



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Matemática

**Representação combinatória e algébrica das  
permutações na análise do problema de rearranjo  
de genomas por reversões**

por

**Thaynara Arielly de Lima**

Brasília

2010

Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Matemática

**Representação combinatória e algébrica das permutações na  
análise do problema de rearranjo de genomas por reversões**

por

**Thaynara Arielly de Lima**

*Dissertação apresentada ao Departamento de Matemática da Universidade de  
Brasília, como parte dos requisitos para obtenção do grau de*

**MESTRE EM MATEMÁTICA**

Brasília, 2010.

Comissão Examinadora:

---

Prof. Mauricio Ayala-Rincón - Orientador (MAT/UnB)

---

Profa. Daniele da Silva B. Martins Neto - Membro (MAT/UnB)

---

Profa. Maria Emília M. T. Walter - Membro (CIC/UnB)

---

Dedicatória

---

*À Lenita.*

---

## Agradecimentos

---

Agradeço a Deus, pelas oportunidades concedidas, pela família que tenho e pelas pessoas que me rodeiam.

Agradeço à minha família, em especial, aos meus avós Lenita e Sebastião, à minha tia Rosana e à minha mãe Rosemeire, pela educação, pelo respeito, pelo apoio em todos os momentos e por não poupar esforços para tornar meu caminho menos penoso sempre.

Agradeço ao professor Mauricio, pela paciência, dedicação e pelos ensinamentos, não só em Lógica Formal, Teoria da Prova ou no Problema de Rearranjo de Genomas, mas por ensinar sobre caráter e sobre a postura forte que devemos manter diante de certas situações.

Agradeço às professoras Daniele e Maria Emília por terem aceito o convite de participar da banca e por darem tão valiosas sugestões que contribuíram para que o trabalho fosse consideravelmente melhorado. À professora Maria Emília agradeço também pelas sugestões de referências bibliográficas e pela disposição em promover seminários a fim de discutir o assunto estudado.

Agradeço aos amigos Sunamita, Marcelo, Tarcísio, Kélem, Dylene e Arianny pelo apoio e pelo incentivo dados desde a graduação.

Agradeço às amigas Adriana, Kaliana e Mariana pelos conselhos, pelo carinho, pelo 101 e pelos ótimos momentos que dividimos nesses dois anos de convivência.

Agradeço ao meu amigo Bruno Nunes pelas piadas (apesar do meu senso de humor ser muito mais refinado que o dele), pela confiança e pelo respeito que temos um para

com o outro.

Agradeço aos amigos Bruno César, Wembesom, Laura, Daniele, João Paulo, Ana Paula, Daniel, Eduardo, Elias, Ricardo Assunção, André e Simone pelo companheirismo e pela ajuda nas disciplinas.

Agradeço à minha amiga Andréia pela força que tem me dado nesses últimos meses, pela amizade sincera que tem dedicado e por sempre estar disposta a ajudar no que preciso.

Agradeço aos amigos Francisca, Larissa, Luz, Luciana, Vagner, Tonires, Miguel, Magno e Flávia pela agradável companhia.

Agradeço ao meu amigo e namorado Leonardo, pelos ótimos momentos juntos, pela tranquilidade que tenta me passar quando estou à beira de um ataque de nervos, pela doçura e pela paciência que tem comigo (apesar de eu realmente merecer, já que sou uma namorada exemplar e também devoto todos esses sentimentos a ele).

Agradeço aos funcionários do Departamento de Matemática da UnB, Célia, Cláudia, Irene, Eveline, Manoel e Pereira, pelo respeito e cuidado que tem pelos alunos.

Enfim, agradeço a todos que passaram, em algum momento, pela minha vida e que generosamente deixaram, comigo, um pouco de si.



*“E ainda que tivesse o dom de profecia, e conhecesse todos os mistérios e toda a ciência, e ainda que tivesse toda fé, de maneira tal que transportasse os montes, e não tivesse amor, nada seria.”*

*I Coríntios. 11: 2*

---

## Resumo

---

Na genômica comparativa soluções algorítmicas eficientes para o problema da distância de rearranjo de genomas são uma ferramenta importante para o desenvolvimento de software que permite estabelecer relacionamento evolutivo entre organismos, por exemplo para a construção de árvores filogenéticas de organismos. Existem diversas operações sobre palavras que modelam mutações ocorridas nos genes dos seres vivos (e.g. reversões, transposições, troca de blocos, etc.). Restritos à operação de reversão o problema de rearranjo de genomas é  $NP$ -difícil. Sendo assim, é plausível considerar algoritmos de aproximação. O algoritmo conhecido que melhor aproxima a solução do problema de rearranjo via reversões tem raio de 1.375. No seu artigo seminal, Bafna e Pevzner apresentam soluções  $O(n^2)$  de raios de aproximação 1.5 para permutações com sinal e  $\frac{7}{4}$  para permutações sem sinal.

Neste trabalho propõe-se um uso cuidadoso e discriminado entre as representações combinatória (palavras e grafos de pontos de quebra) e algébrica (ciclos de permutações) das permutações, que contribuirão para analisar com precisão e de maneira adequada diversas características do problema da distância de reversão e das soluções apresentadas por Bafna e Pevzner.

**Palavras-chave:** Rearranjo de genomas, reversões, permutações, grafo de pontos de quebra.

---

## Abstract

---

Efficient algorithmic solutions to the problem of genome rearrangements in comparative genomics are an important tool for the development of software allowing one to establish the evolution link between organisms, for instance the construction of phylogenetic trees. There are several string operations modelling mutations occurring inside genes (e.g. reversals, transpositions and block interchange, etc.). When restricted to the reversal operation, the problem of genome rearrangement is NP-hard. Thus, polynomially bounded approximated algorithms are considered to be admissible solutions. The best known approximated algorithm to solve the rearrangement problem through reversals has approximation ratio of 1.375. In their seminal paper, Bafna and Pevzner presented  $O(n^2)$  solutions of ratio 1.5 for signed permutations and  $\frac{7}{4}$  for unsigned permutations.

This work proposes a careful discrimination between the combinatorial (strings and breakpoint graphs) and the algebraic (cycles of permutations) representations of permutations to analyse precisely, and in an adequate way, many of the problem characteristics and solutions presented by Bafna and Pevzner.

**Keywords:** Genome rearrangement, reversals, permutations, breakpoint graphs.



---

## Índice

---

<b>Introdução</b>	<b>1</b>
<b>1 Fundamentos</b>	<b>6</b>
1.1 Problema biológico molecular . . . . .	6
1.2 Representação de permutações como palavras e ciclos . . . . .	12
<b>2 Uso adequado das notações para permutação na solução de Bafna e Pevzner</b>	<b>16</b>
2.1 Grafo de pontos de quebra e limitante inferior para a distância de reversão	16
2.2 Permutação de Gollan e diâmetro de reversão . . . . .	23
2.3 Limitante para o valor esperado da distância de reversão . . . . .	34
2.4 Aproximação para a distância de reversão . . . . .	39
<b>3 Algoritmos para o problema de rearranjo via reversões</b>	<b>45</b>
3.1 Permutações com sinal . . . . .	45
3.2 Um algoritmo de aproximação para ordenação por reversões . . . . .	49
3.3 Uma melhor aproximação para ordenação por reversões . . . . .	55
<b>4 Conclusões e Trabalhos Futuros</b>	<b>61</b>
<b>Referências Bibliográficas</b>	<b>64</b>

---

## Introdução

---

Evolução biológica diz respeito às modificações ocorridas nos organismos através dos tempos. Essas mudanças são evolutivas se passadas de geração a geração, via material genético.

Teorias evolucionistas surgiram combatendo as idéias criacionistas e fixistas, que propunham que os seres eram criações divinas e que a cadeia formada pelos seres vivos era estática desde a criação.

Um dos pioneiros na tentativa de explicar cientificamente o processo pelo qual ocorre a evolução foi Jean-Baptiste Lamarck[11]. Lamarck propôs que os organismos vão desenvolvendo características de acordo com suas necessidades e que determinados órgãos podem se atrofiar caso não sejam necessários. Esta ficou conhecida como primeira lei de Lamarck ou lei do uso e desuso. A segunda lei de Lamarck sugere que características adquiridas pelo uso são transmitidas de geração a geração. Esta é conhecida como lei da herança dos caracteres adquiridos. O biólogo alemão Weissman[11] refutou as leis de Lamarck, realizando um experimento que consistia em cortar caudas de ratos durante várias gerações. Weissman notou que os filhotes continuavam a nascer com caudas; assim, verificou que a ausência de caudas, adquirida por esses ratos, não foi uma característica transmitida a outras gerações.

O cientista britânico Charles Darwin ficou conhecido por propor uma teoria de como se dá a evolução por meio de seleção natural[11]. O conceito básico de seleção natural é que características favoráveis que são hereditárias tornam-se mais comuns em gerações futuras de uma população de organismos, e que características desfavoráveis

que são hereditárias tornam-se menos comuns. Em seu livro, "A origem das espécies", Darwin preconizou a idéia de evolução a partir de um ancestral comum, por meio de seleção natural.

A fusão da teoria de evolução das espécies por meio de seleção natural de Charles Darwin, da genética como base para a herança biológica de Gregor Mendel e a genética populacional deu origem ao Neo-Darwinismo ou síntese moderna. Essencialmente, a síntese moderna introduziu a conexão entre duas importantes descobertas: as unidades de evolução (genes) com o mecanismo de evolução (seleção natural). Pela síntese moderna, a variação genética em populações surge por meio de mutações e recombinação genética. A evolução consiste das modificações na frequência dos alelos entre uma e outra geração, por meio de deriva genética<sup>1</sup>, fluxo gênico<sup>2</sup> e seleção natural. Assim, a formação de espécies ocorre de maneira gradual quando populações são isoladas reprodutivamente, por exemplo, por barreiras geográficas.

Projetos de sequenciamento de genomas geraram uma grande quantidade de informação sobre o DNA de diversos organismos. Isto auxilia em pesquisas no sentido de mapear e comparar genomas. A Matemática e a Computação aparecem, então, como ferramentas para o desenvolvimento e automatização de processos que auxiliem no armazenamento e análise dos dados obtidos.

Recentemente, softwares são desenvolvidos com o intuito de construir árvores filogenéticas, a partir de métodos de rearranjos de genomas, entre organismos. GRAPPA, MGR, e MGRA são exemplos desses softwares [1]. O GRAPPA, criado em 2001 por Moret et al. foi desenvolvido para analisar rearranjos a partir da análise de distância de pontos de quebra, inicialmente feita por Sankoff e Blanchette. Em 2002, Bourque e Pevzner usaram a distância genômica para desenvolver o MGR. O MGRA foi um algoritmo desenvolvido para a reconstrução de árvores filogenéticas e auxilia no estudo da história de rearranjo de genomas de sete mamíferos: humanos, chimpanzés, macacos, camundongos, ratos, cachorros e gambás. A árvore filogenética abaixo, adaptada de [1], foi obtida a partir desse algoritmo.

Os métodos utilizados para se efetuar a comparação entre genomas levam em consideração eventos que são possíveis na natureza e que constituem fatores relevantes no processo evolutivo. Essa comparação leva em conta ainda a lei da Parcimônia, que supõe que o processo evolutivo é econômico, o que significaria dizer que o caminho aparentemente mais simples que pode ter sido tomado durante a evolução de qualquer

---

<sup>1</sup>Deriva genética corresponde à variação do fundo genético das populações como consequência do acaso. Um pequeno número de indivíduos, chamados fundadores, ao deslocarem-se para outro habitat podem transportar apenas uma parte do fundo genético da população inicial. Assim, certos genes podem ser fixados ou eliminados de uma população por simples acaso, e não por seleção natural.

<sup>2</sup>Fluxo gênico é a transferência de genes de uma geração para outra.

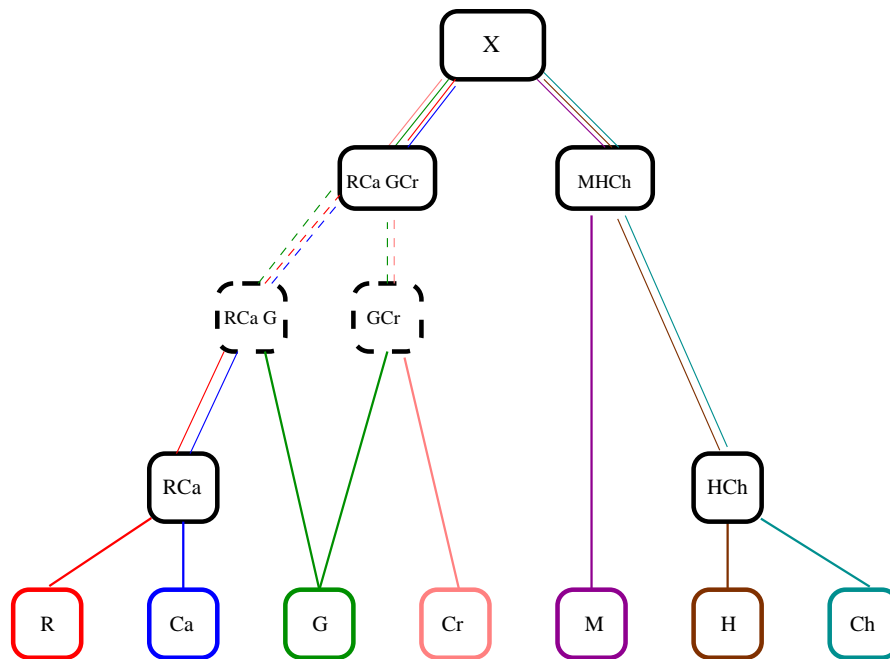


Figura 1: Árvore filogenética, construída a partir do software MGRA, dos mamíferos: rato(R), camundongo(Ca), gambá(G), cachorro(Cr), macaco(M), humano(H), chimpanzé(Ch). Note que a ramificação correspondente ao gambá é dúbia; as ramificações representadas por linhas pontilhadas apresentam possíveis variações do ancestral(Figura extraída de [1], pág.948)

grupo corresponde ao processo real.

Sendo assim, é plausível determinar o menor número de eventos mutacionais necessários para se transformar um genoma em outro.

E é isso que motiva o problema de rearranjo: dados dois genomas, ou parte de genomas, vistos como uma lista ordenada de genes e um conjunto de possíveis operações, o problema de rearranjo consiste em encontrar o número mínimo dessas operações que transformam um genoma em outro.

