# Aesthetic Representation of De Bruijn Graphs Based on DNA Base Sequences

Moon-ki Park

Department of Mathematics, Kwangwoon University

6030-33-8277-01:Advanced Seminar on Combinatorics

Prof. Sang-Mok Kim

December 15, 2024

**Aesthetic Representation of De Bruijn Graphs Based on DNA Base Sequences**


## Abstract

De Bruijn graphs are mathematical constructs used in combinatorics and computational biology that gracefully depict sequence overlaps. The theoretical characteristics of De Bruijn graphs are examined in this work, with a focus on DNA sequence representation. We highlight their aesthetic qualities, which come from symmetry and cyclic patterns, and we show 2D and 3D representations as a synthesis of artistic appeal and scientific accuracy.

Aesthetic visualization also makes it easier to understand complex facts. This method helps to better comprehend and interpret data by providing a visual representation of De Bruijn graphs, one of the complex data structures.


## 1. Introduction

De Bruijn graphs serve as a cornerstone in the mathematical modeling of sequence overlaps. Their utility spans fields such as graph theory, computational biology, and data visualization. DNA base sequences, composed of the alphabet {A, C, G, T}, present a natural application of De Bruijn graphs due to their inherent need for efficient sequence overlap representation.

The objectives of this paper are threefold:

1. To provide a rigorous mathematical definition of De Bruijn graphs.
2. To analyze their structural properties with a focus on symmetry and cyclic patterns.
3. To demonstrate the aesthetic potential of De Bruijn graphs through 2D and 3D visualizations.

# 2. Mathematical Foundation of De Bruijn Graphs

## 2.1 Definition

A De Bruijn graph B(k, N) is defined over an alphabet of size k and a sequence length N. It comprises:

- Vertices: All possible strings of length N over the given alphabet.
- Edges: Directed edges between vertices, representing overlaps of N-1 characters.

## 2.2 Properties

Vertex Count: The total number of vertices in a De Bruijn graph is given by: $|V| = k^{N}$ where k is the alphabet size, and N is the length of the sequences.

Edge Count: The total number of edges is expressed as: $|E| = k^{N+1}$. Each edge corresponds to a string of length N+1, ensuring complete overlap coverage.

## 2.3 Example: B(2, 2)

For an alphabet {0, 1} and sequence length N=2:

- Vertices: {00, 01, 10, 11}
- Edges: Each edge represents a length-3 string (e.g., ( 00 → 01)).

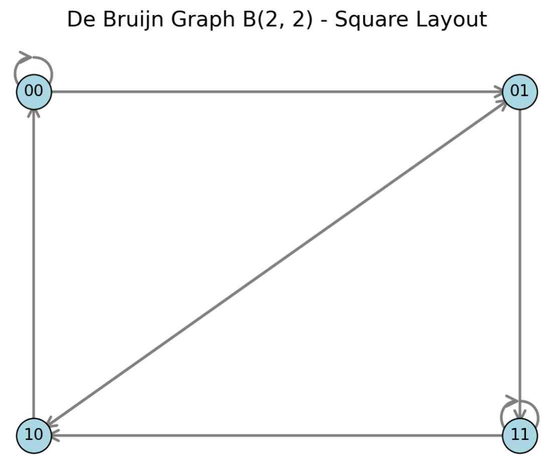



**Figure1.** Example graph of B(2, 2)

Eulerian Property: De Bruijn graphs are Eulerian, ensuring the existence of a path traversing all edges exactly once. This property is critical for sequence reconstruction tasks in computational biology.

# 3. Aesthetic Properties of De Bruijn Graphs

## 3.1 Symmetry and Cyclic Patterns

The intrinsic symmetry of De Bruijn graphs emerges as a manifestation of their structural rigor and mathematical elegance. This symmetry is characterized by the following principles:

### 3.1.1. Balanced Vertex Degree:

Every vertex within the De Bruijn graph exhibits a uniform degree of connectivity, with exactly  incoming and  outgoing edges. This equilibrium reflects a harmony that is mathematically exact and visually symmetrical, ensuring each state (vertex) maintains structural integrity within the graph.

