arXiv:2512.20311v1 [cs.CG] 23 Dec 2025

基于动机持久上同调的可解释图特征算法

Yoshihiro Maruyama

名古屋大学信息学部，日本
maruyama@i.nagoya-u.ac.jp

摘要。 我们提出了色多项式持久性算法（CPA），这是一种通过图形超平面排列计算加权图的持久上同调特征的事件驱动方法；图形超平面排列是计算几何中的经典对象。我们建立了严格的复杂度结果：CPA 在最坏情况下具有指数复杂度，在树宽参数下是固定参数可解的，并且对于树、圈和串并联图等常见图族，其复杂度接近线性。最后，我们通过对类分子图结构开展受控实验，展示了该算法在实践中的适用性。

关键词： 计算几何 · 持久同调 · 图

1 引言

持久同调（PH）已成为拓扑数据分析的基石，为从数据中提取多尺度拓扑特征提供了坚实的数学框架 [10, 11, 35, 3]。它能够稳健地概括结构信息，因此已成功应用于计算几何、机器学习、生物学和网络科学等众多领域。这些进展使持久同调既成为一种理论范式，也成为分析复杂数据结构的实用工具。

然而，标准的持久同调流程只捕捉过滤过程中同调群的秩，因而舍弃了有关空间的更精细代数信息。在许多情形下，尤其是过滤来自代数或组合构造时（例如，由超平面排列或图上的构造所产生的空间），我们可以通过记录能够细化底层不变量的附加结构，来丰富经典条形码。本文针对图开发了这一思想的专门版本，并探讨其算法层面的影响。

我们介绍色多项式持久性算法（*CPA*），这是一种用于分析加权图的事件驱动方法。CPA 沿着自然的阈值过滤处理图——按权重顺序逐条添加边——并在每一步计算：（i）一个概括图在该阈值下整体结构的多项式；（ii）一个跃迁，即插入新边时发生的上同调变化。该算法仅在实际“事件”（边插入）发生时进行计算，因此概念上简单且计算效率高。

这一方法依赖于图论和排列理论中的两个经典结果。第一个结果将图 H 的色多项式 $\chi_H(q)$与其相关超平面排列 $[25, 26, 8, 1]$, 的补空间的庞加莱´多项式联系起来，使图着色能够直接刻画拓扑不变量。事实上，在图形排列所涉及的霍奇–泰特情形中，同一恒等式通过直接特化 $E(M(H);\, u,\ v) = \chi_H(uv)$。即可得到霍奇–德利涅多项式 $E(M(H);\, u,\ v)$。第二个要素是删除–收缩恒等式，它描述了添加或收缩一条边时这些不变量如何变化 $[8]$。结合起来，这些结果为 CPA 奠定了可计算的组合基础：每次更新都归结为计算色多项式并应用代数修正。我们给出了色多项式持久性算法的复杂度保证：最坏情况下为指数时间；对树宽而言是固定参数可解（FPT）；对于树、环和串并联图等常见图族，则接近线性 $[14, 24, 2, 5]$。

具体而言，问题设定如下。设 $G = (V, E, w)$ 是一个边权重互不相同的有限加权图 $t_1 < \cdots < t_m$。记 $H_j := G_{\le t_j}$ 为阈值子图，并令 e_j 为第 j 步加入的唯一一条边。对于每个 H_j，我们考虑图形超平面排列 $\mathcal{A}(H_j) \subset \mathbb{C}^{|V|}$ 及其补集 $M(H_j)$。目标是在所有阈值下计算 $M(H_j)$ 的霍奇–德利涅多项式 $E(M(H_j); u, v)$，以及加入 e_j 所产生的跃迁。在霍奇 – 泰特情形下，这等价于对每个阈值求色多项式 $\chi_{H_j}(uv)$；在算法上，我们只计算收缩后的子式 $\chi_{H_j/e_j}(q)$，并通过删除–收缩递推关系 $E_j(u, v) = E_{j-1}(u, v) - \chi_{H_j/e_j}(uv)$ 更新 E_j。

我们解决该问题的方法可概括如下。我们利用上述两个恒等式，将问题转化为离散的事件驱动计算。(i) chromatic→Poincar´e 恒等式（命题 2）将配置补空间的拓扑转化为着色的组合结构，因此在每个阈值处，我们得到 $E(M(H_j); u, v) = \chi_{H_j}(uv)$。(ii) 删除 – 收缩跳跃恒等式（定理 1 ）表达了添加一条边所导致的变化；在E–多项式层面， $E(M(H_{j-1});\, u, v) - E(M(H_j);\, u, v) = E(M(H_j/e_j);\, u, v)$。因此，在算法上，我们只需处理事件（边插入）：在第 j 步，我们计算 $\chi_{H_j/e_j}(q)$，并通过删除-收缩递推关系$E_j(u, v) = E_{j-1}(u, v) - \chi_{H_j/e_j}(uv)$ 更新 E_j，同时应用条形码 ζ 更新这一生成函数，以紧凑的形式编码所有跃迁。由此得到的 CPA 程序具有可证明的保证：最坏情况下为指数复杂度，在树宽方面固定参数可解，并且在树、环和串并联图上近乎线性。文中还提供了基于类分子图的实验验证。

