# How to Identify Structural Plagiarism

Structural Criteria Beyond Textual Similarity

Kaifan Xie

2026-02-05

**Abstract**

Textual plagiarism detection has become increasingly automated, yet a large class of intellectual reuse remains largely invisible: the reuse of generative structure. In many technical and theoretical works, originality does not primarily reside in wording, but in how problems are formulated, how reasoning is organised, and how conclusions are generated. This article proposes a framework for identifying *structural plagiarism*, defined as the un-attributed reuse of non-inevitable generative structure. We introduce a formal representation of research structure, define structural equivalence and non-inevitability, and present operational criteria that allow experts to assess whether observed alignment is more plausibly explained by inheritance than by independent generation.

## 1 Introduction

Most existing plagiarism detection systems focus on surface-level similarity: string overlap, syntactic patterns, or citation matching. These approaches implicitly assume that text is the primary carrier of intellectual contribution. While this assumption may be adequate for narrative or discursive writing, it becomes increasingly fragile in technical and theoretical research.

In mathematics, theoretical computer science, and many areas of systems research, the core intellectual labour lies elsewhere: in how a problem is formalised, which constraints are imposed, how intermediate results are organised, and why conclusions take a particular form. Two works may share little or no textual similarity, yet exhibit striking alignment in their underlying generative mechanisms.

This phenomenon is commonly described informally as "copying the structure but not the text". Despite its intuitive recognisability among experts, it lacks a clear formal treatment. As a result, such cases are often dismissed as "coincidental similarity" or "shared background", even when the alignment involves highly specific and non-standard design choices.

This article does not approach structural plagiarism as a moral accusation or legal claim. Instead, it frames the issue as a problem of explanation: given a set of observed structural alignments, is independent generation still the simplest and most plausible model, or does structural inheritance provide a lower-cost explanation?

## 2 Formal Framework

To reason about structural plagiarism, we first require a representation that separates surface expression from generative structure. We model research works at a level of abstraction sufficient to capture their core intellectual mechanisms.

**Definition 1** (Research Structure)**.** *A research work $W$ is represented as a triple*

$$W := (P, R, C),$$

*where:*

- *$P$ denotes the* problem structure*, including formalisation choices, variable selection, and imposed constraints;*
- *$R$ denotes the* reasoning structure*, including assumptions, intermediate lemmas, ordering of arguments, and dependency relations;*
- *$C$ denotes the* conclusion structure*, including result form, strength hierarchy, and boundary conditions.*

This decomposition reflects a common pattern in technical research: the same result may be reachable through different problem formulations, different reasoning paths, and different ways of presenting conclusions.

**Definition 2** (Structural Abstraction)**.** *For each level $L \in \{P, R, C\}$, let*

$$\alpha_L : L \to \widehat{L}$$

*map a surface representation to an abstract structural skeleton. The abstraction $\alpha_L$ ignores symbol names, phrasing, and superficial reordering, while preserving dependency and generative relations.*

Intuitively, $\alpha_L$ extracts "what depends on what" and "what generates what", discarding stylistic or notational variation.

**Definition 3** (Structural Equivalence)**.** *Two works $W_1$ and $W_2$ are* structurally equivalent *at level $L \in \{P, R, C\}$, denoted $W_1 \equiv_L W_2$, if there exists an isomorphism*

$$\phi_L : \alpha_L(L_1) \to \alpha_L(L_2)$$

*that preserves dependency and generative relations.*

Structural equivalence does not require identical wording or notation. It asserts that the same abstract structure is being realised in two works.

**Definition 4** (Non-Inevitable Structure)**.** *A substructure is* non-inevitable *if it admits viable alternatives that are standard or widely accepted, and if its selection reflects author-specific modelling or reasoning choice rather than logical necessity.*

Non-inevitability is central: alignment in inevitable components (e.g. basic definitions or standard lemmas) carries little evidentiary weight, whereas alignment in non-inevitable components is highly informative.

**Definition 5** (Structural Plagiarism)**.** *A work $W_2$ exhibits* structural plagiarism *with respect to $W_1$ at level $L$ if a non-inevitable substructure of $\alpha_L(L_1)$ appears within $\alpha_L(L_2)$ without explicit structural attribution.*

**Proposition 1** (Multi-Level Convergence)**.** *If structural plagiarism holds at two or more distinct levels $L \in \{P, R, C\}$, the explanatory cost of the independent-generation hypothesis increases multiplicatively.*

**Proposition 2** (Redundancy Amplification)**.** *Structural equivalence concentrated in redundant or suboptimal components carries greater evidentiary weight than equivalence concentrated in forced core steps.*

# 3 Operational Criteria for Structural Plagiarism

The formal definitions above specify what constitutes structural plagiarism in principle. In practice, however, experts rarely have complete information or the ability to explicitly construct the isomorphism $\phi_L$. The following criteria function as observable signals that operationalise the formal system.

## 3.1 Criterion 1: Non-Inevitable Problem-Structure Alignment

If a problem admits multiple reasonable formulations, yet two non-citing works select the same non-standard decomposition, constraint scheme, or variable reduction, this alignment constitutes a strong structural signal.

Such choices are rarely forced. When they coincide, coincidence becomes an expensive explanation.

## 3.2 Criterion 2: Isomorphic Ordering of Reasoning Steps

When key assumptions, lemmas, and intermediate results appear in the same non-forced order, the ordering itself acts as a structural invariant.

Independent derivations tend to differ in ordering unless constrained by necessity.

## 3.3 Criterion 3: Replication of Suboptimal Reasoning Detours

Optimal reasoning paths are often predictable. Suboptimal yet stable detours, such as unnecessary intermediate constructions, typically reflect individual preference. Their replication strongly suggests inheritance.

## 3.4 Criterion 4: Synchronous Reproduction of Pathological Constructions

Highly specific or pathological constructions have low prior probability under independent design. Their repeated appearance with identical structural roles is therefore difficult to explain without reuse.

## 3.5 Criterion 5: Shared Preservation of Redundant Assumptions

Redundant assumptions or removable steps are the least likely to align by chance. Shared redundancy is thus among the strongest signals of structural reuse.

## 3.6 Criterion 6: Non-Natural Alignment of Result Forms

When multiple result formulations are available, repeated selection of the same non-most-natural form indicates inheritance of generative constraints, not merely similarity of outcomes.

## 3.7 Criterion 7: Multi-Criteria Convergence

A single criterion may admit coincidence. When several criteria hold simultaneously, especially at the same structural level, independent generation rapidly loses plausibility.

### 3.8 Criterion 8: Robustness to Textual and Symbolic Variation

If structural equivalence persists after rewriting, renaming, or reordering, the similarity must reside in structure rather than expression.

## 4 Relation Between Formal System and Operational Criteria

The formal system specifies *what* structural plagiarism is. The criteria specify *how* it manifests under realistic observational constraints.

Each criterion corresponds to a partial invariant of the abstract skeleton $\widehat{L}$. Together, they allow experts to assess explanatory adequacy: whether independent generation or structural inheritance provides the simpler account.

## 5 Conclusion

Structural plagiarism operates at the level of generative mechanisms, not surface text. As paraphrasing, symbolic variation, and automated rewriting tools become increasingly common, textual similarity alone is no longer a reliable proxy for originality.

By shifting attention from wording to structure, the framework presented here offers a principled way to reason about intellectual reuse without relying on intent, ethics, or legal adjudication. Its goal is modest but essential: to make structural similarity visible, discussable, and analyzable.

## Keywords

Structural plagiarism; research structure; structural equivalence; non-inevitable structure; generative mechanisms; originality