Tonal Structures in Western Classical Music: A Directed Graph Analysis of Harmonies Across Compositional Eras

Farish Firstian Erifiawan
Informatics Engineering
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Jl. Ganesha No. 10, Bandung, Indonesia 40132
13525058@std.stei.itb.ac.id - notfaris2007@gmail.com

***Abstract*—Chord progressions in Western classical music encode compositional intent that shifts across historical periods. This paper models these progressions as directed weighted graphs — nodes represent diatonic scale degrees (I through vii°), edges denote harmonic transitions, and weights reflect transition frequency. Three composers are analyzed: Bach (Baroque), Beethoven (Classical), and Tchaikovsky (Romantic). Methods include degree centrality, cycle detection, SCC decomposition, and degree distribution comparison. Results reveal measurable structural differences across eras, with later Romantic harmony showing greater graph density relative to the V–I-dominant structures of Baroque counterpoint.**

***Index Terms*—chord progressions, discrete mathematics, graph theory, harmonic analysis, music theory, strongly connected components, Western classical music**

I. INTRODUCTION

All music, regardless of genres and traditions, is structured sound. This underlying structure consists of multiple parts: rhythm, melody, form, and harmony, to name a few.

This paper analyzes harmonic structure in Western classical music, specifically in three pieces from differing compositional eras, using graph theory to model harmonic relationships and their frequency of occurrence. Western classical pieces are chosen because:

1) they offer a wide range of tonal movement and variety, and
2) commonly accepted understanding of music are rooted in Western classical traditions.

The pieces analyzed are as follows:

- "Prelude No. 2", from *The Well-Tempered Clavier, Book 2*, by Johann Sebastian Bach (Baroque);
- *Piano Sonata No. 14 "Moonlight"* — II. Allegretto, by Ludwig van Beethoven (Classical); and
- *The Seasons* — II. February: Carnival, by Pyotr Ilyich Tchaikovsky (Romantic).

By knowing how the harmony "flows" in these pieces, inferences can be made of stylistic decisions these composers made and gain a deeper understanding (and perhaps appreciation) of these pieces and eras.

II. BACKGROUND

In order to understand the analysis process, this section presents a quick primer on harmony and graph theory.

A. Harmony

Harmony refers to the simultaneous combination of pitches into chords and the arrangement of those chords into musical passages. Roughly speaking, harmony plays a central role in shaping the character of a piece of music. It sets the mood, the tension and release.

At its most basic level, this combination is described through intervals — the relationship in pitch between two notes. Stacking two intervals, specifically thirds, produces the triad, which is considered the foundational unit of harmony in Western tonal music. A triad is composed of three notes, each referred to as the root, third, and fifth. Within a given key, triads can be built on each of the seven scale degrees, producing seven distinct chords that form the harmonic vocabulary of that key.

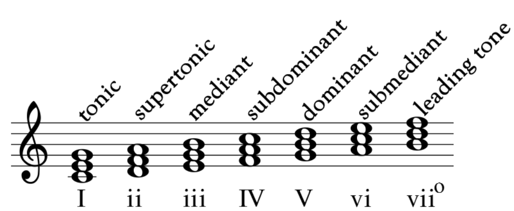

Fig. 1. The seven scale degrees in C major with their respective triads and Roman numeral notation.
Source: Wikimedia Commons, CC Attribution-Share Alike 3.0 Unported

These chords are conventionally labelled using Roman numeral notation, where the numeral indicates the scale degree on which the chord is built (I for the first degree, IV for the fourth degree, etc.) and the case of the numeral denotes the chord's "quality":

- uppercase for major triads,
- lowercase for minor triads,

- lowercase with a degree symbol for diminished triads, and
- uppercase with a plus symbol for augmented triads

Different scales produce different sequences of triads. The major key, for instance, yields the sequence I, ii, iii, IV, V, vi, vii°, whereas the harmonic minor key yields i, ii°, III+, iv, V, VI, vii°. While a given key establishes a default set of diatonic triads composers could utilize in their songwriting, they may introduce chords borrowed from the parallel key in a technique known as modal mixture. They can also manipulate these triads into inversions, slash chords, suspended chords, seventh chord, half-diminished chords, and much more.

There are 3 functional groupings of these triads: tonic, dominant, and subdominant. Each groups are associated with different degrees of harmonic stability and tension. As with the case of triads, different scales produce different sets of these groups.