A fim de se obter um modelo matemático para o problema, atribuímos aos genes, que ocorrem nos cromossomos que serão comparados, números inteiros. Quando a orientação desses genes<sup>3</sup> é conhecida os números recebem sinais; o sinal (+) indica que esses fragmentos possuem a mesma orientação, enquanto o sinal (-) indica que os genes possuem orientações contrárias. Assim, o problema de rearranjo é equivalente

<sup>3</sup>Orientação de um gene diz respeito à fita da dupla hélice onde o gene está localizado. Sendo assim, genes ocorrendo na mesma fita têm mesma orientação (+) e genes ocorrendo em fitas diferentes têm orientações contrárias.

ao seguinte: dadas duas sequências numéricas e uma operação, determine o número mínimo de eventos necessários para se transformar uma sequência em outra.

Existem várias operações que são admitidas nesse processo. Dentre elas estão, reversões, transposições, transreversões e troca de blocos que atuam no mesmo cromossomo.

Informações relevantes sobre a complexidade do problema de rearranjo, bem como algoritmos que fornecem raios de aproximação para a sua resolução já estão presentes na literatura [2, 10, 5, 8, 17, 16].

Reversões atuam nas sequências revertendo parte de seus fragmentos. Quando essas sequências possuem sinal, o sinal dos elementos do bloco invertido também é trocado. Caprara [7] mostrou que o problema de reversão sem sinais é NP-difícil. Kececioglu e Sankoff [15] desenvolveram um algoritmo que possui um raio de aproximação de 2. Bafna e Pevzner [4] melhoraram essa aproximação para  $\frac{7}{4}$ . Um algoritmo proposto por Berman et al. [5] reduziu esse fator para 1.375. Para sequências com sinal Bader [2] propôs um algoritmo exato que resolve o problema de reversões com sinal em tempo linear.

Outra importante operação é a transposição. Ela atua cortando parte da sequência de genes e colando-a em outra parte do cromossomo. A complexidade de ordenação por transposições ainda é um problema em aberto. O algoritmo de melhor desempenho para esse problema foi dado por Elias e Hartman [10] que fornece uma aproximação de 1.375 em tempo subquadrático.

Para a troca de blocos, que consiste em destacar dois fragmentos de um mesmo cromossomo e trocar suas posições, Christie [8], além de introduzir essa operação, mostrou que o problema pode ser resolvido em tempo subquadrático.

Pode-se ainda mesclar essas operações. Para transposições com reversões, por exemplo, que destaca dois fragmentos de um mesmo cromossomo, troca-os de posição e inverte sua ordem, Mira e Meidanis [17] e Lin et al. [16], encontraram um algoritmo exato que resolve o problema em tempo subquadrático.

A tabela abaixo, adaptada de [18] resume as complexidades e algoritmos conhecidos para solucionar o problema de rearranjo.

Neste trabalho, estaremos interessados na solução do problema de rearranjo via reversões.

Os resultados principais, que estão presentes nesse texto, encontram-se em [4]. Em [4] propõe-se abordagens aproximadas para resolver o problema de rearranjo restrito a reversões de raio no mínimo  $\frac{7}{4}$  para reversões sem sinal.

Como já dissemos, existem resultados que são mais satisfatórios quanto ao raio de aproximação dos algoritmos propostos. Porém, nossa maior preocupação foi em caracterizar reversões e usar essa caracterização de maneira cuidadosa ao longo do texto, do ponto de vista algébrico.

Tabela 1: Operações e complexidades para o problema de rearranjo de genomas

Operação de rearranjo	Complexidade	Tipo do algoritmo
Reversão sem sinais	NP-difícil	Aproximação de 1.375
Reversão com sinais	$O(n)$	Exato
Transposição	Não conhecido	Aproximação de 1.375
Reversão com sinais + Transposição	Não conhecido	Aproximação de 2
Troca de blocos	$O(n^2)$	Exato
Troca de blocos + Reversão	$O(n^2)$	Exato

Neste trabalho são apresentados, em detalhe, os métodos propostos em [4], mantendo uma disciplina de separação entre as representações das permutações como ciclos, que são as usuais algebricamente, e palavras, que se mostram adequadas do ponto de vista biológico, mas não são sempre a melhor escolha no desenvolvimento das provas.

No primeiro capítulo, fazemos um esboço do problema biológico molecular da distância de reversão, contextualizando-o, e propomos uma discriminação das representações de permutações como ciclos e palavras. Essa diferenciação proposta, importante para tornar mais claras e precisas as demonstrações dos resultados, é o ponto de maior relevância no trabalho. No artigo, *Genome Rearrangement and Sorting by Reversals* [3] não fica clara a diferença entre permutações como ciclos e palavras, que são os objetos utilizados para se caracterizar os genes matematicamente.

No segundo capítulo, define-se grafo de pontos de quebra proposto por Bafna e Pevzner [3], uma ferramenta importante que será utilizada em grande parte das provas dos resultados. Discute-se ainda sobre permutações de Gollan, que são as permutações que requerem o maior número de reversões para serem ordenadas, e apresenta-se resultados sobre o valor esperado da distância de reversão.

No terceiro capítulo, são apresentados algoritmos que fornecem um raio aproximado para a resolução do problema de distância por reversões. Inicialmente, considera-se permutações com sinais e propõe-se um algoritmo com raio de aproximação de  $\frac{3}{2}$  para ordenar essas permutações. Em seguida, usa-se esse algoritmo para obter outros que ordenam permutações sem sinais com raio de no mínimo  $\frac{7}{4}$ .

O último capítulo é destinado a listar algumas situações em que se detecta inconsistência no uso indiscriminado das duas representação (ciclos e palavras) em [3]. Destaca-se também pontos em que existem ganhos com o uso da representação de permutações como ciclos ou palavras, de acordo com o que for mais conveniente. Por fim, sugerimos trabalhos futuros.

# CAPÍTULO 1

---

## Fundamentos

---

### 1.1 Problema biológico molecular

Em todo ser vivo dois tipos de moléculas presentes nas células são essenciais. Estas são proteínas e ácidos nucleicos [14].

Proteínas são compostos orgânicos de alto peso molecular, formadas pelo encadeamento de aminoácidos. Representam cerca de 50 a 80% do peso seco da célula sendo, portanto, o composto orgânico mais abundante de matéria viva.

As proteínas podem ser agrupadas em várias categorias de acordo com a sua função. De uma maneira geral, elas desempenham nos seres vivos funções estrutural, enzimática, hormonal, de defesa, nutritivo, coagulação sanguínea e transporte.

Como já dissemos, ao lado das proteínas, os ácidos nucleicos formam um importante tipo de molécula no organismo. Estes são responsáveis por codificar e armazenar a informação hereditária. Logo, permitem a transmissão dessas informações de uma geração a outra.

Assim como as proteínas, os ácidos nucleicos consistem de uma cadeia formada por entidades menores: os nucleotídeos. Um nucleotídeo é dado por um açúcar, um fosfato e uma base.

As bases distinguem os aspectos dos nucleotídeos. Conseqüentemente, podemos escrever um ácido nucleico como uma seqüência de bases ao longo de uma cadeia de polinucleotídeos.

Basicamente, existem dois tipos de ácidos nucleicos. Estes são os ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA). Os prefixos desoxirribo e ribo referem-se aos açúcares desoxirribose e ribose presentes no DNA e no RNA, respectivamente.

No DNA estão contidas as bases adenina (A), citosina (C), guanina (G) e timina (T). No RNA a timina é substituída pela base uracila (U).

Uma molécula de DNA consiste basicamente de uma fita dupla de cadeias de nucleotídeos. Essas duas longas fitas se entrelaçam formando uma dupla hélice. Bases ao longo dessas duas faixas ligar-se-ão aos pares a um átomo de hidrogênio. Relacionam-se aos pares, adenina e timina, citosina e guanina. Esses pares de bases são ditos complementares. Sendo esses pares complementares, a sequência de bases de uma fita pode ser deduzida diretamente da sequência de bases da outra fita.

Em organismos vivos, existem dois tipos de células: células procariontes e eucariontes. Nas células eucariontes o DNA está encapsulado no núcleo, enquanto que nas procariontes este encontra-se disperso no citoplasma. Organismos multicelulares, como animais e plantas, são eucariontes; seres unicelulares, como por exemplo bactérias, são, em geral, procariontes.

A sequência de aminoácidos de proteínas é codificada a partir de informações codificadas no DNA. Um segmento de DNA que leva à produção de uma cadeia polipeptídica e inclui regiões que antecedem e que seguem a região codificadora é dito um gene. Uma molécula de DNA contendo vários genes é chamada cromossomo. A coleção de informação genética em uma célula é dita seu genoma.

Em células eucariontes o DNA não é capaz de deixar o núcleo. Assim, a síntese de proteínas no citoplasma ocorre com a ajuda de certas moléculas chamadas ribossomos formadas a partir de RNA.

O processo de transformar informação genética codificada em uma sequência de nucleotídeos em uma sequência de aminoácidos é feito em, basicamente, duas etapas: transcrição e tradução.

A transcrição ocorre no interior do núcleo das células e consiste na síntese de uma molécula de mRNA (RNA Mensageiro) a partir da leitura da informação contida numa molécula de DNA. Este processo inicia-se pela ligação de um complexo enzimático à molécula de DNA, o RNA-polimerase. Esta enzima desfaz a dupla hélice, destruindo as ligações de hidrogênio que ligam as bases complementares das duas cadeias, afastando-as. O RNA-polimerase, inicia a síntese de uma molécula de mRNA de acordo com a complementaridade das bases nitrogenadas. Por exemplo, se na cadeia do DNA o nucleotídeo for a adenina(A), o RNA - polimerase liga o mRNA ao nucleotídeo uracila(U). Quando a leitura termina, a molécula mRNA separa-se da cadeia do DNA, e esta restabelece as ligações de hidrogênio reconstituindo a dupla hélice. Mas nem todas as sequências da molécula do DNA codificam aminoácidos. O RNA sintetizado sofre uma maturação antes de abandonar o núcleo. Algumas porções do RNA transcrito, vão



ser removidas (íntrons), e as porções não removidas (éxons), ligam-se entre si, formando assim um mRNA maturado. O RNA que sofre este processo de exclusão de porções, é designado de RNA pré-mensageiro. No final do processo, o mRNA é constituído apenas pelas sequências que codificam os aminoácidos de uma proteína, podendo assim migrar para o citoplasma e passar a informação do DNA para os ribossomos.

Nos ribossomos, a informação armazenada no mRNA é transformada em uma sequência de aminoácidos. Este processo é chamado de tradução. Neste ponto, uma tripla de bases nitrogenadas ao longo do mRNA, chamada códon, codifica um determinado aminoácido. Os diferentes códons determinam o tipo, o número e a posição dos aminoácidos na cadeia polipeptídica. Outro tipo de RNA, o RNA transportador (tRNA), desloca-se para o citoplasma, onde se liga a aminoácidos, levando-os aos pontos de síntese protéica. Numa determinada região, a molécula de tRNA apresenta uma tripla especial de nucleotídeos, o anticódon, que é complementar a um códon do mRNA. Observa-se que existem  $4^3 = 64$  diferentes códons, mas apenas 20 diferentes aminoácidos são usados na síntese de proteínas. Então, alguns aminoácidos são codificados por vários códons. Dentre os códons possíveis, três não especificam aminoácidos, e referem-se a sinais de terminação da síntese de uma cadeia de aminoácidos. Esses códons são chamados de códons de parada (*stop codons*). Um tRNA com seu anticódon correspondente liga-se ao mRNA, e libera seu aminoácido. À medida que estes se juntam, uma cadeia de aminoácidos é construída. A síntese da proteína é encerrada quando os ribossomos encontram um códon de parada no mRNA.

Diferentes argumentos podem ser usados para justificar a evolução das espécies. Dentre eles citamos anatomia comparada, argumentos embriológicos, bioquímicos, análise de proteínas e comparação entre moléculas de DNA.

A anatomia comparada consiste em analisar diferentes espécies e observar se estas apresentam estruturas semelhantes ou membros que possuem a mesma função. Características semelhantes apoiam idéias evolucionistas e indicam uma possível origem comum entre espécies.

O cientista alemão Ernest Haeckel, ao comparar os embriões das diferentes espécies, pôde observar que, num estado muito prematuro, os embriões são extremamente semelhantes. À medida que estes vão se desenvolvendo, vão perdendo a sua semelhança com os embriões de outras espécies, sendo mais diferentes quanto menor for o seu grau de parentesco. Estas observações levaram Haeckel [11] a concluir que o embrião de determinada espécie passa por todos os estados adultos dos seus antepassados. Se considerarmos que a ontogenia é a história do desenvolvimento embrionário até ao estado adulto de um indivíduo, e a filogenia é a história evolutiva de determinada espécie, poderíamos afirmar que Haeckel considerou que a ontogenia recapitula a filogenia. Esta é a chamada lei da recapitulação. Mais tarde, esta lei foi reescrita por von Baer [11], o qual lhe acrescentou que o embrião passava por todos os estados em-

brionários dos seus antepassados, e não pelos estados adultos, como Haeckel afirmou. Esta lei passou a chamar-se lei biogenética. Deste modo, argumentos embriológicos nos permitem estabelecer um certo grau de parentesco entre duas espécies. Quanto mais afastados filogeneticamente estiverem dois indivíduos mais curtas são as fases ontogênicas comuns. Assim, os seres mais aparentados apresentam fases ontogênicas semelhantes mais longas.

Ao analisar os componentes químicos das várias espécies podemos notar que quanto mais semelhante for a sua constituição química, maior o grau de relacionamento evolutivo que pode ser estabelecido. Pode-se considerar como argumento bioquímico: o fato de todos os seres vivos serem constituídos principalmente pelas biomoléculas: lipídeos, glicídeos, ácidos nucleicos, dentre outras; a existência de DNA e RNA e sua intervenção na síntese proteica; a universalidade do código genético, o qual é constituído, para quase todos os seres, por cinco bases, cujas combinações podem dar origem a diferentes proteínas; o fato de existirem 20 aminoácidos nos seres vivos.

Como já vimos, as diferentes proteínas são agrupamentos de aminoácidos arranjados com uma determinada ordem, a qual está definida no DNA. Assim, uma proteína que tenha determinada ordem de aminoácidos será diferente de uma outra que possua uma outra ordem. Frequentemente encontram-se em diferentes espécies proteínas de mesma estrutura, mas com alguns aminoácidos diferentes. Isto ocorre devido a mutações que o DNA sofreu, as quais são lidas pelo RNA, sendo este quem vai transmitir a informação para que a proteína seja sintetizada. Deste modo, e uma vez que cada três bases do DNA vão dar origem a um aminoácido, uma alteração de uma dessas bases, ou de mais, poderá dar origem a um novo aminoácido. Isto leva a crer que as proteínas tiveram origem num ancestral, e que se foram alterando, dependendo da evolução sofrida pelos indivíduos. A partir disto pode-se afirmar que quanto maior for a semelhança entre as proteínas de indivíduos de diferentes espécies, maior é também o grau de parentesco.

