ITEM #254 - Two-Ways CCC over Starmap Spaces: A General Comparative Study Engine for Structural Signals

Conversation: Two-Ways Graph CCC

20260202

Authors: Sizhe Tan & GPT-Obot

DBM-COT ITEM #254 (EN)

Two-Ways CCC over Starmap Spaces: A General Comparative Study Engine for Structural Signals

1. Motivation and Positioning

Modern comparative studies across science, engineering, and society increasingly face a common limitation:
most real-world signals are **structural**, **contextual**, and **multi-dimensional**, yet prevailing analytical tools remain dominated by **one-dimensional statistics** or **flattened feature vectors**.

Within the DBM framework, *Starmap spaces*—including **GraphStarmap**, **SequenceStarmap**, and **ImageStarmap**—form a unified representational substrate for world models. These spaces preserve relational structure, compositional context, and locality, while remaining compatible with systematic IR generation.

This ITEM proposes that:

Two-Ways CCC over Starmap Spaces constitutes a general, powerful, and automatable comparative study engine for detecting structural signals across domains.

The approach unifies:

- Differential Trees
- Two-Phase Search

- CCC (Common Concept Core)
- Metric Distance
  into a single comparative methodology operating over structural spaces rather than scalar features.

---

2. Conceptual Overview

At its core, **Two-Ways CCC** transforms comparative analysis from:

"Do two groups differ on average?"
into
"Which structural patterns systematically distinguish these groups, and how?"

The method operates over **families of CCC fragments**, not single measurements, and explicitly contrasts **in-group** and **out-group** structure using a symmetric, two-directional test.

This symmetry is essential:

- One-way correlation is insufficient for signal validation.
- Two-way contrast enforces discriminative power and guards against spurious structure.

---

3. Algorithmic Framework

3.1 Structural Space and Inputs

Let each sample be represented as a Starmap object:

- Gi∈{GraphStarmap,SequenceStarmap,ImageStarmap}G_i \in \{\text{GraphStarmap}, \text{SequenceStarmap}, \text{ImageStarmap}\}Gi ∈{GraphStarmap,SequenceStarmap,ImageStarmap}

Each sample is associated with:

- RHS attributes yiy_iyi (binary or multi-class)
- Optional metadata for stratification or control

3.2 Anchor-Driven Structural Localization

Within the global Starmap space:

1. **Anchor Generation**
   - Manually defined (domain knowledge)
   - Automatically mined (hotspots, centrality, anomaly)

    - Hybrid strategies are encouraged
2. **Two-Phase Search**
    - Phase-1: Efficient candidate occurrence retrieval
    - Phase-2: Metric refinement and pruning
3. **Local Area Extraction**
    - Anchor occurrences induce localized structural areas
    - Areas preserve topology, order, and relational context

3.3 CCC Family Construction

From extracted areas:

- Compute **CCC fragments** (subgraphs, motifs, structural patterns)
- Retain evidence chains linking:
  anchor → occurrence → area → CCC fragment

Each sample yields a **CCC family**, not a single core.

3.4 Two-Ways Comparative Test

Partition samples into:

- In-group I\mathcal{I}I
- Out-group O\mathcal{O}O

For each candidate CCC signal sss:

- **In-group coverage**

  Cin(s)=Pr(s∈Fi|i∈I)C_{in}(s) = \Pr(s \in F_i \mid i \in \mathcal{I})Cin(s)=Pr(s∈Fi|i∈I)

- **Out-group leakage**

  Cout(s)=Pr(s∈Fi|i∈O)C_{out}(s) = \Pr(s \in F_i \mid i \in \mathcal{O})Cout(s)=Pr(s∈Fi|i∈O)

Signal strength is derived from contrast functions such as:

- Difference: Cin−CoutC_{in} - C_{out}Cin−Cout
- Log-ratio: log Cin+ϵCout+ϵ\log\frac{C_{in}+\epsilon}{C_{out}+\epsilon}logCout+ϵCin+ϵ
- Distance-weighted or stability-weighted variants

A **Signal CCC** is defined as a structural pattern that:

- Recurs consistently within the in-group
- Is suppressed or structurally distant in the out-group
- Remains stable across sampling and anchor perturbations

---

4. Structural Generalization of Classical Statistics

Two-Ways CCC represents a direct generalization of classical statistics into structural space:

Classical Statistics	Structural Counterpart
Mean	Signal CCC Prototype
Std. Deviation	Metric Distance to Signal CCC
Distribution	CCC Family Distribution
Outlier	Structural Anomaly Fragment

This transition enables **comparative studies over graphs, sequences, images, and hybrid IRs**, without collapsing structure into brittle scalar summaries.

