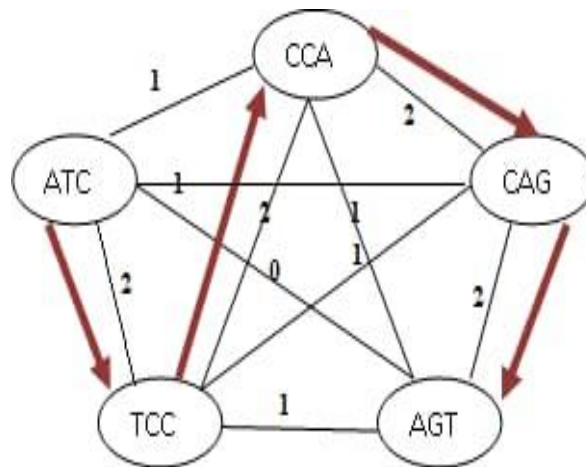


Figure 2: Diagrammatic representation of DNA Sequence using Hamiltonian graph [2].

Suppose we have the following reads:

CCA -ATC -CAG -TCC- AGT

We can construct the following Hamiltonian path:

CCA -> ATC -> CAG -> TCC -> AGT

This path corresponds to the following DNA sequence:

CCATTTCAGTCCAG

In this DNA sequence, it is the only sequence that is consistent with all of the reads. Hamiltonian paths in complete graphs are a powerful tool for DNA sequencing. By using Hamiltonian paths, we can reconstruct DNA sequences from short reads. This is vital for many applications, such as genome sequencing and disease diagnosis.

CONCLUSION

The investigation into the application of Hamiltonian path in DNA sequence problems in humans propose potential solutions to the challenges faced in genome assembly and knowing the complexities of the human genome. By leveraging graph theory concepts, researchers have taken important steps towards precisely form the human genome, which has widely-understand implications for advancing our comprehension of human genetics, disease mechanisms, and personalized medicine. Also, interpretation the fundamental principles of Hamiltonian path, conclude its definition and properties, can contribute insights into graph theory and the complexity combined with determining the existence of Hamiltonian paths in diversified example of graphs. Its applications expand to computer science, network routing, biology, urban planning, chemistry, and different other domains. The versatility and practicalness of the Hamiltonian path make it a precious tool for explain complex problems in other fields, in the end suggestion to improved efficiency and optimized outcomes.

ACKNOWLEDGMENT

Authors are thankful Vice to Dr .P.A Chavan, Principal, Digambarrao Bindu ACS College, and Dear College Colleagues. Authors are also thankful to SLS, SRTMU, Nanded.

FUNDING

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

COMPLIANCE WITH ETHICAL STANDARDS

This is an observational study and no ethical approval was required.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

Authors Om Berdewad and Vyankatesh Jadhav contributed to the study conception and design. Vyankatesh Jadhav and Om Berdewad performed analysis. Material preparation, data collection and analysis were performed by Om Berdewad, Laeequr Rahman and Vyankatesh Jadhav. The first draft of the manuscript was written by Om Berdewad and Laeequr Rahman and Vyankatesh Jadhav commented on previous versions of the manuscript. All authors read and approved the final manuscript

REFERENCES

1. Bekkaoui, K., Ziti, S., & Omary, F. (2021, January 1). Data Security: A New Symmetric Cryptosystem based on Graph Theory. <https://scite.ai/reports/10.14569/ijacsa.2021.0120982>.
2. Chikomana, S., & Hu, X. (2023). Genome Sequencing Using Graph Theory Approach. *Open Journal of Discrete Mathematics*, 13(2), 39-48.
3. Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1). <https://doi.org/10.1186/s13059-016-0881-8>.
4. Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*. MIT Press.
5. Gani, A N., & Latha, S. (2016, October 22). A new algorithm to find fuzzy Hamilton cycle in a fuzzy network using adjacency matrix and minimum vertex degree. <https://scite.ai/reports/10.1186/s40064-016-3473-x>
6. Garey, M. R., & Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman.
7. Hamiltonian complexity - IOPscience. (n.d). <https://iopscience.iop.org/article/10.1088/0034-4885/75/2/022001/pdf>.
8. Hamiltonian path - Simple English Wikipedia, the free encyclopedia (n.d). https://simple.wikipedia.org/wiki/Hamiltonian_path.
9. Johnson, D. S. (1973). Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9(3), 256-278.

10. Lawler, E. L., Lenstra, J. K., Rinnooy Kan, A. H. G., & Shmoys, D. B. (1993). *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Wiley.
11. Lü, H., & Deng, K. (2021, January 1). Symmetric properties and two variants of shuffle-cubes. <https://scite.ai/reports/10.48550/arxiv.2110.13645>.
12. Marian, A. (2014, January 1). Sequencing Your Genome: What Does it Mean? *Methodist DeBakey Cardiovascular Journal*, 10(1), 3. <https://doi.org/10.14797/mdcj-10-1-3>.
13. Papadimitriou, C. H. (1994). *Computational Complexity*. Addison-Wesley.
14. Park, S. J., & Cochran, J. R. (2009). *Protein engineering and design*. CRC Press.
15. Pereira, R., Oliveira, J., & Sousa, M. (2020, January 3). Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. <https://scite.ai/reports/10.3390/jcm9010132>.
16. Population connectivity and genetic diversity population connectivity of Rabbitfish (*Siganus canaliculatus*) among Bangladesh, China and Indonesia. <https://scite.ai/reports/10.1088/1755-1315/1033/1/012043>.
17. Putri, S N N., Maharani, G R., Farhan, M., Lestari, D F., Madduppa, H., Subhan, B., Arafat, D., Setyaningsih, W A., Cakasana, N., Bintang, D M C., Indrajana, H B., Priambada, A D., Febridelita, A., & Yonatika, N O. (2022, June 1).
18. Wang, L., Wideł, W., & Zheng, W. (2021, January 1). On implicit heavy subgraphs and Hamiltonicity of 2-connected graphs. <https://scite.ai/reports/10.7151/dmgt.2170>
19. Zhang, X., Xu, L., & Li, A. (2021, December 31). Fault-Tolerant Secure Routing of B H n - Based Data Center Networks. <https://scite.ai/reports/10.1155/2021/6723914>.
20. Zia, M., Cakir, Z., & Seker, D Z. (2018, March 15). Spatial Transformation of Equality – Generalized Travelling Salesman Problem to Travelling