### 3.1.2. Cyclic Structure:

The graph inherently embodies cyclic behavior, encapsulating all possible overlaps between sequences of length . The Eulerian circuit—a path that traverses each edge exactly once—represents the ultimate form of continuity and recurrence, providing an efficient and complete representation of the graph's combinatorial landscape. This cyclic symmetry resonates deeply with the philosophical concept of eternal recurrence, where patterns repeat infinitely within a bounded yet continuous system.

## 3.2 Mathematical Aesthetics

De Bruijn graphs exemplify principles of visual and mathematical beauty that transcend their functional purpose:

Symmetry: The graph's vertices and edges form regular, predictable patterns. This property aligns with the mathematical pursuit of order and balance, reflecting an intrinsic harmony in both form and function.

Continuity: The smooth transition between states (vertices) ensures that every path flows seamlessly across the graph. This continuous structure mirrors natural processes, such as the progression of DNA sequences, emphasizing the elegance of uninterrupted transitions.

Repetition: The recurrence of patterns, as seen in overlapping sequences, underpins the graph's combinatorial completeness. This repetition is not merely structural but also aesthetic, as it reinforces the visual coherence of the graph's layout.

In essence, the De Bruijn graph serves as a bridge between abstract mathematical ideals and tangible visual representation, offering a unique perspective on the interplay between structure, symmetry, and continuity.

# 4. Visualization Techniques

The DNA nucleotide sequences {A, C, G, T} are used to create a 4-ary span-2 graph, which is a De Bruijn graph with overlaps of length two. The relevant amino acids are visually represented by the edges in both the 2D and 3D renderings.

## 4.1 2D Visualization

A 2D representation of B(4, 2) ensures clarity in depicting:

- Cyclic overlaps.

- Balanced degree properties.

