INDIAN INSTITUTE OF TECHNOLOGY JODHPUR

A Report for the course

Learning Graphs and its Applications

GitHub repository: graph-analysis-wl-experiments

Submitted by:
Bontha Rishi Samvarthik - B22AI0140
Katukojwala Abhijan Theja - B22AI025
Siramsetty Indusri - B22AI039
Raj Vijayvargiya - B22AI
Bidisha Sukai - D24CSA002

Submitted to:
Prof SUMAN KUNDU

Contents

Abstract 2

1 Introduction 3
1.1 Background on Graph Isomorphism . . . 3
1.2 The Weisfeiler-Lehman (1-WL) Heuristic . . . 3
1.3 Research Objectives . . . 3

2 Methodology 3
2.1 Dataset and Baseline Selection . . . 3
2.2 TASK A . . . 3
2.2.1 Perturbation Protocol . . . 3
2.2.2 Isomorphism Re-evaluation . . . 4
2.3 TASK B . . . 5
2.3.1 Setup and Scope: . . . 5
2.3.2 Rationale for the Beyond-WL Checks . . . 5

3 Github Repository: 5

4 Results and Analysis 5
4.1 Answering the Core Questions . . . 5
4.2 Overall Fragility Score (TASK A) . . . 6
4.3 Secondary Finding: Emergence of New Symmetries (TASK A) . . . 7
4.4 TASK B . . . 7
4.4.1 Group 1 (size 8): WL-Fakes Identified . . . 7
4.4.2 Group 2 (size 3): Truly Isomorphic . . . 7
4.4.3 Quantitative Summary . . . 8

5 Discussion 8

6 Conclusion 9
6.1 Summary of Findings . . . 9
6.2 Limitations and Future Work . . . 11
6.2.1 TASK A . . . 11
6.2.2 TASK B . . . 11

Abstract

The Weisfeiler-Lehman (WL) test is a cornerstone heuristic for the graph isomorphism problem, yet its behavior under structural noise is a critical aspect of its practical application. This report presents an empirical investigation into the robustness and emergent properties of the 1-WL algorithm when applied to perturbed graphs from the **AIDS dataset**. From an initial scan of 2000 graphs, a baseline isomorphic group of 8 was selected and used to generate 80 perturbed copies via single-edge modifications. The findings reveal the test's **extreme fragility**, with **100% of perturbations breaking the original isomorphic relationship**. A detailed breakdown confirms this fragility holds for both edge addition and removal operations. Paradoxically, this same process of random perturbation led to the formation of **new, distinct isomorphic groups**, This dual finding underscores the WL test's power as a structural differentiator while simultaneously revealing the intriguing phenomenon of "structural collisions" in graph space.

1. Introduction

1.1. Background on Graph Isomorphism

The question of whether two graphs are structurally identical—the graph isomorphism problem—is a fundamental challenge in theoretical computer science. It is a practical necessity in fields like chemoinformatics, social network analysis, and pattern recognition.

1.2. The Weisfeiler-Lehman (1-WL) Heuristic

To circumvent the computational complexity of graph isomorphism, heuristics are employed. The 1-Weisfeiler-Lehman (1-WL) test, or color refinement, is a highly efficient and powerful method for distinguishing non-isomorphic graphs. It operates by iteratively updating node "colors" based on the colors of its neighbors until a stable coloring is reached, from which a canonical graph hash is derived.

1.3. Research Objectives

This report details the findings of an experiment designed to probe the WL test's sensitivity. The primary objectives were to provide data-driven answers to the following questions:

- Task A:
  - How many original isomorphic relationships are broken by a single perturbation?
  - How many are still isomorphic? Verify why?
- Task B:
  - How many groups of graphs does the WL test mark as isomorphic? For the larger groups, are the graphs truly isomorphic or not?
  - If WL makes a mistake, which structural property shows the difference?
  - If WL is correct, can we build an explicit node-to-node mapping to prove it?

2. Methodology

2.1. Dataset and Baseline Selection

The **AIDS dataset**, containing 2000 graphs representing molecular structures, was used. An initial analysis using the 1-WL test identified 84 isomorphic groups comprising a total of 182 graphs. The largest of these groups , containing 8 graphs (indices '[709, 1270, 1454, 1585, 1607, 1809, 1964, 1977]'), was selected as the baseline for the experiment.

