Reductions on Double Occurrence Words

Ryan Arredondo
rarredon@mail.usf.edu

University of South Florida
Tampa, Florida

03/08/2013

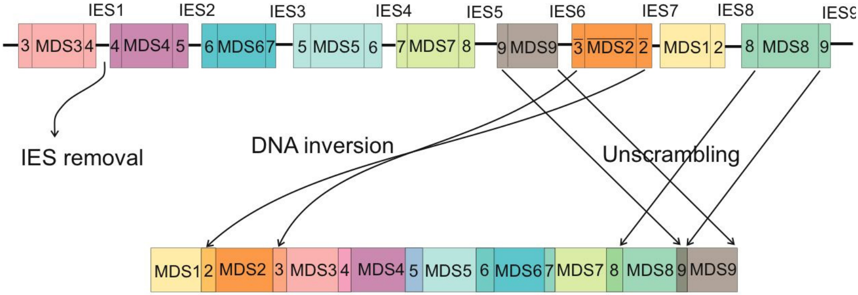
Table of Contents

1. Biological background
2. Double occurrence words
3. Reduction of a double occurence word
4. Nesting index of a double occurrence word
5. Chord diagrams
6. Circle graphs
7. Questions and Conjectures

Biological motivation

Figure: A ciliate under microscope

Genome rearrangement



- Segments of ciliate DNA are eliminated or rearranged. [1]

[1] Prescott, D.M. and Greslin, A.F., Scrambled actin I gene in the micronucleus of *Oxytricha nova*. Dev. Genet. (1992), 13: 66-74.

Model for rearrangement

- 4-valent rigid vertex graphs with endpoints (assembly graphs) are used to model genome rearrangement [2]

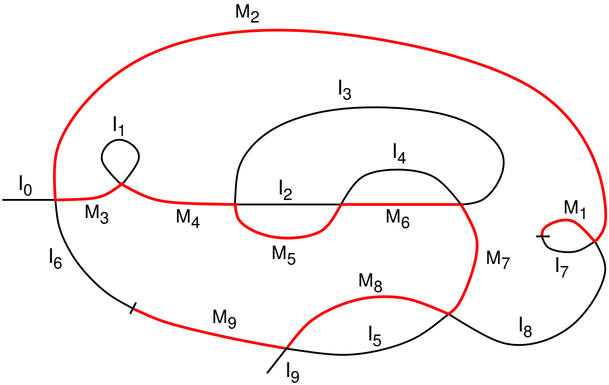


Figure: Assembly graph model for Actin I gene in *Oxytricha Nova*

[2] A. Angeleska, N. Jonoska, M. Saito, L.F. landweber, RNA-Guided DNA Assembly, J. of Theoretical Biology (2007) 248:706-720.

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. [3]

[3] W.-J. Chang, P.D. Bryson, H. Liang, M.K. Shin, L.F. Landweber. The evolutionary origin of a complex scrambled gene. Proc. Natl. Acad. Sci. (2005) 102(42):15149-15154

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. [3]
- Goal:
  1. Double occurrence words from assembly graphs

[3] W.-J. Chang, P.D. Bryson, H. Liang, M.K. Shin, L.F. Landweber. The evolutionary origin of a complex scrambled gene. Proc. Natl. Acad. Sci. (2005) 102(42):15149-15154

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. [3]
- Goal:
  1. Double occurrence words from assembly graphs
  2. Define words that relate to the patterns observed

[3] W.-J. Chang, P.D. Bryson, H. Liang, M.K. Shin, L.F. Landweber. The evolutionary origin of a complex scrambled gene. Proc. Natl. Acad. Sci. (2005) 102(42):15149-15154

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. [3]
- Goal:
  1. Double occurrence words from assembly graphs
  2. Define words that relate to the patterns observed
  3. Define reduction operations on double occurrence words

[3] W.-J. Chang, P.D. Bryson, H. Liang, M.K. Shin, L.F. Landweber. The evolutionary origin of a complex scrambled gene. Proc. Natl. Acad. Sci. (2005) 102(42):15149-15154

