Product-Simplicial Complexes on a Word Graph

Lina Fajardo Gómez, Margherita Maria Ferrari, Nataša Jonoska, Masahico Saito

4202 E. Fowler Ave Tampa, FL 33620, USA

Introduction

The ciliate *Oxytricha trifallax* undergoes a major rearrangement process during the development of a somatic macronucleus (MAC) from a germline micronucleus (MIC). Repetitive sequences, called pointers indicate where a MIC segment will be placed in a MAC nanochromosome. These rearrangement processes can be modeled by labeling pointers in MIC segments and studying the sequences of these labels. The labels yield double occurrence words (DOWs): strings where each symbol appears exactly twice.

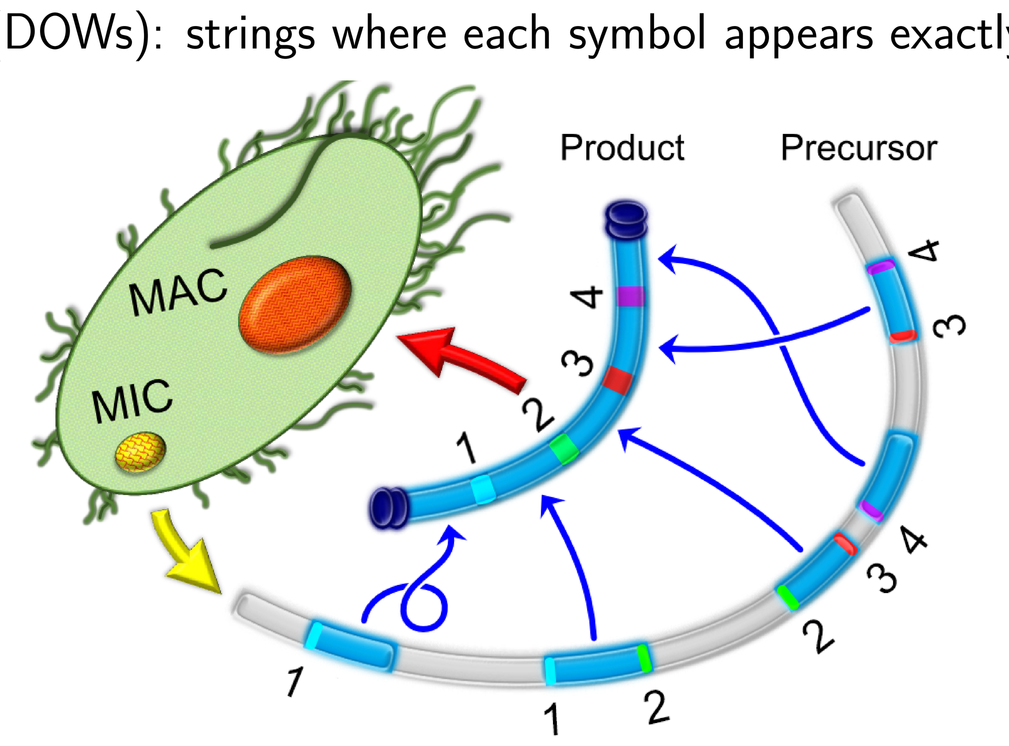


Figure 1: Pointers form double occurrence words (DOWs). In this example, the DOW is 11223434.

Double Occurrence Word Reduction

A DOW is in ascending order if its symbols are labeled by order of appearance. Two DOWs are equivalent if they have the same ascending order representation.

Ascending Order Equivalence

311223 ∼ 122331

3 1 1 2 2 3
1 2 2 3 3 1

uu is a repeat word in w if $w = z_1 u z_2 u z_3$. uu^R is a return word in w if $w = z_1 u z_2 u^R z_3$. In both of these cases, we say that w reduces to $z_1 z_2 z_3$. Repeat and return words describe over 90% of DNA rearrangement in *Oxytricha trifallax* [1].

