arXiv:2305.05818v1 [math.CO] 10 May 2023

有向图的乘积单纯复形的贝蒂数及其在词约简中的应用

Lina F ajardo G´omez，Margherita Maria Ferrari，Nataˇsa Jonoska，Masahico Saito

摘要

我们提出定制的胞腔复形，特别是乘积单纯复形，以分析由有向图构成的数据。这些复形通过粘接单形乘积构成的胞腔而建立，适用于研究无环有向图；本文将这类图称为一致定向图。我们探究第一和第二贝蒂数可能取值的范围，以及生成非平凡同调的有向环类型。我们将这些工具应用于与双出现词约简相关的有向图；双出现词与某些纤毛虫物种的DNA重组过程有关。我们研究词操作对这些图的同调所产生的影响。

1 引言

近年来，拓扑数据分析（TDA）已广泛应用于生物学及其他科学领域。它试图通过连通性、闭环和高维空洞等性质，捕捉给定数据集的底层结构。拓扑不变量可用于检测空间的此类“拓扑特征”。通常，待分析的数据集是欧几里得空间中的点集，而该空间先验上是离散的。为捕捉此类数据集的拓扑形状，人们根据数据点之间的邻近关系构造单纯复形。

在一些生物学现象中，例如 DNA重组过程，数据集以图的形式表示，由顶点和边组成。例如， [16], 基因相互作用模式以图表示，其中顶点代表基因，边表示两个基因如何相互作用，例如彼此交叉。本文提出一种新颖的基于图的模型，用于描述某些纤毛虫物种中的基因组重排。基因组装途径可以表示为双出现词（DOW）中的子词模式删除；这类词中的每个符号恰好出现两次[19]。对重组模式进行建模的迭代子词删除可以用一种称为词图的图表示：其顶点为DOW，若一个词可通过模式删除由另一个词得到，则二者之间存在一条有向边。因此，针对这类有向图数据开展拓扑数据分析的方法值得研究。

常见拓扑数据分析的另一个不足之处在于，其构造复形时往往仅以单纯复形为基础。在我们用于基因组装的词图模型中，为研究这一生物过程的拓扑方面，同时填入正方形和三角形似乎很自然。鉴于这种具体的生物情境，我们受到启发，采用更一般的定制胞腔复形来研究数据集的拓扑性质。本文旨在将其具体应用于我们的词图模型，因此提出使用乘积单纯复形对有向图开展拓扑研究。

因此，本文的重点有两个方面：（1）定义特定的胞腔复形，称为乘积单纯复形；这种复形通过附加单形的乘积从有向图构造而成，用于对图的输出进行拓扑分析；（2）应用积单纯复形同调，通过词图研究基因组装过程的复杂性，这些词图对某些种类纤毛虫中的 DNA 重排进行建模。

本文的组织结构如下。在第2 节中，我们为有向图定义乘积单纯形胞腔复形。在第3 节中，我们描述同调群的一些非平凡生成元，并

构造贝蒂数可以任意大的有向图。第4 节介绍 DOW、其约简路径和词图，而第5 节研究词操作对词图及其相关拓扑不变量的影响。

2 有向图的乘积单纯同调

拓扑数据分析通过研究数据集所在空间的不变量，来确定该空间的内在形状。例如， n维贝蒂数记作 β_n，是一个拓扑不变量，对应于维数为 n的“孔洞”数量。为了研究DNA重组，我们不使用单形作为胞腔，而是定义一个乘积单纯复形。本节回顾图论中的基本概念，并介绍本文后续讨论中使用的符号。

2.1 带源点和目标点的有向图

有向图（简称digraph） G是一个四元组（V, E, ι, τ），其中 V是一个有限集合， $E \subseteq V \times V \setminus \{(x, x) \mid x \in V\}$，且 ι, τ是从 E到 V的映射。 V和 E的元素分别称为顶点和边；对于给定的有向图 G，我们用 $V(G)$和 $E(G)$表示其顶点集和边集。映射 ι 表示每条边的源点（也称初始顶点），类似地， τ 表示目标点（也称终止顶点）。根据定义可知，所讨论的图没有自环或平行边。源点为 u 、目标点为 v的边 e 记作 [1]，其中 $e = [u, v]$。 G的源点是满足以下条件的顶点 s：对于每条边 $e \in E$，都有 $\tau(e) \neq s$。类似地， G 的目标点是满足以下条件的顶点 t：对于每条边 $e \in E$，都有 $\iota(e) \neq t$。

为便于记号表示，我们将有向图记作 $G = (V, E)$，并在上下文明确时省略映射 ι 和 τ。关于此处未明确回顾的图论基础定义，请参阅 [11]。

定义 2.1（弱有向、一致定向）。若连通有向图 G 具有唯一的源点 s 和唯一的目标点 t，则称其为弱有向的。此外，如果 G 是弱有向的且没有（有向）圈，则称 G 为一致定向的。在弱有向图 G 中，我们用 $s(G)$ 和 $t(G)$ 分别表示唯一的源点和目标点。

例 2.2。 图 1 展示了一个弱有向但并非一致定向的有向图，因为它有一个源点（以绿色标示）、一个目标点（以红色标示）以及一个有向环。

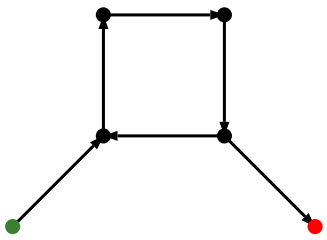


图 1：一个弱有向但并非一致定向的有向图。源点以绿色标示，目标点以红色标示。

我们采用 [24] 中对笛卡尔积的定义。

[1]为了区分两个集合的笛卡尔积中的有序对与有向边，当有序对 (u, v) 对应一条边时，我们用 $[u, v]$ 表示它，因此 $[(u_1, v_1), (u_2, v_2)]$ 表示顶点 (u_1, v_1) 和 (u_2, v_2) 之间的有向边 $(u_1, v_1) \to (u_2, v_2)$。

定义 2.3（笛卡尔积，素图）。设 $G_1 = (V_1, E_1)$ 和 $G_2 = (V_2, E_2)$ 为有向图。 G_1 和 G_2 的笛卡尔积是有向图 $G = G_1 \square G_2 = (V, E)$，其中 $V = V_1 \times V_2$，且 $[(u_1, v_1), (u_2, v_2)] \in E$，当且仅当 $u_1 = u_2$ 且 $[v_1, v_2] \in E_2$，或 $v_1 = v_2$ 且 $[u_1, u_2] \in E_1$。

有向图 G 称为关于笛卡尔积是素的，如果 $G = G_1 \square G_2$ 蕴含 G_1 或 G_2 是只有一个顶点的平凡图，记作 Δ^0。文献 [13] 证明了，连通有向图关于笛卡尔积的素因子分解是唯一的。

注意， $G_1 \square G_2 \cong G_2 \square G_1$，根据 $[(u_1, v_1), (u_2, v_2)] \in E(G_1 \square G_2)$ 当且仅当 $[(v_1, u_1), (v_2, u_2)] \in E(G_2 \square G_1)$，因此笛卡尔积在同构意义下是可交换的。此外，正如 [18]，中所证明的，它也满足结合律。因此，在处理多于两个图的笛卡尔积时，我们写作 $G = ((\ldots ((G_1 \square G_2) \square G_3) \ldots) \square G_k)$，而不加嵌套括号。同样，对于 $1 \le i \le k\}$，我们将 $V(G) = \{((v_1, \ldots, v_{k-1}) v_k) \mid v_i \in V(G_i)$ 写作不带嵌套括号的形式。

笛卡尔积 $G_1 \square G_2$ 为 G_2中的每个顶点生成一个 G_1的副本（在 [18]中称为 G_1-层），反之亦然。可以证明，笛卡尔积 $G_1 \square G_2$ 是一致定向的，当且仅当两个因子都是一致定向的 [12]。根据结合律，这一结果可推广到多个因子的笛卡尔积。

2.2 有向图的乘积单纯复形

图中的顶点和边分别可视为 0 维和 1 维胞腔，因此，通过在特定子图处附加高维胞腔，可以赋予图更丰富的结构。

本节介绍 [20]针对有向图提出的乘积单纯复形的主要定义。其他复形也曾在 [14, 15]中得到研究，例如 p-路径复形，即所谓的 p-路径复形。我们的动机来源于一个生物过程：连接两个顶点的两条长度为 2 的路径对应于 DNA重组的两条相互独立的通路。对于这样的路径对，我们附加实心正方形面，使这两条通路被视为等价。

定义 2.4（单纯有向图）。 n维单纯有向图记为 Δ^n，它是一个顶点为 $V(\Delta^n) = \{v_0, v_1, \ldots, v_n\}$、边为 $E(\Delta^n) = \{[v_i, v_j] \mid 0 \le i < j \le n\}$ 的有向图。

注意， Δ^n 的边的源点和目标点由顶点集 $V(\Delta^n) = \{v_0, v_1, \ldots, v_n\}$ 上的全序诱导。因此，Δ^n 是一致定向的，其源点为 v_0，目标点为 v_n。此外，顶点上的全序保证，由 $V(\Delta^n)$ 的任意子集诱导的 Δ^n 的所有子图也都是单纯有向图。另外，具有相同顶点数的任意两个单纯有向图作为有向图都是同构的。一般而言， Δ^n 可由 Δ^{n-1} 添加新顶点 v_n 以及边 $[v_i, v_n]$（其中 $i = 0, \ldots, n-1$）得到。

在其他文献中，单纯有向图也称为有向团 [21, 23] 或传递竞赛图 [8]。单纯有向图在笛卡尔积意义下是素的。得益于唯一的素分解， [13], 下文所述的乘积单纯形胞腔定义良好。

定义 2.5（乘积单纯形胞腔）。一个 N 维乘积单纯形胞腔 P 是由单形 $\prod_{i=1}^{k} \Delta_i^{n_i} = \Delta_1^{n_1} \times \cdots \times \Delta_k^{n_k}$ 的乘积构成的 N 胞腔，其中 $n_i > 0$ 对所有 $1 \le i \le k$ 均成立，且 $N = \sum_{i=1}^{k} n_i$。其 1-骨架是单纯有向图的笛卡尔积。也就是说，一个形如以下形式的图。

$$\bigsquare_{i=1}^{k} \Delta_i^{n_i} = \Delta_1^{n_1} \square \cdots \square \Delta_k^{n_k}.$$

为简洁起见，我们将 P称为一个乘积单纯形 N-胞腔。当 $k = 1$ 时，我们称 P为一个 N-单纯形，并将其记为 Δ^N。

由于图的笛卡尔积在同构意义下满足交换律，我们假设在乘积单纯形胞腔 $P = \square_{i=1}^{k} \Delta_i^{n_i}$ 中，因子的顺序满足 $n_1 \ge n_2 \ge \cdots \ge n_k$。

从此以后，除非另有说明，单纯形仅指具有单个单纯有向图因子的乘积单纯形胞腔。鉴于单纯有向图因子与单纯形之间的对应关系，我们滥用记号，写作 $P = \square_{i=1}^{k} \Delta^{n_i}$ 而不是 $\prod_{i=1}^{k} \Delta_i^{n_i}$。还需注意，由于 P 的定义中各因子的维数可能不同，因此必须使用下标。不过，为简洁起见，在不会引起混淆的情况下，我们省略下标。