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. [3]
- Goal:
    1. Double occurrence words from assembly graphs
    2. Define words that relate to the patterns observed
    3. Define reduction operations on double occurrence words
    4. Define the nesting index as a measurement of complexity for scrambled genes

[3] W.-J. Chang, P.D. Bryson, H. Liang, M.K. Shin, L.F. Landweber. The evolutionary origin of a complex scrambled gene. Proc. Natl. Acad. Sci. (2005) 102(42):15149-15154

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. [3]
- Goal:
  1. Double occurrence words from assembly graphs
  2. Define words that relate to the patterns observed
  3. Define reduction operations on double occurrence words
  4. Define the nesting index as a measurement of complexity for scrambled genes
  5. Observe how the nesting index relates to other representations of double occurence words (e.g. chord diagrams and circle graphs)

[3] W.-J. Chang, P.D. Bryson, H. Liang, M.K. Shin, L.F. Landweber. The evolutionary origin of a complex scrambled gene. Proc. Natl. Acad. Sci. (2005) 102(42):15149-15154

Table of Contents

1. Biological background
2. Double occurrence words
3. Reduction of a double occurence word
4. Nesting index of a double occurrence word
5. Chord diagrams
6. Circle graphs
7. Questions and Conjectures

Double occurrence words

- A *double occurrence word* is a word over a finite alphabet $\Sigma = \{1, 2, 3, \ldots, n\}$ in which each symbol appears exactly twice

Double occurrence words

- A *double occurrence word* is a word over a finite alphabet $\Sigma = \{1, 2, 3, \ldots, n\}$ in which each symbol appears exactly twice

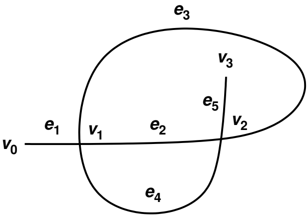
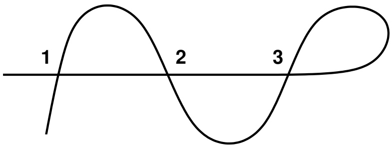


Figure: Assembly graph represented by word $v_1 v_2 v_1 v_2$ or 1212

Patterns translated

- A *return word* is a word of the form

$$a_1 a_2 \cdots a_n a_n \cdots a_2 a_1, \qquad \mathrm{a}_i \in \Sigma \text{ for all } i, \text{ and } a_i \neq a_j \text{ for } i \neq j$$



(a) 123321 is a return word

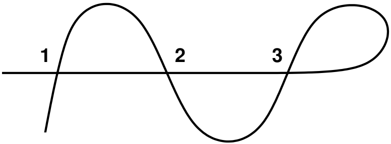
Patterns translated

- A *return word* is a word of the form

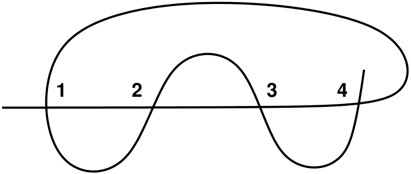
$$a_1 a_2 \cdots a_n a_n \cdots a_2 a_1, \qquad a_i \in \Sigma \text{ for all } i, \text{ and } a_i \neq a_j \text{ for } i \neq j$$

- A *repeat word* is a word of the form

$$a_1 a_2 \cdots a_n a_1 a_2 \cdots a_n, \qquad a_i \in \Sigma \text{ for all } i, \text{ and } a_i \neq a_j \text{ for } i \neq j$$



(a) 123321 is a return word



(b) 12341234 is a repeat word

Table of Contents

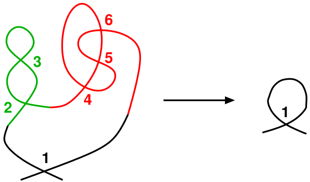
1. Biological background
2. Double occurrence words
3. Reduction of a double occurence word
4. Nesting index of a double occurrence word
5. Chord diagrams
6. Circle graphs
7. Questions and Conjectures