Repeat and Return Words

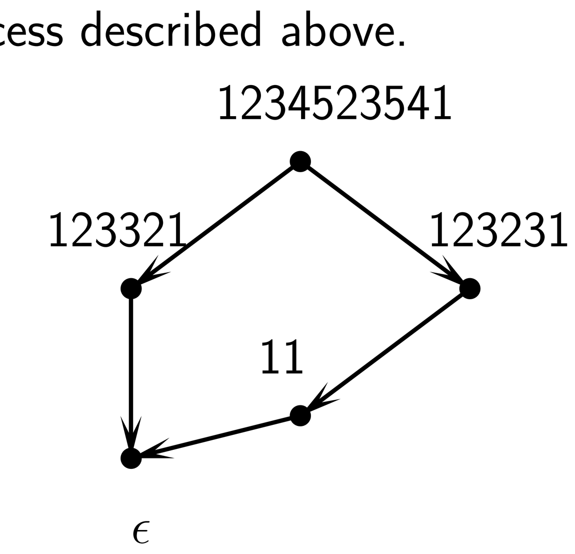
2323 is a repeat word in 1234523541
4554 is a return word in 1234523541.

Word Reduction Graph

We can create a directed graph where vertices are words and an edge indicates that one word can be reduced to another through the process described above.

Reduction of a Word

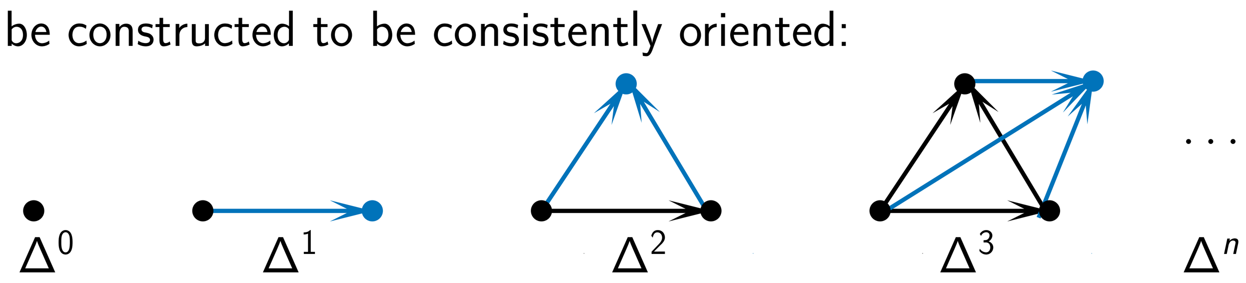
1234523541 → 145541 ∼ 123321
1234523541 → 123231
123231 → 11
123321 → ϵ
11 → ϵ



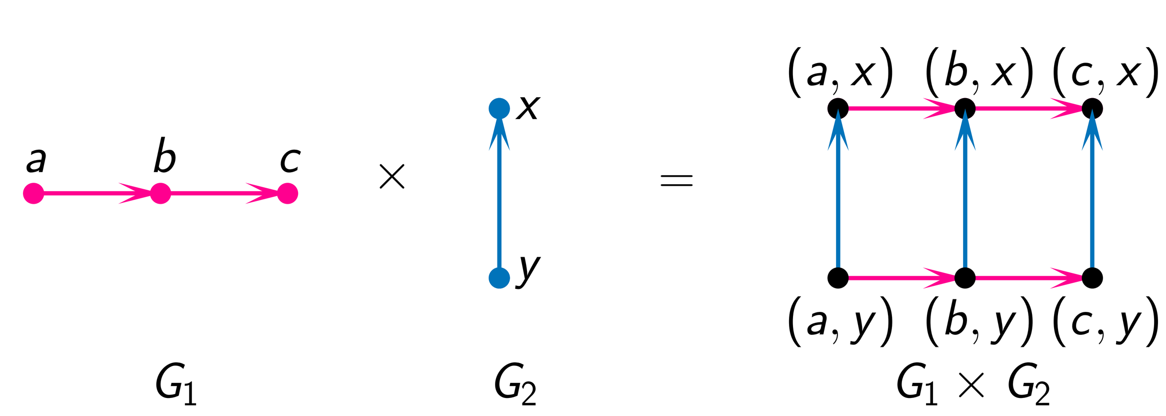
Consistently Oriented Graphs

Directed graphs with a single source and a single target are said to be consistently oriented:

Simplices can be constructed to be consistently oriented:



Cartesian Product of Simplices



$((u_1, v_1), (u_2, v_1)) \in E$ whenever $(u_1, u_2) \in E(G_1)$. $((u_1, v_1), (u_1, v_2)) \in E$, whenever $(v_1, v_2) \in E(G_2)$. Up to relabeling of vertices, this product is associative.