我们可以像对标准单形那样，自然地定义单纯有向图的几何实现。借助这一几何实现，乘积单纯形 N-胞腔可以视为$\mathbb{R}^N$ 的子集，并继承其乘积拓扑。

例 2.6. 乘积单纯形胞腔不仅包括单形，还包括立方体（由 Δ^1反复作乘积得到），以及三棱柱（由任意胞腔与 Δ^1作乘积得到）。图 2 展示了所有乘积单纯形 3-胞腔的 1-骨架。

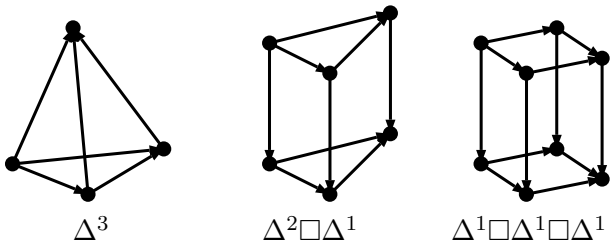


图 2：与乘积单纯形 3-胞腔的 1-骨架相对应的有向图。

乘积单纯形胞腔这一概念最初由 Kozlov 在[20]中针对无向图的直积提出。我们选择笛卡尔积，以获得一种兼具单纯复形和立方复形某些特征、同时胞腔仍保持一致定向的复形。由乘积单纯形胞腔构成的复形也用于结理论[5, 6]。

定义 2.7（乘积单纯形胞腔的定向）。设 Δ^n 是一个顶点集为 $V(\Delta^n) = \{v_0, v_1, \ldots, v_n\}$ 且 $s(\Delta^n) = v_0$ 的单纯形。设 $N(s(\Delta^n))$ 是源点的邻居集合，$V(\Delta^n) \setminus s(\Delta^n)$。$\Delta^n$ 的一个定向是 $N(s(\Delta^n))$ 的各个排列所构成的等价类（相差偶置换）。一个有向单纯形记作 $[\Delta^n]$，是附带定向的单纯形。

设 $P = \square_{i=1}^{k} \Delta^{n_i}$ 是一个乘积单纯形胞腔，且 $s(P)$ 是 P 的源点。设 $v_{i,\,0} = s(\Delta^{n_i})$ 和 $N_i = \{(v_{1,\,0}, v_{2,\,0}, \ldots\ldots, v_{i,\,j}, \ldots\ldots, v_{k,\,0}) \mid v_{i,\,j} \in N(s(\Delta^{n_i}))\ 1 \le j \le n_i\}$。给每个 N_i 赋予一个全序后，N_1，……，N_k 上的全序便确定了 $N(s(P))$ 的一个全序。P 的一个定向是 $N(s(P))$ 的这类全序所构成的等价类（相差偶置换）。附带定向的乘积单纯形胞腔 P 称为有向乘积单纯形胞腔，记作 $[P]$。

例 2.8（定向的乘积单纯形胞腔）。设 Δ^2 是具有 $V(\Delta^2) = \{a, b, c\}$ 的 2-单纯形，Δ^1 是具有 $V(\Delta^1) = \{x, y\}$ 的 1-单纯形。设 $P = \Delta^2\square\Delta^1$。则 $s(P) = (a, x)$ 和 $N_1 = \{(b, x)、(c, x)\}$ 以及 $N_2 = \{(a, y)\}$。邻居集合的顺序 N_1, N_2 在 $N(s(P))$ 上诱导出顺序 $((b, x), (c, x), (a, y))$，而 N_2, N_1 则诱导出顺序 $((a, y), (b, x), (c, x))$。若 N_1 的顺序满足 (c, x) 排在 (b, x) 之前，则顺序 N_2, N_1 诱导出顺序 $((a, y), (c, x), (b, x))$。前两种顺序赋予该乘积单纯形胞腔相同的定向，而第三种顶点顺序则定义了一个定向相反的胞腔。

定义 2.9（面、面、边界集）。如果存在 $n_i' \le n_i\ \forall i = 1、\ \ldots k$，使得 $P' = \square_{i=1}^{k} \Delta^{n_i'}$，我们就称 P' 是 N 维乘积单纯形胞腔 $P = \square_{i=1}^{k} \Delta^{n_i}$ 的一个面。P 的面是其 $(N-1)$ 维面。P 的边界集记作 ∂P，由它的所有面组成。

定义 2.10（乘积单纯形胞腔复形）。给定一个有向图 $G =$（V，E），我们可以通过以下粘合过程归纳地构造一个与 G 相关联的乘积单纯形胞腔复形 Γ，记作 $\Gamma(G)$：

- 令 $\Gamma_{(0)}$ 为 G的所有顶点的集合。

- 令 $\Gamma^{(1)} = \Gamma^{(0)} \cup E$。也就是说，将 G 的所有边添加到 $\Gamma^{(0)}$ 中，方法是将所有面为 G顶点的积单纯形 1-胞腔附加到当前复形上。
- 令 $\Gamma^{(N-1)}$ 表示前 $N-1$ 步所创建的复形。令 P 为一个乘积单纯形N胞腔。如果存在映射 $\phi : \partial P \to \Gamma^{(N-1)}$，使得 ∂P 与 $\phi(\partial P)$ 同胚，ϕ 保持每条边的定向，并且将 ϕ 限制在 P 的每个面上仍为同胚，则将 P 加入 $\Gamma^{(N)}$。

例 2.11 (乘积单纯形胞腔复形)。图 3 展示了一个乘积单纯形胞腔复形 $\Gamma^{(3)}$ 的构造，该复形包含维数至多为 3 的胞腔，其中包括七个正方形和一个立方体（其六个面计入这些正方形中）。有向图 G 对应于 $\Gamma^{(1)}$，它通过收集 G的顶点和边得到。

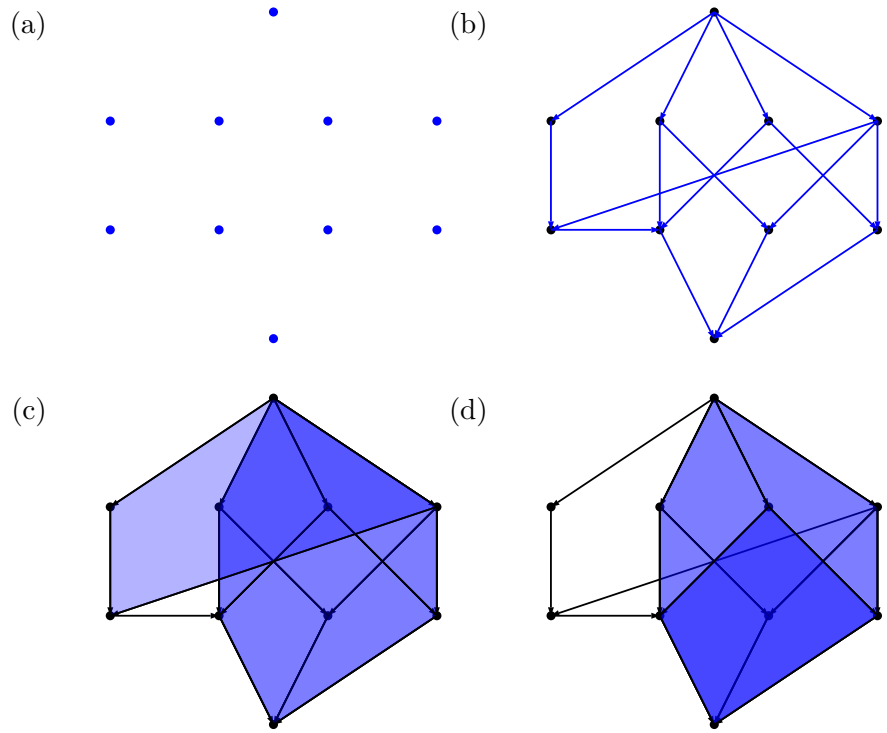


图 3：随着胞腔被加入复形，0维的面显示在 (a) 中，1维的面显示在 (b) 中，2维的面显示在 (c) 中，3维的面显示在 (d) 中。

2.3 有向图的乘积单纯形同调群

本节介绍乘积单纯形胞腔的边界算子，据此可以计算乘积单纯形胞腔复形的同调群。虽然乘积单纯复形作为 CW 复形的一个特例，具有链复形结构，但为了便于计算，我们将利用乘积法则给出乘积单纯形胞腔上的显式边界算子。

给定乘积单纯形胞腔复形 Γ，在 Γ 上的乘积单纯形 N-链群记作 $C_N(\Gamma)$，它是由 Γ 中定向的 N 维乘积单纯形胞腔生成的自由阿贝尔群。其典型元素——乘积单纯形 N-链——是 N 维乘积单纯形胞腔的有限形式线性组合，系数为整数。在单纯形与链的笛卡尔积中，乘积对和满足分配律；也就是说

$$\Delta^{n_0} \square \left(\sum_{i=1}^{k} \Delta^{n_i} \right) = \sum_{i=1}^{k} (\Delta^{n_0} \square \Delta^{n_i}).$$

回顾一下，单形的边界算子定义如下

$$\partial_n(\Delta^n)=\sum_{i=0}^{n}(-1)^i[v_0,v_1,\ldots,v_{i-1},\hat{v_i},v_{i+1},\ldots,v_n],$$

其中，$\hat{}v_i$ 表示从单纯形中删除了顶点 v_i。为简洁起见，在计算中我们用 $[\hat{}v_i]$ 表示单纯形 $[v_0,v_1,\ldots,v_{i-1},\hat{}v_i,v_{i+1},\ldots,v_n]$。

定义 2.12（边界算子）。设 $P=\square_{i=1}^{k}\Delta^{n_i}$ 是一个 prodsimplicial N-胞腔，其中 $N=\sum_{i=1}^{k}n_i$。对于 $1\le i\le k$，我们使用记号

$$\langle\overline{\partial_{n_i}\Delta^{n_i}}\rangle=\Delta^{n_1}\square\Delta^{n_2}\square\cdots\square\Delta^{n_{i-1}}\square\partial_{n_i}\left(\Delta^{n_i}\right)\square\Delta^{n_{i+1}}\square\cdots\square\Delta^{n_k}.$$

对于 $1\le i<j\le k$，令

$$\langle\overline{\partial_{n_i}\Delta^{n_i}\partial_{n_j}\Delta^{n_j}}\rangle=\Delta^{n_1}\square\cdots\square\partial_{n_i}\left(\Delta^{n_i}\right)\square\cdots\square\partial_{n_j}\left(\Delta^{n_j}\right)\square\ldots\square\Delta^{n_k}$$

N维边界算子作用于 P 定义为

$$\partial_N(P)=\partial_N\left(\underset{i=1}{\overset{k}{\square}}\Delta^{n_i}\right)=\sum_{i=1}^{k}(-1)^{\alpha(i)}[\overline{\partial_{n_i}\Delta^{n_i}}]$$