从图算法的视角来看，CPA 将图上的持久性重新表述为对色多项式进行一系列组合更新，并利用树分解上的删除-收缩和动态规划（DP）。这使拓扑汇总直接与 Tutte/色多项式工具箱及参数化复杂性相联系，从而能够进行有原则的最坏情况分析，同时在分子骨架中常见的稀疏、低树宽输入上实现实际的近线性行为 $[14, 24, 2]$。对于化学图机器学习，CPA

可提供确定且易于解释的全局环和圈敏感特征（通过 E 多项式和跳跃类实现）。这些特征可补充广泛使用的局部描述符或消息传递描述符，例如 ECFP 风格环形指纹（扩展连接指纹，即基于图的分子描述符，它将局部原子—键邻域编码为固定长度的位向量）、Weisfeiler–Lehman 核以及神经消息传递。与这些局部描述符不同，CPA 编码了能够反映环大小及相关图族的超平面排列/色结构 [30, 12, 9, 34, 23]。这些特征可直接作为独立输入，也可作为额外通道加入图神经网络。由于它们由事件驱动且受到代数控制，因此对阈值缩放具有稳定性，并适用于缓存、增量更新以及与经典图分解流程集成。

本文其余部分组织如下。第 2 节介绍数学预备知识。第 3 节给出 CPA 算法并证明正确性定理。第 4 节证明复杂度定理，并介绍针对特殊类别的加速方法。第 5 节提供实验验证。

2 数学预备知识与符号

假设图均为有限简单图。全文中，我们记 $n := |V|$ and$m := |E|$。

定义 1（阈值链）。加权图是一个简单图 $G = (V, E)$，并配有权函数 $w : E \to \mathbb{R}$。假设所有边权重互不相同，则可将它们排序为 $t_1 < \cdots < t_m$，其中 $m = |E|$。阈值链是由以下子图组成的序列

$$H_j := G_{\le t_j}, \quad j = 1, \dots, m,$$

其中 H_j 包含所有权重不超过 t_j 的边，而 e_j 表示在第 j 步加入的唯一一条边。

我们用 $O^*(\cdot)$ 表示时间/空间界，其中忽略关于 n 和 m 的多项式因子。记号 tw(H) 表示图 H 的树宽。对于树宽有界的图，可以通过在树分解 [24, 5] 上进行动态规划（DP），在 $f(w)$ poly(n) 时间内计算色多项式，并使用 $g(w)$ poly(n) 空间，其中 $w =$ tw(H) 且 f, g 是可计算函数。

2.1 从图形超平面排列到庞加莱多项式

给定一个具有 $n = |V|$ 的（简单）图 $H = (V, E)$，其在 $\mathbb{C}^n$ 中的图形超平面排列为

$$\mathcal{A}(H) := \{\, x_i - x_j = 0 : \{i, j\} \in E \,\}, \qquad M(H) := \mathbb{C}^n \setminus \bigcup_{H_{ij} \in \mathcal{A}(H)} H_{ij}.$$

令 MHS 表示有理混合霍奇结构的范畴，K_0(MHS) 表示其格洛腾迪克环 [26, 8]（不熟悉 Hodge 理论的读者可以放心地

略过以下背景知识；在我们的图设置中，只需知道 E–多项式在 uv 处取值后即为色多项式）。对于复代数簇 X，霍奇–德利涅多项式（又称 E–多项式）定义为

$$E(X;u,v) \;=\; \sum_{k\ge 0}(-1)^k \sum_{p,q} h^{p,q}\big(H_c^k(X;\mathbb{Q})\big)\, u^p v^q,$$

它可经由一个环同态 $E:\, K_0(\text{混合霍奇结构}) \to \mathbb{Z}[u,v]$.[1] 在霍奇–泰特情形下（即当霍奇数除非满足 $p=q$ 时均为零），所有贡献都位于对角线 $p=q$ 上。因此，$E(X;u,\ v)$ 仅取决于乘积 uv，而对于图形超平面排列，它可由 $E(M(H);\ u,\ v) = \chi_H(uv)$ 直接给出。

命题 1（图形补空间的霍奇–泰特纯性）。对于每个$k \ge 0$, $H^k(M(H);\,\mathbb{Q})$ 均为类型 $(k,\ k)$ 的纯霍奇–泰特结构（除非满足 $p=q=k$，否则所有霍奇数均为零）。等价地，

$$E(M(H);u,v) = (uv)^n\, P_{M(H)}\!\Big(-\tfrac{1}{uv}\Big) = \sum_{k=0}^{n} (-1)^k\, b_k(H)\,(uv)^{\,n-k},$$