Product-Simplicial Cells

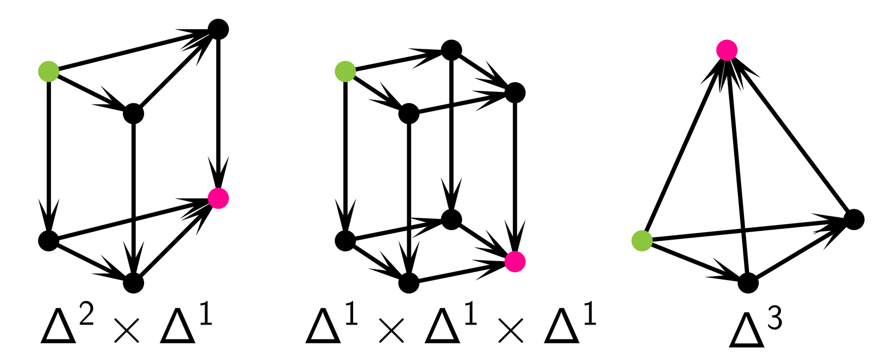
We define product-simplicial cells as Cartesian products of consistently oriented simplices:

$$G = \Delta^{n_1} \times \cdots \times \Delta^{n_k}$$

The dimensions of the factors can be added to obtain the dimension of the resulting cell: $\dim(G) = \sum_{j=1}^{k} n_j$.

Lemma

Cartesian products of consistently oriented graphs are consistently oriented.



Homology Groups

An N-chain is a linear combination of N-dimensional simplices. A prodsimplicial cell complex is a collection Γ of prodsimplicial cells such that the nonempty intersection of any two prodsimplicial cells in Γ is in the union of their face sets. We define a boundary operator on prodsimplicial cells by

$$\partial_N : \prod_{j=1}^{k} \Delta^{n_j} \mapsto \sum_{j=1}^{k} (-1)^{\alpha(j)} [\overline{\partial_{n_j} \Delta^{n_j}}]$$

where $N = \sum_j n_j$, $\alpha(j) = \sum_{l=1}^{j-1} n_l + j$ and $[\overline{\partial_{n_i} \Delta^{n_i}}]$ denotes the product $\Delta^{n_1} \times \Delta^{n_2} \times \cdots \times \Delta^{n_{i-1}} \times \partial_{n_i}(\Delta^{n_i}) \times \Delta^{n_{i+1}} \times \cdots \times \Delta^{n_k}$. This operator can be linearly extended to N-chains.

Proposition

The boundary operator ∂ applied to prodsimplicial cells, as defined, satisfies $\partial^2 = 0$.

The Nth homology group can be defined as:

$$H_N = \frac{\ker \partial_N}{\operatorname{im} \partial_{N+1}} \cong \underbrace{\mathbb{Z} \oplus \mathbb{Z} \cdots \oplus \mathbb{Z}}_{n = \operatorname{rank}(H_N)} + \underbrace{\mathbb{Z}_{{p_1}^{s_1}} \oplus \mathbb{Z}_{{p_2}^{s_2}} \cdots \oplus \mathbb{Z}_{{p_m}^{s_m}}}_{\text{torsion group}}$$

The rank of H_N is also known as the Nth Betti number, denoted β_N, and it counts the N-dimensional 'holes' in the complex.