其中，$\alpha(i)=\sum_{\ell=1}^{i-1}n_\ell$ 是第 i个因子之前各因子的维数之和。

由于边界算子只能作用于定向链（$[\Delta^n]$ 和 $[P]$），而不能作用于图（Δ^n 和 G），为简洁起见，我们省略方括号。

注意，给定一个乘积单纯形胞腔复形 Γ, ∂_N 定义了一个群同态 $\partial_N:C_N(\Gamma)\longrightarrow C_{N-1}(\Gamma)$。我们验证上述定义的算子确实定义了一个链复形。回忆一下，对于顶点为 $\{v_0,v_1,\ldots,v_{n+1}\}$ 的单纯形 Δ^{n+1}，有 $(\partial_n\circ\partial_{n+1})(\Delta^{n+1})=0$。

命题 2.13（乘积单纯形胞腔上的 $\partial^2=0$ ）。设 $P=\square_{i=1}^{k}\Delta^{n_i}$ 是一个乘积单纯形（$N+1$）胞腔，其中 $N+1=\sum_{i=1}^{k}n_i$。则 $(\partial_N\circ\partial_{N+1})(P)=0$。

证明。根据定义 2.12，我们有

$$\begin{aligned}(\partial_N\circ\partial_{N+1})(P)&=(\partial_N\circ\partial_{N+1})\left(\underset{i=1}{\overset{k}{\square}}\Delta^{n_i}\right)\\&=\partial_N\left(\sum_{i=1}^{k}(-1)^{\alpha(i)}[\overline{\partial_{n_i}\Delta^{n_i}}]\right)\\&=\sum_{j<i}(-1)^{\alpha(i)}(-1)^{\alpha(j)}[\overline{\partial_{n_j}\Delta^{n_j}\partial_{n_i}\Delta^{n_i}}]+\sum_{j>i}(-1)^{\alpha(i)}(-1)^{\alpha(j)-1}[\overline{\partial_{n_i}\Delta^{n_i}\partial_{n_j}\Delta^{n_j}}]\\&+(\partial_{n_i-1}\circ\partial_{n_i})(\Delta^{n_i}).\end{aligned}$$

项 $(\partial_{n_i-1}\circ\partial_{n_i})(\Delta^{n_i})$ 为零。注意，当 $j>i$ 时，第 i 项的维数减小了一，因此对 $j>i$ 求和时 (-1) 的指数为 $\alpha(j)-1$。

由于这一差值为 1，两个求和（分别对 $j<i$ 和 $j>i$ 求和）包含相同的项但符号相反，因此最终表达式等于零。 □

利用该边界算子，我们可以计算有向图的乘积单纯复形的环群、边界群、同调群和贝蒂数。

3 一致定向图的贝蒂数和生成元

我们考察第一同调群和第二同调群的生成元，并研究有向图的乘积单纯复形中贝蒂数的可能取值。我们采用记号 $\beta_n(\Gamma(G))$ 表示与有向图 n关联的乘积单纯复形的第 G个贝蒂数，其中 $\Gamma(G)$ 是通过定义 2.10 中定义的粘合过程构造的。为简洁起见，我们滥用记号，改写为 $\beta_n(G)$。

众所周知， β_0 表示连通分支的数量。由于我们研究的是由单个连通分支组成的一致定向有向图，因此对于所有感兴趣的有向图，我们取 $\beta_0(G) = 1$ 。

根据定义，一致定向图在图论意义下不包含任何有向环。在下文中，在不会引起混淆的情况下，我们使用术语 n-有向环来指代环群中的元素。

3.1 第一同调群的生成元

图 4 展示了定义在三个顶点上的两个连通有向图。图 4(a) 是形如 $(v_0, v_1, v_2 = v_0)$ 的闭路径，并构成一个 1-环；而如图 4(b) 所示，单纯有向图不对 β_1作出贡献。尽管形如 $(v_0, v_1, .., v_n = v_0)$ 的闭路径会对 β_1 作出贡献，但它们在一致定向图中不存在，因此我们在寻找同调群生成元时将其排除。

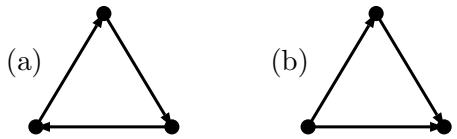


图 4：三个顶点上的连通有向图只有两种可能。

四个顶点上的一致定向图只有两种，如图 5 所示。其中一种是由一条边和一条长度为三的路径组成的并，如图5(a) 所示，它形成一个与圆 S^1 同胚的复形。另一种是形如 $\Delta^1 \square \Delta^1$ 的正方形胞腔，如图 5(b) 所示。我们将观察结果总结如下。

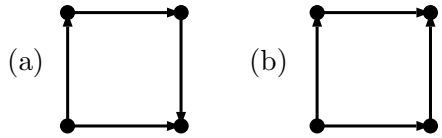


图 5：四个顶点上的所有一致定向图。

引理 3.1. 设 $G = (V, E)$ 其中 $V = \{v_0, v_1, v_2, v_3\}$ 且 $E = \{[v_0, v_3], [v_0, v_1], [v_1, v_2], [v_2, v_3]\}$，如图 5(a) 所示。则 $\beta_1(G) = 1$ 且 $\beta_n(G) = 0$ 对所有 $n > 2$。

我们可以通过如下方式，从源点到目标点连接长度为 3 的边不相交路径，构造出具有任意大 β_1 值的图。

引理 3.2. 对于所有 $k \geq 0$，存在一个图 G ，使得 $\beta_1(G) = k$ 且 $\beta_n(G) = 0$ ，其中$n > 2$。

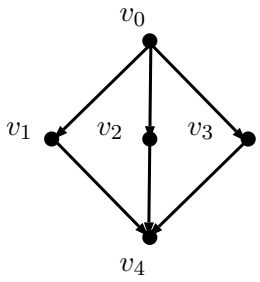


图7：其乘积单纯复形由三个正方形、六条边和五个顶点组成的一致定向图。

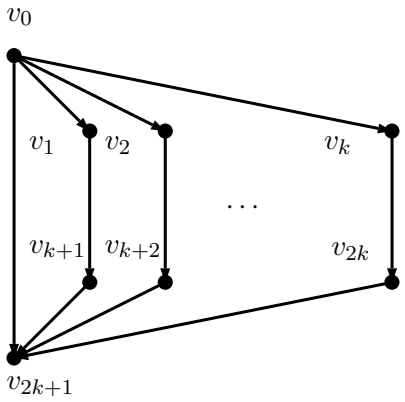


图6：与 $[v_0, v_{2k+1}]$“平行”延伸的 k 条长度为 3 的不相交路径。

证明。 令 G 为图 6 中的图6。那么，作为一个拓扑空间， G 同伦等价于由 k 个圆组成的楔和，因此 $\beta_1(G) = k$ 且 $\beta_n(G) = 0$ 对于 $n > 2$。 □

引理3.3. 对于一致定向图，所有表示第一同调群中非平凡元素的 1-环，都由具有共同源点和目标点 $p_1 = (s = u_0, u_1, \ldots, u_m = t)$ 和 $p_2 = s = v_0, v_1, \ldots, v_n = t)$ 的路径组成，其中 $\{u_1, u_2, \ldots, u_{m-1}\} \cap \{v_1, v_2, \ldots, v_{n-1}\} = \varnothing$。

3.2 第二同调群的生成元

本节将重点讨论顶点标记为 v_i 的有向图，因此我们将记号 $[v_0, v_1, \ldots, v_n]$ 保留给单形，并将其他低维乘积单纯形胞腔标记为顶点序列。例如， $[v_0, v_1, v_2, v_3]$ 表示一个四面体，而 $v_0v_1v_2v_3$ 表示一个正方形。

在2维情况下，找出所有对应的乘积单纯复形具有非平凡2-圈的图，并不那么简单。这里我们给出几个最简单的例子。

为了考察第二同调群的生成元，我们考虑至少有五个顶点的一致定向图，因为所有具有三或四个顶点的一致定向图的贝蒂数已如图所示5 和6所示。通过检查可以发现，具有五条边和五个顶点的连通一致定向图只有引理3.3中所示的路径对。对于一个具有六条边和五个顶点且满足 $\beta_2(G) = 1$的例子，我们考虑由三个正方形组成的图，如图7所示。

引理3.4。 设 G 如图 7所示。则 $\beta_1(G) = 0$， $\beta_2(G) = 1$ 以及 $\beta_n(G) = 0$ 对所有 $n > 2$均成立。

证明。我们根据 G 的顶点，明确描述每个链群。注意，正方形按照圈进行标记：从源顶点开始，沿其字典序最低的邻顶点依次标记。与该图相关联的乘积单纯复形中的链群 C_n 描述如下，其生成元为：

$$
\begin{aligned}
C_0 &= \langle [v_0], [v_1], [v_2], [v_3], [v_4] \rangle \\
C_1 &= \langle [v_0, v_1], [v_0, v_2], [v_0, v_3], [v_1, v_4], [v_2, v_4], [v_3, v_4] \rangle \\
C_2 &= \langle v_0 v_1 v_4 v_2, v_0 v_2 v_4 v_3, v_0 v_1 v_4 v_3 \rangle .
\end{aligned}
$$

在该图中，任意两条长度为 2 的路径都会构成一个一致定向的正方形，因此所得的乘积单纯复形与球面同胚。由此可得，对于所有 $n > 2$ ， $\beta_1(G) = 0, \beta_2(G) = 1$ 和 $\beta_n(G) = 0$ ，正如所需。□

除此之外，我们还给出另一个具有五个顶点的图，它的第二同调群非平凡，并且也产生一个 2-圈。

引理 3.5. 设 G 为图 8(a) 中的图。则 $\beta_1(G) = 0, \beta_2(G) = 1$ 和 $\beta_n(G) = 0$对于所有 $n > 2$。

证明。该与该图相关联的乘积单纯复形中的链群如下 ：

$$
\begin{aligned}
C_0 &= \langle [v_0], [v_1], [v_2], [v_3], [v_4], [v_5] \rangle \\
C_1 &= \langle [v_0, v_1], [v_0, v_2], [v_1, v_3], [v_1, v_4], [v_2, v_3], [v_2, v_4], [v_3, v_5], [v_4, v_5] \rangle \\
C_2 &= \langle v_0 v_1 v_3 v_2, v_0 v_1 v_4 v_2, v_1 v_3 v_5 v_4, v_2 v_3 v_5 v_4 \rangle .
\end{aligned}
$$

这四个正方形彼此连接，构成一个与球面同胚的复形，由此即可得出结论。

□

注 3.6. 在图 7 或图 8中，都可以为一致定向的正方形添加连接其相对顶点的边（将其分成两个单纯有向图）。这样便得到一些图，其复形与球面同胚，并具有不同类型的非平凡多面体 2-循环。例如，可以修改图 8(a)，得到图 8(b)。