Ao longo dos tempos, o DNA vai sofrendo alterações a nível de suas bases. Já vimos que a dupla cadeia de DNA está ligada através de pontes de hidrogênio, que se estabelecem entre as bases de nucleotídeos. Estas ligações apenas acontecem entre bases complementares. Deste modo, se ocorre uma mutação numa das cadeias, a ponte de hidrogênio entre o nucleotídeo mutado e o da outra cadeia não vai se estabelecer. É através da análise da quantidade de DNA que não se emparelha que se pode calcular a diferença existente entre dois DNAs.

Um método biológico molecular que determina o grau de semelhança genética entre combinações de sequências de DNA é a hibridização DNA-DNA. Esta técnica geralmente é usada para determinar a distância genética entre duas espécies.

Como medidas para comparar sequências de DNA usamos o número de mutações locais necessárias para se transformar uma sequência em outra. Dentre as mutações locais estão deleção, inserção e substituição de um nucleotídeo por outro.

Na natureza também ocorrem outros tipos de mutações em porções maiores do DNA, tais como cortar parte de um cromossomo e colocá-la na sequência em uma outra posição ou reverter parte da sequência do DNA. Podemos assumir que mutações desse tipo apenas mudam a ordem dos genes no DNA sem alterá-los ou destruí-los. A esse tipo de mutações damos o nome de rearranjo de genomas.

Podemos usar o número de tais rearranjos de genomas como medida da distância entre dois genomas.

Para determinadas espécies, a sequência de DNA codificada por genes simples geralmente são quase idênticas. Logo, uma comparação das sequências de genes não irá ajudar muito a detectar as diferenças no DNA desses organismo. Por outro lado, a ordem dos genes geralmente varia. E isso constitui um fator relevante para examinarmos a diferença entre dois organismos [6].

Diferentes tipos de transformações podem alterar a ordem dos genes. Dentre elas podemos citar reversões, transposições, troca de blocos, translocações, fusão e fissão.

Dada uma sequência de genes, uma reversão inverte parte dessa sequência. Uma transposição corta parte dessa sequência de genes e cola em outra parte do cromossomo. Uma troca de blocos destaca dois fragmentos de um mesmo cromossomo e os troca. Na fissão ocorre um corte entre dois genes de um cromossomo formando, assim, dois novos cromossomos. Essas operações são aplicadas em um mesmo cromossomo. A esse tipo de transformação damos o nome de transformações intracromossomais.

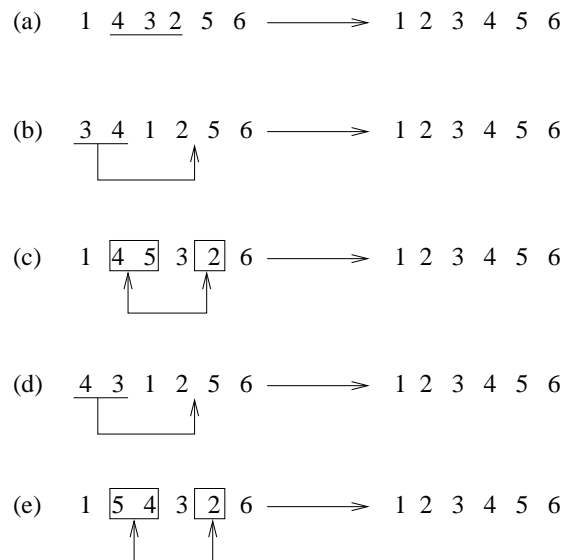


Figura 1.1: A figura mostra como agem as operações de reversão(a), transposição(b), troca de blocos(c), transposição + reversão(d) e troca de blocos + reversão(e)

A translocação permuta blocos de dois cromossomos em suas extremidades. A fusão

um e dois diferentes cromossomos transformando-os em um.

Translocação e fusão já são ditas operações intercromossomais. Isto porque elas atuam em cromossomos distintos.

Duas ou mais operações podem ocorrer conjuntamente.

De uma maneira geral, o problema de rearranjo de genomas consiste em encontrar um número mínimo de operações (reversão, transposição, troca de blocos, etc) para se transformar um genoma em outro. Neste trabalho, o foco é a operação de reversão.

Em um caso especial, por exemplo, em genomas mitocondriais, reversões ocorrem essencialmente como única forma de rearranjo de genomas. Neste caso, para determinarmos a distância entre dois genomas tentamos encontrar o mínimo de reversões que transformam um em outro.

Primeiramente, tratamos o caso em que cada genoma consiste exatamente de um cromossomo. Vamos assumir adicionalmente que não existem múltiplas ocorrências do mesmo gene dentro de um genoma. Podemos então descrever a ordem dos genes como uma sequência numérica onde cada elemento da sequência identifica-se com um gene particular. Assim, ao compararmos dois organismos, o mesmo elemento da sequência é atribuído a genes homólogos nesses seres.

Uma reversão de um fragmento de genes desses organismos corresponde à reversão de uma parte da sequência numérica associada. Logo, podemos formalizar o problema da distância de reversão como o problema de encontrar o número mínimo de reversões que transformam uma sequência na outra.

Caso conheçamos a orientação dos genes podemos modelar o problema levando em conta essa orientação. Relacionamos um elemento negativo na sequência numérica se o gene é lido da direita para a esquerda e um elemento positivo caso contrário. Veja por exemplo, na Figura 1.2 uma distância de reversão computada sobre permutações com sinais.

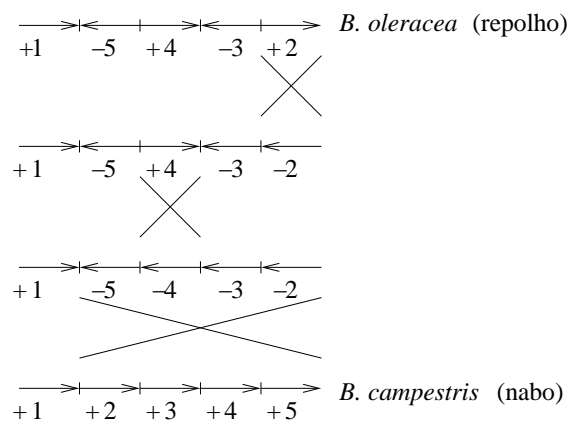


Figura 1.2: Distância de reversão com sinais (Exemplo tomado de [6], pág. 240)

## 1.2 Representação de permutações como palavras e ciclos

Sejam  $X$  um conjunto não vazio e  $S_X = \{ \alpha : X \rightarrow X \text{ bijetiva} \}$ .  $S_X$  munido da operação composição de funções é um grupo. Esse é dito o grupo das permutações do conjunto  $X$ . Se  $X = \{1, 2, \dots, n\}$  o grupo é denominado simétrico e denotado por  $S_n$ .

Uma permutação  $\alpha \in S_n$  é chamada de  $r$ -ciclo se existem elementos distintos  $a_1, a_2, \dots, a_r \in \{1, 2, \dots, n\}$  tais que  $\alpha(a_1) = a_2, \alpha(a_2) = a_3, \dots, \alpha(a_{r-1}) = a_r, \alpha(a_r) = a_1$  e  $\alpha(j) = j, \forall j \in \{1, 2, \dots, n\} \setminus \{a_1, \dots, a_r\}$ ; tal  $r$ -ciclo será denotado por  $(a_1 a_2 \dots a_r)$ . O número  $r$  é chamado o comprimento do ciclo.

Um dos maiores cuidados que tivemos no trabalho foi preservar a diferença entre a noção de palavra e permutação, que são os objetos que representam os genes e mutações de um organismo na abordagem algébrica.

No artigo *Genome Rearrangements and Sorting by Reversals*, escrito por Bafna e Pevzner, essa diferença não fica bem clara. Isso gerava grande dificuldade na leitura e compreensão do que é apresentado no texto.

A fim de facilitar essa leitura e dar um maior rigor matemático à formalização do problema, propomos a unificação da notação dos objetos estudados.

Consideremos a ordem dos genes em um organismo representada por uma palavra  $\pi = \pi_1 \pi_2 \dots \pi_n$ ,  $\pi_i \in \{1, 2, \dots, n\}$ ,  $1 \leq i, j \leq n$  e  $\pi_i \neq \pi_j, \forall i \neq j$ .

**Definição 1.1 (Permutações associadas a palavras).** Consideremos a palavra  $\pi = \pi_1 \pi_2 \dots \pi_n$ . Dizemos que  $\pi(j) := \pi_j$ . A permutação em  $S_n$  associada à palavra  $\pi$  mapeia  $\pi_j$  em  $j$ , para todo  $1 \leq j \leq n$ .

**Exemplo 1.1.** À palavra 2 3 4 5 1 6 associamos a permutação  $(2 \ 1 \ 5 \ 4 \ 3) \in S_6$

**Definição 1.2 (Palavras associadas a permutações).** Seja  $\sigma \in S_n$ .

- (a) associa-se a  $\sigma$  uma palavra de símbolos pertencentes ao conjunto  $\{1, 2, \dots, n\}$ , denotada por  $\sigma_{str}$ , de forma que  $\sigma_{str}(i) = j$ , onde  $\sigma(j) = i$ ;
- (b) dada uma palavra  $\pi_{str}$  e uma permutação  $\sigma$ , comete-se o abuso notacional de tomar  $\pi_{str}\sigma$  por  $(\pi\sigma)_{str}$ , onde  $\pi_{str}\sigma$  é definida como uma palavra de comprimento  $n$  tal que  $\pi_{str}\sigma(i) = \pi_{str}(j)$ , onde  $\sigma_{str}(i) = j$ .

**Definição 1.3. (Reversão)** Uma permutação  $\rho \in S_n$  é denominada *reversão* se é da forma  $(i \ j)(i+1 \ j-1) \dots (i + \lfloor \frac{j-i}{2} \rfloor + 1 \ i + \lceil \frac{j-i}{2} \rceil + 1)$  para  $1 \leq i < j \leq n$ .

**Observação 1.1.** As notações  $\lfloor k \rfloor$  e  $\lceil k \rceil$  correspondem, respectivamente, ao maior inteiro menor que  $k$  e ao menor inteiro maior que  $k$ .

**Observação 1.2.** A inversa de uma reversão  $\rho$ ,  $\rho^{-1}$ , é ela mesma.

Isto segue do fato de  $\rho$  ser uma transposição de ciclos disjuntos.

**Observação 1.3.** Uma reversão  $\rho$  permuta elementos do intervalo  $[i, j]$ ; isto é, para  $\pi \in S_n$  temos que  $\pi_{str}\rho(i) = \pi_{str}(j)$ ,  $\pi_{str}\rho(j) = \pi_{str}(i)$ , etc.

**Exemplo 1.2.** Sejam  $\pi = (2\ 4\ 3)(5\ 7)$  e  $\rho = (1\ 6)(2\ 5)(3\ 4)$  permutações em  $S_7$ . Observe que  $\pi_{str} = 1\ 3\ 4\ 2\ 7\ 6\ 5$  e  $\pi_{str}\rho = 6\ 7\ 2\ 4\ 3\ 1\ 5$ .

A subpalavra no intervalo  $[1, 6]$  de  $\pi_{str}$  é revertida em  $\pi_{str}\rho$ .

Pelas definições (1.1) e (1.2) vemos que é possível associar a cada permutação de  $S_n$  uma palavra e vice-versa. A permutação identidade, por exemplo, é representada pela palavra  $id_{str} = 1\ 2\ 3\ \dots\ n$ .

Logo, é possível identificarmos o conjunto  $S = \{\pi_{str} = \pi_1\ \pi_2\ \dots\ \pi_n \mid \pi_i \in \{1, 2, \dots, n\}\}$  com  $S_n$ . Temos então a liberdade de usarmos, na demonstração dos resultados, permutações ou palavras, de acordo com o que nos for mais conveniente.

**Definição 1.4.** Sejam  $\pi$  e  $\sigma$  duas permutações dadas e  $\rho_1, \rho_2, \dots, \rho_t$  uma série de reversões tais que  $\pi\rho_1\rho_2\dots\rho_t = \sigma$  e  $t$  é mínimo. Dizemos que  $t$  é a *distância de reversão* entre  $\pi$  e  $\sigma$ .

**Proposição 1.1.** A distância de reversão entre  $\pi$  e  $\sigma$  é igual a distância entre  $\sigma^{-1}\pi$  e  $id$ .

*Demonstração.* Seja  $t$  a distância de reversão entre  $\pi$  e  $\sigma$ . Então existem reversões  $\rho_1, \rho_2, \dots, \rho_t$ , tais que

$$\pi\rho_1\rho_2\dots\rho_t = \sigma \quad (1.1)$$

onde  $t$  é o número mínimo de reversões que transformam  $\pi$  em  $\sigma$ . De (1.1) segue que  $\sigma^{-1}\pi\rho_1\rho_2\dots\rho_t = id$ .

Suponhamos, por absurdo, que existe  $k < t$  tal que  $k$  seja a distância entre  $\sigma^{-1}\pi$  e  $id$ . Então existem reversões  $\theta_1, \theta_2, \dots, \theta_k$  tais que

$$\sigma^{-1}\pi\theta_1\theta_2\dots\theta_k = id \quad (1.2)$$

e  $k$  é o número mínimo de reversões que transformam  $\sigma^{-1}\pi$  em  $id$ .

Por (1.2), temos que  $\pi\theta_1\dots\theta_k = \sigma$ , o que contraria o fato de  $t$  ser a distância entre  $\pi$  e  $\sigma$  já que  $k < t$ .

E o resultado é verificado. □

Logo, ordenar uma permutação  $\pi$  por reversões equivale ao problema de encontrar a distância de reversão entre  $\pi$  e a  $id$ ; ou ainda, encontrar o número mínimo de reversões que transformam  $\pi_{str}$  em  $id_{str}$ . Chamaremos essa distância de  $d(\pi)$ .

Note que neste ponto é interessante analisarmos o problema de rearranjo por reversões a partir de permutações, já que isso nos permite fazer uso das propriedades de composição de funções e existência de inversa e identidade; noções estas que não estão bem definidas para palavras.

**Proposição 1.2.** Dada uma permutação  $\pi \in S_n$  é sempre possível atingir a permutação identidade por meio de reversões.

*Demonstração.* Seja  $\pi_{str} = \pi_1 \pi_2 \dots \pi_n$ .

A prova será por indução no número de elementos ordenados em  $\pi_{str}$ .

Digamos que o elemento 1 da palavra encontra-se na posição  $j$  ( $\pi_j = 1$ ). A reversão  $\rho_1 = (1 \ j)(2 \ j-1) \dots (1 + \lfloor \frac{j-1}{2} \rfloor + 1 \ 1 + \lceil \frac{j-1}{2} \rceil + 1)$  é tal que

$$\pi_{str}\rho_1 = 1 \ \pi_{21} \ \pi_{31} \ \dots \ \pi_{n1}$$

Suponhamos que existam reversões  $\rho_1, \rho_2, \dots, \rho_t$ ,  $t \leq k$ , que ordenem os  $k$ ,  $1 \leq k < n$ , primeiros elementos de  $\pi_{str}$ .