其中 $n=|V|$ 且 $b_k(H) =$秩 $H^k(M(H);\,\mathbb{Q})$。

这一事实使我们能够只考虑 uv 对角线，并直接从色多项式确定 E 多项式。

2.2 色多项式 ⇒庞加莱/E–多项式

我们介绍两个基本概念。[2]

定义 2（色多项式）。对于图 H，色多项式 $\chi_H(q)$ 是统计 H 使用 q 种颜色进行正常顶点着色的数量的多项式。

定义 3（庞加莱´多项式）。对于拓扑空间 X，庞加莱多项式为

$$P_X(t) := \sum_{k\ge 0} b_k(X)\, t^k, \qquad b_k(X) = \operatorname{rank} H^k(X;\mathbb{Q}).$$

[1] $H_c^k(X;\,\mathbb{Q})$ 表示 X 上第 k个具有有理系数的紧支撑上同调群，并配备混合霍奇结构。整数 $h^{p,q}(H_c^k(X;\,\mathbb{Q}))$ 是其霍奇数，即（p,q）-分次部分的维数。变量u,v 是霍奇双次数的形式标记，而因子（-1）^{k} 反映了对上同调次数取交错和，与欧拉示性数类似。在霍奇–泰特情形（图形排列满足此条件）下，所有非零项都位于对角线 $p=q$上，因此 $E(X;\ u,\ v)$ 仅取决于乘积 uv。在这种情况下，可以通过 $E(M(H);u,v) = \chi_H(uv)$，直接从色多项式得到 E–多项式。

[2] 初次接触排列理论的读者或许只需关注结论：图形排列补空间的 Poincar´e 多项式可以直接根据图的色多项式计算得出。

对于图 H，色多项式 $\chi_H(q)$ 不仅编码了适当 q–着色的数量，还决定了图形排列补空间的拓扑。具体而言，Orlik–Solomon 公式用 χ_H 表示 $M(H)$ 的 Poincar´e 多项式。在图形排列的 Hodge–Tate 情形下，这还可进一步简化为一个直接的恒等式：Hodge–Deligne$E$–多项式可通过特化 $E(M(H);\,u,\,v) = \chi_H(uv)$。因此，色多项式提供了获取图形排列补空间的贝蒂数（进而获取其上同调）的完整途径。

命题 2（色多项式 $\Rightarrow$ Poincar´e/E；图的情形）。若 $n = |V|$，则

$$P_{M(H)}(t) \;=\; (-t)^n\,\chi_H\!\left(-\tfrac{1}{t}\right), \qquad E(M(H);u,v) \;=\; \chi_H(uv).$$

因此，$M(H)$ 的贝蒂数由 χ_H 的系数决定。例如，对于圈 C_n，$\chi_{C_n}(q) = (q-1)^n + (-1)^n(q-1)$，因此 $E(M(C_n);\,u,v)$显式依赖于 n。

对于连通的 H，超平面排列 $\mathcal{A}(H) \subset \mathbb{C}^n$ 的秩为 $n-1$。其特征多项式满足

$$\chi_{\mathcal{A}(H)}(q) = \frac{\chi_H(q)}{q} \quad [25,26].$$

Orlik–Solomon 公式给出 $P_{M(H)}(t) = (-t)^{n-1}\,\chi_{\mathcal{A}(H)}(-1/t)$，因此$P_{M(H)}(t) = (-t)^n\,\chi_H\!\left(-\frac{1}{t}\right)$.于是指数 n来自环境空间$\mathbb{C}^n$，尽管本质秩为 $n-1$。

2.3 删除–收缩跳跃恒等式（图的特化）

向 H 添加一条边 e 以构成 $H' := H \cup \{e\}$时，超平面排列补空间上同调的变化可以用删除–收缩来描述。[3] 我们在紧支撑上同调和混合霍奇结构的 Grothendieck 群中陈述这一结果；在我们的 Hodge–Tate 情形下，经 Hodge–Deligne 映射推送后，即得到相应的 E–多项式结论。

定理 1（删除-收缩跳跃恒等式）。设 H 为一个（简单）图，令 $H' := H \cup \{e\}$ 表示通过添加一条边 e得到的图，并令 H/e 表示其收缩。紧支撑上同调中存在一个自然的长正合序列

$$\cdots \to H_c^k(M(H')) \to H_c^k(M(H)) \to H_c^{k-1}(M(H/e)) \to H_c^{k+1}(M(H')) \to \cdots$$

与混合霍奇结构相容。因此，在 *Grothendieck* 群K_0(混合霍奇结构)中，

$$\sum_{k\in\mathbb{Z}} (-1)^k [H_c^k(M(H))] \;-\; \sum_{k\in\mathbb{Z}} (-1)^k [H_c^k(M(H'))] \;=\; \sum_{k\in\mathbb{Z}} (-1)^k [H_c^{k-1}(M(H/e))].$$