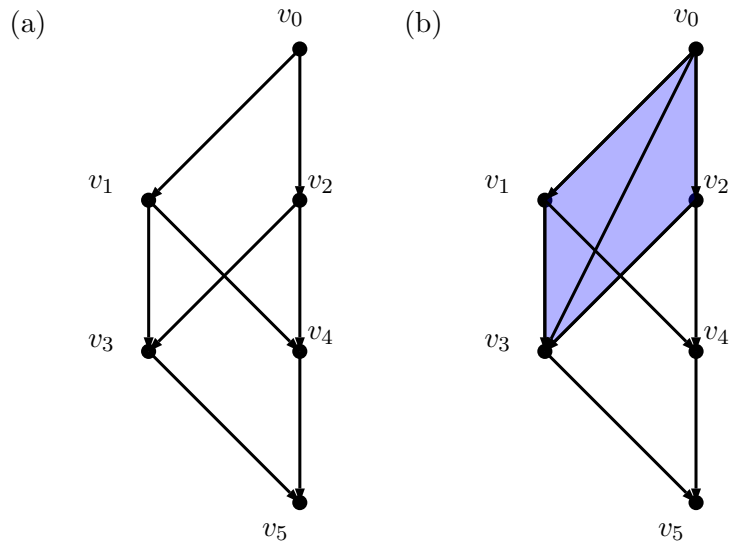


图 8：具有非平凡 2-圈的图。

3.3 贝蒂数组合的可实现性

利用前面结果中的一些图，我们构造出阶更大的图，使其具有特定的同调群。为证明这一结果，我们使用下面这个由 Mayer-Vietoris 序列推出的引理 [17]。

引理 3.7. 设 G 和 H 是一致定向图，使得 $G\cup H$ 是一致定向的，$\Gamma(G\cup H)=\Gamma(G)\cup\Gamma(H)$，$H_0(\Gamma(G\cap H))=\mathbb{Z}$，且对于 $n\geq 1$，$H_n(\Gamma(G\cap H))=0$。那么

$$H_n(\Gamma(G\cup H))\cong H_n(\Gamma(G))\oplus H_n(\Gamma(H))$$

对于 $n\geq 1$。

因此，我们能够将任意数量的生成图“粘合”在一起，从而实现贝蒂数的任意组合。

推论3.8. 存在一个图 G，使得 $\beta_1(G)=0$，$\beta_2(G)=k$，并且 $\beta_n(G)=0$ 对于所有$n>2$。

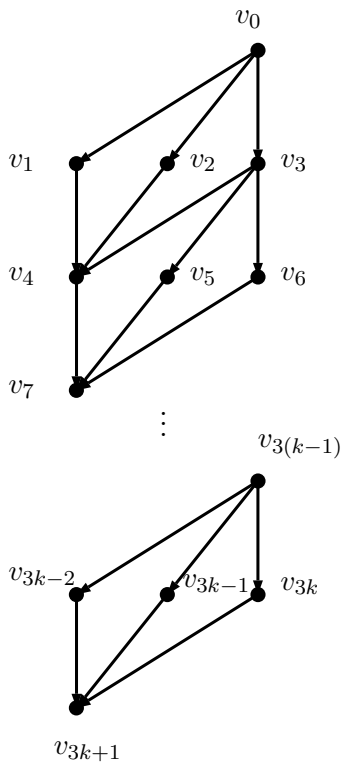


图9：Gra ph，其对应的乘积单纯复形的第二同调秩为 k .

证明。设 G 如图 9所示。那么，作为一个拓扑空间，G 等价于沿一条边粘合的球面的楔和（其同调群是平凡的）。因此，我们有$\beta_1(G)=0$，$\beta_2(G)=k$ 以及 $\beta_n(G)=0$ 对所有 $n>2$均成立。 □

推论 3.9。设 G如引理 *3.2*所述，其顶点集为 $\{v_0,v_1,\ldots,v_{2k+1}\}$ 且 $\beta_1(G)=k$；再设 H如推论 *3.8*所述，其顶点集为 $\{u_0,u_1,\ldots,u_{3\ell+1}\}$ 且 $\beta_2(H)=\ell$，使得 $v_0=u_{3k+1}$且 G，并且 H 仅在顶点 v_0处相交。

令 $G\cup H=(V(G)\cup V(H)\ E(G)\cup E(H))$。那么 $G\cup H$是一致定向的，并满足 s

$$\beta_1(G\cup H)=k$$

$$\beta_2(G \cup H) = \ell$$
$$\beta_n(G \cup H) = 0 \quad for\ n \geq 3.$$

证明。 容易验证， $G \cup H$ 的唯一源点和目标点分别是u_0 和 v_{2k+1}，因此 $G \cup H$ 是一致定向的。根据假设， $H_0(\Gamma(G \cap H)) = \mathbb{Z}$ 且$H_n(\Gamma(G \cap H)) = 0$ 对于 $n \geq 1$ 因此结论由引理 3.7得出。 □

因此，对于任意正整数 $\mathbb{Z}^k$ 和 $\mathbb{Z}^\ell$，都可以构造出其第一和第二同调群分别同构于 k 和 ℓ 的一致定向图。该构造在顶点数量上并非最简。例如，图 10 采用了与引理 3.2 类似的构造，且使用更少的顶点和边，实现了许多相同的 k值。

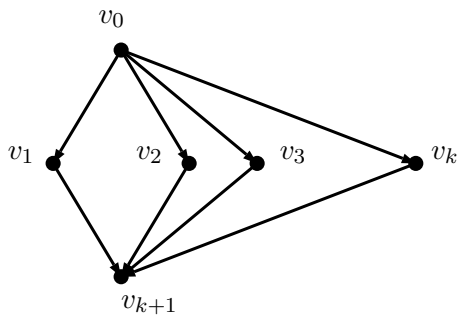


图 10：包含 $\binom{k-1}{2}$ 2-圈的有向图。

引理 3.10. 设 G 是一个有向图，其顶点为 $V(G) = \{v_0, v_1, \ldots, v_{k+1}\}$，边由路径集合 (v_0, v_i, v_{k+1}) 定义，其中 $0 < i < k+1$。那么，对于所有 $n > 2$，均有 $\beta_1(G) = 0$、 $\beta_2(G) = \binom{k-1}{2}$，以及 $\beta_n(G) = 0$ 。

4 双出现词的词图

本节聚焦于特定的生物分子过程，其中会出现一致定向图，并可应用乘积单纯复形。

在某些纤毛虫物种（如 Oxytricha trifallax）的体细胞核发育过程中，可以观察到大规模的重排过程。重组由DNA片段两侧的短DNA重复序列引导，这些片段会发生重排，其排列顺序也由这些重复序列决定。这些短重复序列可以用一系列双出现词来建模，即每个符号都出现两次的词 [7, 22, 25]。特别是，这些物种中超过90%的DNA 重排都可以描述为反复删除重复词与返回词的过程 [4]。重复词与返回词是词中正方形和回文因子的推广，在语言理论中具有研究价值 [2, 4, 19]。本节将介绍与双出现词相关的图，这些图可用于建模这些DNA重排过程 [3]。

4.1 双重出现词

我们称有序可数的符号集 Σ 为字母表。字母表 Σ 上的一个词是形如 $w = a_1 a_2 \ldots a_n$ 的有限序列，其中 $a_i \in \Sigma$，其长度记作 $|w|$，为 n。我们用 Σ^* 表示字母表 Σ上所有词的集合，包括记作 ϵ的空词。组成词 w 的所有符号构成的集合记作 $\Sigma[w]$。词 $w = a_1 a_2 \ldots a_n$ 的逆序词为 $w^R = a_n a_{n-1} \ldots a_2 a_1$。如果存在 w_1、 $w_2 \in \Sigma^*$ 使得 $w = w_1 v w_2$，则词 v 是词 w的因子，记作 $v \sqsubseteq w$。在本文中，对某个 $n \in \mathbb{N}$，我们设 $\Sigma \subseteq [n]$ 。例如， $w = 122313$是字母表$[3] = \{1, 2, 3\}$ 上的一个词，长度为 $|w| = 6$。 $w = 122313$ 的逆序词是词 $w^R = 313221$。

如果词 $w \in \Sigma^*$ 中的每个符号在 Σ 中出现的次数为零次或两次，则称其为双出现词（DOW）。我们用 Σ_{DOW} 表示字母表 Σ上所有 DOW 的集合。类似地，如果每个符号在 Σ 中出现一次或不出现，则称 w 为单出现词（SOW）。字母表 Σ 上的 SOW 集合记作 Σ_{SOW}。由于长度为 n 的 DOW 使用 $n/2$ 个不同符号，我们称 DOW w 的大小为 $|w|/2$。将 Σ_{DOW} 限制为大小小于或等于 n的 DOW 时，我们用 $\Sigma_{DOW}^{\leq n}$表示该集合。

如果 $w \in \Sigma^*$ 满足 $a_1 = \min(\Sigma[w])$，且每个符号首次出现时都是此前所有符号中最大者的直接后继，则称该词按升序排列。例如，词 $w_1 = 122313$ 是按升序排列的 DOW，而 DOW $w_2 = 133212$ 则不是。

我们称 w_1 和 w_2 为升序等价，并记作 $w_1 \sim w_2$，如果存在一个从 Σ 出发的双射，它在 Σ^* 上诱导出一个态射 f，使得 $f(w_1) = w_2$。词 $w_1 = 122313$ 和 $w_2 = 133212$ 通过以下双射映射等价： $1 \mapsto 1,\ 3 \mapsto 2,\ 2 \mapsto 3.$ 。由于升序词在这种等价关系下是唯一的，因此我们将升序词视为由关系 $\sim$所确定的等价类的代表元。

从现在起，我们只考虑 DOW 的等价类；为避免混淆，在不会引起歧义时，我们将直接用词表示其等价类。以下定义见 [9]。

定义 4.1（重复词、返回词）。设 $x, y, z \in \Sigma^*$ 和 $u \in (\Sigma \setminus \Sigma[w])_{SOW}$。我们称

- 词 uu 是 $w = xuyuz$ 中的重复词，并且词 xyz 可由 w 通过重复删除得到，记作 $d_u(w) = xyz$。此时，我们称 u 为 w中的重复因子。
- 词 uu^R 是 $w = xuyu^Rz$ 中的返回词，并且词 xyz 可由 w 通过返回删除得到，记作 $d_u(w) = xyz$。此时，我们称 u 为 w中的返回因子。

重复词或返回词 uu 或 uu^R，其中 $|u| = 1$，称为平凡的。若不存在其他重复（分别为返回）因子 v，使得它们在 w 中包含 u 且满足 $|v| > |u|$，则称词 uu（分别为 uu^R）是 w 中的最大重复（分别为返回）词。沿用 [1], ，我们用 M_w^{DOW}表示 w中最大重复（uu）词或返回（uu^R）词的集合。此外，我们将 w 中重复因子或返回因子的集合定义为 $M_w^{SOW} = \{u \sqsubseteq w \mid uu \in M_w^{DOW} \text{ 或 } uu^R \in M_w^{DOW}\}$。

引理 4.2. *[1]* 设 w 是一个大小为 n 的 DOW。对于每个 $x \in \Sigma[w]$，存在唯一的 $u \in M^{SO}_{\ w}{}^{W}$ 使得 $x \in \Sigma[u]$。

最大重复词或返回词的集合可能包含平凡重复词或返回词，如例 4.4 所示。按升序排列时，某些因子可能等价，因此 M_w^{DOW}（分别为 M_w^{SOW}）中的元素始终写作 DOW（分别为 SOW），而不是它们的等价类，从而 w 中的每个符号都会出现在 M_w^{SOW}中的某个词里。我们关注最大重复词与返回词，并给出以下定义。该定义改编自 [1] 和 [19]中所述的所谓“模式约化”过程。