The movement between these groups, particularly the resolution from the dominant to the tonic, forms cadences. Cadences give a sense of direction and articulate the overall structure of a musical piece, and it is precisely this directionality that the present paper seeks to formalize using directed graphs.

B. *Graph theory*

In discrete mathematics, a graph is a structure consisting of a set of vertices (referred to as nodes in this paper) and a set of edges connecting pairs of nodes. When edges have a direction, indicating that a relationship flows from one node to another rather than mutually, the graph is said to be directed. This directionality makes directed graphs a natural fit for modeling chord progressions, since harmonic changes from one chord to the next is inherently asymmetric: a V–I resolution and a I–V motion represent distinct musical events, not the same relationship traversed twice.

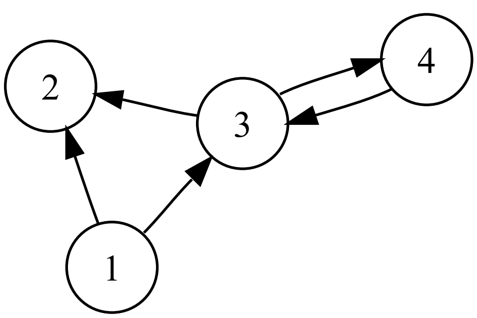

Fig. 2. A directed graph.
Source: Wikimedia Commons

Edges may additionally carry a weight, a numerical value associated with the strength or frequency of that relationship. In this paper, where harmonic motion is modelled to a directed graph, the edge from chord A to chord B can be assigned a weight equal to the number of times that specific transition occurs within a corpus, allowing common progressions to be distinguished from rare ones.

Several graph-theoretic measures are used to characterize the structure of such a graph.

1) Degree centrality counts the number of edges incident to a node, identifying which chords function as structural hubs or central points.
2) A cycle is a path that returns to its starting node, corresponding to recurring harmonic loops.
3) A strongly connected component (SCC) is a maximal subset of nodes in which every node is reachable from every other, revealing clusters of chords bound by mutual reachability.

Together, these measures provide a quantitative basis for comparing harmonic structure across compositions and eras.

III. METHODOLOGY

The harmonic structure of each piece is formalized as a directed weighted graph

$$G = (V, E, w) \tag{1}$$

where V is the set of nodes, $E \subseteq V^2$ is the set of directed edges, and $w : E \rightarrow \mathbb{Z}^+$ is a weight function assigning a positive integer to each edge.

Each node $v \in V$ represents a distinct diatonic chord, identified by its Roman numeral scale degree. Chord inversions and added tones (as found in slash and seventh chords, for instance) are abstracted away; two chord annotations that share the same scale degree and quality are mapped to the same node regardless of voicing.

A directed edge $E \in (u, v)$ is introduced for each consecutive chord pair in the annotated sequence, where u denotes the preceding or "source" chord and v denotes the succeeding or "destination" chord. The weight $w(u, v)$ records the total number of times a particular transition occurs within the piece, such that more frequent progressions correspond to heavier edges. Self-loops, which arise when a chord annotation is repeated across adjacent beats without harmonic change, are excluded from the model as they do not represent harmonic motion.

Chromatic and non-diatonic chords encountered during annotation — including augmented sixth chords (Italian, French, and German variants) and cadential six-four figures — are normalized to their nearest diatonic functional equivalent prior to graph construction. Augmented sixth chords are mapped to IV, reflecting their pre-dominant function, while cadential six-four figures are mapped to V, reflecting their dominant prolongation role. These normalization decisions are documented as methodological constraints and their implications are discussed in Section VII.

IV. DATA COLLECTION AND PROCESSING

The annotation for the three pieces are drawn from the When-in-Rome dataset, a collection of Roman numeral analyses compiled and maintained by Gotham et al.

Annotations in the When-in-Rome dataset are provided in the RomanText format, a plain-text encoding that specifies the key context and Roman numeral label for each chord at the beat level. All three annotations used in this study are sourced from the Digital and Cognitive Musicology Lab (DCML) and have undergone manual verification by domain experts, ensuring annotation quality sufficient for structural analysis.

Roman numeral sequences were extracted from each annotation file using the music21 Python library. The resulting chord counts per piece are:

1) 108 chords for Bach,
2) 96 chords for Beethoven, and
3) 256 chords for Tchaikovsky.