Reductions

- Two reduction operations

Reductions

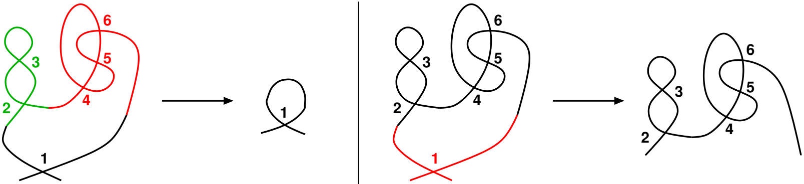
- Two reduction operations
  1. Removal of all maximal repeat and return words



(a) **123324564561** $\longrightarrow$ **11**

Reductions

- Two reduction operations
  1. Removal of all maximal repeat and return words
  2. Removal of a letter

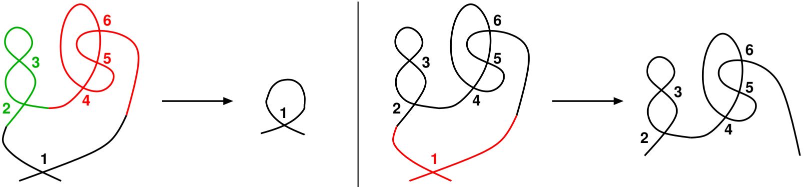


(a) **123324564561** $\longrightarrow$ **11**

(b) **123324564561** $\longrightarrow$ **2332456456**

Reductions

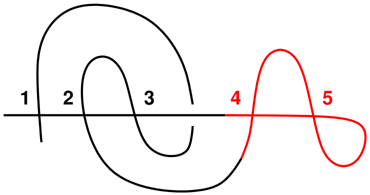
- Two reduction operations
  1. Removal of all maximal repeat and return words
  2. Removal of a letter



(a) **123324564561** $\longrightarrow$ **11** (b) **123324564561** $\longrightarrow$ **2332456456**

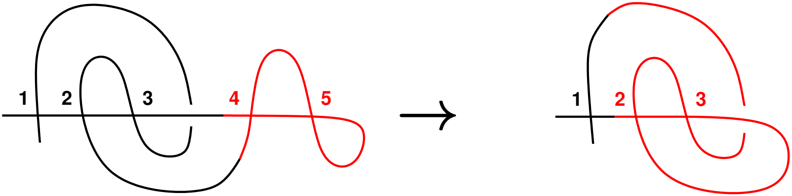
- A reduction $\mathcal{R}$ of a word w is a sequence of words $(w = w_0, w_1, w_2, \ldots, w_n = \epsilon)$ where w_{i+1} is obtained from w_i for $0 \leq i < n$ by one of the two reduction operations.