[3] 不熟悉 Hodge 理论的读者可以跳到定理 1；在我们的图论情形中，这仅表示添加一条边会以受控的方式改变不变量，而这种变化由一个收缩子图描述。

应用霍奇–德利涅态射 $E:\ K_0$(混合霍奇结构) $\to \mathbb{Z}[u,v]$ 可得

$$E(M(H);u,v)\ -\ E(M(H');u,v)\ =\ E(M(H/e);u,v).$$

因此，沿着阈值链 $H_{j-1}\xrightarrow{+e_j}H_j$，每一步的跃迁仅取决于收缩 H_j/e_j。

这是对超平面配置补空间经典删除–收缩正合序列的直接特化（见 [25, 26, 8]）。

引理 1（分量乘法性）。若 $H=\bigsqcup_\ell H^{(\ell)}$ 是不交并，则 $E(M(H);u,v)=\prod_\ell E(M(H^{(\ell)});u,v)$。

证明。 这直接由 $M(H)\cong\prod_\ell M(H^{(\ell)})$ 以及 E 的乘法性得出。

3 图上的色多项式持久性算法（CPA）

本节给出针对加权图的具体事件驱动过程（*CPA* 算法），并提供正确性证明。我们依赖第 §2 节中的两个要素：chromatic→Poincar´e 恒等式（命题 2）和删除–收缩跳跃恒等式（定理 1）。

在每个阈值步骤 j 中，算法维护每个阈值的 E-多项式E_j 以及跃迁类 Δ_j（按照定理 1 的规定）。每一步的复杂度主要取决于对收缩子图色多项式χ_{H_j/e_j}的计算；对所有阈值求和后，即可得到后文所示的最坏情况界和固定参数界。

设 $\Delta_j\in K_0$(混合霍奇结构) 表示步骤 j 处的跃迁。条形码ζ函数 是有限欧拉乘积 $Z_G(T):=\prod_{j=1}^{m}(1-T^j)^{-\Delta_j}\ \in\ 1+T\,K_0$(混合霍奇结构)$[[T]]$，利用 K_0(混合霍奇结构) 上的标准幂结构构造；沿 Hodge–Deligne 映射$E:\ K_0$(混合霍奇结构) $\to\mathbb{Z}[u,v]$ 进行推送，可得到具有乘法汇总 [13] 的紧凑数值摘要。

算法 1（CPA）如下，依据的是第 2 节回顾的恒等式。

定理 2（正确性）。对于每个阈值 t_j，算法 *1* 返回

$$E_j(u,v)=E(M(H_j);u,v)\quad and\quad \Delta_j(u,v)=E(M(H_j/e_j);u,v)\,,$$

因此，每个事件对应的跃迁等于删除–收缩跳跃恒等式所规定的类。

证明。我们对 j 进行归纳，证明 $E_j(u,v)\ =\ E(M(H_j);\,u,v)$，并证明$\Delta_j(u,v)\ =E(M(H_j/e_j);\,u,v)$。

基础情形 $j=0$。</*style*>令 为一个具有 个顶点的无边图。于是*()*，因此根据命题 *2*，*(();*，$v)\ =\chi_{H_0}(uv)\ =(uv)^n$。算法将 $E_0(u,\ v)=(uv)^n$ 初始化，因此 $E_0=E(M(H_0);\,u,v)$。

算法 1CPA 算法：图上的事件驱动动机持久性

输入： 加权图 $G=(V,\ E,\ w)$，阈值为 $t_1<\cdots<t_m$
输出： 逐阈值 E 多项式 $E_j(u,\ v)$；跳跃类 Δ_j；条形码ζ函数 $Z_G(T)$

$Z_G(T)\leftarrow 1$ ▷ 初始化
$E_0(u,v)\leftarrow (uv)^{|V|}$ ▷ 含 $|V|$ 个顶点的无边图
对于 $j=1$ 到 m **执行** ▷ 仅在事件发生时处理
  $H_j\leftarrow G_{\le t_j}$；令 e_j 为第 j步添加的边
  色多项式步骤： 计算$\chi_{H_j/e_j}(q)$ ▷ 删除-收缩或基于树宽的动态规划
  跃迁步骤： $\Delta_j(u,v)\leftarrow \chi_{H_j/e_j}(uv)$
  Betti/E 更新： $E_j(u,v)\leftarrow E_{j-1}(u,v)-\Delta_j(u,v)$
  ζ函数： $Z_G(T)\leftarrow Z_G(T)\cdot(1-T^j)^{-\Delta_j}$
结束循环
返回 $\{E_j(u,\ v)\}_{j=1}^m$, $\{\Delta_j\}_{j=1}^m$，以及 $Z_G(T)$

归纳步骤。</style>假设 $(,\ v)=E(M(H_{j-1});\ u,\ v)$。在第 j 步，算法计算 $\chi_{H_j/e_j}(q)$，根据命题 2，我们有

$$E(M(H_j/e_j);u,v)=\chi_{H_j/e_j}(uv).$$

定理 1 在此发挥作用：它给出了精确的递推关系

$$E(M(H_{j-1});u,v)-E(M(H_j);u,v)=E(M(H_j/e_j);u,v).$$

将归纳假设 $E_{j-1}=E(M(H_{j-1}))$ 代入此恒等式，可知更新规则

$$E_j(u,v)\ \leftarrow\ E_{j-1}(u,v)-\chi_{H_j/e_j}(uv)$$

会产生 $E_j(u,\ v)=E(M(H_j);\ u,\ v)$。最后，根据定义，跃迁为 $\Delta_j(u,\ v)=\chi_{H_j/e_j}(uv)$，因此 $\Delta_j=E(M(H_j/e_j);\ u,\ v)$。