The disparity in chord counts reflects differences in piece length, tempo, and annotation granularity rather than harmonic complexity per se, and is acknowledged as a limitation when comparing absolute edge weights across composers. Normalization decisions applied during the extraction process are described in Section III.

V. ANALYSIS

A. Degree Centrality

In-degree and out-degree centrality were computed for each graph using raw edge counts weighted by transition frequency. A chord's in-degree reflects the number of distinct chords that precede it, while its out-degree reflects the number of distinct chords it moves toward. Weighted degree values additionally account for transition frequency, yielding a measure of harmonic traffic rather than mere connectivity.

Across all three composers, V emerges as consistent high in-degree nodes, reflecting its role as a common harmonic destination in voice-leading-driven motion. Bach's graph exhibits V as the dominant out-degree hub (weighted degree 10), consistent with the functional directionality characteristic of Baroque counterpoint. In Beethoven's graph, I dominates both weighted in-degree (36) and out-degree (35), suggesting a tonic-centric harmonic language in which the tonic chord serves simultaneously as the primary point of departure and arrival. Tchaikovsky's graph shows V as the dominant hub in both directions (weighted in-degree and out-degree of approximately 85), indicating an exceptionally strong dominant gravity relative to the other composers.

B. Cycle Detection

Simple cycles were enumerated for each graph using depth-first traversal. The total cycle counts are:

1) 99 for Bach,
2) 15 for Beethoven, and
3) 142 for Tchaikovsky.

Short cycles of length two — representing immediate chord alternations — were identified across all three composers, with ii $\rightarrow$ V and I $\rightarrow$ V being the most recurrent two-node cycles. The comparatively low cycle count in Beethoven's graph reflects the sparse connectivity of the Allegretto's harmonic language, in which a small set of chords governs the majority of transitions.

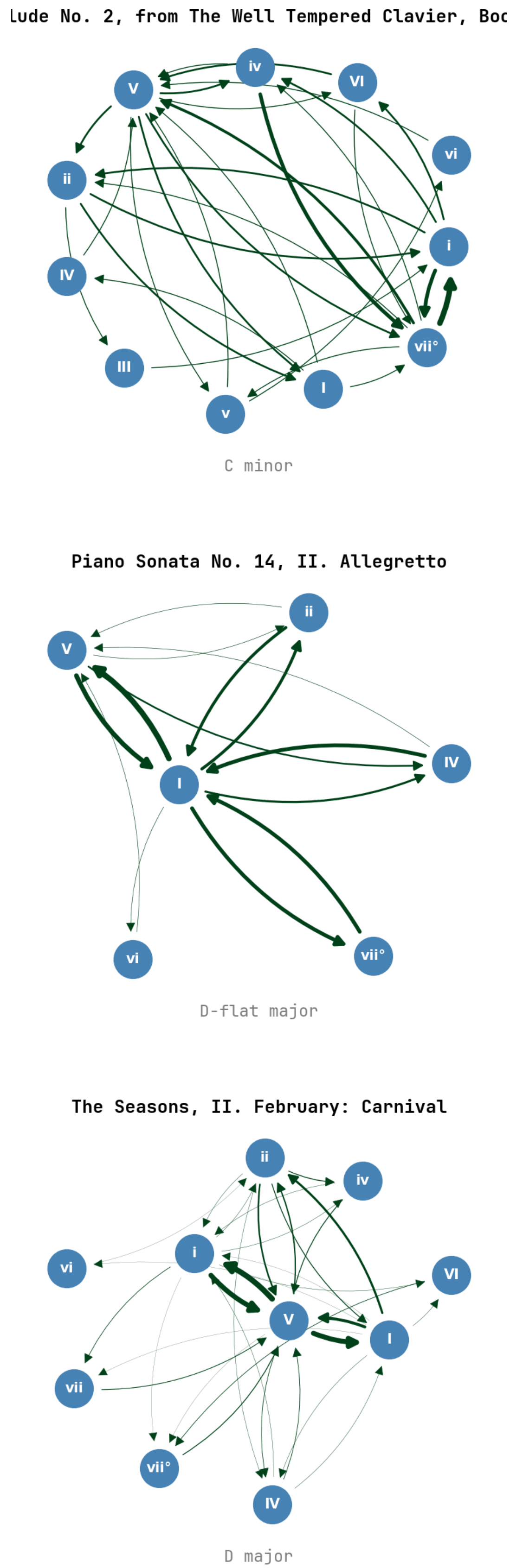

Fig. 3. Graph models of the harmonic movements of the sample pieces. The heaviness of an edge is indicated by its thickness.