定义 4.3（后继、前驱）。集合 $D(w) = \bigcup_{u \in M_w^{SOW}} \{v \mid v \text{ 按升序排列，而 } v \sim d_u(w)\}$ 称为 w 的直接后继集合。如果存在词序列 $w = w_1、w_2、\ldots、w_n = w'$，使得对于某个 $u_i \in M_{w_i}^{SOW}$ 的取值，$w_i \sim d_{u_i}(w_{i-1})$，则称 w' 是 w 的后继，且 w 是 w' 的前驱。注意，空词 ϵ 是所有词的后继。

例 4.4. 设 $w = 1234523541$。w 中的极大重复词或返回词集合为 $M_w^{DOW} = \{11, 2323, 4554\}$和 $M_w^{SOW} = \{1, 23, 45\}$。由于 $d_1(w) = 23452354 \sim 12341243$且 $d_{23}(w) = 145541 \sim 123321$，我们得到 w 的直接后继集合为 $D(w) = \{12341243, 123321, 123231\}$。我们可以继续从后继中删除子词，如下所示。12341243、123321 和 123231 的极大重复因子或返回因子为

$$M_{12341243}^{SOW} = \{12, 34\}$$
$$M_{123321}^{SOW} = \{123\}$$

$$M_{123231}^{SOW} = \{1, 23\}$$

使其删除后得到

$$d_{12}(12341243) = 3443 \sim 1221 \qquad d_{123}(123321) = \epsilon \qquad d_1(123231) = 2323 \sim 1212$$
$$d_{34}(12341243) = 1212 \qquad D(123321) = \{\epsilon\} \qquad d_{23}(123231) = 11$$
$$D(12341243) = \{1221, 1212\} \qquad D(123231) = \{1212, 11\}.$$

再次重复这一过程，得到 $D(1212) = D(1221) = D(11) = \{\epsilon\}$。因此，我们可以说， w 的所有后继构成集合

$$\{12341243, 123321, 123231, 1221, 1212, 11, \epsilon\}.$$

4.2 词图

关于图论的基本定义，请读者参阅 [11]；关于有向图、源点和目标点的定义，请参阅第 2.1 节。

定义 4.5（全局词图）。大小为 n 的双出现词的全局词图 $G_n = (V, E)$ 是如下定义的图：

- $V(G_n) = \Sigma_{DOW}^{\leq n}/_{\sim}$;
- $E(G_n) = \bigcup_{w \in V} E_w$，其中 $E_w = \{[w, v] \mid v \in D(w)\}$。

对于 G_n 中的顶点 w，我们将以 w 为根的词图记为 G_w，定义为全局词图的诱导子图，其顶点包括 w 及其所有后继。

根据构造， G_w 不包含任何有向环，并且有唯一的源点 w 和唯一的目标点 ϵ，因此词图是一致定向的。

本节中的图均由计算机生成。虽然我们在展示中省略了字符，但为便于阅读，顶点上的标签以逗号分隔。

例 4.6. 大小为 2 的全局词图 G_2 如图 11所示，其中以 1122 为根的词图以蓝色突出显示。顶点集为 $V(G_2) = \Sigma_{DOW}^{\leq 2}/_{\sim} = \{\epsilon, 11, 1221, 1212, 1122\}$。以 $w = 1234523541$为根的词图（其后继在例 4.4 中计算）如图 12 所示。

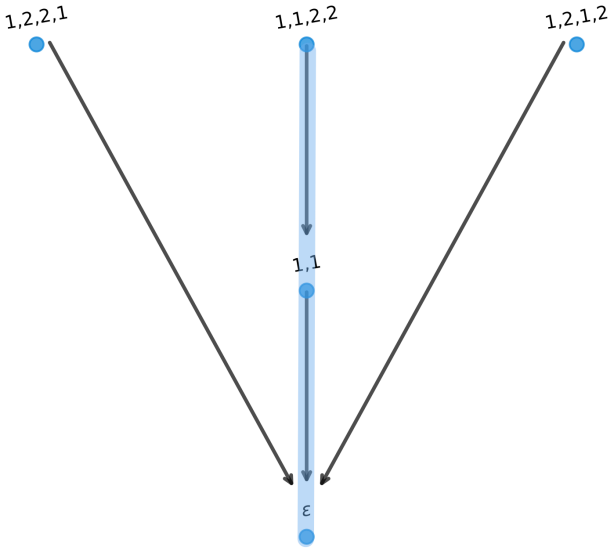


图 11：大小为 2 的全局词图。

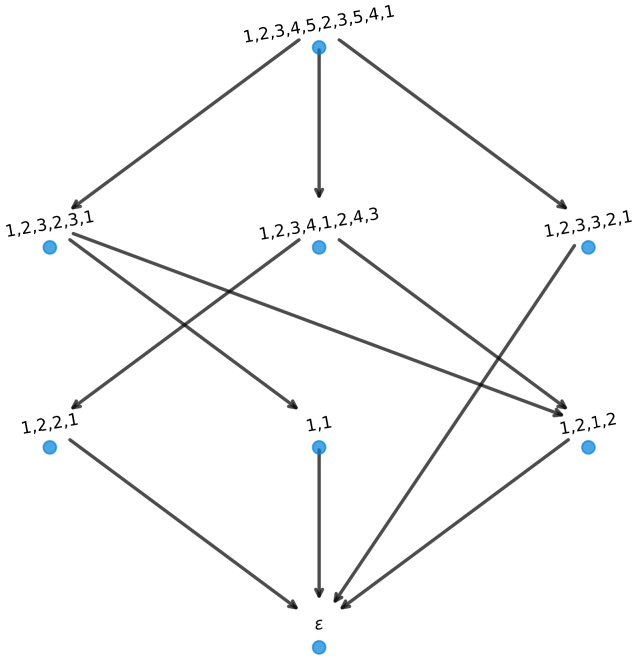


图 12：以 1234523541 为根的词图。

4.3 DOW 操作及其对词图的影响

4.3.1 导致同构词图的操作

我们给出一些结果，讨论插入、替换或反转不会影响 DOW 的词图的情形。与 $w \in \Sigma_{DOW}$ 相关联的乘积单纯复形是与 G_w 相关联的乘积单纯复形。我们考察能够产生这样的 DOW 类的操作：这些 DOW 的复形具有相似的拓扑性质。

为了找出给定 DOW w 的所有前驱，我们考虑 w 的所有可能插入，直至升序等价。在 [9] 中，能够产生等价 DOW、从而得到同构词图的插入已得到刻画。这里我们给出关于那些彼此不升序等价、但能产生同构词图的 DOW 的结果。

定义 4.7。设 $w, w' \in \Sigma_{DOW}$ 满足 $\Sigma[w] \cap \Sigma[w'] = \varnothing$。我们将 w 与 w' 的串接定义为与 ww' 升序等价的 DOW。

P命题 4.8。设 $w \in \Sigma_{DOW}$，且 $u \in M_w^{SOW}$。设 $v \in (\Sigma \setminus \Sigma[w])_{SOW}$，并设

$$w' = xu_1vu_2yu_1vu_2z$$
$$w'' = xu_1vu_2yu_2^R v^R u_1^R z$$

其中 u_1、$u_2 \in \Sigma^*$ 满足 $u = u_1u_2$。那么 $G_w \cong G_{w'} \cong G_{w''}$。

证明。在这种情况下，$u_1vu_2u_1vu_2$ 是 w'中的极大重复词。注意，$M_{w'}^{SOW} = (M_w^{SOW} \setminus \{u\}) \cup \{u_1vu_2\}$，但 $d_u(w) = d_{u_1vu_2}(w')$。因此，w 和 w' 的词图同构：$G_{w'} \cong G_w$。uu^R 是极大返回词的情况类似。□

如果 $w \in \Sigma_{DOW}$ $w^R \sim w$，我们就称其为回文词。例如，DOW $w = 123231$是回文词，因为 $w^R = 132321 \sim 123231$。对于回文词 w，w 和 w^R 都属于同一个升序等价类，因此对应的词图相同。因此，反转操作不会影响所得到的乘积单纯复形。

引理 4.9. 令 $w \in \Sigma_{DOW}$。则 $G_w \cong G_{w^R}$。

证明。注意，u 是 w 中的重复词（分别为回返词）当且仅当 u^R 是 w^R 中的重复词（分别为回返词）。因此，如果 v 是 u 的后继，对于每条边 $[u, v] \in G_w$，我们都有一条对应的边 $[u^R, v^R] \in E(G_{w^R})$。这个边之间的双射诱导出图之间的同构。 □

例 4.10。设 $w = 122133$。那么 $w^R = 331221 \sim 112332$。以 122133 和 112332 为根的词图是同构的，如图 13所示。

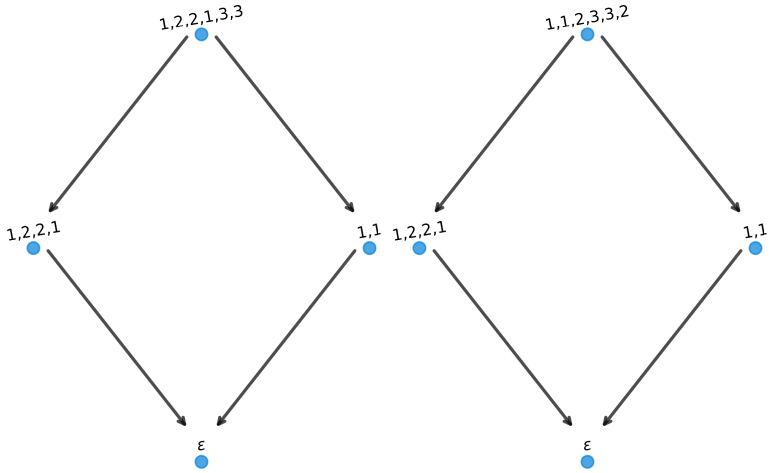


图 13：DOW 112332 与 122133^R 升序等价，因此 G_{112332} 与 G_{122133} 同构。

本节余下部分，我们考察用返回词 uu^R 替换重复词 uu 对词图的影响。在某些情况下，我们可以用 w 中的一个极大返回词替换 DOW w 中的一个最大重复词，而不影响词图。以下性质似乎在这一情境中发挥着重要作用。

定义 4.11（平方重复词或回返词）。设 $w \in \Sigma_{DOW}$ 且设 $u \in M_w^{SOW}$。若存在 $v \in M_w^{SOW}$ 使得 $v \neq u$，但 $|v| = |u|$，则称 u 是 w 的平方因子。若一个 DOW w 不含平方因子，则称其为无平方因子的。

注意，如果 w 无平方因子且 $uu \in M_w^{DOW}$，那么对于所有满足 $|u| = |v|$ 的 $v \in M_w^{SOW}$，都有 vv、$vv^R \notin M_w^{DOW}$。由于 w 中没有重复因子或返回因子出现多次，因此对于 $u \in M_w^{SOW}$，所有删除 $d_u(w)$ 都会得到互不相同的 DOW。特别地，w 的所有子词也都无平方因子。

例 4.12（平方因子、无平方因子词）。设 $w = 12123434$。那么 $M_w^{DOW} = \{12, 34\}$，且 12 是 w 中的一个平方重复词。设 x、y、$z \in \Sigma_{SOW}$ 的长度互不相同。那么 $w' = xyxzyz$ 无平方因子。

定义 4.13（互素）。给定两个 DOW w 和 w'，其中 $\Sigma[w] \cap \Sigma[w'] = \varnothing$，如果所有形如 uv 的词（其中 $u \in V(G_w)$ 和 $v \in V(G_{w'})$）按升序排列后互不相同，我们就称 w 与 w' 互素。

根据互素词的定义，词 w 和 w' 互素，当且仅当对于所有的 $u, u' \in V(G_w)$ 以及所有的 v，$v' \in V(G_{w'})$，若 $u \neq u'$ 和 $v \neq v'$，则 $uv \not\sim u'v'$。

例 4.14（互素词）。词 $w = 12234143$ 的后继为 $\{1221, 112332, 123132, 11, \epsilon\}$，词 $w' = 5678978956$ 的后继为 $\{5656, 789789, \epsilon\}$. 可以验证，所有拼接都各不相同，因此 w 和 w' 互素。

引理 4.15. 设 w' 是 w 的一个后继。那么 $G_{w'}$ 是 G_w 的一个诱导子图。 .