因此，通过归纳法，对于所有 j，每个阈值的不变量 E_j 以及每个事件的跃迁 Δ_j 都能正确计算。

4 复杂度分析

我们分析算法 1 的端到端成本。全文中，$n:=|V|$ and$m:=|E|$。由于假设边权重互不相同，阈值链恰好包含 m 个步骤。我们使用 $O^*(\cdot)$ 来省略关于 n and m 的多项式因子。主要成本在于沿阈值链计算收缩子图色多项式χ_{H_j/e_j}；动机层更新的计算量与阈值数量和上同调次数呈线性关系。计算 χ_H 在最坏情况下是 #P 难题，但可以通过树宽固定参数可解（FPT）的动态规划，以及适用于标准图族的闭式递推关系来完成，从而在这些图类上实现近线性行为。

$T_\chi(H)$ 和 $S_\chi(H)$ 分别表示计算 χ_H 所需的时间和空间。我们将使用以下标准界。

– 删除-收缩。经典递归 $\chi_H(q) = \chi_{H\setminus e}(q) - \chi_{H/e}(q)$ 配合记忆化，可得 $T_\chi(H) = O^*\big(2^{|E(H)|}\big)$ 和 $S_\chi(H) = O^*\big(2^{|E(H)|}\big)$。 – 树宽动态规划。若 $\mathrm{tw}(H) \le w$，则可在树分解上运行动态规划，以 $f(w)\,\mathrm{poly}(n)$ 时间和 $g(w)\,\mathrm{poly}(n)$ 空间计算 χ_H（其中 f、g 是可计算函数）[24, 5]。 – 动机/E–更新。将 χ_H 转换为 $P_{M(H)}$ 和 $E(M(H); u, v)$，并组合跃迁/ζ函数，每一步需要 $O(d_E)$ 次环运算，其中 $d_E \le n$ 为非零次数的数量。

定理 3（复杂度）。设 $H_j := G_{\le t_j}$ 为阈值链，e_j 为第 j 步添加的边。则：

*1.*最坏情况（删除-收缩）。

$$T_{\mathrm{total}} \;=\; O^*\Big(\sum_{j=1}^{m} 2^{\,j}\Big) \;=\; O^*(2^m), \qquad S_{\mathrm{total}} \;=\; O^*(2^m).$$

2. 树宽参数化可固定参数处理。若每个 H_j 的 $\mathrm{tw}(H_j) \le w$，则

$$T_{\mathrm{total}} \;=\; O\big(m\, f(w)\,\mathrm{poly}(n)\big), \qquad S_{\mathrm{total}} \;=\; O\big(g(w)\,\mathrm{poly}(n)\big).$$

*3.*特殊图类。对于树和环，χ 的显式闭式蕴含$T_{\mathrm{total}} = O(m+n)$。对于串并联图，线性时间分解递推可得到相同的界。

在所有情形下，动机层更新（E–组装和 ζ 乘积）需要$O(m \cdot d_E)$ 次环运算。

证明。 我们用 $O^*(\cdot)$ 来省略关于 $n := |V|$ 和 $m := |E|$ 的多项式因子。对于每个阈值 j，主要开销是计算 χ_{H_j/e_j}；其他所有工作将在最后汇总。

(1) 最坏情况。 记忆化的删除-收缩得到 $T_\chi(H_j) = O^*(2^{|E(H_j)|})$ 和 $S_\chi(H_j) = O^*(2^{|E(H_j)|})$。同一次记忆化求值会涵盖 χ_{H_j/e_j}，因为它会遍历 H_j 的所有子式；或者，单独调用满足 $T_\chi(H_j/e_j) \le T_\chi(H_j)$。由于$|E(H_j)| = j$，每个阈值的开销为 $O^*(2^j)$ 时间。求和可得

$$T_{\mathrm{total}} = O^*\Big(\sum_{j=1}^{m} 2^{j}\Big) = O^*(2^m), \qquad S_{\mathrm{total}} = O^*(2^m).$$

(2) 树宽参数化可固定参数处理。如果每个 H_j 都满足 $\mathrm{tw}(H_j) \le w$，那么在宽度为w的树分解上运行动态规划，可在 $f(w)\,\mathrm{poly}(n)$时间内计算χ_{H_j}，并使用 $g(w)\,\mathrm{poly}(n)$空间 [24, 5]。由于树宽具有小图单调性，χ_{H_j/e_j} 也满足相同的界。遍历所有 m阈值后，可得

$$T_{\mathrm{total}} = O(m\, f(w)\,\mathrm{poly}(n)), \qquad S_{\mathrm{total}} = O(g(w)\,\mathrm{poly}(n)).$$