Isto é,

$$\pi_{str}\rho_1\rho_2 \dots \rho_t = 1 \ 2 \ \dots \ k \ \pi_{(k+1)k} \ \pi_{(k+2)k} \ \dots \ \pi_{nk}$$

Consideremos que o elemento  $k+1$  da palavra encontra-se na  $l$ -ésima posição,  $l > k$  ( $\pi_{lk} = k+1$ ). Se necessário (i.e., caso  $l \neq k+1$ ), aplicando a reversão

$$\rho_{t+1} = (k+1 \ l)(k+2 \ l-1) \dots (k+1 + \lfloor \frac{l-(k+1)}{2} \rfloor + 1 \ k+1 + \lceil \frac{l-(k+1)}{2} \rceil + 1)$$

temos que

$$\pi_{str}\rho_1\rho_2 \dots \rho_{t+1} = 1 \ 2 \ \dots \ k \ k+1 \ \pi_{(k+2)(k+1)} \ \pi_{(k+3)(k+1)} \ \dots \ \pi_{n(k+1)}$$

Assim, por indução segue que dada uma permutação  $\pi$  é sempre possível transformá-la na permutação identidade.  $\square$

Notamos que, na prova da proposição anterior, desenvolvida neste trabalho e não vista demonstrada em outro texto, a representação de permutações como palavras nos permite observar claramente quais e como podemos aplicar as reversões adequadas para se chegar à identidade. Como consequência, as reversões de  $S_n$  são geradores do grupo.

Isso garante que dadas permutações  $\pi$  e  $\sigma$  é sempre possível transformar uma na outra por meio de reversões. E disso segue que o problema de distância por reversões está bem definido.

**Proposição 1.3.** Seja  $\pi$  uma permutação de  $S_n$ . A distância de reversão  $d(\pi)$  equivale à distância  $d(\pi^{-1})$ , onde  $\pi^{-1}$  é a inversa de  $\pi$ .

*Demonstração.* Seja  $t$  a distância entre  $\pi$  e a  $id$ . Então existem reversões  $\rho_1, \rho_2, \dots, \rho_t$  tais que

$$\pi \rho_1 \rho_2 \dots \rho_t = id$$

De onde,

$$\pi^{-1} \rho_t \rho_{t-1} \dots \rho_1 = id \tag{1.3}$$

Da Equação (1.3) segue que existem  $t$  reversões que transformam  $\pi^{-1}$  em  $id$ .

Mostremos que não é possível fazer essa transformação com um número menor de reversões.

Suponhamos, por contradição, que a distância entre  $\pi^{-1}$  e  $id$  seja  $s$ ,  $s < t$ .

Então, existem reversões  $\theta_1, \theta_2, \dots, \theta_s$  tais que

$$\pi^{-1} \theta_1 \theta_2 \dots \theta_s = id$$

Logo,

$$\pi \theta_s \theta_{s-1} \dots \theta_1 = id$$

O que contraria o fato de  $d(\pi) = t$ .

□



---

### Uso adequado das notações para permutação na solução de Bafna e Pevzner

---

Neste capítulo, são apresentados os resultados do artigo de Bafna e Pevzner [3] incluindo provas detalhadas nas quais discrimina-se com precisão quando cada uma das notações para permutações (como palavras e ciclos de permutação) deve ser utilizada.

#### 2.1 Grafo de pontos de quebra e limitante inferior para a distância de reversão

Nesta seção iremos introduzir os conceitos de ponto de quebra e adjacência de uma permutação. A partir disso daremos os passos para a construção do grafo de ponto de quebra, uma importante ferramenta no estudo do problema de distância por reversões.

**Definição 2.1.** Dizemos que  $i \sim j$  se  $|i - j| = 1$  e  $i \not\sim j$  caso contrário.

Seja  $\pi \in S_n$  uma permutação dada e  $\pi_{str} = \pi_1 \pi_2 \dots \pi_n$  sua palavra. Estendamos  $\pi_{str}$  adicionando às extremidades esquerda e direita, respectivamente,  $\pi_0 = 0$  e  $\pi_{n+1} = n + 1$ .

**Definição 2.2.** Um par de elementos consecutivos  $\pi_i$  e  $\pi_{i+1}$ ,  $0 \leq i \leq n$ , de  $\pi_{str}$  é dito *adjacente* se  $\pi_i \sim \pi_{i+1}$  e um *ponto de quebra* se  $\pi_i \not\sim \pi_{i+1}$ .

**Exemplo 2.1.** Na palavra  $\pi_{str} = 0\ 1\ 5\ 2\ 3\ 4\ 6\ 7$  os pares  $(1, 5)$ ,  $(5, 2)$ ,  $(4, 6)$  são pontos de quebra, enquanto que  $(0, 1)$ ,  $(2, 3)$ ,  $(3, 4)$  e  $(6, 7)$  são adjacentes.

**Definição 2.3. (Grafo de pontos de quebra)** Seja  $\pi$  uma permutação sobre o conjunto  $\{1, 2, \dots, n\}$  e considere  $0\ \pi_{str}\ n + 1$ . Define-se o *grafo de pontos de quebra* de  $\pi$  como  $G(\pi) := G(\{0, 1, \dots, n + 1\}, E)$ , onde cada vértice  $i$  tem rótulo  $\pi_{str}(i)$  se  $1 \leq i \leq n$  e os vértices  $0$  e  $n + 1$  têm rótulos  $0$  e  $n + 1$ , respectivamente. O conjunto das arestas  $E$  está dividido em arestas de pontos de quebra (pretas) e de desejo (cinza). Entre os vértices  $i$  e  $i + 1$  para  $i = 0, 1, \dots, n$ , incluímos uma aresta preta se e somente se os rótulos de  $i$  e  $i + 1$  não são consecutivos. Além disso, existe uma aresta cinza entre os vértices  $i$  e  $j$  se e somente se  $|i - j| > 1$  e os rótulos de  $i$  e  $j$  não são consecutivos.

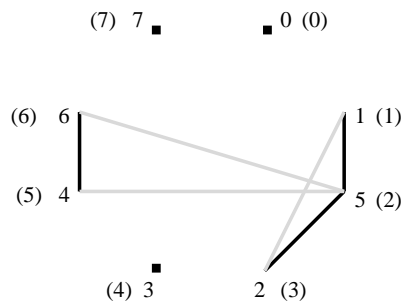


Figura 2.1: Grafo de pontos de quebra de  $\pi = (2\ 3\ 4\ 5) \in S_6$

Observe que os vértices estão dispostos no sentido horário. Na figura, estes estão representados entre parênteses ao lado dos seus respectivos rótulos.

No que se segue, a não ser que seja importante para a compreensão, iremos omitir a representação dos vértices nos grafos de pontos de quebra.

**Definição 2.4. (Ciclos alternados)** Seja  $G(V, E)$  um grafo. Um *ciclo* em  $G(V, E)$  é uma seqüência de vértices  $x_1, x_2, \dots, x_m = x_1$  tal que para todo  $i \leq m$  a aresta  $(x_i, x_{i+1}) \in E$ . Em um grafo de pontos de quebra  $G(\pi)$ , para alguma permutação  $\pi$ , um ciclo é dito *alternado* se as cores das arestas alternam-se.

**Definição 2.5. (i-ciclo alternado)** Seja  $G(V, E)$  um grafo. Diremos que um ciclo em  $G(V, E)$  é um *i-ciclo* se este possui um total de  $i$  arestas.

**Observação 2.1.** Os ciclos alternados no grafo de pontos de quebra de uma permutação possui comprimento par.

No que se segue consideraremos ciclos de decomposição em  $G(\pi)$ . Vamos denotar por  $c(\pi)$  o número máximo de ciclos alternados disjuntos em  $G(\pi)$ . Denotemos, ainda,

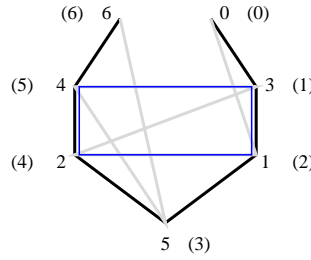


Figura 2.2: Um 4-ciclo alternado no grafo de pontos de quebra de  $\pi = (5\ 3\ 1\ 2\ 4)$

por  $b(\pi)$  o número de arestas pretas em  $G(\pi)$ ; essas representam o número de pontos de quebra em  $0 \leq \pi_{str} \leq n + 1$ .

Dada uma reversão arbitrária  $\rho$ , abreviemos  $G(\pi), G(\pi\rho), b(\pi\rho), b(\pi), c(\pi\rho)$  e  $c(\pi)$  por  $G, G', b', b, c'$  e  $c$ , respectivamente.

Façamos ainda  $\Delta b = \Delta b(\pi, \rho) = b - b'$  e  $\Delta c = \Delta c(\pi, \rho) = c' - c$ .

**Teorema 2.1.** *Para toda permutação  $\pi$  e reversão  $\rho$  temos  $\Delta b(\pi, \rho) + \Delta c(\pi, \rho) \leq 1$ .*

*Demonstração.* Toda reversão remove ou adiciona no máximo dois pontos de quebra.

Logo, temos cinco possíveis valores para  $\Delta b$  ( $\Delta b \in \{-2, -1, 0, 1, 2\}$ ).

No que se segue, vamos assumir que  $\rho$  reverte um fragmento de  $\pi_{str}$  que começa em  $i$  e termina em  $j$ .

As arestas tracejadas nas figuras correspondem a possíveis caminhos, no grafo, para se chegar de um vértice a outro.

- **Caso 1:**  $\Delta b = 2$ ;

Se  $\Delta b = 2$  então uma permutação  $\rho$  transforma  $\pi_{str}$  do tipo  $\dots j' i \dots j i' \dots$  em uma palavra  $(\pi\rho)_{str}$  do tipo  $\dots j' j \dots i i' \dots$  onde  $j' \sim j, i' \sim i, j' \not\sim i$  e  $i' \not\sim j$ .

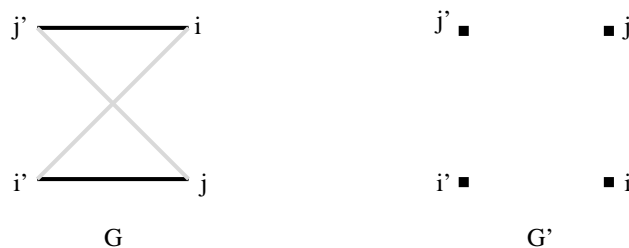


Figura 2.3: Caso 1 da prova do Teorema 2.1

Pontos de quebra  $j', i$  e  $j, i'$  correspondem a um 4-ciclo alternado em  $G(\pi)$ , como mostra a figura acima. A reversão  $\rho$  simplesmente remove este ciclo de  $G(\pi)$ . Dessa maneira, toda decomposição cíclica de  $G'$  em  $c'$  ciclos induz uma decomposição em ciclos de  $G$  em  $c' + 1$  ciclos adicionando o 4-ciclo alternado mostrado acima. Logo,  $c \geq c' + 1$  implicando  $\Delta c \leq -1$  e  $\Delta b + \Delta c \leq 1$ .

• **Caso 2:**  $\Delta b = 1$ ;

Se aplicando uma permutação  $\rho$  temos  $\Delta b = 1$  então duas possibilidades devem ser analisadas:

i)  $\rho$  cria uma nova adjacência (Figura 2.4);

Neste caso  $\pi_{str}$  tem a forma  $\dots j' i \dots j i' \dots$  e  $(\pi\rho)_{str} \dots j' j \dots i i' \dots$ , com  $i \sim i', j' \not\sim i, i' \not\sim j$  e  $j' \not\sim j$ .

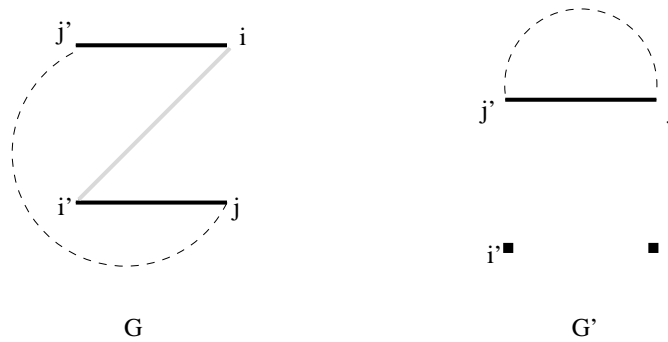


Figura 2.4: Caso 2.i da prova do Teorema 2.1

Note que todo ciclo alternado em  $G'$  contendo a aresta  $(j', j)$  induz um ciclo alternado em  $G$  contendo as arestas  $(j', i)$ ,  $(i, i')$  e  $(i', j)$ . Consequentemente, toda decomposição cíclica de  $G'$  em  $c'$  ciclos induz uma decomposição de  $G$  em no mínimo  $c'$  ciclos.

ii)  $\rho$  transforma um ponto de quebra em uma adjacência (Figura 2.5).

$\pi_{str}$  tem a forma  $\dots j' i \dots j i' \dots$  e  $(\pi\rho)_{str}$  é dado por  $\dots j' j \dots i i' \dots$ , onde  $i \sim i', j \sim j', i \sim j'$  e  $i' \not\sim j$ .

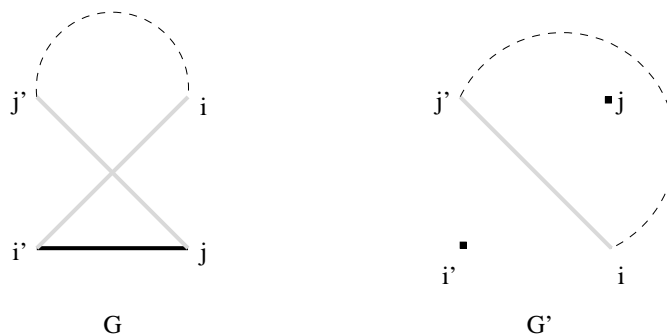


Figura 2.5: Caso 2.ii da prova do Teorema 2.1

Todo ciclo alternado em  $G'$  contendo a aresta  $(j', i)$  induz um ciclo alternado em  $G$  contendo as arestas  $(j', j)$ ,  $(j, i')$  e  $(i, i')$ .

Logo, uma decomposição cíclica de  $G'$  em  $c'$  ciclos induz uma decomposição de  $G$  em no mínimo  $c'$  ciclos.

Em ambos os casos, (i) e (ii),  $c \geq c'$ . Disso segue que  $\Delta c \leq 0$  e  $\Delta b + \Delta c \leq 1$ .

• **Caso 3:**  $\Delta b = 0$ ;

Como  $\Delta b = 0$  então  $b = b'$ .