---

5. Why This Engine Is Powerful

The power of Two-Ways CCC derives from five properties:

1. **Structure-Preserving**
   No forced vectorization or semantic flattening.
2. **Symmetric and Discriminative**
   True signals must pass both in-group enrichment and out-group exclusion.
3. **Automatable**
   Anchor generation, search, and comparison can be systematized.
4. **Interpretable**
   Each signal carries a concrete structural explanation chain.
5. **Domain-Agnostic**
   Applicable to biology, finance, software systems, social networks, and beyond.

---

6. Engineering Considerations

Key engineering controls include:

- Anchor explosion suppression
- Area and fragment budgets
- Stop-rules for convergence and stability
- Evidence chain size limits

These controls ensure scalability without sacrificing signal fidelity.

---

7. Application Scope

Potential applications include:

- Functional DNA and regulatory signal detection
- Financial fraud and behavioral risk modeling
- Software fault localization and dependency analysis
- Social and organizational structure comparison
- Scientific knowledge graph validation

In all cases, the task is reframed as:

Structural signal detection via two-way comparative CCC analysis.

---

8. Conclusion

Two-Ways CCC over Starmap Spaces elevates comparative study from numeric testing to **structural signal discovery**.

It provides DBM with a general-purpose engine capable of transforming expert-driven comparative reasoning into **systematic, explainable, and scalable AI computation**.

This marks a critical step toward automated structural intelligence.

DBM-COT ITEM #254（中文）

双向 CCC 在 Starmap 空间上的通用对照研究引擎

1. 动机与定位

当代各领域的对照研究正在遭遇一个共同瓶颈：

真实世界中的关键信号往往是结构性的、情境相关的、多维的，而主流方法仍停留在一维统计量或扁平特征之上。

在 DBM 体系中，**GraphStarmap / SequenceStarmap / ImageStarmap** 构成了统一的世界模型表达空间。
它们既保留了结构、关系与局部性，又能通过规范化 IR 自动生成。

本 ITEM 提出并确立：

双向 CCC 在 Starmap 空间上的计算，是一种通用、强力、可自动化的结构信号对照研究引擎。

---

2. 方法总览

Two-Ways CCC 的核心转变是：

从“数值是否不同”，
升级为“哪些结构模式在组内稳定出现、并能区分组间差异”。

该方法：

- 以 **CCC 族群** 而非单一结果为分析对象
- 通过 **双向对照**（In-group / Out-group）验证结构信号的判别性
- 天然避免单向相关性带来的错判

---

3. 算法框架

3.1 Starmap 空间与输入

- 每个样本：Graph / Sequence / Image Starmap
- RHS Attributes：用于分组或标签定义

3.2 Anchor 驱动的结构定位

在全局 Starmap 空间中：

1. 生成 Anchors（人工 / 自动 / 混合）
2. 通过 Two-Phase Search 定位 occurrences
3. 由 occurrences 切割局部结构区域（Areas）

3.3 CCC 族群生成

- 在局部 Areas 内计算 CCC fragments
- 保留完整解释链：
  Anchor → Occurrence → Area → CCC Fragment

每个样本不再是一个点，而是一个 **CCC 族群分布**。

3.4 双向 CCC 对照测试

将样本分为：

- In-group
- Out-group

对任一候选 CCC：

- 计算组内覆盖率
- 计算组外泄露率
- 通过差值、比值或距离函数评估信号强度

Signal CCC 必须同时满足：

- 组内稳定出现
- 组外显著稀少或结构上遥远
- 对采样与 anchor 扰动具有稳定性

---

4. 对经典统计的结构化推广

传统统计	结构空间对应
均值	信号 CCC 原型
标准差	到信号 CCC 的度量距离
分布	CCC 族群形态
异常点	结构异常片段

这使得对照研究第一次能够在**结构空间**中完成，而非被迫降维。

---

5. 为什么它是“强力”的

- 保留结构，不压扁世界
- 双向判别，防止伪相关
- 全流程可系统化
- 结果可解释、可审计
- 与领域无关，可迁移

---

6. 工程与规模控制

- Anchor 爆炸抑制
- Area / CCC 预算
- 收敛与稳定性 Stop Rules
- 解释链大小控制

确保该方法既“强”，又“可用”。

---

7. 应用前景

- DNA 功能信号与调控机制

- 金融欺诈与行为模式
- 软件系统结构与故障根因
- 社会网络与组织行为
- 科学知识图谱与因果结构

---

8. 总结

双向 CCC 在 Starmap 空间上的通用对照研究引擎，
标志着 DBM 从数值智能迈向 **结构智能** 的关键一步。

它将长期依赖专家经验的比较研究，转化为
系统化、可解释、可扩展的结构计算智能。

DBM-COT ITEM #254 — Section 9 (EN)

9. Evolution Layer Amplification: From Structural Statistics to Structural Scientific Method

9.1 Why an Evolution Layer is the Missing Multiplier

Two-Ways CCC over Starmap spaces already upgrades classical scalar statistics into **structural statistics**:
signal prototypes, metric deviations, and CCC family distributions.

However, comparative studies in real science and engineering are not single-pass computations.
They are iterative:
hypothesis → measurement design → counterexample pressure → refinement.