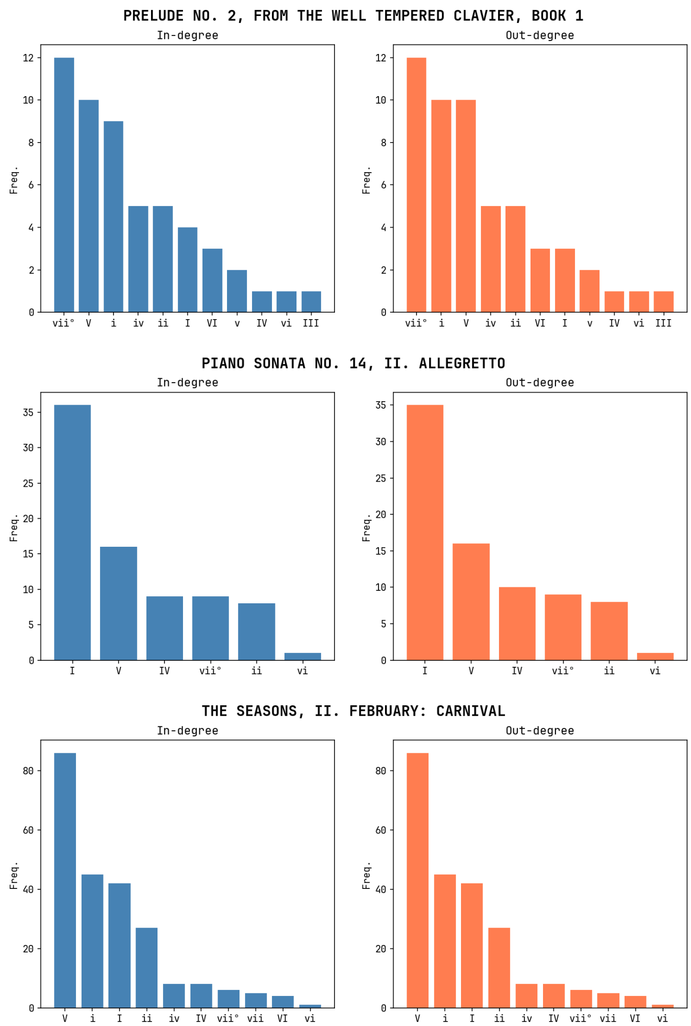

Fig. 4. Degree distribution (weighted) of the sample pieces. Notice that for Bach, the top in- and out-degree chord is different for the weighted and nonweighted calculations.

C. Strongly Connected Components

Strongly connected component decomposition was applied to each directed graph. All three graphs yield a single SCC spanning the entire node set, indicating that every chord in each piece is mutually reachable from every other chord via directed harmonic paths. This result holds across all three compositional eras and suggests that tonal harmony, regardless of period, maintains a fully traversable harmonic space — a property consistent with the functional interdependence of diatonic chords within a key.

D. Degree Distribution

Weighted degree distributions were computed and plotted for each composer. Bach's distribution is relatively gradual, with vii°, V, and i leading in-degree and no single chord dominating overwhelmingly. Beethoven's distribution is the most concentrated, with I accounting for a disproportionate share of both in- and out-degree weight — a steep drop-off following the leading chord is visible in the distribution plot. Tchaikovsky's distribution is the most skewed, with V exhibiting a weighted degree far exceeding all other chords, followed by i and I at roughly half its value.

VI. Cross-Era Structural Comparison

The three graphs, taken together, reveal a progressive structural evolution in Western tonal harmony across the Baroque, Classical, and Late Romantic periods. This section synthesizes the per-composer findings from Section V into a comparative analysis along four dimensions: graph density, harmonic centralization, cyclical complexity, and modal orientation.