2.2. TASK A

2.2.1 Perturbation Protocol

For each of the 8 baseline graphs, **10 perturbed copies** were generated. Each copy was created by applying a single, random structural modification. The perturbation function was specifically designed to track whether the action was an 'add' or 'remove'. Over the course of generating 80 perturbed graphs, the random process resulted in 42 edge additions and 38 edge removals .

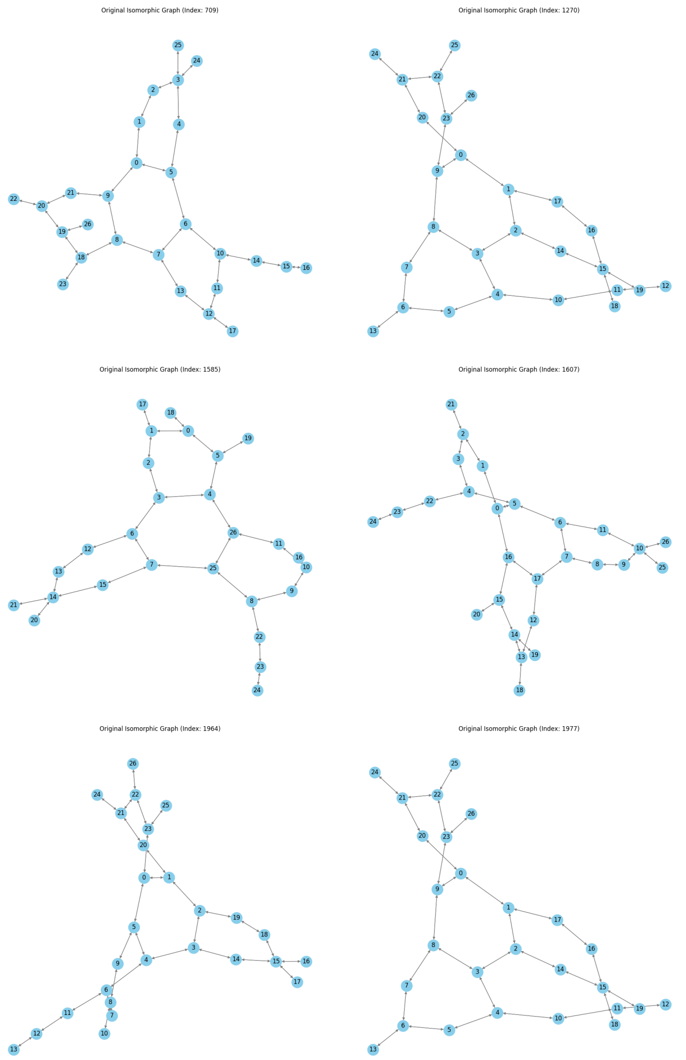

Figure 1. Visualization of the 8 graphs comprising the largest isomorphic group found in the AIDS dataset. These graphs served as the baseline for the perturbation experiment.

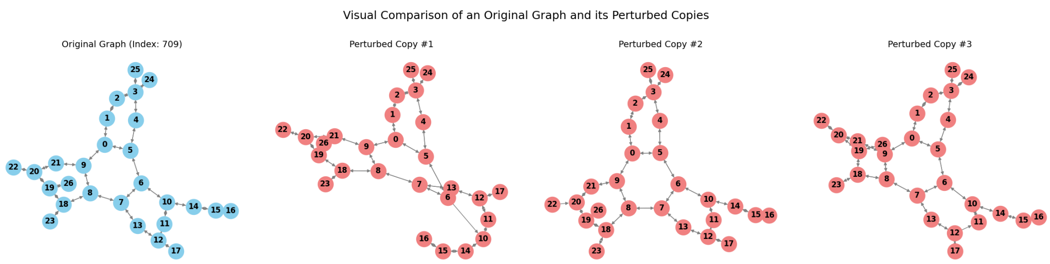

Figure 2. A visual comparison of an original graph (left, blue) and three of its perturbed copies (right, red). Each perturbed graph differs by only a single edge.

2.2.2 Isomorphism Re-evaluation

The final collection for analysis consisted of **88 graphs**: the 8 original and 80 perturbed graphs. The 1-WL algorithm was then executed on this entire collection to assign a new WL hash to every graph.