Example 1



(a) $w_0 =$ **1234554231**

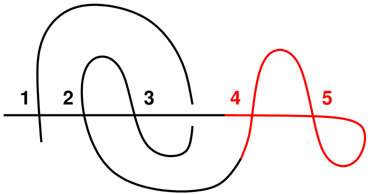
Example 1



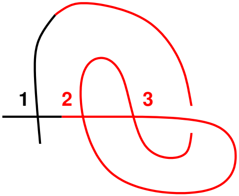
(a) $w_0 = \mathbf{1234554231}$

(b) $w_1 = \mathbf{123231}$

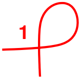
Example 1



(a) $w_0 = \mathbf{1234554231}$

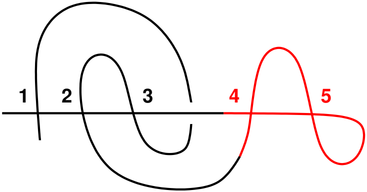


(b) $w_1 = \mathbf{123231}$

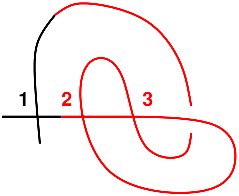


(c) $w_2 = \mathbf{11}$

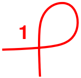
Example 1



(a) $w_0 = $ **1234554231**



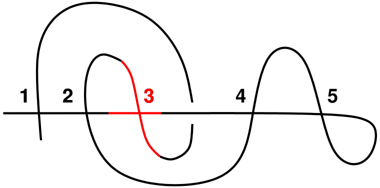
(b) $w_1 = $ **123231**



(c) $w_2 = $ **11**

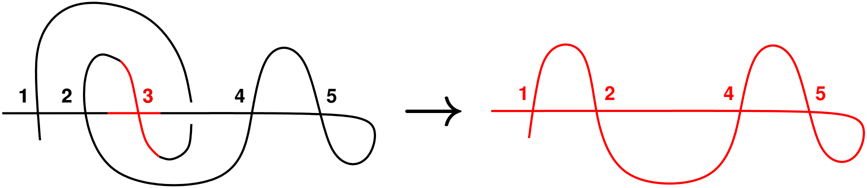
(d) $w_3 = \epsilon$

Example 2



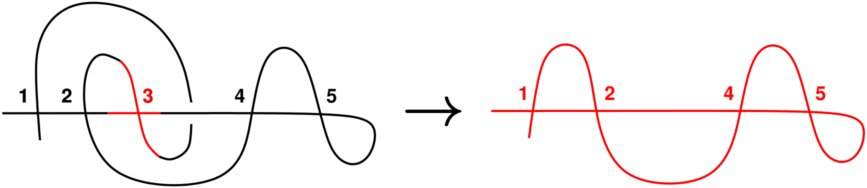
(a) $w_0 =$ **1234554231**

Example 2



(a) $w_0 = \mathbf{1234554231}$ (b) $w_1 = \mathbf{124421}$

Example 2



(a) $w_0 = $ **1234554231**

(b) $w_1 = $ **124421**

↙

(c) $w_2 = \epsilon$

Table of Contents

1 Biological background

2 Double occurrence words

3 Reduction of a double occurence word

4 Nesting index of a double occurrence word

5 Chord diagrams

6 Circle graphs

7 Questions and Conjectures

Nesting index

- *Nesting index* of a word w is the least number of reduction operations that reduce w to ϵ
- More formally,

$$\mathrm{NI}(w) = \min\{n : (w = w_0, w_1, \ldots, w_n = \epsilon) \text{ is a reduction of } w\}$$

	Example 1	Example 2
w_0	1234554231	1234554231
w_1	123231	12455421
w_2	11	ϵ
w_3	ϵ	N/A

Table: Reductions of $w = 12344321$

- No known efficient algorithm to compute Nesting index

A more complex case

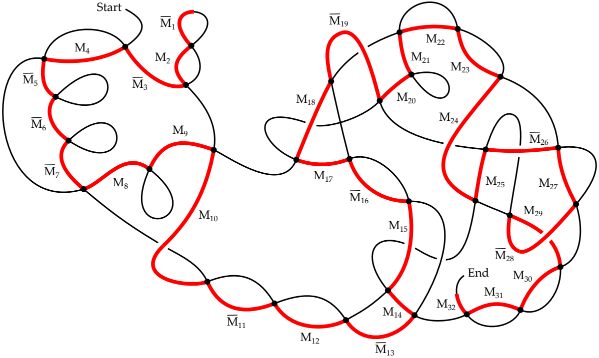


Figure: Assembly graph for word 1, 2, 3, 4, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 11, 18, 12, 19, 20, 18, 21, 7, 22, 23, 16, 24, 25, 26, 3, 27, 27, 28, 28, 2, 1, 29, 30, 30, 29, 5, 26, 25, 24, 17, 23, 22, 6, 21, 31, 31, 8, 9, 10, 20, 19, 13, 14, 15

- Nesting index = 15

Table of Contents

1 Biological background

2 Double occurrence words

3 Reduction of a double occurence word

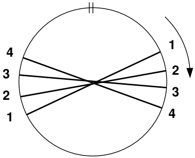
4 Nesting index of a double occurrence word

5 Chord diagrams

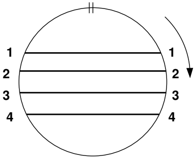
6 Circle graphs

7 Questions and Conjectures

Chord diagrams



(a) $w = 12341234$



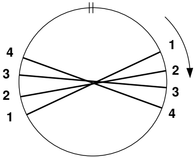
(b) $w = 12344321$

Chord diagrams

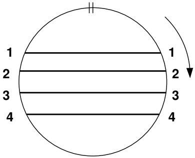
(a) $w = 12341234$ (b) $w = 12344321$

- If w is a repeat word, then every pair of chords intersect.
- If w is a return word, then no pair of chords intersect.