A. Graph Density

Graph density, defined as the ratio of realized edges to the maximum possible edges in a directed graph of the same size, measures how fully connected a harmonic space is. Beethoven's graph yields the highest density at 0.467, followed by Tchaikovsky at 0.417 and Bach at 0.273. The high density of Beethoven's graph is notable given its small node set of six chords — the Allegretto's harmonic language is narrow but internally dense, with most chord pairs connected by at least one directed edge. Bach's lower density reflects a larger node set with more sparse inter-chord connectivity.

TABLE I
Overview of the graph models' metrics.

	Bach	**Beethoven**	**Tchaikovsky**
Node(s)	11	6	10
Edge(s)	30	14	32
Top in-degree	V	I	V
Top out-degree	V	I	I
Cycle(s)	99	15	163
SCC count	1	1	1
Density	0.273	0.467	0.356

B. Harmonic Centralization

The identity of the dominant hub chord shifts across eras. In Bach's graph, V functions as the primary out-degree hub, directing harmonic motion outward toward multiple destinations — a pattern consistent with the dominant-driven voice leading of Baroque counterpoint. In Beethoven's graph, I assumes both the highest in-degree and out-degree, acting as the central axis around which the entire harmonic structure rotates. In Tchaikovsky's graph, V reclaims hub status with considerably greater weight than in Bach, suggesting an intensification of dominant gravity in the Late Romantic period relative to the Baroque.

C. Cyclical Complexity

The cycle counts reveal a non-linear relationship between era and harmonic cyclical complexity. Tchaikovsky's graph produces the highest cycle count at 142, followed by Bach at 99 and Beethoven at 15. Rather than reflecting a monotonic increase across eras, this pattern suggests that cyclical complexity is more strongly governed by graph connectivity and piece structure than by historical period alone. Beethoven's sparse graph naturally yields fewer cycles, while Tchaikovsky's denser connectivity and larger node set generate a substantially richer cycle space.

D. Modal Orientation

The three pieces differ in their modal orientation, which is reflected in the composition of their node sets and dominant transitions. Bach's graph contains both major and minor scale degree chords owing to the C minor tonality of the prelude, with i and v present alongside their major counterparts. Beethoven's graph is predominantly major-mode, with minor-mode chords appearing only marginally. Tchaikovsky's graph is the most modally mixed, with i, V $\rightarrow$ i, and ii $\rightarrow$ V transitions all prominent, reflecting the D major tonality of the piece punctuated by frequent minor-mode inflections. This modal variability across the corpus further illustrates the expanded harmonic vocabulary of the Late Romantic period relative to earlier eras.

VII. DISCUSSION

A. Implications

The results of this examination demonstrate that directed weighted graphs provide a meaningful and quantifiable framework for characterizing harmonic structure in Western classical music, or even any musical work that can be described using the Western theory of music. The single-SCC finding across all three composers establishes a baseline property of tonal harmony: regardless of era, the diatonic harmonic system maintains full mutual reachability among its constituent chords. This structural invariant suggests that the fundamental connectivity of tonal harmony is preserved even as surface-level stylistic features.

The shift in hub chord identity from V in Bach to I in Beethoven and back to V in Tchaikovsky raises an interpretively interesting question about the relationship between formal structure and harmonic centrality, although this cannot be fully generalized due to the very limited size of the corpus. The Allegretto's tonic-centric hub likely reflects the constraints of its ternary dance form, in which phrases repeatedly depart from and return to I, rather than a broader claim about Classical-era harmony. This highlights an important methodological point: graph-theoretic measures capture piece-level harmonic tendencies, which are jointly determined by compositional era, genre, and formal structure. Cross-era comparisons must therefore be interpreted with awareness of these confounding factors.

The progressive increase in node set size from Beethoven (6) to Tchaikovsky (9) to Bach (11) — noting that Bach's larger set reflects the chromatic scope of fugal writing rather than Late Romantic expansion — is consistent with the broader musicological narrative of harmonic vocabulary expansion across Western music history, even if the ordering here does not map cleanly onto a linear timeline, but once again, this proposition cannot be taken fully as a definitive conclusion.

B. Limitations

Several limitations of the present study warrant acknowledgment. First, the corpus is restricted to a single piece per composer, which constrains the generalizability of the findings. Harmonic structure varies across genres, movements, and compositional periods within a single composer's output, and the pieces selected may not be representative of each composer's broader harmonic language.

Second, the normalization of chromatic chords — specifically the mapping of augmented sixth chords to IV and cadential six-four figures to V — introduces a degree of analytical simplification. These mappings reflect standard functional interpretations but collapse chromatic distinctions that may be analytically significant, particularly in the Late Romantic corpus.