(3) 特殊图类。对于树和环，色多项式有显式闭式： $\chi_T(q) = q(q-1)^{n-1}$ 以及 $\chi_{C_n}(q) = (q-1)^n + (-1)^n(q-1)$。因此，每个阈值处的色多项式步骤耗时 $O(1)$，沿整个链的耗时为 $O(m)$；加上 $O(n)$初始化后，总耗时为 $O(m+n)$。对于串并联图，线性时间分解递推可沿固定的串并联分解计算 χ，每次边插入只需进行常数时间更新。因此，沿整个链的色多项式阶段耗时同样为 $O(m+n)$。

动机层开销。给定 χ_H，构造 $E(M(H);\, u,\, v) = \chi_H(uv)$ 需要 $O(d_E)$ 次环运算，其中 d_E 是最高非零次数（$d_E \le n$）。组装每个事件的跃迁 $\Delta_j(u,\, v) = E(M(H_j/e_j);\, u,\, v)$ 并更新 ζ 因子 $Z_G(T) \leftarrow Z_G(T)\cdot(1-T^j)^{-\Delta_j}$，每一步需要 $O(d_E)$。若将 $Z_G(T)$ 截断至 T^m，每次更新仍需 $O(d_E)$。对所有阈值求和可得动机层成本为 $O(m\, d_E)$。因此，动机层仅在阈值和次数方面带来线性开销，与前文确立的色多项式求值成本相比可忽略不计。

在实践中，相邻的阈值图仅相差一条边，因此跨步骤记忆化可降低常数。同样，若 H_j 分解为多个连通分量，则 χ_{H_j} 和 $E(M(H_j))$ 可乘法分解。这两点在实验中都能缩短运行时间，但不会影响定理 3 的渐近复杂度。

我们在此明确后续实验中使用的逐步开销。对于树，色多项式有闭式，因此每次阈值更新中，色多项式步骤的开销为 $O(1)$，动机层更新的开销为 $O(d_E)$，总开销为 $O(m+n)$。对于环，χ_{C_n} 的闭式同样使色多项式子程序每步的开销为 $O(1)$，而动机层更新的开销为 $O(d_E)$。对于串并联图，沿分解计算色多项式，在整个阈值链上总共需要 $O(m+n)$ 时间，动机层更新仍为线性开销。

我们的 O^* 省略了递归/动态规划表管理、从 χ_H 到 $P_{M(H)}$ 和 E的转换，以及ζ 乘积所产生的多项式因子。环运算按 $\mathbb{Z}[u,v]$（或在 uv对角线上）计数，次数为 $\le d_E \le n$。系数增长适中，且主要由色多项式阶段决定。

计算瓶颈完全在于图组合部分（χ_H）。一旦χ_H 可用，动机层的开销很小，并且随阈值数量和上同调次数线性增长。

5 实验验证

我们在一个受控的“环大小识别”任务上评估 CPA 算法，该任务用于区分由 5 条边和 6 条边构成的环。我们将其与一种轻量级的 1-骨架持久同调基线进行比较，该基线基于沿阈值链的 b_0/b_1 轨迹构建。图的特殊化使我们能够通过命题 2，根据色多项式计算每个阈值处的 E 多项式，并通过删除–收缩恒等式（定理 1）计算逐事件跃迁量。

数据与过滤。 我们生成两个平衡类别的无标签环图：$\mathcal{C}_5 = \{C_5\}$ 和 $\mathcal{C}_6 = \{C_6\}$，每类各 30 个实例。每个实例都采用先生成生成树 的阈值调度：除一条边外，其余所有边的权重均设为 $[0.1, 0.3]$，而唯一一条闭合环的边权重设为 0。95。此设计使基线在两种环大小下都恰好观察到最后一个事件发生的一次 H_1-出生，从而暴露出一个已知的盲点。

方法。基线（持久同调，*1*-骨架）。我们在（归一化）阈值处记录 $b_0(t)$ 和 $b_1(t)$，并汇总标准的尺度不变统计量：曲线下面积 AUC(b_0)、AUC(b_1)、最终 b_1、 b_1 次跃迁的数量，以及 b_1 的归一化出生时间。 我们的方法（*CPA*，图的）。 在每个阈值子图 H_j 上，我们计算 $P_{M(H_j)}(t) = (-t)^{|V|}\chi_{H_j}(-1/t)$，并据此得到贝蒂向量（从而通过 $t \mapsto uv$ 得到 E多项式）。对于森林（所有最终步骤之前的步骤），我们使用闭式形式 $P_{M(H_j)}(t) = (1+t)^{|E(H_j)|}$；对于终端环 $H_m = C_n$，我们使用 $P_{M(C_n)}(t) = (1+t)^n - t^{n-1}(1+t)$；参见 §2。我们通过串接各阈值处的贝蒂向量（填充至固定形状）进行向量化，并采用 1-近邻（欧几里得距离）和留一法（LOO）评估 [4]。