Novamente, duas situações são possíveis:

- i)  $\pi_{str} = \dots j' j \dots i i' \dots$ , onde  $j' \not\sim j$ ,  $i' \not\sim i$  e  $(\pi\rho)_{str} = \dots j' i \dots j i' \dots$ , onde  $j' \not\sim i$  e  $i' \not\sim j$  (Figura 2.6).

Note que, na pior das hipóteses, ciclos alternados contendo as arestas  $(j', i)$  e  $(i', j)$  em  $G'$  poderão se converter em um único ciclo em  $G$  contendo as arestas  $(j', j')$ ,  $(i', i)$  e  $(i, j')$ .

Logo,  $c' - c \leq 1$ , de onde  $\Delta b + \Delta c \leq 1$ .

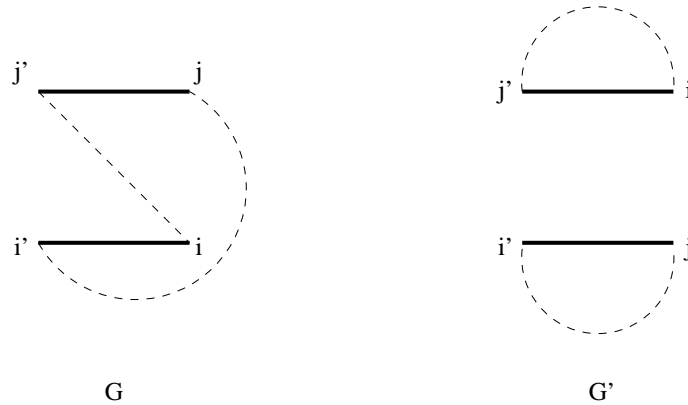


Figura 2.6: Caso 3.i da prova do Teorema 2.1

- ii)  $\pi_{str} = \dots j' j \dots i i' \dots$ , onde  $j \not\sim j'$ ,  $i \sim i'$  e  $(\pi\rho)_{str} = \dots j' i \dots j i' \dots$ , onde  $i' \not\sim j$ ,  $i \sim j'$  (Figura 2.7).

Neste caso, ciclos alternados em  $G'$  contendo as arestas  $(i, i')$  e  $(j, i')$ , como mostra a Figura 2.6, induz uma decomposição em  $G$  em um ciclo contendo as arestas  $(j, j')$  e  $(j', i)$ .

Logo,  $c' - c \leq 1$ . E disso segue que  $\Delta b + \Delta c \leq 1$ .

**Observação 2.2.** O caso em que  $\Delta b = 0$  e  $\Delta b' = 0$  não se dá.

De fato, se  $\Delta b = 0$  então  $\pi_{str} = \dots j' j \dots i i' \dots$ , onde  $j \sim j'$  e  $i \sim i'$ . Suponhamos que exista uma reversão  $\rho$  tal que  $(\pi\rho)_{str} = \dots j' i \dots j i' \dots$ , com  $j' \sim i$  e  $j \sim i'$ . Assim,

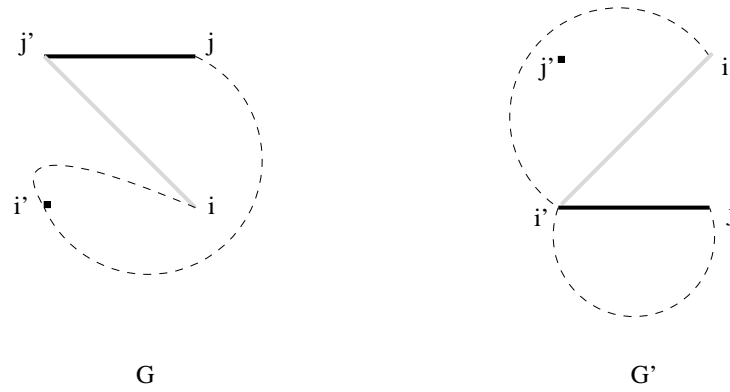


Figura 2.7: Caso 3.ii da prova do Teorema 2.1

$$j = j' + 1 \text{ ou } j = j' - 1 \text{ e } i' = i + 1 \text{ ou } i' = i - 1; \quad (2.1)$$

e também

$$i = j' + 1 \text{ ou } i = j' - 1 \text{ e } i' = j + 1 \text{ ou } i' = j - 1. \quad (2.2)$$

Se  $j = j' + 1$  então  $i = j' - 1$ , por (2.2); isto implicaria, por (2.1)  $i' = j' - 2$ . Mas  $i' = j + 1 = j' + 2$  ou  $i' = j - 1 = j'$ , por (2.2); o que é uma contradição.

Analogamente, se  $j = j' - 1$  então, por (2.2),  $i = j' + 1$ ; isto nos daria  $i' = j' + 2$ . Mas, por (2.2),  $i' = j + 1 = j'$  ou  $i' = j - 1 = j' - 2$ . O que também é uma contradição.

Disso segue a observação.

• **Caso 4:**  $\Delta b = -1$ .

Se  $\rho$  é tal que  $\Delta b = -1$  então:

- i)  $\pi_{str} = \dots j' j \dots i i' \dots$  e  $(\pi\rho)_{str} = \dots j' i \dots j i'$ , onde  $i \sim i'$ ,  $j \sim j'$ ,  $i \sim j'$  e  $i' \not\sim j$  (Figura 2.8).

Ciclos em  $G'$  que contêm as arestas  $(j', j)$ ,  $(j, i')$  e  $(i, i')$  induzem um ciclo em  $G$  que contém a aresta  $(j', i)$ .

- ii)  $\pi_{str} = \dots j' j \dots i i' \dots$ , onde  $j' \sim j$  e  $i \not\sim i'$ . A reversão  $\rho$  aplicada a  $\pi_{str}$  nos fornece  $(\pi\rho)_{str} = \dots j' i \dots j i' \dots$ , onde  $i \not\sim j'$  e  $i' \not\sim j$  (Figura 2.9).

Uma decomposição cíclica de  $G'$  que contém as arestas  $(j', i)$ ,  $(j', j)$  e  $(j, i')$  nos dá um decomposição em  $G$  de um ciclo que contém a aresta  $(i', i)$ .

Em ambos os casos  $c' - c \leq 2$ .

E o resultado segue.

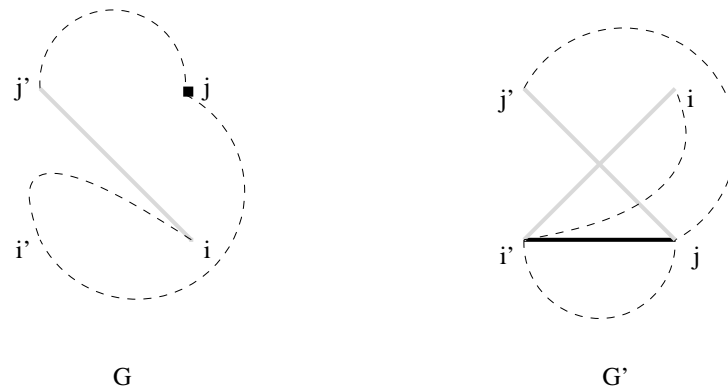


Figura 2.8: Caso 4.i da prova do Teorema 2.1

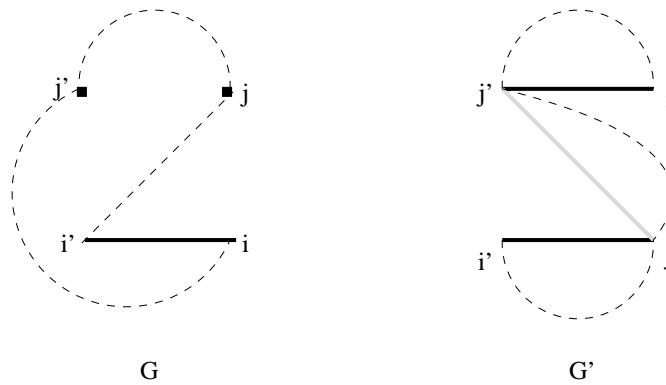


Figura 2.9: Caso 4.ii da prova do Teorema 2.1

• **Caso 5:**  $\Delta b = -2$ .

Neste caso, a permutação  $\rho$  transforma duas adjacências em pontos de quebra. Então,  $\pi_{str} = \dots j' j \dots i i' \dots$ , onde  $i \sim i'$  e  $j \sim j'$ ; e  $(\pi\rho)_{str} = \dots j' i \dots j i' \dots$ , onde  $i \not\sim j'$  e  $i' \not\sim j$  (Figura 2.10).

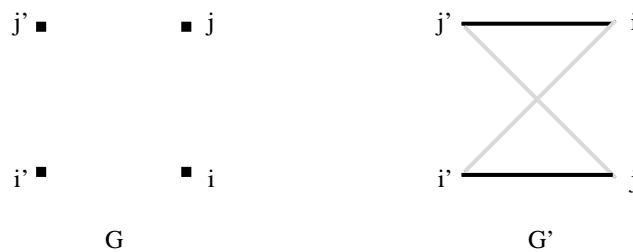


Figura 2.10: Caso 5 da prova do Teorema 2.1

Uma decomposição cíclica em  $G$  em ciclos que contém as arestas  $(j', i)$ ,  $(i, i')$ ,  $(i', j)$  e  $(j, j')$  induz um único ciclo em  $G$  contendo os vértices  $j', j, i$  e  $i'$ .

Logo,  $c' - c \leq 3$ . De onde  $\Delta b + \Delta c \leq 1$ .

A análise desses casos conclui a prova do teorema.  $\square$

A partir do Teorema 2.1 será possível obter uma limitação inferior para a distância de reversão.

**Teorema 2.2.** *Para toda permutação  $\pi$ ,  $d(\pi) \geq b(\pi) - c(\pi)$ .*

*Demonstração.* Seja  $\rho_k, \rho_{k-1}, \dots, \rho_1$  a menor série de reversões que transformam  $\pi = \pi_k$  na permutação identidade  $id$ . Denotemos  $\pi_{i-1} = \pi_i \rho_i$  para  $i = 1, \dots, k$ .

$$\begin{aligned} d(\pi_i) &= d(\pi_{i-1}) + 1 \\ &\geq d(\pi_{i-1}) + \Delta b(\pi_i, \rho_i) + \Delta c(\pi_i, \rho_i) \\ &= d(\pi_{i-1}) + (b(\pi_i) - b(\pi_{i-1})) + (c(\pi_{i-1}) - c(\pi_i)) \end{aligned}$$

A desigualdade acima é dada pelo Teorema 2.1.

Lembrando que  $d(id) = b(id) = c(id) = 0$ , vale que para todo  $1 \leq i \leq k$ ,

$$\begin{aligned} d(\pi_i) - (b(\pi_i) - c(\pi_i)) &\geq \\ d(\pi_{i-1}) - (b(\pi_{i-1}) - c(\pi_{i-1})) &\geq \dots \geq \\ d(id) - (b(id) - c(id)) &= 0 \end{aligned}$$

Logo,  $d(\pi) \geq b(\pi) - c(\pi)$ .  $\square$

## 2.2 Permutação de Gollan e diâmetro de reversão

Neste ponto iremos tratar a conjectura de Gollan sobre diâmetro de reversão do grupo simétrico.

Para o problema de ordenação por reversões, Gollan conjecturou que o máximo da distância de reversão é  $n - 1$  e que apenas uma permutação,  $\gamma_n$ , e seu inverso,  $\gamma_n^{-1}$  requerem  $n - 1$  reversões para serem organizadas.

**Definição 2.6.** O *diâmetro de reversão* do grupo simétrico de ordem  $n$  é definido como  $d(n) = \max_{\pi \in S_n} d(\pi)$ .

Em notação cíclica a permutação de Gollan é definida como segue:

$$\gamma_n = \begin{cases} (n \ n-2 \ n-4 \ \dots \ 1 \ 2 \ 4 \ 6 \ \dots \ n-1), & n \text{ ímpar} \\ (n-1 \ n-3 \ n-5 \ \dots \ 1 \ 2 \ 4 \ 6 \ \dots \ n), & n \text{ par} \end{cases}$$



**Observação 2.3.** A permutação  $\gamma_n$  tem

$$(\gamma_n)_{str} = \begin{cases} 3 \ 1 \ 5 \ \dots \ n-5 \quad n \quad n-3 \ n-1, & n \text{ ímpar} \\ 3 \ 1 \ 5 \ \dots \ n-1 \ n-4 \quad n \quad n-2, & n \text{ par} \end{cases}$$

como palavra associada se  $n > 2$ ,  $\gamma_1 = 1$  e  $\gamma_2 = 2 \ 1$  se  $n = 1$  e  $n = 2$ , respectivamente.

**Observação 2.4.** Observe que a permutação  $\gamma_n^{-1}$  é dada por

$$\gamma_n^{-1} = \begin{cases} (n-1 \ n-3 \ n-5 \ \dots \ 4 \ 2 \ 1 \ \dots \ n-2 \ n), & n \text{ ímpar} \\ (n \ n-2 \ n-4 \ \dots \ 4 \ 2 \ 1 \ \dots \ n-3 \ n-1), & n \text{ par} \end{cases}$$

e tem como palavra associada

$$(\gamma_n^{-1})_{str} = \begin{cases} 2 \ 4 \ 1 \ \dots \ n-1 \ n-4 \quad n \quad n-2, & n \text{ ímpar} \\ 2 \ 4 \ 1 \ \dots \ n-5 \quad n \quad n-3 \ n-1, & n \text{ par} \end{cases}$$

**Exemplo 2.2.**  $\gamma_8 = (7 \ 5 \ 3 \ 1 \ 2 \ 4 \ 6 \ 8)$  que tem  $(\gamma_8)_{str} = 3 \ 1 \ 5 \ 2 \ 7 \ 4 \ 8 \ 6$  como palavra associada.

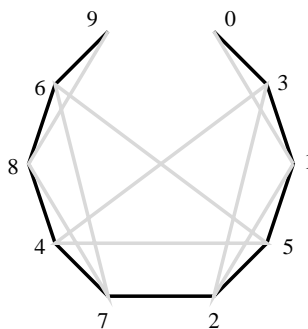


Figura 2.11: Grafo de pontos de quebra para  $G(\gamma_8)$

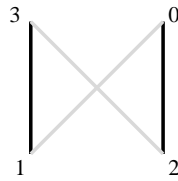
Em [3] não fica clara a definição de permutação de Gollan. No capítulo 4 aponta-se a dubiedade da definição proposta por Bafna e Pevzner em [3].

**Lema 2.1.** Todo ciclo alternado em  $G(\gamma_n)$  contém o vértice de rótulo 1 ou 3(ou ambos).

*Demonstração.* Os casos em que  $n \leq 5$  são tratados especificamente.

- $n = 1$ ;

Se  $n = 1$  então  $G(\gamma_1)$  não contém nenhum ciclo. O resultado segue por vacuidade.