Third, the model operates exclusively on diatonic triads within a single key context. Passages involving secondary dominants are reduced to their base Roman numeral, discarding information about tonicization and local key context that may be relevant to a fuller account of harmonic structure.

C. Future Work

Several extensions to the present model are worth pursuing. Expanding the corpus to include multiple pieces per composer and per era would substantially improve the robustness of cross-era comparisons. Incorporating secondary dominants and chromatic chords as distinct nodes — rather than normalizing them to diatonic equivalents — would yield a richer and more harmonically precise graph representation. The active versus passive node distinction introduced in graph theory literature may also prove useful for differentiating structurally load-bearing chords from those that appear only as transient passing harmonies.

VIII. CONCLUSION

This paper has presented a directed weighted graph model of chord progressions in Western classical music, applied to three annotated pieces spanning the Baroque, Classical, and Late Romantic periods. By representing chords as nodes and harmonic transitions as weighted directed edges, the model renders the structure of tonal harmony amenable to formal graph-theoretic analysis.

Four analytical methods were applied: degree centrality, cycle detection, strongly connected component decomposition, and degree distribution comparison. The results reveal both invariant properties and era-dependent structural differences. The single strongly connected component found in all three graphs suggests that full harmonic reachability is a stable property of tonal harmony across periods. In contrast, graph density, hub chord identity, cycle count, and degree distribution all vary measurably across composers, reflecting the evolving harmonic priorities of each era.

Bach's graph exhibits a moderately connected structure with V as the primary out-degree hub, consistent with dominant-driven Baroque voice leading. Beethoven's graph is the densest relative to its size, with I functioning as the central harmonic axis — a pattern attributable in part to the formal constraints of the Allegretto's ternary structure. Tchaikovsky's graph is the most cyclically complex and exhibits the strongest dominant gravity of the three, with V dominating both in- and out-degree by a substantial margin.

Taken together, these findings support the claim that directed graph models provide a principled and computationally tractable basis for the comparative analysis of harmonic structure. The approach complements existing music-theoretic frameworks by offering a representation that is both formally precise and amenable to quantitative cross-era comparison. Future work expanding the corpus and relaxing the diatonic normalization constraints outlined in Section VII would further strengthen the analytical scope of this framework.

IX. ATTACHMENTS

[1] Analysis Jupyter Notebook

ACKNOWLEDGEMENTS

The author would like to thank the contributors of the When-in-Rome dataset and the Digital and Cognitive Musicology Lab (DCML) for making their Roman numeral annotations publicly available under open licenses. Harmonic sequence extraction, graph construction, and data processing and visualizations were performed using the music21, NetworkX, matplotlib, and Pandas open-source libraries. This work was completed as a course requirement for IF1220 Discrete Mathematics at Bandung Institute of Technology.

REFERENCES

[1] J. Miller, V. Nicosia, and M. Sandler, "Discovering Common Practice: Using Graph Theory to Compare Harmonic Sequences in Musical Audio Collections," in Proceedings of the 8th International Conference on Digital Libraries for Musicology, in DLfM '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 93–97. doi: 10.1145/3469013.3469025.

[2] D. Tymoczko, M. Gotham, M. S. Cuthbert, and C. Ariza, "The Romantext Format: A flexible and standard method for representing Roman numeral analyses," in Proceedings of the 20th International Society for Music Information Retrieval Conference, 2019, pp. 123–129.

[3] R. J. Trudeau, Introduction to graph theory, Dover ed. in Dover books on advanced mathematics. New York, NY: Dover Publ, 1993.

[4] J. D. Lomas and H. Xue, "Harmony in Design: A Synthesis of Literature from Classical Philosophy, the Sciences, Economics, and Design," She Ji: The Journal of Design, Economics, and Innovation, vol. 8, no. 1, pp. 5–64, 2022, doi: 10.1016/j.sheji.2022.01.001.

[5] E. Nuutila and E. Soisalon-Soininen, "On finding the strongly connected components in a directed graph," Information Processing Letters, vol. 49, no. 1, pp. 9–14, Jan. 1994, doi: 10.1016/0020-0190(94)90047-7.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 19 Juni 2026

Farish Firstian Erifiawan - 13525058