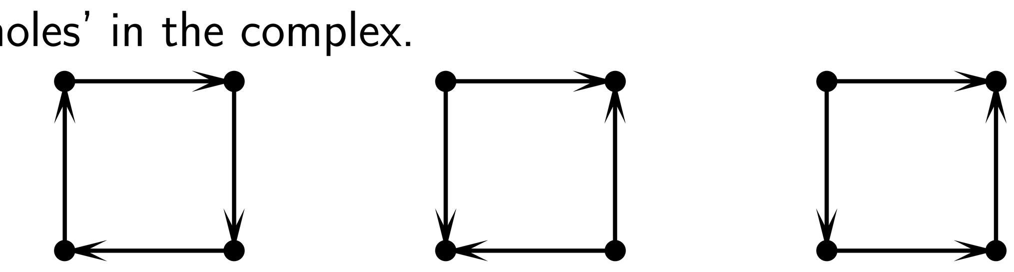


Figure 2: Holes in two dimensions correspond to cells that are either not consistently oriented or not Cartesian products.

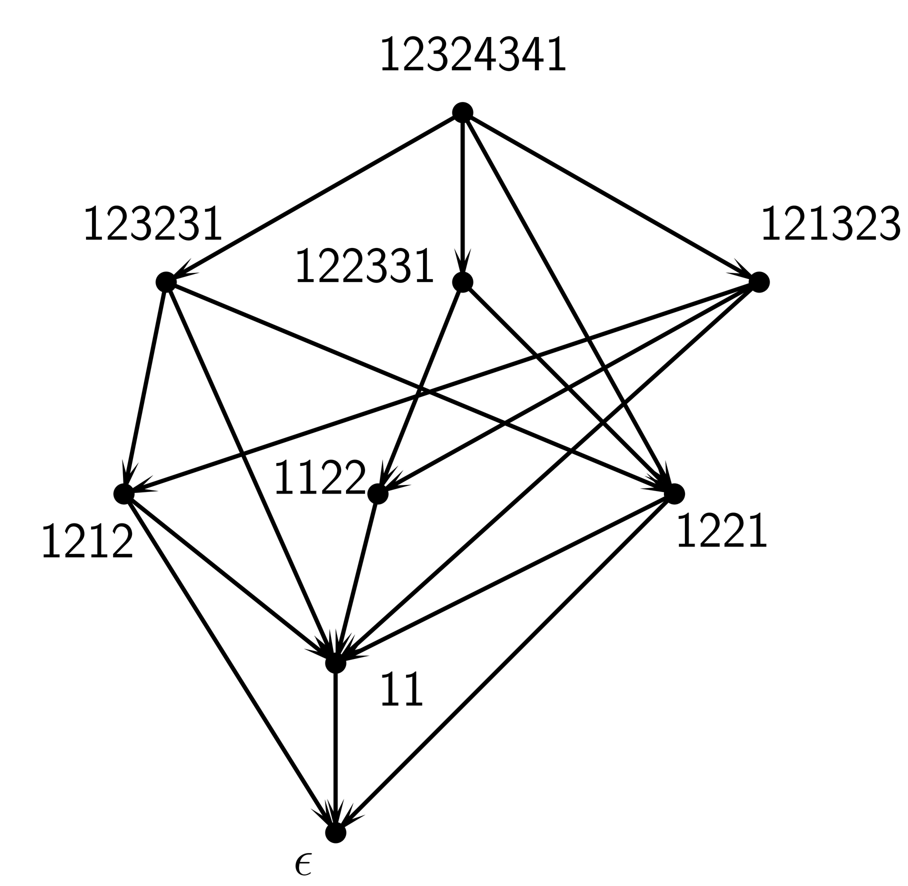


Figure 3: The cell complex induced by the reduction graph for the pointers on Contig5288.0 of *Oxytricha trifallax* has Betti numbers $\beta_0 = 1$, $\beta_1 = 0$, $\beta_2 = 3$ and $\beta_n = 0$ for $n \geq 3$.

Future Work

Homology groups can be used to describe a geometric signature of the DNA scrambling process, which may be used to compare different species with similar rearrangement mechanisms on an evolutionary timeline.

References

[1] J. Burns, D. Kukushkin, X. Chen, L. Landweber, M. Saito, N. Jonoska. *Recurring Patterns Among Scrambled Genes in the Encrypted Genome of the Ciliate Oxytricha trifallax.* Journal of Theoretical Biology 410 pp. 171-180 (2016).