引理 4.16。如果 w 与 w' 互素，则 w' 与 w 互素。

证明。假设 w' 与 w 不互素，并令 $u, u' \in V(G_w)$（其中 $u \neq u'$）和 $v, v' \in V(G_{w'})$（其中 $v \neq v'$）满足 $uv = u'v'$。注意，由于 $\epsilon \in V(G_w) \cap V(G_{w'})$，若存在非空 DOW $x \in V(G_w) \cap V(G_{w'})$，则 w 不可能与 w'互素，因为 $\epsilon x = x\epsilon = x$。不失一般性，令 $|u| > |u'|$。那么，对于某个 DOW x（因为 $u \in \Sigma_{DOW}$）和 $v = xv'$，有 $u' = ux$，从而 $M_v^{SOW} = M_{v'}^{SOW} \cup M_x^{SOW}$。此外，对于所有 $y \in M_{v'}^{SOW}$，若 $v_1 = d_y(v)$，则 $v = xv_1'$，其中 $v_1' = d_y(v')$。归纳地，若 v通过删除 $y_1, y_2, \ldots, y_k$ 化简为 $v_1', v_2', \ldots, \epsilon$，那么 v通过删除 $y_1, y_2, \ldots, y_k$化简为 x。也就是说，$x \in V(G_w)$。类似地，若 $u' = ux$，则有 $x \in V(G_{w'})$，因此 w' 与 w不互素。 □

鉴于这种对称性，我们不说 w 与 w'互素，而说 w 与 w' 互素。

在以下示例中，$w = xyz$ 与 u 是互素词，而将 uu（分别将 uu^R）插入 w 后得到的词 w'（分别为 w''）是无平方因子词。在一个例子中，替换得到 $G_{w'} \cong G_w$；而在另一个例子中则没有。我们推测，无平方因子和互素是重复词与返回词的替换保持词图不变的必要条件。

例 4.17 (替换得到同构词图)。设 $x = 12$，$y = 345$，$z = 54312$和 $u = 6789$。那么 $w' = xuyuz \sim 123456789345987612$和 $w'' = xuyu^Rz \sim 123456789543987612$。这里，$uu$ 和 xyz 互素，并且两次插入都得到无平方因子词，其中 $G_{w'} \cong G_{w''}$如图14所示。

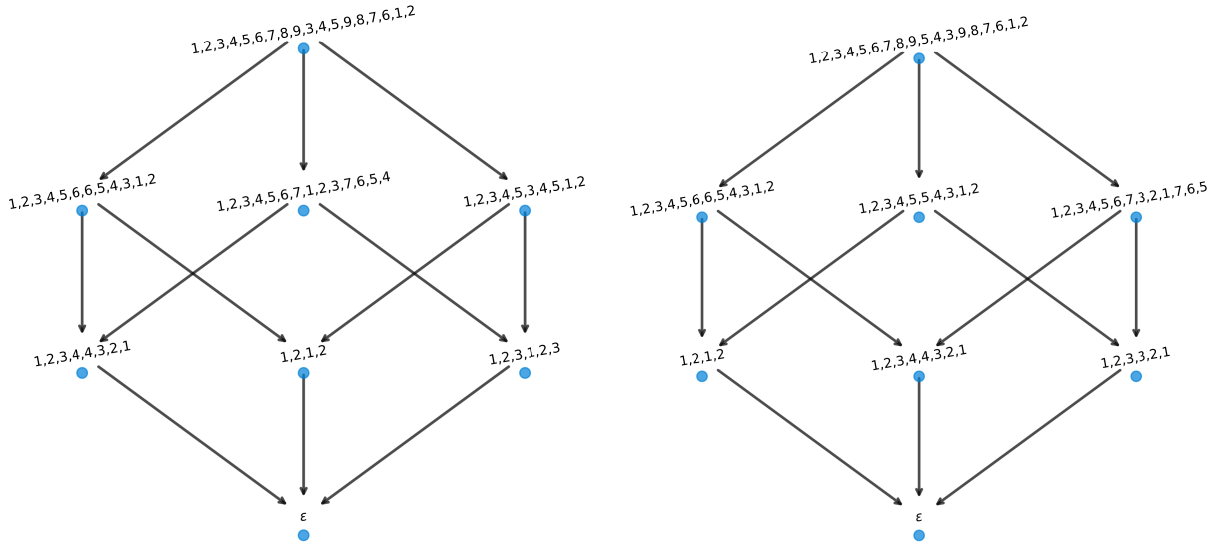


图 14：$w' = xuyuz$（左）和 $w'' = xuyu^Rz$（右）的词图同构，因为它们都对应于立方体。

例4.18（替换不会产生同构词图）。设 $x = 1$，y =12345，$z = 5432$ 和 $u = 67$。那么 $w' = xuyuz \sim 12314567237654$和 $w'' = xuyu^Rz \sim$12314567327654。这里有 uu 与 xyz 互素，所得的插入结果均无平方因子，但 $G_{w'} \ncong G_{w''}$ 如图 15所示。

4.3.2 加倍效应

在现有词 w末尾串接一个重复词 uu（分别为返回词 uu^R），会在 G_{wuu}（分别为 G_{wuu^R}）中生成两个与 G_w 同构的子图。除非施加附加条件，否则这两个子图可能并不互不相交，以下示例将对此进行讨论。

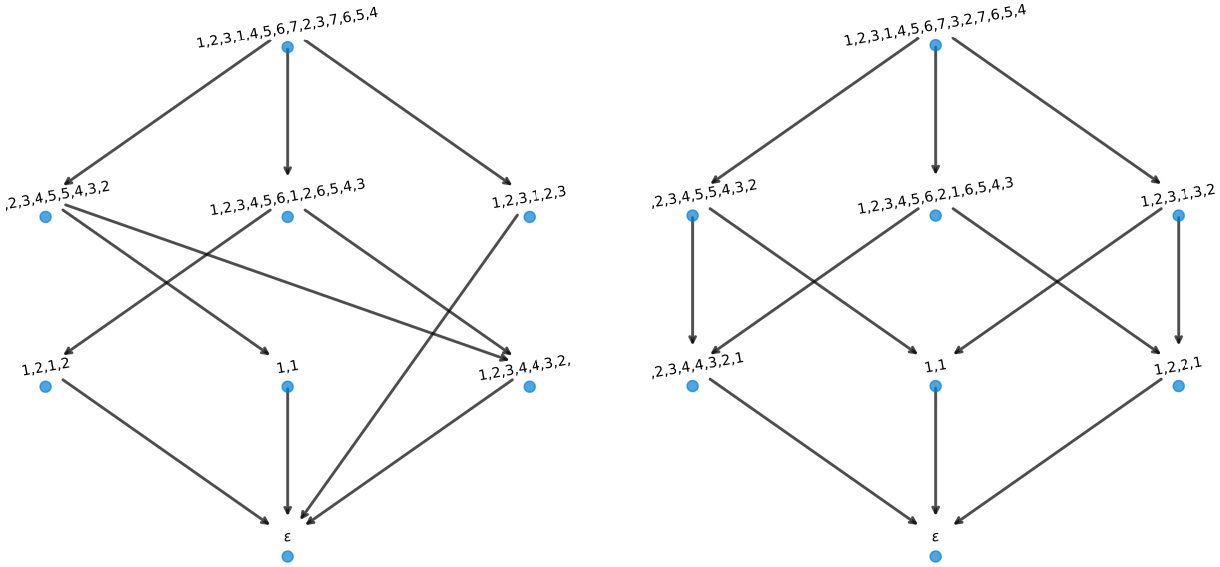


图15: $w' = xuyuz$ （左）和 $w'' = xuyu^R z$ （右）的词图不同构 .

引理 4.19. 设 $w \in \Sigma_{DOW}$ 和 $uu \in (\Sigma \setminus \Sigma[w])_{DOW}$ 互素。则存在两个不同（但可能不互不相交）的 G_{wuu} 的子图 G_1 和 G_2，它们与 G_w 同构。类似地，G_{uuw} 有两个与 G_w 同构的子图。此外，我们有 $G_{wuu} \cong G_{uuw}$。

证明。无妨只对 G_{wuu} 证明该结论。我们断言，对于每个$v \in V(G_w)$，词 vuu 都与某个 $v' \in V(G_{wuu})$ 升序等价。事实上，如果 v 是通过反复删除从 w 得到的，使得

$$v \sim d_{x_1}(d_{x_2}(d_{x_3}(\cdots d_{x_n}(w))\cdots))$$

其中对所有 $1 \le i \le n$都有 $x_i \not\sim u$，那么 $v' \sim vuu$ 是通过以下方式从 wuu 得到的

$$vuu \sim d_{x_1}(d_{x_2}(d_{x_3}(\cdots d_{x_n}(wuu))\cdots)).$$

令 $\iota : V(G_w) \to V(G_{wuu})$ 为顶点的包含映射，并令 G_1 为 G_w 在 G_{wuu}中的像所诱导的图。令 G_2 为 G_{wuu} 中由形如 $v' \sim vuu$ （其中 $v \in V(G_w)$）的顶点集合生成的诱导子图。定义 $f : V(G_1) \to V(G_2)$，其中 $v \mapsto v' \sim vuu$。由于上述对应关系， f 是顶点集之间的双射。注意，若 $[v_1, v_2] \in G_w$，则对某个 $x \in M_{v_1}^{SOW}$有$v_2 \sim d_x(v_1)$。这成立当且仅当 $v_2uu \sim d_x(v_1uu)$ 或 $[v_1uu, v_2uu] \in G_{wuu}$，因此 f 在保持顶点邻接关系的同时，在边集之间诱导出一个双射。由此可得 $G_1 \cong G_2$。 □

例 4.20. 与 12132344 对应的词图中有两个同构于五边形的子图 G_{121323}，如引理 4.19所述；它们并不互不相交。这一点可从图 16中看出，其中两个子图以不同颜色突出显示。