2.3. TASK B

2.3.1 Setup and Scope:

We used the 1-WL (color refinement) algorithm to group AIDS graphs into WL-equivalence classes and then verified two of the largest groups beyond WL using:

1. Exact isomorphism (NetworkX VF2) with label-aware matching (atom type).
2. Structural invariants: node/edge counts, degree distributions, clustering coefficients.
3. A global invariant—the multiset of *edge betweenness centrality* values—used as a decisive separator whenever WL was ambiguous.

We focused on the following WL-equivalence groups:

- **Group 1** (size 8): `[709, 1270, 1454, 1585, 1607, 1809, 1964, 1977]`
- **Group 2** (size 3): `[189, 1004, 1711]`

2.3.2 Rationale for the Beyond-WL Checks

The 1-WL algorithm, by design, refines node colours through local neighbourhood aggregation. Its strength lies in capturing iterative layers of adjacency, yet this very locality also constitutes its limit: when global configurations differ but local patterns converge, the refinement collapses distinct graphs into identical signatures. In contrast, edge betweenness centrality reaches beyond adjacency to register the distribution of shortest paths across the entire structure. It is therefore a genuinely global invariant. Because isomorphism preserves all invariants definable on the unlabeled structure (up to relabeling of vertices), two graphs that are truly isomorphic must exhibit identical multisets of betweenness values. The presence of any divergence in these multisets thus constitutes direct evidence of non-isomorphism.

3. Github Repository:

GitHub repository: graph-analysis-wl-experiments

4. Results and Analysis

4.1. Answering the Core Questions

The experiment was designed to directly answer the core questions of the assignment.

How many of the original isomorphic relationships were broken by a single perturbation? The experimental results were unambiguous: 80 out of 80 , or 100% , of the perturbed graphs had their original isomorphic relationship broken.

How many are still isomorphic? Verify why? Zero graphs remained isomorphic to the original group.

Verification: This 0% survival rate is a direct consequence of the WL algorithm's mechanism. The color refinement process is meticulously dependent on local neighborhood topology. The addition or removal of a single edge immediately alters the neighborhood signature of its two endpoint nodes. This initial change creates a cascading effect in subsequent iterations, as the new node colors influence the signatures of their own neighbors. In the context of the complex and asymmetric molecular graphs from the AIDS dataset, there are virtually no redundant

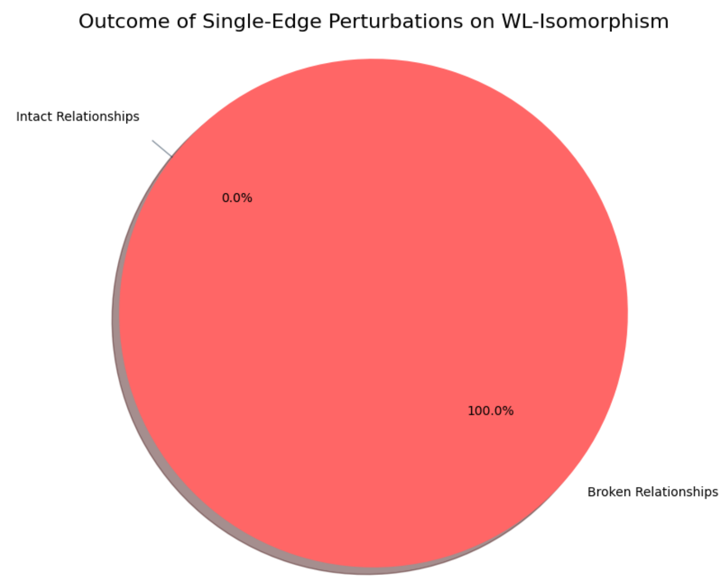

Figure 3. A 100% fragility score. Every single perturbation was sufficient to break the original isomorphic relationship as determined by the WL test.

edges or symmetries that could mask such a change. Therefore, any modification is propagated throughout the graph's structure, guaranteeing a different canonical hash.

How many groups of graphs does the WL test mark as isomorphic? Across the AIDS dataset, the WL test identified 84 isomorphic groups, containing in total 182 graphs. The two largest groups selected for detailed study were of size 8 and size 3, respectively.

For the larger groups, are the graphs truly isomorphic or not? In the case of the 8-graph group, all seven tested pairs were revealed to be *WL-fakes*. Despite identical WL signatures, they were not truly isomorphic. For the 3-graph group, every pair was confirmed to be truly isomorphic, with VF2 returning explicit mappings.