结果。 表 1 总结了准确率；CPA 特征能完美区分 C_5 和 C_6，而持久同调基线的表现接近随机猜测。

表 1。 环大小识别（留一法 1-近邻）。运行时间是我们机器上每个图的运行时间。

方法	准确率（留一法）	平均时间 / 图（毫秒）
持久同调基线（1-骨架）	0.55	0.042
CPA（逐阈值 E特征剖面）	**1.00**	0.143

我们计算了配对的麦克尼马尔统计量（经连续性校正），比较我们的方法与基线 [22]。设 b 为基线预测正确而 *CPA* 预测错误的案例数， c 为基线预测错误而 *CPA* 预测正确的案例数。我们得到$b = 0$、 $c = 27$，从而得到 $\chi^2 \approx 25.04$，证实了显著改进。

在此设置中，基于 1-骨架的持久同调基线只记录了两个类别中同一个圈都在最后一个事件处诞生，因此大多数汇总结果都一致。相比之下，CPA 特征会在每个阈值处读取色彩／排列结构。对于环， $\chi_{C_n}(q) = (q-1)^n + (-1)^n(q-1)$，这意味着 $P_{M(C_n)}(t) = (1+t)^n - t^{n-1}(1+t)$。最终事件之前的子图都是森林，其中 $P(t) = (1+t)^{|E(H_j)|}$。因此， E-轮廓（进而包括贝蒂向量）沿整个链都取决于 n，从而清晰地区分 C_5 与 C_6。

请注意，绝对运行时间会因环境而异；尽管如此，定性模式（完美分离，且 CPA 的开销可忽略不计）仍然稳健。

6 结论

我们介绍了*CPA* 算法，这是一种事件驱动过程，可沿图阈值过滤，通过删除-收缩和色多项式计算每个事件的跃迁以及每个阈值对应的 E-多项式

，用于图形排列补空间。在图形（Hodge–Tate）情形下，这些量可简化为离散的组合量，并具有可计算的代数控制；计算瓶颈在于求值 χ_H，这一问题在最坏情况下是困难的，但在树宽方面是固定参数可解的，而且对基本图族存在闭式。

在机器学习中，CPA 提供了一组轻量、可插拔的全局 描述符，可补充局部的消息传递表示。具体而言：

– **特征通道**。 每个阈值对应的 E–概况（每个事件的贝蒂向量）和跳跃类（收缩补集）提供了固定长度且考虑顺序的摘要。截断的条形码–ζ系数通过乘法汇总，提供了更紧凑的替代方案。这些特征可以与学习嵌入拼接，用作图级协变量，或作为独立特征输入经典模型（支持向量机/核函数）。

– **归纳偏置与可解释性**。由于这些通道是通过 χ_H 和删除-收缩定义的，因此具有明确的环/圈敏感性，能够编码消息传递中往往表现不足的全局形状信息。其系数可通过组合结构（着色、极小项）加以解释，从而实现清晰的归因：哪些边插入（事件）以及哪些收缩影响了决策。

– **稳定性与效率**。 事件驱动设计仅取决于边权重的顺序；保持顺序的重缩放不会改变输出，而不会交换事件顺序的小幅扰动也不会产生影响。对于化学及其他域中常见的稀疏、低树宽图，动态规划和闭式解使离线预计算十分高效；训练时间开销几乎为零（从缓存读取特征）。

– **即插即用集成**。 在 GNN 流程中，CPA 特征可以：(i) 与图级读出拼接；(ii) 作为条件信号（例如 FiLM/门控）注入，引导消息传递进入具有圈感知能力的模式；或 (iii) 用作辅助自监督任务的目标（根据局部视图预测 E–轮廓），以规范表征。

具体而言，在分子性质预测和检索中，CPA 提供与环系统和环状骨架相关的确定性、可解释描述符，而这些结构是反应性和物理化学行为的关键决定因素。它们补充了 ECFP/Weisfeiler–Lehman指纹和神经消息传递，能够捕捉随环大小和组成而变化的超平面排列/色结构。由于 CPA 由事件驱动且具有代数性质，因此支持全数据集缓存、快速消融实验（切换事件/子式）以及反事实分析（编辑一条边；重新计算单个步骤）。

总之，本文开发的算法使霍奇精化的事件驱动持久性在图上切实可行；它将代数拓扑、排列理论与图算法联系起来，从而为图结构数据（如化合物）提供稳定、可解释且计算高效的特征，可直接应用于现代机器学习工作流程。我们还计划开发本方法的变体，用于逻辑、范畴论与机器学习交叉领域的应用 [15–21]。

参考文献

1. V. Baranovsky 和 R. Sazdanovic。 *Graph homology and graph configuration spaces*。J. Homotopy Relat. Struct., 7:223–235, 2012。2. H. L. Bodlaender。寻找小树宽树分解的线性时间算法。 *SIAM J. Comput.* 25(6):1305–1317, 1996。3. D. Cohen-Steiner、H. Edelsbrunner 和 J. Harer。持久图的稳定性。*Discrete Comput. Geom.* 37(1):103–120, 2007。4. T. M. Cover 和 P. E. Hart。最近邻模式分类。 *IEEE Trans. Inf. Theory* 13 1 21 27, 1967 ( ): – 。