4.3.3 乘积效应

在本小节中，我们研究连接词的词图。

定理 4.21. 设 w 与 $w' \in \Sigma_{DOW}$ 互素。则有 $G_{ww'} \cong G_w \square G_{w'}$。

证明。我们首先注意到，根据串接的定义， $G_{ww'}$ 中的每个顶点都可以分解为 u_iv_j，其中 $u_i \sim u \in V(G_w)$ 且 $v_j \sim v \in V(G_{w'})$。我们定义映射 $f : V(G_w \square G_{w'}) \to V(G_{ww'})$，其中 $(u, v) \mapsto u_iv_j$。根据假设，由于 w和 w'互素，

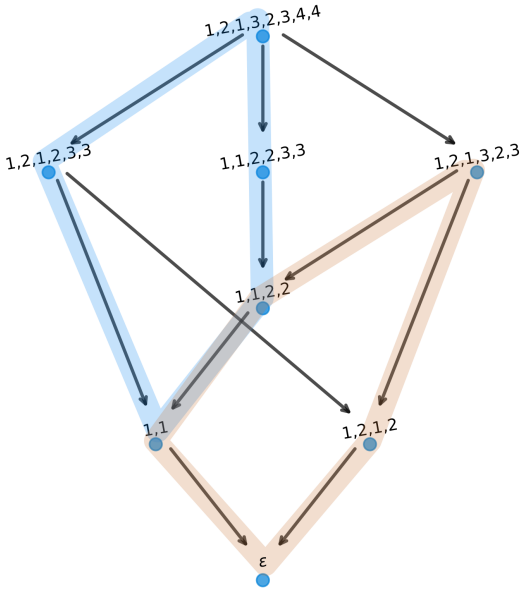


图 16：词 12132344 的词图，其中两个高亮子图同构于 G_{121323}。

$u \neq u'$ 和 $v \neq v'$ 意味着对于所有 $u, u' \in V(G_w)$ 以及所有 $v, v' \in V(G_{w'})$，都有 $uv \not\sim u'v'$。这使得 f 是单射。满射性源于这样一个事实：ww' 的后继只能由 w 的后继与 w' 的后继串接而成。因此，f 是顶点集合之间的双射。

我们稍微滥用一下记号，将 $G_{ww'}$ 的顶点写作 uv，其中 $\Sigma[u] \cap \Sigma[v] = \varnothing$，并按照上述分解，令 $u \sim u' \in V(G_w)$ 和 $v \sim v' \in V(G_{w'})$。令 $[(u_i, v_j), (u'_i, v'_j)] \in E(G_w \square G_{w'})$。那么，要么 $u_i = u'_i$ 且对于某个 $x \in M^{SOW}_{v_j}$ 有 $v'_j \sim d_x(v_j)$，要么 $v_j = v'_j$ 且对于某个 $y \in M^{SOW}_{u_i}$ 有 $u'_i \sim d_y(u_i)$。注意，由于 w 和 w' 不共享任何符号，我们可以写成 $M^{DOW}_{ww'} = M^{DOW}_{w} \sqcup M^{DOW}_{w'}$。我们观察到，对于 $u_i v_j, u'_i v'_j \in V(G_{ww'})$，存在一条边 $[u_i v_j, u'_i v'_j] \in E(G_{ww'})$ 当且仅当 $u_i = u'_i$ 且 $[v_j, v'_j] \in E(G_{w'})$，或者 $[u_i, u'_i] \in E(G_w)$ 且 $v_j = v'_j$。这意味着，$[(u_i, v_j), (u'_i, v'_j)] \in E(G_w \square G_{w'})$ 当且仅当 $[f((u_i, v_j))\ f((u'_i, v'_j))] \in E(G_{ww'})$，这表明 $G_{ww'} \cong G_w \square G_{w'}$，如所需。 □

当 w' 是重复词或返回词，且其词图由两个顶点和连接它们的一条边构成时，有如下推论。

注释 4.22。注意，若 $w \in \Sigma_{DOW}$ 和 $u \in \Sigma_{SOW}$ 满足：对于所有 $v \in V(G_w)$，都有 $uu \not\sim v$，则 $G_{wuu} \cong G_w \square \Delta^1$。类似地，若对于所有 $v \in V(G_w)$，都有 $uu^R \not\sim v$，则 $G_{wuu^R} \cong G_w \square \Delta^1$。

5 词图的贝蒂数与生成元

在证明了任意图的一般结果后，我们现在聚焦于与 DOW 的词图对应的一致定向图，以研究贝蒂数的可实现性。注意，对于 DOW w 和 w'，若 $w \in M^{DOW}_{w'}$，则 $G_w \leq G_{w'}$。因此，为了在 DOW 的词图中寻找第一同调群和第二同调群的生成元，我们转而考察大小最小且对应图具有特定 β_1 和 β_2 值的词。

我们使用了一个 Python 脚本（可在 https://github.com/fajardogomez/dowgraphs 获取），用来为任意有向图 G 构造相应的乘积单纯复形并计算贝蒂数。

5.1 词图的贝蒂数

让我们考察 DOW 的词图中大小最小的非平凡 1-循环。我们断言，对于 DOW w， G_w 满足 $H_1(\Gamma(G_w)) \cong \mathbb{Z}$，且所有其他同调群都是平凡的：

$$1. \quad 121323, \qquad\qquad 2. \quad 122331.$$

对应的词图如图17所示。

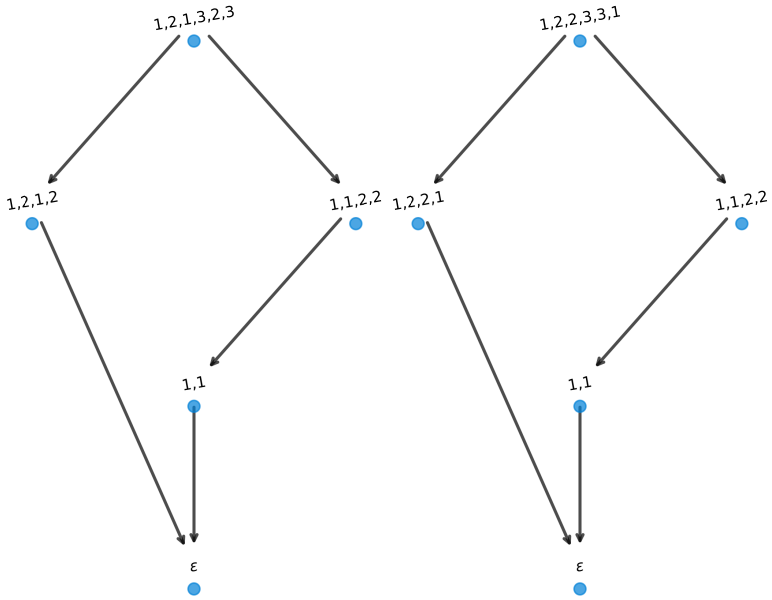


图17：DOW 121323 和 122331 对应的词图同构，且其乘积单纯复形中包含非平凡的 1-循环。

该结果由引理 3.3 得出，因为这些图由从源点到目标点的平行路径组成，路径长度分别为 2 和 3，共有五条边，形成一个五边形。能在第一同调群中产生生成元且符号数最少的 DOW 是 121323 和 122331。

请注意，作为正方形的 1-环不能成为第一同调群的生成元，因为词图（DOW 图）中的每个正方形都由两条长度为 2 的路径组成，因此围成一个 2-胞腔。在我们考察的所有例子中，表示$n \geq 6$ 的非平凡 1-循环的 n边形都可以化简为五边形，因此我们猜想所有非平凡的 1-循环都与五边形同调等价。

例5.1. 设 u_1、 u_2 和 α 是单出现词，且 $\Sigma[u_1] \cap \Sigma[u_2] \cap \Sigma[\alpha] = \varnothing$。以下词模式会形成五边形，并将词图121323和122331推广为： $\alpha u_1 u_1^R u_2 u_2^R \alpha^R$和 $u_1 \alpha u_1 u_2 \alpha u_2$ ，条件是 $|u_1| = |u_2|$。

另一方面，第3.2 节讨论的所有表示非平凡第二同调生成元的多面体类型，都可以在词图中实现。

以 12323414 为根的词图对应的复形与引理 3.5 中的图的复形同胚。参见图 18 可验证这一点。其他词图包含与引理 3.10 和引理 3.4 中所述子图同构的子图。例如，以 12323144 为根的词图同时包含这两种子图。

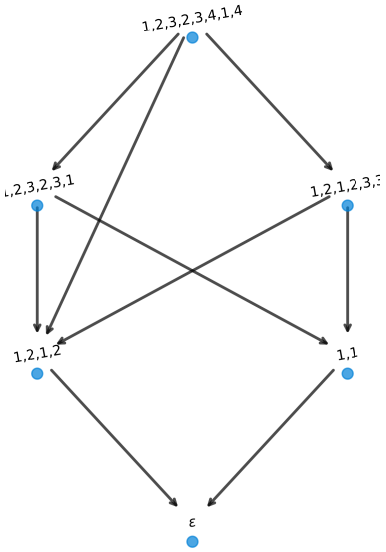


图18：以12323414为根的词图。

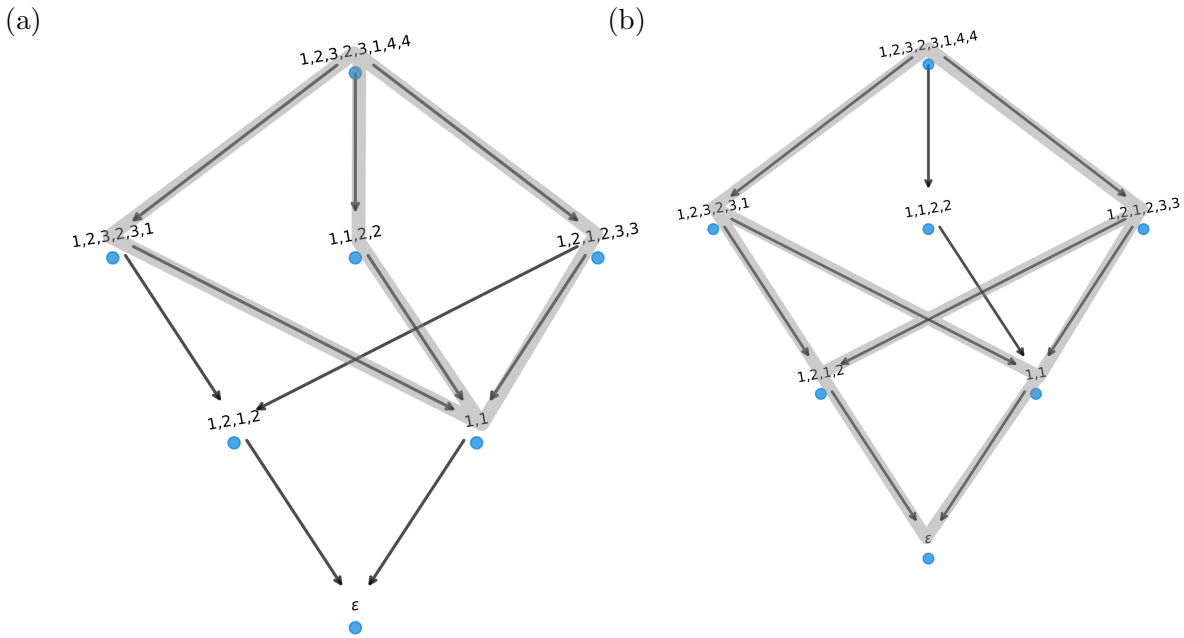


图19：（a）以12323144为根的词图，其中一个与图7中的子图同构的子图已高亮显示。（b）以12323144为根的词图，其中一个与图8（a）中的子图同构的子图已高亮显示。