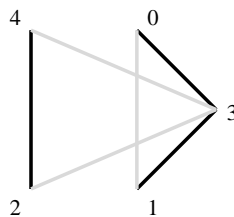
Figura 2.12:  $G(\gamma_2)$ 

- $n = 2$ ;

Neste caso  $G(\gamma_2)$  será o 4-ciclo dado na figura. Logo, o único ciclo alternado em  $G(\gamma_2)$  conterá os vértices de rótulos 1 e 3 (Figura 2.12).

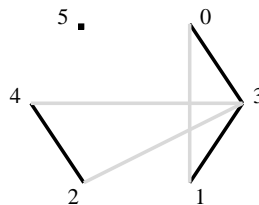
- $n = 3$ ;

Assim como no item anterior,  $G(\gamma_3)$  possui apenas um ciclo alternado. Este ciclo contém os vértices 1 e 3 (Figura 2.13).

Figura 2.13:  $G(\gamma_3)$ 

- $n = 4$ ;

Neste caso,  $G(\gamma_4)$  também contém um único ciclo alternado. E o resultado é válido (Figura 2.14).

Figura 2.14:  $G(\gamma_4)$ 

- $n = 5$ .

Note que qualquer ciclo alternado que parta do vértice de rótulo 0, 1 ou 3 deve conter o vértice de rótulo 1 ou 3 (Figura 2.15).

Observemos o caso em que o ciclo alternado parte do vértice de rótulo 5. Se deste vértice parte uma aresta preta, então temos que 5 liga-se ao vértice de rótulo 1, e neste caso o resultado segue, ou 5 liga-se ao vértice de rótulo 2 e logo 2 liga-se a 1 ou 3 por uma aresta cinza. Consideremos que do vértice 5 parta uma aresta cinza. Analisemos, por exemplo, o caso em que 5 liga-se a 6. Logo, 6 deve-se ligar a 4 por uma aresta preta. Se 4 chega a 3 por uma aresta cinza, o resultado é verificado; caso 4 volte a 5 por uma aresta cinza, este vértice deve-se ligar a 1 ou 2 por uma aresta preta. Se o segundo caso ocorre, 2 deve-se ligar a 1 ou 3 por uma aresta cinza.

Fazendo uma análise análoga para os outros vértices, concluiremos que qualquer ciclo alternado em  $G(\gamma_5)$  contém os vértices 1 ou 3.

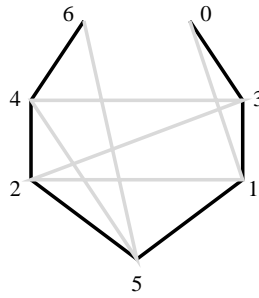


Figura 2.15:  $G(\gamma_5)$

Para  $n \geq 5$  seja  $i$  o vértice de rótulo ímpar mínimo de um ciclo alternado  $X$  em  $G(\gamma_n)$ .

Considere a sequência  $i, j, k$  de rótulos consecutivos em  $X$  onde  $(i, j)$  representa uma aresta preta e  $(j, k)$  uma aresta cinza.

Se  $i > 5$  então  $j = i - 3$  ou  $j = i - 5$  e  $k = j + 1$  ou  $k = j - 1$ , implicando  $k$  ímpar e  $k < i$ , o que contraria a minimalidade de  $i$ .

Se  $i = 5$  então  $j = 2$  e  $k = 1$  ou  $3$ , o que também contraria a minimalidade de  $i$ .

Logo, segue que  $i = 1$  ou  $i = 3$ . □

**Teorema 2.3 (Conjectura de Gollan).** Para todo  $n$ ,  $d(\gamma_n) = d(\gamma_n^{-1}) = n - 1$ .

*Demonstração.* Consideremos quatro casos iniciais.

- se  $n = 1$ ,  $\gamma_1 = (1)$  e nenhuma reversão precisa ser feita para que  $(\gamma_1)_{str}$  chegue a  $id_{str}$ ;
- se  $n = 2$ ,  $\gamma_2 = (1\ 2)$  e  $(\gamma_2)_{str} = 2\ 1$  e apenas uma reversão é necessária para se chegar a  $id_{str}$ ;

- se  $n = 3$ ,  $\gamma_3 = (3 \ 1 \ 2)$  e  $(\gamma_3)_{str} = 3 \ 1 \ 2$  e são necessárias duas reversões para se chegar a  $id_{str}$ .
- se  $n = 4$ ,  $\gamma_4 = (3 \ 1 \ 2 \ 4)$  e  $(\gamma_4)_{str} = 3 \ 1 \ 4 \ 2$  e são necessárias três reversões para se chegar a  $id_{str}$ .

Para  $n \geq 4$  particionamos o conjunto de vértices de  $G(\gamma_n)$  em  $V_l = \{0, 1, 3\}$  e  $V_r$ . Todo ciclo alternado em  $G(\gamma_n)$  contém o vértice de rótulo 1 ou 3, pelo lema anterior. Além disso, observe que não existe ciclo alternado em  $V_l$ .

Dessas observações segue que todo ciclo alternado deve conter ao menos duas arestas entre  $V_l$  e  $V_r$ . Como o corte  $(V_l, V_r)$  contém 4 arestas  $((1, 2), (1, 5), (3, 2), (3, 4))$ , o número máximo de ciclos alternados disjuntos em um ciclo de decomposição  $G(\gamma_n)$  é  $\frac{4}{2} = 2$ .

Do Teorema 2.2,  $d(\gamma_n) \geq b(\gamma_n) - c(\gamma_n) \geq n + 1 - 2 = n - 1$ .

Por outro lado, em [3] é dito que podemos encontrar em [19] um algoritmo que ordene, na pior das hipóteses, uma palavra de tamanho  $n$  em  $n - 1$  passos. Logo,  $d(\gamma_n) \leq n - 1$ .

Disso, segue que  $d(\gamma_n) = n - 1$ .

Já vimos que uma permutação e sua inversa têm a mesma distância de reversão.

Portanto,  $d(\gamma_n) = d(\gamma_n^{-1}) = n - 1$ .

Isso conclui a prova do teorema.  $\square$

**Definição 2.7.** Definamos a permutação de Gollan sobre o conjunto  $\{1, 2, \dots, n + 1\}$ , denotada por  $\gamma_n^+$ , por:

$$\gamma_n^+ = \begin{cases} (n + 1 \ n - 1 \ n - 3 \ \dots \ 2 \ 3 \ 5 \ \dots \ n - 2 \ n), & n \text{ ímpar} \\ (n \ n - 2 \ n - 4 \ \dots \ 2 \ 3 \ 5 \ \dots \ n - 1 \ n + 1), & n \text{ par} \end{cases}$$

**Teorema 2.4 (Conjectura forte de Gollan).** Para todo  $n$ ,  $\gamma_n$  e  $\gamma_n^{-1}$  são as únicas permutações que requerem, no mínimo,  $n - 1$  reversões para serem ordenadas.

*Demonstração.* Definamos  $P_n \equiv \{\pi \mid \pi \in S_n \text{ e } d(\pi) = n - 1\}$ .

Pelo Teorema 2.3 temos que  $\{\gamma_n, \gamma_n^{-1}\} \subseteq P_n$ .

Nossa estratégia será mostrar que  $P_n \subseteq \{\gamma_n, \gamma_n^{-1}\}$ . A prova é por indução na distância de reversão da permutação.

Notemos que  $\gamma_1 = (1) = \gamma_n^{-1}$  e nenhuma reversão precisa ser feita para se chegar à identidade.

Considere  $\pi \in P_n$ . Seja  $\rho_\pi$  a reversão que leva  $n$  ao fim da palavra, isto é,  $\pi_{str}\rho_\pi = \pi'_{str} n$ , onde  $\pi'$  é uma permutação de  $\{1, \dots, n - 1\}$  e  $\pi'_{str} n$  representa a concatenação de  $\pi'_{str}$  e  $n$ . Como  $\pi' \in S_{n-1}$  segue que  $d(\pi') \leq n - 2$ . Logo, por hipótese de indução  $\pi'$  é  $\gamma_{n-1}$  ou  $\gamma_{n-1}^{-1}$ .

Defina  $A = \{\pi | \pi \rho_\pi = \gamma_{n-1}\}$  e  $B = \{\pi | \pi \rho_\pi = \gamma_{n-1}^{-1}\}$ . Pode-se observar que  $P_n \subseteq A \cup B$ . Da mesma maneira defina  $\rho'_\pi$  como a reversão que coloca 1 no início da palavra, isto é,  $\pi_{str} \rho'_\pi = 1 \pi'_{str}$ , onde  $\pi'$  é uma permutação de  $\{2, 3, \dots, n\}$ . Como antes,  $d(\pi') \leq n - 2$ , o que implica que  $\pi'$  é  $\gamma_{n-1}^+$  ou  $(\gamma_{n-1}^+)^{-1}$ . Definamos  $C = \{\pi | \pi \rho'_\pi = \gamma_{n-1}^+\}$  e  $D = \{\pi | \pi \rho'_\pi = (\gamma_{n-1}^+)^{-1}\}$ . Então,  $P_n \subseteq C \cup D$  e, conseqüentemente,  $P_n \subseteq (A \cup B) \cap (C \cup D)$ .

Afirmamos que

$$A \cap C = \emptyset, A \cap D = \{\gamma_n\}, B \cap C = \{\gamma_n^{-1}\} \text{ e } B \cap D = \emptyset \quad (2.3)$$

De onde  $P_n \subseteq \{\gamma_n, \gamma_n^{-1}\}$ .

Vamos verificar que as afirmações em (2.3) são válidas.

- $A \cap C = \emptyset$ ;

Suponhamos, por contradição, que  $A \cap C \neq \emptyset$ . Então existe uma permutação  $\pi$  tal que

$$\pi \rho_\pi = \gamma_{n-1} \text{ e } \pi \rho'_\pi = \gamma_{n-1}^+$$

Disso segue que

$$\pi = \gamma_{n-1} \rho_\pi \text{ e } \pi = \gamma_{n-1}^+ \rho'_\pi$$

Logo,

$$\begin{aligned} \gamma_{n-1} \rho_\pi &= \gamma_{n-1}^+ \rho'_\pi \\ \gamma_{n-1} &= \gamma_{n-1}^+ \rho'_\pi \rho_\pi \end{aligned}$$

Isso significa que existem reversões  $\rho_\pi$  e  $\rho'_\pi$  que transformam  $(\gamma_{n-1})_{str} n$  em  $1 (\gamma_{n-1}^+)_{str}$ .

Dois casos devem ser analisados:

- i)  $n$  é par;

Neste caso a permutação de Gollan é dada por

$$\gamma_{n-1} = (n-1 \ n-3 \ \dots \ 1 \ 2 \ 4 \ \dots \ n-2)$$

Temos então que:

$$\begin{aligned} (I) \quad (\gamma_{n-1})_{str} n &= 3 \ 1 \ 5 \ \dots \ n-1 \ n-4 \ n-2 \ n \\ (II) \quad 1 (\gamma_{n-1}^+)_{str} &= 1 \ 4 \ 2 \ 6 \ \dots \ n \ n-3 \ n-1 \end{aligned}$$

Para organizarmos a palavra em (II) segundo a palavra em (I) devemos colocar 1 e  $n$  nas segunda e última posições, respectivamente.

As únicas reversões que nos fornecem o desejado são  $\rho_\pi = (1\ 2)$  e  $\rho'_\pi = (n-2\ n)$ . Mas aplicando  $\rho_\pi \rho'_\pi$  à 1  $(\gamma_{n-1}^+)_{str}$  temos, por exemplo, 4 na primeira posição; o que não ocorre em  $(\gamma_{n-1})_{str}$   $n$ .

Portanto, não existem duas únicas reversões que transformam  $\gamma_{n-1}$  em  $\gamma_{n-1}^+$ .

ii)  $n$  é ímpar.

A permutação de Gollan é

$$\gamma_{n-1} = (n-2\ n-4\ \dots\ 1\ 2\ 4\ \dots\ n-3\ n-1)$$

De onde

$$(I)\quad (\gamma_{n-1})_{str}\ n = 3\ 1\ 5\ \dots\ n-5\ n-1\ n-3\ n$$

$$(II)\quad 1\ (\gamma_{n-1}^+)_{str} = 1\ 4\ 2\ 6\ \dots\ n-4\ n\ n-2$$

Como no caso anterior queremos encontrar duas reversões  $\rho_\pi$  e  $\rho'_\pi$  que transformam a palavra em (II) na palavra em (I).

Para que coloquemos, em (II), 1 e  $n$  nas posições que se encontram em (I) devemos aplicar  $\rho_\pi = (1\ 2)$  e  $\rho'_\pi = (n-1\ n)$ ; mas essas reversões deixam, por exemplo, 4 na primeira posição; o que não acontece em  $(\gamma_{n-1})_{str}$   $n$ .

Então, segue por (i) e (ii) que  $A \cap C = \emptyset$ .

- $B \cap D = \emptyset$ ;

Como no caso anterior suponhamos  $B \cap D \neq \emptyset$ . Então existem uma permutação  $\pi$  e reversões  $\rho_\pi$  e  $\rho'_\pi$  tais que

$$\gamma_{n-1}^{-1} = (\gamma_{n-1}^+)^{-1} \rho'_\pi \rho_\pi$$

i)  $n$  é par;

Temos então que

$$\gamma_{n-1}^{-1} = (n-2\ n-4\ \dots\ 4\ 2\ 1\ \dots\ n-5\ n-3\ n-1)$$

e

$$(I)\quad (\gamma_{n-1}^{-1})_{str}\ n = 2\ 4\ 1\ \dots\ n-5\ n-1\ n-3\ n$$

$$(II)\quad 1\ ((\gamma_{n-1}^+)^{-1})_{str} = 1\ 3\ 5\ 2\ \dots\ n-4\ n\ n-2$$

As reversões que, em (II) organizam 1 e  $n$  como em (I) são  $\rho_\pi = (1\ 3)$  e  $\rho'_\pi = (n-1\ n)$ . Mas aplicando aplicando  $\rho_\pi$  e  $\rho'_\pi$  temos que, por exemplo, 5 fica na primeira posição em  $1\ ((\gamma_{n-1}^+)^{-1})_{str}$ , o que não ocorre em  $(\gamma_{n-1}^{-1})_{str}\ n$ .

Portanto, não conseguimos transformar  $1\ ((\gamma_{n-1}^+)^{-1})_{str}$  em  $(\gamma_{n-1}^{-1})_{str}\ n$  aplicando apenas duas reversões.

ii)  $n$  é ímpar.