Chord diagrams



(a) $w = 12341234$

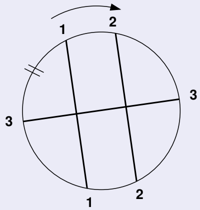


(b) $w = 12344321$

- If w is a repeat word, then every pair of chords intersect.
- If w is a return word, then no pair of chords intersect.

And now for something completely different...

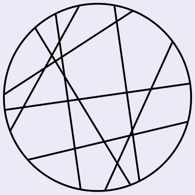
Word $w = 123213$ has chord diagram



Chord diagrams and nesting index

Theorem

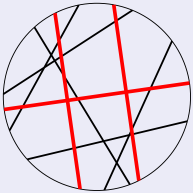
Every reduction of w must use the second reduction operation if and only if the chord diagram of w has the following sub-chord diagram.



Chord diagrams and nesting index

Theorem

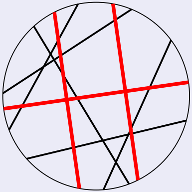
Every reduction of w must use the second reduction operation if and only if the chord diagram of w has the following sub-chord diagram.



Chord diagrams and nesting index

Theorem

Every reduction of w must use the second reduction operation if and only if the chord diagram of w has the following sub-chord diagram.



Corollary

Let $2 \leq n \leq m$ be integers. If a word w has a chord diagram with the following as a sub-chord diagram, then $\mathrm{NI}(w) \geq n+1$.

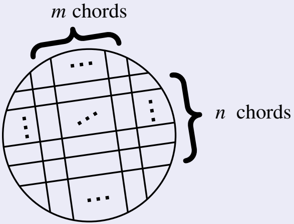


Table of Contents

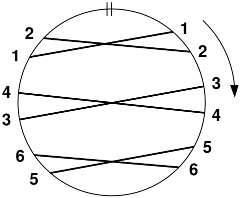
1. Biological background
2. Double occurrence words
3. Reduction of a double occurence word
4. Nesting index of a double occurrence word
5. Chord diagrams
6. Circle graphs
7. Questions and Conjectures