DBM's evolution layer turns Two-Ways CCC from a structural analyzer into an **iterative structural scientific method engine**.

In DBM terms, the engine no longer outputs only a *Signal CCC Package*, but also an *Experiment Proposal*:

- where to look next (anchors),
- how to cut next (area policies),
- when to stop next (stop-rule tuning),
- how to score next (contrast/stability weighting),

- how to stratify next (avoid confounding).

9.2 What Gets Evolved (Four Primary Evolving Objects)

1. **Anchor policies (observation coordinate system)**
   - evolve from domain/statistical anchors into *max-discriminative*, *min-confusion*, *max-stability* anchors.
2. **Area cutting + stop-rule contracts (scalability and signal-to-cost)**
   - evolve budgets and early-stop thresholds to prevent explosion while preserving signal fidelity.
3. **Signal scoring functions (from diff to structural tests)**
   - evolve contrast functions: log-ratio, distance-weighted, stability-weighted, strata-aware aggregation.
4. **Signal CCC archive (mechanism catalog)**
   - stable signals become reusable templates and priors,
     accelerating future comparative studies and improving interpretability.

9.3 Selection Pressure: Fitness as "Structural Reproducibility × Discriminability ÷ Cost"

Evolution requires a fitness definition that reflects DBM values:

- **Discriminability** (Two-Ways contrast between in/out groups)
- **Stability** (top signals reproducible under resampling/anchor perturbations)
- **Cost** (budgets, time, explosion penalties)

A minimal fitness function can be:

Fitness=α·Separability+β·Stability−γ·CostFitness = \alpha \cdot Separability + \beta \cdot Stability - \gamma \cdot CostFitness=α·Separability+β·Stability−γ·Cost

where separability is typically derived from top Signal CCC scores per stratum.

9.4 Minimal Runtime Integration Contracts

The evolution layer can be minimal and still powerful with three slots:

- **EvolutionPolicy**: propose the next EngineConfig and/or anchor rules based on previous results.
- **FitnessEvaluator**: compute fitness from engine outputs (including per-stratum signals).
- **SignalArchive**: store best configurations and stable signal templates as reusable priors.

This completes the loop:
run → score → propose → archive → run …,
transforming comparative analysis into iterative, self-improving structural research.

DBM-COT ITEM #254 — Section 9（中文）

9. 演化层加持：从结构统计到结构科学方法引擎

9.1 为什么演化层是“缺失的乘法器”

Two-Ways CCC 已经把一维统计（mean/std/dist）推广到了结构统计：
信号 CCC 原型、到信号的度量偏差、CCC 族群分布。

但真实世界的对照研究从来不是一次性计算，而是迭代过程：
假设 → 实验/测量设计 → 反例压力测试 → 修正与收敛。

DBM 的演化层让 Two-Ways CCC 从“结构分析器”升级为：
可迭代的结构科学方法引擎。

也就是说，引擎输出的不只是 Signal CCC Package，
还会输出下一轮的 Experiment Proposal（实验计划）：

- 下一轮看哪里（anchors）
- 下一轮怎么切（areas）
- 下一轮何时收手（stop rules）
- 下一轮怎么评分（contrast/stability 权重）
- 下一轮怎么分层（strata，避免混杂）

9.2 演化对象：四类最关键可演化变量

1. **Anchor 策略的演化（观察坐标系）**
   从人工/统计热点，演化为最大区分、最小混淆、最大稳定的 anchors。
2. **Area Cutting + Stop-Rule 合约的演化（规模化关键）**
   动态学习预算、早停阈值，在控制爆炸的同时保住信号。
3. **评分函数的演化（从差值到结构检验）**
   发展 log-ratio、distance-weight、stability-weight、strata-aware 聚合等。

4. **Signal CCC Archive 的演化（机制模板库）**
   稳定信号沉淀为可复用模板与先验，反哺下一轮搜索与解释。

9.3 选择压力：Fitness = 可区分 × 可复现 ÷ 成本

演化必须定义 fitness，且符合 DBM 的价值观：

- 可区分（Two-Ways 对照差异）
- 可复现（重采样/扰动下稳定）
- 成本（预算、耗时、爆炸惩罚）

最小形式可写为：

Fitness=α·Separability+β·Stability−γ·CostFitness = \alpha \cdot Separability + \beta \cdot Stability - \gamma \cdot CostFitness=α·Separability+β·Stability−γ·Cost

9.4 最小运行时插槽：三件套即可闭环

- **EvolutionPolicy**：基于上一轮结果提出下一轮 EngineConfig / anchors 策略
- **FitnessEvaluator**：计算 fitness（支持按 stratum）
- **SignalArchive**：沉淀最佳配置与稳定信号模板，作为先验反哺

形成闭环：
run → score → propose → archive → run → …
把“对照分析”升级为“可自我改进的结构研究引擎”。