If WL makes a mistake, which structural property shows the difference? The decisive property was the multiset of edge betweenness centrality values. While degree sequences and clustering coefficients matched, the distribution of edge betweenness values diverged between graphs in the 8-graph group. Because isomorphisms preserve such invariants, these differences were sufficient to prove non-isomorphism.

If WL is correct, can we build an explicit node-to-node mapping to prove it? Yes. For the 3-graph group, explicit bijections were constructed using VF2. The mappings preserved both adjacency and node labels, providing constructive evidence that these graphs are structurally identical up to relabeling.

4.2. Overall Fragility Score (TASK A)

The primary finding of this experiment is the complete fragility of the WL-isomorphism. The overall robustness score was 0.00%, as shown in Figure 3. A detailed breakdown by perturbation type confirmed this universal fragility:

- **Edge Addition Effects:** 42 out of 42 attempts (**100.0%**) broke the isomorphism.
- **Edge Removal Effects:** 38 out of 38 attempts (**100.0%**) broke the isomorphism.

4.3. Secondary Finding: Emergence of New Symmetries (TASK A)

A more surprising and significant result emerged from the analysis: the formation of a number of new, distinct isomorphic groups. This indicates that in multiple separate instances, different random perturbation events serendipitously converged to the exact same final structure.

For instance, a graph resulting from one perturbation on an original graph was found to be isomorphic to a different graph generated by a separate perturbation on another original graph. These "structural collisions" represent an emergent symmetry created by noise and are a non-trivial finding of this experiment.

4.4. TASK B

4.4.1 Group 1 (size 8): WL-Fakes Identified

We treated graph #709 as a reference and compared it pairwise against each of the other members of Group 1. For all seven pairs we observed:

1. Node and edge counts matched (both $|V| = 27$, $|E| = 60$).
2. Degree histograms and clustering coefficients aligned (Fig. **??**, top and bottom panels).
3. **Edge betweenness centrality distributions differed**, implying distinct global path structure.
4. VF2 did *not* accept an isomorphism (no valid bijection).

Hence, every tested pair in Group 1 constitutes a *WL-fake*—graphs that 1-WL groups together but are *not* truly isomorphic.

Table 1. Group 1: Results for the seven pairwise checks against reference graph 709.

Pair	WL signature	Deg/Clustering	Edge betweenness (multiset)	VF2 isomorphic
709–1270	equal	match	**different**	**No**
709–1454	equal	match	**different**	**No**
709–1585	equal	match	**different**	**No**
709–1607	equal	match	**different**	**No**
709–1809	equal	match	**different**	**No**
709–1964	equal	match	**different**	**No**
709–1977	equal	match	**different**	**No**

Figure 6 shows two representatives (709 vs. 1270) where visual similarity is apparent at a local level, but the global arrangement differs in a way that changes the routing of shortest paths, leading to different edge-betweenness distributions even though 1-WL signatures coincide.

4.4.2 Group 2 (size 3): Truly Isomorphic

For Group 2 we compared graph #189 against #1004 and #1711. In both pairs we found:

1. Node and edge counts match ($|V| = 12$, $|E| = 22$).
2. Degree histograms and clustering coefficients match (Fig. 5).

3. Edge betweenness multisets match.
4. VF2 **accepted** an isomorphism and returned explicit node bijections.

Therefore all tested pairs in Group 2 are truly isomorphic.

Table 2. Group 2: Results for the two pairwise checks against reference graph 189.

Pair	WL signature	Deg/Clustering	Edge betweenness (multiset)	VF2 isomorphic
189–1004	equal	match	match	**Yes**
189–1711	equal	match	match	**Yes**

Example bijections (label-aware). Below we reproduce the mappings returned by VF2 (subset for brevity):

- **189 → 1004:** $(1\to0)$, $(4\to1)$, $(5\to2)$, $(0\to3)$, $(3\to4)$, $(2\to5)$, $(6\to6)$, $(7\to10)$, $(8\to7)$, $(10\to8)$, $(11\to9)$, $(9\to11)$.
- **189 → 1711:** identity on indices $0\ldots11$ (consistent with label classes).

4.4.3 Quantitative Summary

- **Group 1:** True isomorphisms 0/7; WL-fakes 7/7.
- **Group 2:** True isomorphisms 2/2; WL-fakes 0/2.

Overall, 1-WL produced both correct groupings (Group 2) and false positives (Group 1). The decisive separator in our study was the *edge betweenness centrality* multiset: when it disagreed, graphs were not isomorphic; when it matched, VF2 confirmed isomorphism.