Circle graphs

123456563412

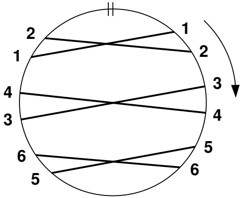
Circle graphs

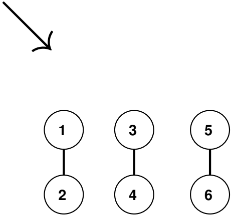
123456563412 →



Circle graphs

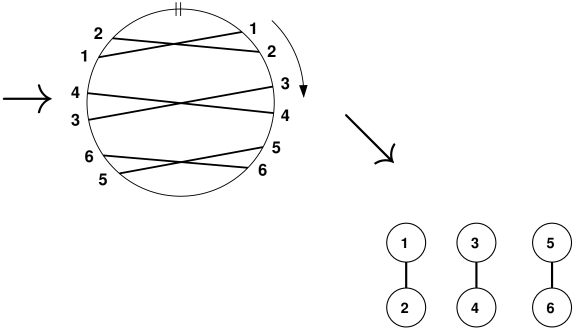
123456563412 →





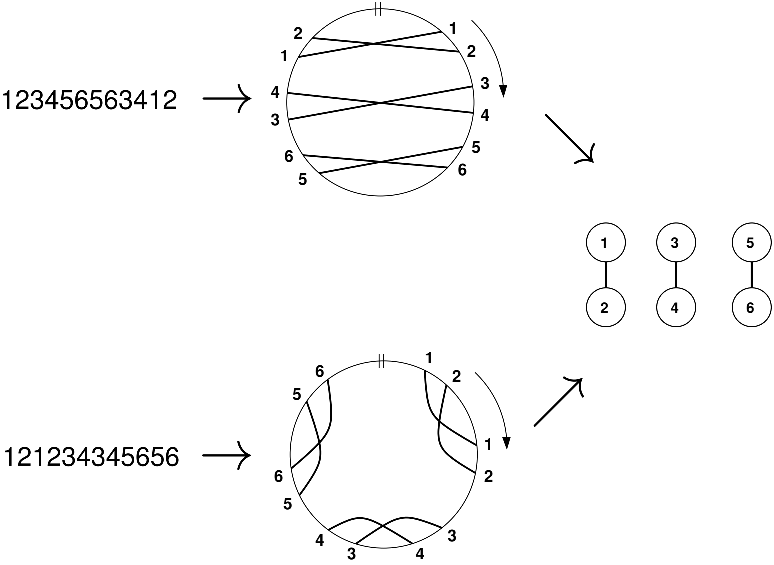
Circle graphs

123456563412

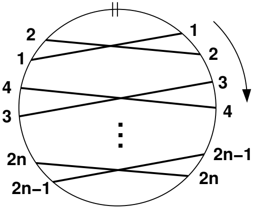


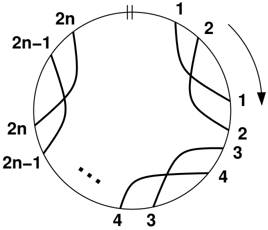
121234345656

Circle graphs



Circle graphs and nesting index





Fact

There exist words w_1 and w_2 over $2n$ letters, that have the same circle graph and

$$\mathrm{NI}(w_1) = n, \qquad \mathrm{NI}(w_2) = 1$$

Table of Contents

1 Biological background

2 Double occurrence words

3 Reduction of a double occurence word

4 Nesting index of a double occurrence word

5 Chord diagrams

6 Circle graphs

7 Questions and Conjectures

Questions and Conjectures

Questions involving cirlce graphs

- Can we characterize words with same circle graph and same nesting index?

Questions and Conjectures

Questions involving cirlce graphs

- Can we characterize words with same circle graph and same nesting index?
- Let S be a set of words with the same circle graph and nesting index. Does there exist N independent of S such that $|S| \leq N$?.

Questions and Conjectures

Questions involving cirlce graphs

- Can we characterize words with same circle graph and same nesting index?
- Let S be a set of words with the same circle graph and nesting index. Does there exist N independent of S such that $|S| \leq N$?.

Minimal realization

How many letters are needed to construct a word with nesting index n?

Questions and Conjectures

Questions involving cirlce graphs

- Can we characterize words with same circle graph and same nesting index?
- Let S be a set of words with the same circle graph and nesting index. Does there exist N independent of S such that $|S| \leq N$?.

Minimal realization

How many letters are needed to construct a word with nesting index n?
Conjecture:

$$n + \#\text{nonzero squares less than } n$$

Nesting index counts and conjectures

Table: Number of words of a given size and nesting index

Size	Nesting Index									
	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	3	0	0	0	0	0	0	0	0	0
3	7	8	0	0	0	0	0	0	0	0
4	17	78	10	0	0	0	0	0	0	0
5	41	424	479	1	0	0	0	0	0	0
6	99	1915	6248	2133	0	0	0	0	0	0
7	239	7914	50247	69879	6856	0	0	0	0	0
8	577	31370	328810	1004642	648065	13561	0	0	0	0
9	1393	122530	1927900	10125920	17081040	5187788	12854	0	0	0
10								2019	0	0
11									4	0
12										0

- First column corresponds to the numerators of continued fraction convergents to $\sqrt{2}$.

Thanks!

Many thanks to Dr. Nataša Jonoska, Dr. Masahico Saito, Jonathan Burns, Sarah Croome, Egor Dolzhenko, Maja Milošević, Jamie Sprecher, and Timothy Yeatman.

For related work please visit `http://knot.math.usf.edu`

This work has been supported in part by NSF Grant DMS #0900671