5.2 后继的贝蒂数

一般而言，如果 u 是 w 的后继，并不能由此推出 $\beta_n(G_w) \geq \beta_n(G_u)$。受 [3]，中工作的启发，我们定义 n-符号缠结弦 为如下词

$$t_n := 1213243 \cdots (n-1)(n-2)n(n-1)n.$$

注意，$d_n(t_n) = t_{n-1}$ 因此 $t_{n-1} \in V(G_{t_n})$ 对任意 $n > 1$。例如，$t_6 =$ 121324354656（其对应的词）是大小为五的缠结绳索 $t_5 = 1213243545$的一个后继。根据自定义 Python 脚本的结果，我们有 $\beta_1(G_{t_6}) = 1$ 以及$\beta_1(G_{t_5}) = 2$。这是出现贝蒂数反转的最小一对连续缠结绳索。下一次反转出现在 β_2 当 $n = 11$时，如表 1所示。

n	t_n	$\beta_1(G_{t_n})$	$\beta_2(G_{t_n})$	$\lvert V(G_{t_n})\rvert$
2	1,2,1,2	0	0	2
3	1,2,1,3,2,3	1	0	5
4	1,2,1,3,2,4,3,4	1	2	8
5	1,2,1,3,2,4,3,5,4,5	2	6	13
6	1,2,1,3,2,4,3,5,4,6,5,6	1	27	21
7	1,2,1,3,2,4,3,5,4,6,5,7,6,7	1	54	34
8	1,2,1,3,2,4,3,5,4,6,5,7,6,8,7,8	1	86	55
9	1,2,1,3,2,4,3,5,4,6,5,7,6,8,7,9,8,9	1	111	89
10	1,2,1,3,2,4,3,5,4,6,5,7,6,8,7,9,8,10,9,10	1	126	144
11	1,2,1,3,2,4,3,5,4,6,5,7,6,8,7,9,8,10,9,11,10,11	1	116	233
12	1,2,1,3,2,4,3,5,4,6,5,7,6,8,7,9,8,10,9,11,10,12,11,12	1	112	377

表1：缠结绳索及其词图的不变量。

寻找后继 w' of W 的条件很有意义，使得对于某个 $n \geq 1$，$\beta_n(G_w) \geq \beta_n(G_{w'})$ 不成立。

6 总结

本文提出了一种通过为给定有向图构造乘积单纯复形来研究有向图拓扑性质的方法。该构造为在可用图描述的生物数据上应用拓扑数据分析提供了一种途径。文中介绍了一个具体示例，即词图，它通过 DOW 对纤毛虫物种中的组装路径进行建模[19]。本文将所提出的乘积单纯复形构造应用于词图，并确定了其贝蒂数。本文还研究了各种操作（例如词的拼接）对词图的影响，并考察了同调生成元的类型。

仍有许多问题尚未解决。引理 3.3 刻画了有向图第一同调群的生成元，而推论3.9 讨论了 β_1 和 β_2各种组合的可实现性。然而，我们并不知道所有词图中贝蒂数 β_1 和β_2 可能取值的所有组合。虽然本文只列举了少数几个特定 DOW 示例，其对应的词图具有非平凡的 1-环和 2-圈，但通过第 4.3.1 节所述的双重出现词操作，还可以得到更多示例。例如，将 121323 与 4545 串接，对应于词图的笛卡尔积（如注 4.22 所示），因此不会影响贝蒂数，所以 $\beta_1(G_{1213234545}) = \beta_1(G_{121323}) = 1$。然而，这些方法无法穷尽所有具有特定贝蒂数的图。值得进一步刻画其他词操作的影响，例如那些会破坏最大重复词或返回词的插入操作，因为这类操作可能改变其词图上复形的同调群。

我们预计，代表非平凡类的 1-环和 2-圈类型列表（例如引理 3.3、引理 3.4和引理 3.5 中的类型）可能并未穷尽。我们在词图中发现了如图 17 所示的五边形，它们是非平凡的 1-环。我们不知道是否还有其他代表非平凡圈的多边形。 `SageMath` [10] 拓扑学软件包提供了一个选项，可用于求出同调群的生成元。不过，输出中的线性组合未必经过优化，无法保证生成元数量最少。例如，词图 t_{10} 的第一同调群的一个生成元被列为九条边的线性组合。

然而，将这些边画成一个诱导子图时便会发现，圈上的顶点之间存在边，因此它与一个由五边形表示的类同调等价。

对表示非平凡二阶同调类的可能多面体进行分类也很有意义。我们在第 3 节只给出了部分答案。由两个共享全部四条边界边的正方形构成的“枕形”等多面体，不能用于构造乘积单纯复形。这里计算的所有词图，只要其对应的乘积单纯复形中存在非平凡的 1-环或 2-圈，就都含有与引理 3.1 至 3.5 以及 3.10 中所示结构同构的子图。如何用最少数量的顶点和边找到一份完备的生成元列表，仍是一个开放问题。

关于一般的同调群，补充说明一点：SageMath输出的是完整的群，而不只是本文所述的贝蒂数。在对至多含 7 个符号的所有 DOW 进行分析时，没有发现任何 DOW 的词图所对应的乘积单纯复形具有挠[12]。由于每种大小的 DOW 数量呈超指数增长，因此，对于较长的 DOW，只对随机生成的 DOW 样本和特殊词类别计算了同调群。一个有趣的发现是，含有十个符号的缠结弦的同调群 t_{10} 在第一（或第二）同调群中具有 2-挠。然而，由于生成元中顶点和边的数量过多，无法找到一种 2-挠生成元的类型。如何找出并刻画那些对应的乘积单纯复形的同调群具有挠的 DOW，仍是一个开放问题。

本文提出的有向图的乘积单纯复形构造，是首次尝试将拓扑数据分析应用于由图组成的数据集，尤其针对称为词图的生物学输出。生物学研究中还获得了其他图输出，值得进一步开发拓扑数据分析，以便在更一般的情形下处理这类输出。更广泛地说，我们提出构造定制的胞腔复形，以研究具体的生物学问题。构造复形时，应指定多面体胞腔，使其在边界算子作用下封闭，并适用于特定的生物学情境。

DOW的词图对应于弦图的约化，我们预期这些结果可应用于与弦图相关的领域。重复词与返回词的约化可以推广到其他重写系统。对于合流重写系统，值得采用类似的复形构造，并利用其同调。

致谢

本研究部分得到了以下资助：NSF DMS-2054321、CCF-2107267、Simons基金会的Simon’s Fellow资助以及W.M. Keck基金会。此外，本研究是在东南数学与生物学中心的支持下开展的。该中心是一个由美国国家科学基金会资助的复杂生物系统数学研究中心（NSF-Simons Research Center for Mathematics of Complex Biological Systems），经费包括美国国家科学基金会资助（DMS-1764406）和Simons基金会资助（594594）。

参考文献

[1] Ryan Arredondo。《双出现词上的约化》。英文。载于：arXiv.org（2013）。ISSN：2331-8422。[2] Ryan C. Arredondo。《用于模拟 DNA 重组的图的性质》。英文。2014。[3] Jonathan Burns 等。《与 DNA 重组相关的带刚性顶点的四正则图》。载于：《离散应用数学》161.10（2013），第 1378–1394 页。ISSN：0166-218X。

[4] Jonathan Burns 等。《纤毛虫 Oxytricha trifallax 加密基因组中基因重排的重复模式》。载于：《理论生物学杂志》410（2016），第 171–180 页。[5] J. Scott Carter、Victoria Lebed 和 Seung Yeop Yang。《棱柱分类空间》。载于：Contemp. Math. 721（2019），第 43–68 页。[6] J. Scott Carter 等。《具有部分群运算的幺群胚的同调》。载于：《太平洋数学杂志》287.1（2017），第 19–48 页。[7] Andre R. O. Cavalcanti 和 Laura F. Landweber。《深入了解生物计算机：解开纤毛虫中的重排基因》。英文。载于：《纳米技术：科学与计算》。自然计算丛书。柏林、海德堡：Springer Berlin Heidelberg，第 349–359 页。ISBN：3540302956。[8] 《有向图类》。英文。第 1 版。2018。数学专著丛书。Cham：Springer International Publishing，2018。ISBN：3-319-71840-1。[9] Daniel A. Cruz 等。《产生等价双出现词的插入》。载于：《基础信息学》171（2020），第 113–132 页。[10] TheSage Developers 等。SageMath，版本 9.0。2020。URL：http://www.sagemath.org.

[11] Reinhard Diestel。《图论》（数学研究生教材）。纽约：Springer，2005。[12] Lina Fajardo G´omez。《利用离散数学方法研究 DNA 重排过程》。博士论文。南佛罗里达大学，2022。[13] Joan Feigenbaum。《有向笛卡尔积图具有唯一的因子分解，且可在多项式时间内计算》。载于：《离散应用数学》15.1（1986），第 105–110 页。[14] Alexander Grigor'yan 等。《路径复形与有向图的同调》。2013。arXiv：1207。2834 [math.CO]。[15] Alexander A. Grigor'yan 等。《路径复形及其同调》。载于：《数学科学杂志》248（2020），第 564–599 页。[16] Mustafa Hajij 等。《基于图的分析在基因片段组织中的应用：以重排基因组为例》。载于：《理论生物学杂志》494（2020），第 110215 页。ISSN：0022-5193。[17] Allen Hatcher。《代数拓扑》。代数拓扑。纽约：剑桥大学出版社，2002。[18] Wilfried Imrich 和 Sandi Klavˇzar。《积图：结构与识别》。纽约：Wiley，2000。ISBN：0471370398。[19] Nataˇsa Jonoska、Lukas Nabergall 和 Masahico Saito。《与 DNA 重排相关的词中的模式与距离》。载于：《Fundamenta Informaticae》154（2017 年 8 月），第 225–238 页。[20] Dimitry Kozlov。《组合代数拓扑》。第 21 卷。纽约：Springer Science & Business Media，2007。[21] Paolo Masulli 和 Alessandro E. P Villa。《有向团复形的拓扑作为网络不变量》。英文。载于：《SpringerPlus》5.1（2016），第 388–388 页。ISSN：2193-1801。

[22] David M. Prescott 和 Arthur F. Greslin。《Oxytricha nova 微核中的重排肌动蛋白 I 基因》。载于：《发育遗传学》13.1（1992），第 66–74 页。 ISSN：0192-253X。[23] Michael W Reimann 等。《神经元团簇被束缚于空腔中，为结构与功能之间缺失的联系提供线索》。英文。载于：《计算神经科学前沿》11（2017），第 48–48 页。 ISSN：1662-5188。[24] Vadim Georgievich Vizing。《图的笛卡尔积》。载于：Vyčisl. Sistemy 9（1963），第 30–43 页。[25] V Talya Yerlici 和 Laura F Landweber。《纤毛虫 Oxytricha 中的程序化基因组重排》。载于：《微生物学综述》2.6（2014）。