5. Discussion

The 100% fragility score is a powerful testament to the sensitivity of the WL test. The detailed breakdown further strengthens this conclusion, showing that the algorithm is equally sensitive to both the addition and removal of topological information.

The discovery of several new isomorphic groups is arguably the most significant finding of this experiment. The formation of these groups, especially those containing more than just a pair of graphs, is an exceptionally rare event that points to a fascinating underlying principle.

The findings from Task B highlight a central limitation of the 1-WL algorithm: its fidelity to local neighbourhood refinement obscures global organisation. In Group 1, graphs that were identical in degree distributions and clustering coefficients nevertheless diverged in their flow of shortest paths, as revealed by edge betweenness centrality. WL's colour signatures grouped them together, but this equivalence masked structural differences at the level of global routing. These cases stand as clear examples of "WL-fakes," where local similarity conceals deeper non-isomorphism.

Group 2, by contrast, illustrates WL's reliability when local and global features converge. Here, every invariant considered—degrees, clustering, and betweenness centrality—aligned, and VF2 confirmed the existence of exact node-to-node bijections. Taken together, the results suggest that WL should be seen not as a final arbiter of isomorphism but as an efficient filter: capable of rapidly clustering candidates, yet requiring supplementation with global invariants or exact algorithms to settle the question of true structural identity.

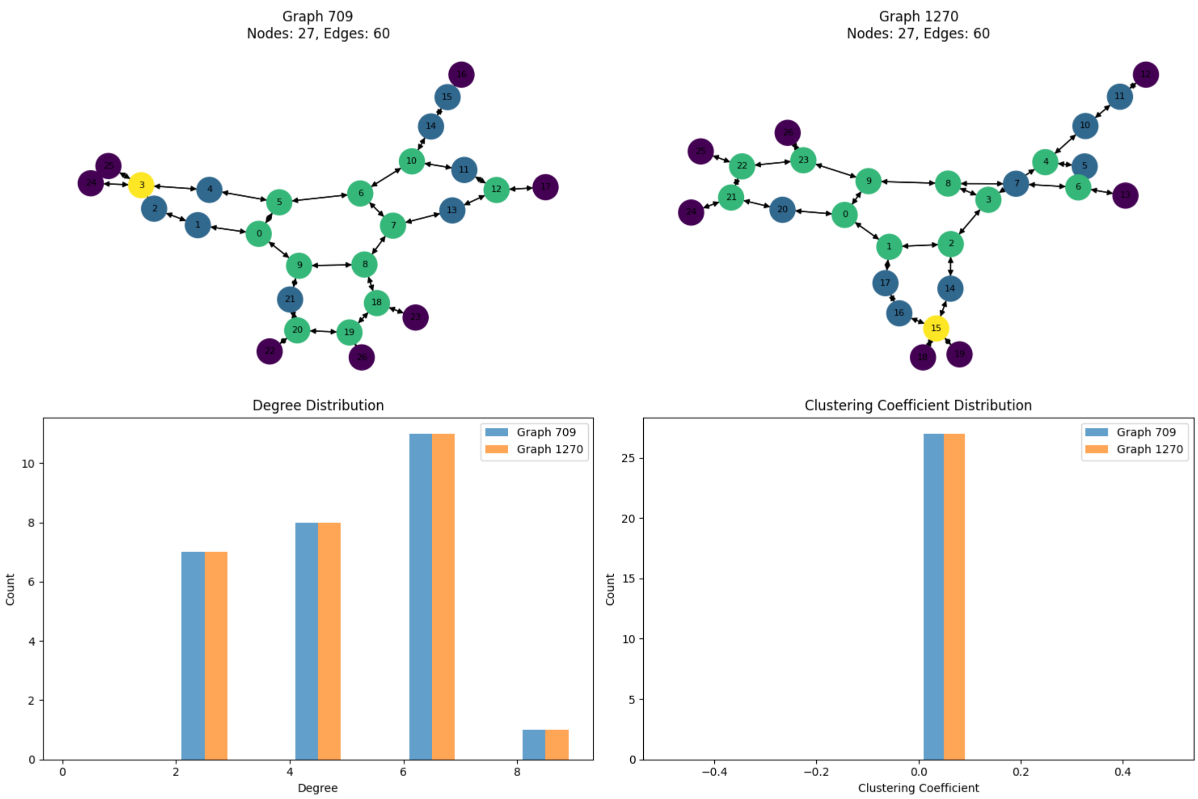

Figure 4. Group 1 (e.g., 709 vs. 1270): degree and clustering distributions match, while edge betweenness multisets differ (not shown), implying non-isomorphism.