Se  $n$  é ímpar segue que

$$\gamma_{n-1}^{-1} = (n-1\ n-3\ \dots\ 6\ 4\ 2\ 1\ \dots\ n-6\ n-4\ n-2)$$

Então

$$\begin{aligned} (I)\quad (\gamma_{n-1}^{-1})_{str}\ n &= 2\ 4\ 1\ \dots\ n-1\ n-4\ n-2\ n \\ (II)\quad 1\ ((\gamma_{n-1}^+)^{-1})_{str} &= 1\ 3\ 5\ 2\ \dots\ n\ n-3\ n-1 \end{aligned}$$

Neste caso,  $\rho_\pi = (1\ 3)$  e  $\rho'_\pi = (n-2\ n)$ ; mas essas levam 5 para a primeira posição em  $1\ ((\gamma_{n-1}^+)^{-1})_{str}$ ; o que não ocorre em  $(\gamma_{n-1}^{-1})_{str}\ n$ .

Assim temos que  $B \cap D = \emptyset$ .

- $A \cap D = \{\gamma_n\}$ ;

Distinguiremos novamente os casos em que  $n$  é par e ímpar.

i)  $n$  é par;

Observemos que

$$\begin{aligned} \gamma_n &= (n-1\ n-3\ \dots\ 1\ 2\ 4\ \dots\ n-4\ n-2\ n) \\ \gamma_{n-1} &= (n-1\ n-3\ \dots\ 1\ 2\ 4\ \dots\ n-6\ n-4\ n-2) \end{aligned}$$

Notemos que para a reversão  $\rho_{\gamma_n} = (n-1\ n)$  vale que

$$\gamma_n = \gamma_{n-1}\rho_{\gamma_n}$$

E disso segue que  $\gamma_n \in A$ .

Agora,

$$\gamma_{n-1}^+ = (n-1\ n-3\ \dots\ 5\ 3\ 2\ \dots\ n-4\ n-2\ n)$$

Logo, para  $\rho'_{\gamma_n}$  temos

$$\gamma_n = \gamma_{n-1}^+ \rho'_{\gamma_n}$$

Disso segue que  $\gamma_n \in D$ .

Portanto, para o caso par  $\gamma_n \in A \cap D$ .

ii)  $n$  é ímpar.

Neste caso,

$$\begin{aligned} \gamma_n &= (n \ n-2 \ \dots \ 1 \ 2 \ 4 \ \dots \ n-5 \ n-3 \ n-1) \\ \gamma_{n-1} &= (n-2 \ n-4 \ \dots \ 1 \ 2 \ 4 \ \dots \ n-5 \ n-3 \ n-1) \end{aligned}$$

Para  $\rho_{\gamma_n} = (n-2 \ n)$  temos que

$$\gamma_n = \gamma_{n-1} \rho_{\gamma_n}$$

E disso vale que  $\gamma_n \in A$ .

Por outro lado,

$$\gamma_{n-1}^+ = (n-1 \ n-3 \ \dots \ 4 \ 2 \ 1 \ \dots \ n-6 \ n-4 \ n-2)$$

Considerando  $\rho'_{\gamma_n} = (1 \ 2)$  temos que

$$\gamma_n = \gamma_{n-1}^+ \rho'_{\gamma_n}$$

De onde temos  $\gamma_n \in D$ .

Logo, para o caso ímpar também vale que  $\gamma_n \in A \cap D$ .

De (i) e (ii) segue que  $\gamma_n \in A \cap D$ .

Agora devemos mostrar que  $\gamma_n$  é o único elemento de  $A \cap D$ .

Suponhamos, por contradição, que exista uma permutação  $\pi \in A \cap D$ , com  $\pi \neq \gamma_n$ .

De  $\pi \in A \cap D$  temos que existem reversões  $\rho_\pi$  e  $\rho'_\pi$  tais que

$$\pi \rho_\pi = \gamma_{n-1} \text{ e } \pi \rho'_\pi = \gamma_{n-1}^+$$

De onde temos

$$\gamma_{n-1} = \gamma_{n-1}^+ \rho'_\pi \rho_\pi \tag{2.4}$$



i)  $n$  é par;

Observemos que

$$(\gamma_{n-1})_{str} = (3 \ 1 \ 5 \ \dots \ n-1 \ n-4 \ n-2 \ n)$$

$$1((\gamma_{n-1}^+)^{-1})_{str} = (1 \ 3 \ 5 \ 2 \ \dots \ n-4 \ n \ n-1)$$

Disso segue que  $\rho'_\pi = (1 \ 2)$  e  $\rho_\pi = (n-1 \ n)$ , o que não pode acontecer já que estamos supondo  $\pi \neq \gamma_n$ .

De fato, se  $\rho_\pi = \rho_{\pi\gamma_n} = (1 \ 2)$  e  $\rho'_\pi = \rho'_{\pi\gamma_n} = (n-1 \ n)$  teríamos:

$$\gamma_n \rho_\pi = \gamma_{n-1} = \pi \rho_\pi$$

E conseqüentemente teríamos

$$\gamma_n = \pi$$

O que gera um absurdo.

ii)  $n$  é ímpar.

No caso em que  $n$  é ímpar temos

$$(\gamma_{n-1})_{str} = (3 \ 1 \ 5 \ \dots \ n-1 \ n-3 \ n)$$

$$1((\gamma_{n-1}^+)^{-1})_{str} = (1 \ 3 \ 5 \ \dots \ n \ n-3 \ n-1)$$

Para que (2.4) seja satisfeita é necessário que  $\rho'_\pi = (1 \ 2)$  e  $\rho_\pi = (n-2 \ n)$ . Argumentando como em (i), disso teríamos que  $\pi = \gamma_n$ . O que novamente nos geraria um absurdo.

Analisados os dois casos, segue que  $A \cap D = \{\gamma_n\}$ .

- $B \cap C = \{\gamma_n^{-1}\}$ .

Primeiramente, mostremos que  $\gamma_n^{-1} \in B \cap C$ .

Novamente, dois casos devem ser levados em consideração.

i)  $n$  é par;

Neste caso,

$$\gamma_n^{-1} = (n \ n-2 \ n-4 \ \dots \ 6 \ 4 \ 2 \ 1 \ \dots \ n-5 \ n-3 \ n-1)$$

$$\gamma_{n-1}^{-1} = (n-2 \ n-4 \ \dots \ 4 \ 2 \ 1 \ \dots \ n-3 \ n-1)$$

Seja  $\rho_{\gamma_n^{-1}} = (n-2 \ n)$ .

Notemos que  $\gamma_n^{-1}(\rho_{\gamma_n^{-1}}) = \gamma_{n-1}^{-1}$ .

Logo,  $\gamma_n^{-1} \in B$ .

Agora,

$$\gamma_{n-1}^+ = (n \ n-2 \ \dots \ 2 \ 3 \ 5 \ \dots \ n-3 \ n-1)$$

Para  $\rho'_{\gamma_n^{-1}} = (1 \ 3)$  temos que

$$\gamma_n^{-1}(\rho'_{\gamma_n^{-1}}) = \gamma_{n-1}^+$$

Logo,  $\gamma_n^{-1} \in C$ .

Segue, das observações feitas acima, que  $\gamma_n^{-1} \in B \cap C$ .

ii)  $n$  é ímpar.

Temos que

$$\gamma_n^{-1} = (n-1 \ n-3 \ n-5 \ \dots \ 6 \ 4 \ 2 \ 1 \ \dots \ n-4 \ n-2 \ n)$$

$$\gamma_{n-1}^{-1} = (n-1 \ n-3 \ \dots \ 6 \ 4 \ 2 \ 1 \ \dots \ n-6 \ n-4 \ n-2)$$

Podemos observar que se  $\rho_{\gamma_n^{-1}} = (n-1 \ n)$  então

$$\gamma_n^{-1}(\rho_{\gamma_n^{-1}}) = \gamma_{n-1}^{-1}$$

Portanto, para o caso  $n$  ímpar,  $\gamma_n^{-1} \in B$ .

Além disso,

$$\gamma_{n-1}^+ = (n-1 \ n-3 \ \dots \ 2 \ 3 \ 5 \ \dots \ n-4 \ n-2 \ n)$$

Para  $\rho'_{\gamma_n^{-1}} = (1 \ 3)$  temos que

$$\gamma_n^{-1}(\rho'_{\gamma_n^{-1}}) = \gamma_{n-1}^+$$

De onde segue que  $\gamma_n^{-1} \in C$ .

Para o caso ímpar também temos que  $\gamma_n^{-1} \in B \cap C$ .

Demonstremos agora a unicidade.

Novamente, suponhamos por contradição, que exista  $\pi \neq \gamma_n^{-1}$  tal que  $\pi \in B \cap C$ .

Então vale que  $\pi\rho_\pi = \gamma_{n-1}^{-1}$  e  $\pi\rho'_\pi = \gamma_{n-1}^+$ .

Logo, temos que

$$\gamma_{n-1}^{-1} = \gamma_{n-1}^+ \rho'_\pi \rho_\pi \quad (2.5)$$

i)  $n$  é par;

Neste caso,

$$\begin{aligned} (\gamma_{n-1}^{-1})_{str} &= 2 \quad 4 \quad 1 \quad \dots \quad n-5 \quad n-1 \quad n-3 \quad n \\ 1 (\gamma_{n-1}^+)_{str} &= 1 \quad 4 \quad 2 \quad 6 \quad \dots \quad n \quad n-3 \quad n-1 \end{aligned}$$

Logo, por (2.5) temos que  $\rho'_\pi = (1 \quad 3)$  e  $\rho_\pi = (n-2 \quad n)$ . O que contraria o fato de  $\pi \neq \gamma_n^{-1}$ .

ii)  $n$  é ímpar.

Analogamente,

$$\begin{aligned} (\gamma_{n-1}^{-1})_{str} &= 2 \quad 4 \quad 1 \quad \dots \quad n-1 \quad n-4 \quad n-2 \quad n \\ 1 (\gamma_{n-1}^+)_{str} &= 1 \quad 4 \quad 2 \quad 6 \quad \dots \quad n-4 \quad n \quad n-2 \end{aligned}$$

Para que valha (2.5), temos que  $\rho'_\pi = (1 \quad 3)$  e  $\rho_\pi = (n-1 \quad n)$ .

Mas isso implica  $\pi = \gamma_n^{-1}$ . O que é uma contradição.

Portanto,  $B \cap C = \gamma_n^{-1}$ .

E isso conclui a prova do Teorema. □

A demonstração das equações em (2.3) não estava presente no artigo de Bafna e Pevzner. Assim, a prova apresentada foi desenvolvida neste trabalho e não transcrita de [3] ou [4].

## 2.3 Limitante para o valor esperado da distância de reversão

Seja  $\pi \in S_n$ . Consideremos o conjunto de ciclos que formam uma decomposição cíclica máxima de  $\pi$  e particionemo-lo de acordo com o comprimento dos ciclos. Denotemos por  $c_i(\pi)$  o número de ciclos alternados de comprimento  $i$  em uma decomposição cíclica máxima.

Temos que

$$c(\pi) = \sum_{i=4}^{2(n+1)} c_i(\pi)$$

**Observação 2.5.** Já observamos que os ciclos alternados em um grafo de pontos de quebra em uma permutação têm comprimento par; sendo assim, no somatório acima  $c_i(\pi) = 0$  para todo  $i$  ímpar. No que se segue, também tomaremos  $c_i(\pi) = 0$  sempre que  $i$  for ímpar.

Para  $k \leq 2(n+1)$ , consideremos ciclos de decomposição máxima cujo comprimento é no mínimo  $k$ . O número de tais ciclos é  $c(\pi) - \sum_{i=4}^{k-1} c_i(\pi)$ .

No grafo de pontos de quebra de uma permutação  $\pi$  para cada aresta preta temos uma aresta cinza de desejo. Logo,  $G(\pi)$  tem exatamente  $2b(\pi)$  arestas. Como os ciclos são disjuntos, o número de arestas para se construir ciclos de comprimento no mínimo  $k$  é  $2b(\pi) - \sum_{i=4}^{k-1} ic_i(\pi)$ .

Disso, segue que para todo  $k \leq 2(n+1)$  vale

$$c(\pi) - \sum_{i=4}^{k-1} c_i(\pi) \leq \frac{1}{k} \left( 2b(\pi) - \sum_{i=4}^{k-1} ic_i(\pi) \right)$$

Logo,

$$c(\pi) \leq \frac{1}{k} \left( 2b(\pi) - \sum_{i=4}^{k-1} ic_i(\pi) \right) + \frac{k}{k} \sum_{i=4}^{k-1} c_i(\pi)$$

De onde,

$$c(\pi) \leq \frac{1}{k} \left( 2b(\pi) + \sum_{i=4}^{k-1} (k-i)c_i(\pi) \right) \quad (2.6)$$

Pelos Teorema 2.2 e Desigualdade (2.6) temos que

$$\begin{aligned} d(\pi) &\geq b(\pi) - \frac{1}{k} \left( 2b(\pi) + \sum_{i=4}^{k-1} (k-i)c_i(\pi) \right) \\ d(\pi) &\geq \left( 1 - \frac{2}{k} \right) b(\pi) - \frac{1}{k} \left( \sum_{i=4}^{k-1} (k-i)c_i(\pi) \right) \end{aligned}$$

Como  $\frac{k-i}{k} < 1$  para todo  $k \leq 2(n+1)$  e para todo  $i = 4, 5, \dots, k-1$ ,

$$d(\pi) \geq \left( 1 - \frac{2}{k} \right) b(\pi) - \left( \sum_{i=4}^{k-1} c_i(\pi) \right) \quad (2.7)$$

O lema a seguir nos fornecerá uma limitação para o valor esperado do número de ciclos de comprimento  $i$  em uma decomposição cíclica máxima. Este número é independente de  $n$ .

**Lema 2.2.**  $E(c_i(\pi)) \leq \frac{2^i}{i}$ .

*Demonstração.* Seja  $\pi \in S_n$ . Consideremos um ciclo alternado  $c_i(\pi)$  em uma decomposição cíclica máxima de  $\pi$ . Digamos que este ciclo tenha comprimento  $i = 2t$ ; ou seja, contém  $t$  arestas pretas. Representemos o conjunto dessas  $t$  arestas por  $\{(x'_t, x_1), (x'_1, x_2), \dots, (x'_{t-1}, x_t)\}$  onde  $x'_j \sim x_j$ .

Consideremos o conjunto de vértices cujos rótulos são  $x_1, x_2, \dots, x_t$ . Afirmamos que, em uma decomposição cíclica máxima, esses vértices são dois a dois distintos.

De fato, suponhamos que  $x_i = x_j$  para algum  $1 \leq i < j < t$ . Então, as arestas  $(x'_i, x_{i+1}), (x'_{i+1}, x_{i+2}), \dots, (x'_{j-1}, x_j = x_i)$  formam um ciclo alternado em  $G(\pi)$ . Dessa forma, o ciclo  $c_i(\pi)$  pode ser quebrado em dois outros. Isso nos fornece uma decomposição cíclica maior do que a inicial, contrariando o fato dessa ser máxima.