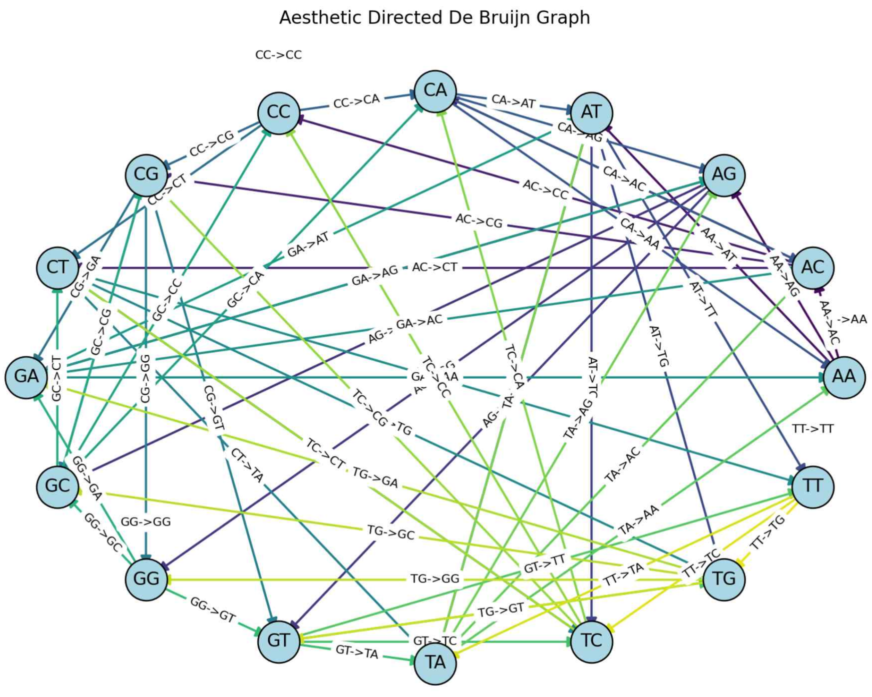
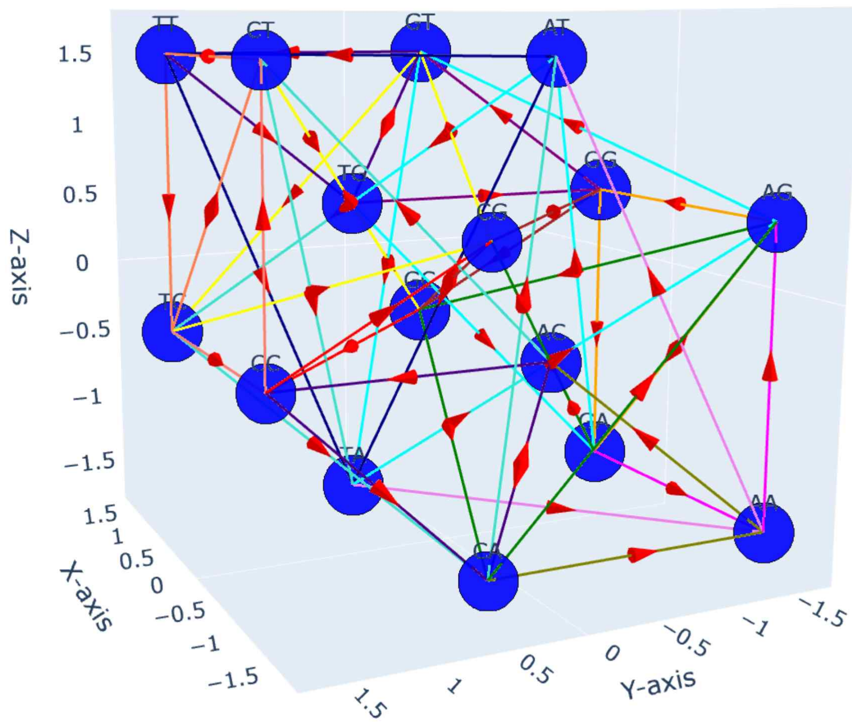



**Figure2.** 2D Graph: Illustrates symmetric patterns.

Analysis: The 2D graph representation showcases the cyclic structure of the De Bruijn graph. Each vertex connects to others based on -character overlaps, and the graph's layout emphasizes symmetry with equal in-degree and out-degree for all nodes. This visual balance reflects the mathematical elegance of the structure.

## 4.2 3D Visualization

Advanced Topology: A 3D visualization introduces depth and perspective, enhancing the perception of higher-dimensional relationships. Gradients and color mapping augment its interpretability and aesthetic appeal.




**Figures3.** 3D Graph: Highlights spatial relationships and cyclic completeness.

Analysis: The 3D graph representation enhances the understanding of higher-dimensional relationships within the graph. Compared to the 2D representation, the 3D structure allows for the visualization of depth and spatial connections, making it easier to identify hierarchical relationships and overlapping sequences. This added dimensionality effectively reveals the graph's intricate topology while preserving its cyclic and symmetric properties.

## 5. Conclusion

This study underscores the dual significance of De Bruijn graphs as mathematical tools and aesthetically pleasing representations. By merging theoretical insights with visual exploration, we have demonstrated their potential to serve as both scientific and artistic instruments. Future work may involve dynamic visualizations and real-time applications in bioinformatics.

## References


1. de Bruijn, N. G. (1946). A combinatorial problem. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen.

2. Bowe, A., Onodera, T., Sadakane, K., & Shibuya, T. (2012). Succinct de Bruijn graphs. In Lecture Notes in Computer Science (Vol. 7534, pp. 225–235). Springer.

3. de Bruijn, N. G. (1946). A combinatorial problem. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen. Series A, 49, 758–764.

4. Johnson, D. M., & Mendelson, N. S. (1970). Planarity properties of the Good-de Bruijn graphs. In Combinatorial Structures and Their Applications (pp. 177–183). Gordon and Breach.