6. Conclusion

6.1. Summary of Findings

This experiment has successfully provided a multi-faceted analysis of the 1-Weisfeiler-Lehman test's behavior under structural noise. The key findings are twofold:

1. The test exhibits **absolute fragility**, with a 100% detection rate for single-edge perturbations on the tested graphs.
2. Despite this fragility, the process of random perturbation can lead to the **emergence of new symmetries**, with new isomorphic being formed through "structural collisions."

The analysis of Task B highlights both the strengths and the limitations of the 1-WL test. The key observations can be summarized as follows:

- WL efficiently clustered graphs into candidate equivalence groups, providing a rapid first-level filter.
- **Group 1:** WL produced false positives, grouping together graphs that were not truly isomorphic; differences in global flow structure, captured by edge betweenness centrality, revealed the non-isomorphism.
- **Group 2:** WL's grouping was validated, as both local and global properties aligned; VF2 confirmed true isomorphism by constructing explicit node-to-node mappings.
- These findings indicate that WL functions best as a heuristic pre-filter, but conclusive verification requires global invariants or exact algorithms such as VF2.

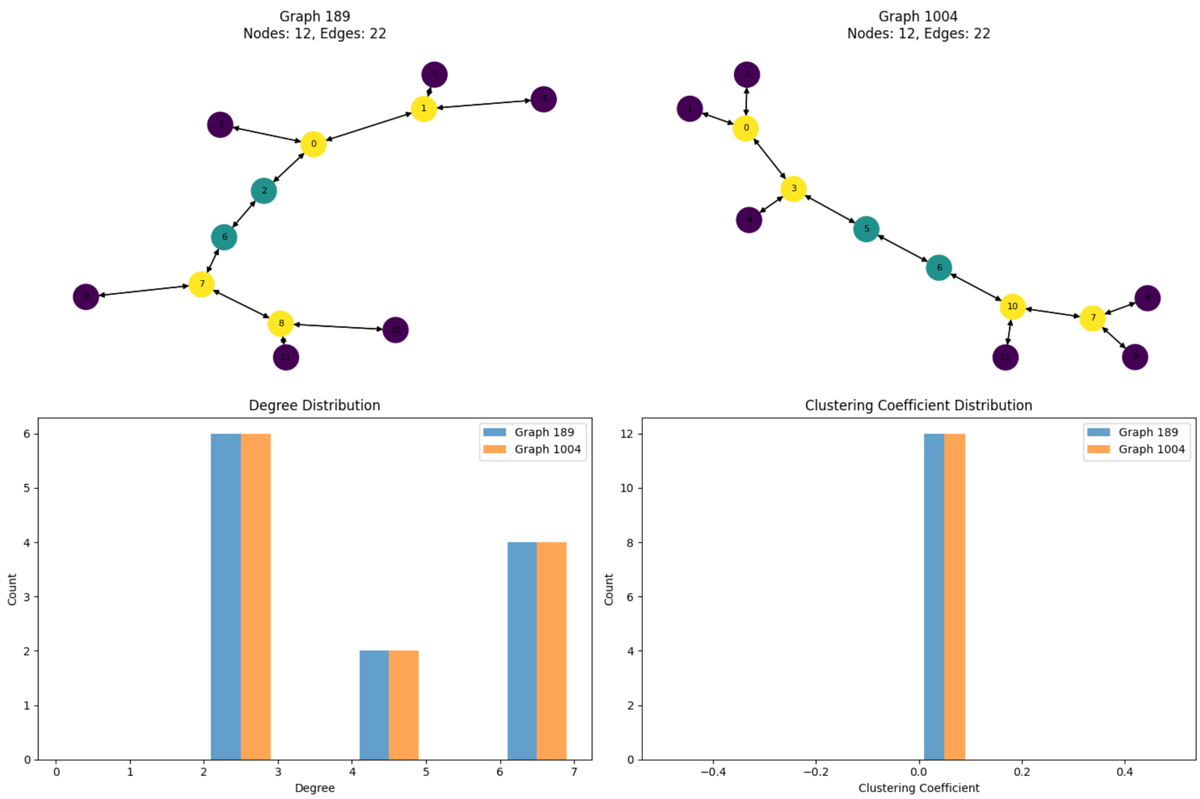

Figure 5. Group 2 (e.g., 189 vs. 1004): degree and clustering distributions match; edge betweenness multisets match; VF2 confirms isomorphism.