5. M. Cygan 等。 *Parameterized Algorithms*。施普林格，2015。6. P. Deligne。Th´eorie de Hodge II。 *Publ. Math. IHÉS* 40:5–57, 1971。7. P. Deligne。Th´eorie de Hodge III。 *Publ. Math. IHÉS* 44:5–77, 1974。8. A. Dimca。 *Hyperplane Arrangements: An Introduction*。施普林格，2017。9. D. K. Duvenaud 等。用于学习分子指纹的图卷积网络。载于 *Proc. NIPS 2015*，第 2224–2232 页，2015。10. H. Edelsbrunner 和 J. L. Harer。 *Computational Topology*。AMS，2010。11. R. Ghrist。条形码：数据的持久拓扑。 *BAMS* 45(1):61–75, 2008。12. J. Gilmer、S. S. Schoenholz、P. F. Riley、O. Vinyals 和 G. E. Dahl。用于量子化学的神经消息传递。载于 *Proc. ICML*，PMLR 70:1263–1272, 2017。13. S. M. Gusein-Zade、I. Luengo 和 A. Melle-Hern´andez。簇的格洛腾迪克环上的幂结构。 *Math. Res. Lett.* 11:49–57, 2004。14. F. Jaeger、D. L. Vertigan 和 D. J. A. Welsh。琼斯多项式与图特多项式的计算复杂性。 *Math. Proc. Camb. Phil. Soc.* 108:35–53, 1990。15. Y. Maruyama。格值逻辑与格值模态逻辑的代数研究。载于 *Proc. ICLA*，LNAI 5378:170–184, 2009。16. Y. Maruyama。无点凸几何的基本结果。 *Ann. Pure Appl. Logic* 161 12 1486 1501, 2010 ( ): – 。

17. Y. Maruyama。自然对偶、模态与余代数。 *J. Pure Appl. Algebra* 216 3 565 580, 2012 ( ): – 。

18. Y. Maruyama。完全 Lambek 超教义：一阶次结构逻辑的范畴语义。载于 *Proc. WoLLIC*，LNCS 8071:211–225, 2013。19. Y. Maruyama。范畴对偶理论：及其在域、凸性与分布单子中的应用。载于 *Proc. CSL*，LIPIcs 23:500–520, 2013。20. Y. Maruyama。认知的符号与统计理论：迈向集成式人工智能。载于 *Proc. SEFM*，LNCS 12524:129–146, 2021。21. Y. Maruyama。通用人工智能的条件。载于 *Proc. AGI*，LNAI 12177:242–251, 2020。22. Q. McNemar。关于相关比例或百分比之差的抽样误差的注记。 *Psychometrika* 12(2):153–157, 1947。23. C. Morris 等。图神经网络的高阶形式：Weisfeiler 和 Leman 的神经网络化方法。载于 *Proc. AAAI* 33:4602–4609, 2019。24. S. D. Noble。计算有界树宽图的图特多项式。*Combinatorics, Probability and Computing* 7 (3):307–321, 1998。25. P. Orlik 和 L. Solomon。超平面补集的组合学与拓扑。 *Invent. Math.* 56:167–189, 1980。26. P. Orlik 和 H. Terao。 *Arrangements of Hyperplanes*。数学及其边界领域成果丛书，第 300 卷。施普林格，1992。27. C. A. M. Peters 和 J. H. M. Steenbrink。 *Mixed Hodge Structures*。数学及其边界领域成果丛书，第 3 辑，第 52 卷。施普林格，2008。

28. D. Rogers 和 M. Hahn. 扩展连接指纹.

J. Chem. Inf. Model. ( ): – . 29.

50 5 742 754, 2010

M. Saito. 可极化霍奇模. *PRIMS* 24(6):849–995, 1988年. 30. N. Shervashidze, P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn 和 K. M. Borgwardt. Weisfeiler–Lehman 图核. *JMLR* 12: 2539–2561, 2011年. 31. A. D. Sokal. 多元 Tutte 多项式（又称 Potts 模型）及其在统计物理学中的应用。载于《2005年组合学综述》, LMS Lecture Note Ser., vol. 327, pp. 173–226. 32. W. T. Tutte. 对色多项式理论的贡献.

Canad. J. : – . 33. D.

Math. 6 80 91, 1954

B. West. *Introduction to Graph Theory* (第2版). 普伦蒂斯·霍尔, 2001年. 34. K. Xu, W. Hu, J. Leskovec 和 S. Jegelka. 图神经网络的能力有多强？载于 *Proc. ICLR*, 2019年. 35. A. Zomorodian 和 G. Carlsson. 计算持久同调.

Discrete Comput. ( ): – .

Geom. 33 2 249 274, 2005