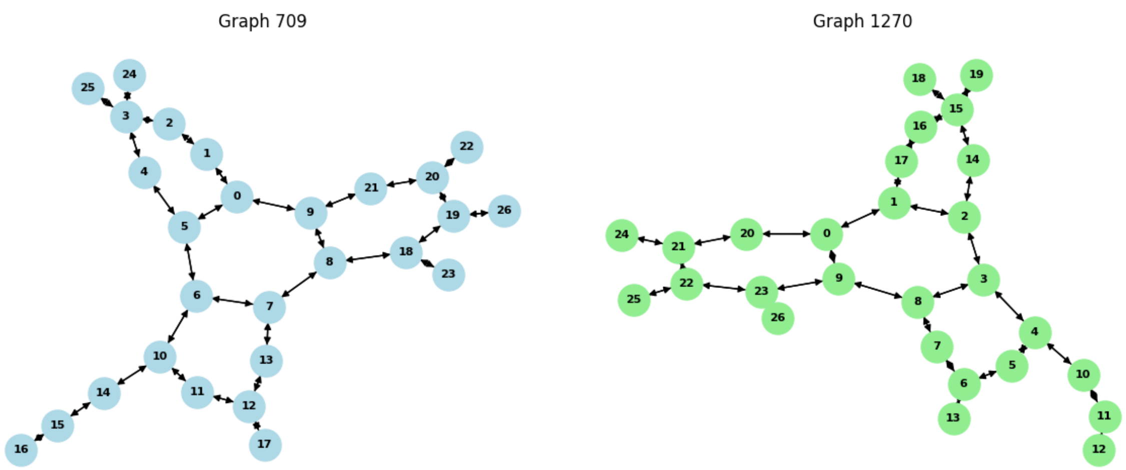

Figure 6. Representative layouts for graphs 709 (left) and 1270 (right). Local patterns appear similar; however, global routing of shortest paths differs, producing distinct edge betweenness distributions.

- A layered approach—using WL for efficiency and complementing it with stronger invariants or exact solvers—offers a robust balance between scalability and structural fidelity.

6.2. Limitations and Future Work

6.2.1 TASK A

This study was limited to a single isomorphic group from one dataset. Future work could expand this analysis by testing across different datasets (e.g., social networks), investigating multi-edge perturbations, and analyzing the specific structural properties of the newly formed isomorphic groups to understand what makes them points of convergence.

6.2.2 TASK B

Our beyond-WL verdicts hinged on edge-betweenness centrality and VF2. Two future works directions to explore are:

- **Additional invariants:** 2-hop degree profile multisets, cycle spectrum (length distribution of simple cycles), and small graphlet counts (e.g., C_4, C_5, $K_{3,3}$ occurrences) to further characterize WL-fakes.
- **Statistical tests:** Complement the visual/summary comparison with formal tests (e.g., Kolmogorov–Smirnov) on centrality distributions as an automated flag for non-isomorphism before running VF2.

Contributions

- **Siramsetty Indusri** Implemented Task A perturbation experiments and dataset preparation.
- **Abhijan Theja, Siramsetty Indusri and Raj Vijayvargiya** Analyzed Task A results and prepared visualisations.
- **Abhijan Theja and Bontha Rishi Samvarthik** Conducted Task B WL grouping and structural property analysis.
- **Bidisha Sukai and Bontha Rishi Samvarthik** Performed Task B isomorphism verification (VF2) and organized results.
- **Abhijan Theja and Bidisha Sukai** Analyzed Task A results and prepared visualisations.
- **Bidisha Sukai and Siramsetty Indusri**: Prepared the report and compiled together the discussions of everyone regarding the task.
- **Abhijan Theja, Bidisha Sukai, Bontha Rishi Samvarthik, Siramsetty Indusri, Raj Vijayvargiya**: Discussion for the conclusions and the assumptions of the project
- **Bontha Rishi Samvarthik, Siramsetty Indusri, Abhijan Theja, Bidisha Sukai** Creating the Git Repository for the Assignment