Observemos que temos  $\frac{n!}{(n-t)!}$  maneiras de escolher o conjunto ordenado  $x_1, x_2, \dots, x_t$ . Uma vez que este conjunto é fixado, para cada  $x_i, i \in \{1, 2, \dots, t\}$ , temos no máximo duas maneiras de tomar  $x'_j$  de modo que  $(x_i, x'_j)$  seja uma aresta no ciclo alternado  $c_i(\pi)$ . Disso segue que obtemos no máximo  $2^t \frac{n!}{(n-t)!}$  ciclos de comprimento  $2t$  em  $G(\pi)$ . Note que cada ciclo de comprimento  $2t$  foi contado  $2t$  vezes. Portanto, um melhor limitante para o número de ciclos de comprimento  $2t$  é  $\frac{2^t}{2t} \frac{n!}{(n-t)!}$ .

Escolhamos um ciclo arbitrário de comprimento  $2t$ . O número de permutações nas quais este ciclo pode ocorrer não excede o número de maneiras de se permutar os  $n - 2t$  vértices restantes de  $\pi$  mais os  $t$  pares que formam o ciclo. Além disso, cada par pode ser invertido nos dando uma ordem diferente; de onde segue que um ciclo arbitrário de comprimento  $2t$  pode ocorrer em no máximo  $2^t(n-t)!$  permutações.

Agora, seja  $E(c_i(\pi))$  o número esperado de ciclos de comprimento  $i$  em uma decomposição cíclica máxima de  $G(\pi)$  e  $p$  a probabilidade de que um ciclo arbitrário de comprimento  $i$  esteja em uma permutação  $\pi$  qualquer. Então,

$$E(c_i(\pi)) = E(c_{2t}(\pi)) = \sum_{\pi \in S_n} c_{2t}(\pi) p_\pi^t$$

Por observações já feitas,  $p_\pi^t \leq \frac{2^t(n-t)!}{n!}$  para toda permutação  $\pi$  em  $S_n$ .

Logo,

$$E(c_i(\pi)) \leq \frac{2^t(n-t)!}{n!} \sum_{\pi \in S_n} c_{2t}(\pi) \leq \frac{2^t(n-t)!}{n!} \frac{2^t}{2t} \frac{n!}{(n-t)!}$$

De onde,

$$E(c_i(\pi)) \leq \frac{2^{2t}}{2t} = \frac{2^i}{i}$$

E o lema é demonstrado.  $\square$

Em uma palavra estendida  $0 \pi_1 \pi_2 \dots \pi_n n + 1$  existe um total de  $2n$  adjacências ordenadas.

Qualquer tal par de adjacências pode ocorrer em  $(n-1)!$  permutações. Então a probabilidade de que uma adjacência ocorra em uma permutação é  $\frac{(n-1)!}{n!} = \frac{1}{n}$ . Disso segue que o valor esperado de adjacências em uma permutação é  $\frac{2n}{n} = 2$ .

Denotemos por  $a(\pi)$  o número de adjacências em uma permutação  $\pi$ . É de fácil observação que

$$b(\pi) + a(\pi) = n + 1$$

Como o operador esperança é linear e o valor esperado de uma constante é a própria constante, segue diretamente da equação anterior que

$$E(b(\pi)) + E(a(\pi)) = n + 1$$

De onde temos que  $E(b(\pi)) = n - 1$ .

No teorema a seguir temos uma limitação inferior para o valor esperado do diâmetro de reversão de  $S_n$ .

**Teorema 2.5.**  $E(d(n)) \geq \left(1 - \frac{4.5}{\log_2 n}\right) n$ .

*Demonstração.* Pela Equação (2.7) e pela monotonicidade da esperança para todo  $k \leq 2(n+1)$ , vale que

$$E(d(\pi)) \geq \left(1 - \frac{2}{k}\right) E(b(\pi)) - \sum_{i=4}^{k-1} E(c_i(\pi))$$

Pelo Lema 2.2

$$E(d(\pi)) \geq \left(1 - \frac{2}{k}\right) (n-1) - \sum_{i=4}^{k-1} \frac{2^i}{i}$$

Disso, temos que

$$E(d(\pi)) \geq \left(1 - \frac{2}{k}\right) (n-1) - \sum_{i=4}^{k-1} 2^i$$

Remarcamos que os somatórios acima correspondem a soma de termos para  $i$  par, já que  $c_i(\pi) = 0$  se  $i$  é ímpar.

Notemos que  $\sum_{i=4}^{k-1} 2^i \leq 2^k - 1$ .

De fato, como os ciclos alternados são de comprimento par  $\sum_{i=4}^{k-1} 2^i$  é uma progressão geométrica de razão  $q = 2^2$ , primeiro termo  $a_1 = 2^4$  e  $j$ -ésimo termo  $a_j = 2^{k-1}$ .

Vamos analisar os casos em que  $k$  é ímpar e par.

- $k$  é ímpar;

Neste caso,  $k-1$  é par e a soma  $S_j$  dos  $j$  primeiros termos da progressão é

$$S_j = \frac{2^4(2^{2j} - 1)}{2^2 - 1} = \frac{16}{3}(2^{2j} - 1) \quad (2.8)$$

Agora,  $a_j = 2^{k-1}$ . Pelo termo geral da progressão geométrica, temos que  $2^{k-1} = 2^4 \cdot 2^{2(j-1)}$

De onde,  $2j + 2 = k - 1$ . Isto implica  $j = \frac{k-3}{2}$ .

Pela Equação (2.8) vale

$$\begin{aligned} \frac{16}{3}(2^{k-3} - 1) &= \frac{16}{3} \left( \frac{2^k}{8} - 1 \right) \\ &= \frac{2}{3} \cdot 2^k - \frac{16}{3} < \frac{2}{3} \cdot 2^k - \frac{2}{3} \\ &= \frac{2}{3}(2^k - 1) < 2^k - 1 \end{aligned}$$

- $k$  é par.

Se  $k$  é par, então  $k - 1$  é ímpar. Logo,  $a_j = 2^{k-2}$ .

Do termo geral da progressão geométrica segue que  $2^{k-2} = 2^4 \cdot 2^{2(j-1)}$ .

Logo,  $2j + 2 = k - 2$ . De onde temos que  $j = \frac{k-4}{2}$ .

$$\begin{aligned} \frac{16}{3}(2^{k-4} - 1) &= \frac{16}{3} \left( \frac{2^k}{16} - 1 \right) \\ &< \frac{16}{3} \left( \frac{2^k}{16} - \frac{1}{16} \right) = \frac{1}{3}(2^k - 1) \\ &< 2^k - 1 \end{aligned}$$

Logo,

$$E(d(\pi)) \geq n - \frac{2n}{k} - \left(1 - \frac{2}{k}\right) - 2^k + 1$$

De onde,

$$E(d(\pi)) \geq n - \frac{2n}{k} - 2^k \quad (2.9)$$

Escolhamos  $k = \log_2 \frac{n}{\log_2 n}$ , para  $n$  suficientemente grande.

Então,

$$2^k = 2^{\log_2 \frac{n}{\log_2 n}} = 2^{\log_2 n - \log_2 \log_2 n} = \frac{2^{\log_2 n}}{2^{\log_2 \log_2 n}} = \frac{n}{\log_2 n}$$

Agora,  $\frac{n}{\log_2 n} \leq n$ .

Logo,  $\log_2 \frac{n}{\log_2 n} \leq \log_2 n$ . Disso segue que  $k \leq \log_2 n$ .

Portanto,  $2^k \leq \frac{n}{k}$ .

Pela Equação (2.9) temos

$$E(d(\pi)) \geq n - \frac{3n}{k} = \left(1 - \frac{3}{\log_2 \frac{n}{\log_2 n}}\right) n \quad (2.10)$$

Afirmamos que se  $n \geq 2^{16}$  então  $\frac{4.5}{\log_2 n} \geq \frac{3}{\log_2 n - \log_2 \log_2 n}$ .

Façamos  $\log_2 n = t$ . Logo, verificar a afirmação acima equivale a mostrarmos que  $\frac{4.5}{t} \geq \frac{3}{t - \log_2 t}$  para  $t \geq 16$ .

Ou ainda, que para  $t \geq 16$ ,

$$\frac{t}{t - \log_2 t} \leq \frac{3}{2}$$

Observe que  $\frac{16}{16 - \log_2 16} = \frac{4}{3} \leq \frac{3}{2}$ . Logo, a desigualdade é válida para  $t = 16$ .

Além disso, a função  $f(t) = \frac{t}{t - \log_2 t}$  é decrescente para todo  $t > 1$ . Disso, segue que  $\frac{4.5}{t} \geq \frac{3}{t - \log_2 t}$  para  $t \geq 16$ .

Portanto, pela Equação (2.10) temos que, para  $n \geq 2^{16}$

$$E(d(\pi)) \geq \left(1 - \frac{4.5}{\log_2 n}\right) n$$

E o resultado segue. □

## 2.4 Aproximação para a distância de reversão

Iniciaremos agora a discussão sobre algoritmos de aproximação para distância de reversão.

O algoritmo proposto nesta seção, procedimento para permutações sem sinal, será de grande valia quando tratarmos de permutações com sinal.

**Definição 2.8. (Faixa)** Seja  $\pi = \pi_1 \pi_2 \dots \pi_n$  uma permutação de  $S_n$ . Definamos uma *faixa de  $\pi$*  como um intervalo  $[i, j]$  tal que os pares  $(\pi_{i-1}, \pi_i)$  e  $(\pi_j, \pi_{j+1})$  são pontos de quebra e não existe nenhum outro ponto de quebra entre eles.

**Definição 2.9.** Uma faixa é dita *crecente* se  $\pi_i < \pi_j$ . Caso contrário, dizemos que é *decrescente*. Uma faixa de um único elemento pode ser tanto crescente quanto decrescente, exceto as faixas formadas por  $\pi_0$  ou  $\pi_{n+1}$  que diremos ser sempre crescente.



Nota-se que uma reversão pode remover no máximo dois pontos de quebra; consequentemente  $d(\pi) \geq \frac{b(\pi)}{2}$ .

O procedimento e o lema que seguem foram estabelecidos por [15].

**Lema 2.3.** Se  $\pi$  é uma permutação com uma faixa decrescente então:

- i)  $\pi$  admite uma reversão que remove no mínimo um ponto de quebra;
- ii) se toda reversão que remove um ponto de quebra de  $\pi$  deixa uma permutação com faixas decrescentes, então  $\pi$  tem uma reversão que remove dois pontos de quebra.

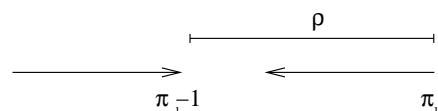
*Demonstração.* i) Consideremos a faixa decrescente de  $\pi$  contendo o menor elemento  $\pi_i$ .

O elemento  $\pi_i - 1$  não pode estar também em uma faixa decrescente, pois isso comprometeria a minimalidade de  $\pi_i$ . Logo,  $\pi_i - 1$  deve estar em uma faixa crescente que fica à direita ou à esquerda de  $\pi_i$ .

Digamos que o elemento  $\pi_i - 1$  esteja numa posição  $l$  da palavra associada à  $\pi$ .

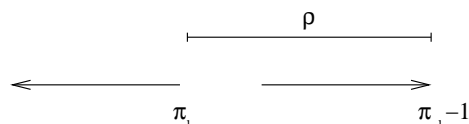
- $\pi_i - 1$  encontra-se em uma faixa crescente à esquerda de  $\pi_i$ ;

Neste caso, basta aplicarmos a reversão que intercambia os elementos do intervalo  $[l + 1, i]$  na palavra associada à permutação para removermos ao menos um ponto de quebra.



- $\pi_i - 1$  encontra-se em uma faixa crescente à direita de  $\pi_i$ .

Neste caso, devemos aplicar a reversão que intercambia o intervalo  $[i + 1, l]$  e dessa forma ao menos um ponto de quebra é removido.



Nas figuras acima, as setas para a direita e esquerda representam faixas crescente e decrescente, respectivamente.

A análise desses casos demonstra o item (i) do lema.

- ii) Como no item anterior consideremos a palavra decrescente de  $\pi$  contendo o menor elemento  $\pi_i$  e digamos que  $\pi_i - 1$  ocupe a  $l$ -ésima posição na palavra.

A faixa crescente contendo  $\pi_i - 1$  deve estar à esquerda da faixa contendo  $\pi_i$ .

De fato, se esta estivesse à direita a reversão que intercambia o intervalo  $[i + 1, l]$  da palavra deixaria uma faixa decrescente. A saber, a faixa em que estão os elementos  $\pi_i$  e  $\pi_i - 1$ . E isto contrariaria a hipótese do lema.

Chamemos de  $\rho_i$  a reversão que inverte os elementos do intervalo  $[l + 1, i]$ . Note-mos que  $\rho_i$  elimina um ponto de quebra, já que deixa contíguos os elementos consecutivos  $\pi_i - 1$  e  $\pi_i$ .

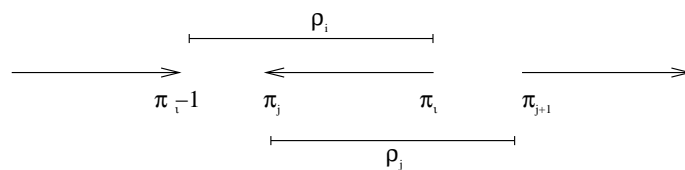
Agora, consideremos a faixa decrescente de  $\pi$  cujo primeiro elemento,  $\pi_j$ , é o maior dentre aqueles que estão em uma faixa decrescente.

O elemento  $\pi_j + 1$  deve estar em uma faixa crescente à direita de uma faixa contendo  $\pi_j$ .

Com razão, digamos que  $\pi_j + 1$  ocupe a posição  $k$  na palavra. Se  $\pi_j + 1$  estivesse em uma faixa decrescente a maximalidade de  $\pi_j$  estaria comprometida. Além disso, se a faixa crescente que contém  $\pi_j + 1$  estivesse à esquerda da faixa que contém  $\pi_j$  a reversão que inverte o intervalo  $[k, j - 1]$  deixaria  $\pi$  com uma faixa decrescente.

À reversão que intercambia os elementos  $[j, k - 1]$  na palavra daremos o nome de  $\rho_j$ .

Tanto as reversões  $\rho_i$  quanto  $\rho_j$  agem sobre o intervalo  $[j, i]$  da palavra. Logo, os intervalos que são revertidos por  $\rho_i$  e  $\rho_j$  se justapõem.



Nossa estratégia será mostrar que esses intervalos não somente se justapõem, mas são os mesmos. Tendo isso em vista, já que a reversão  $\rho_i$  remove um ponto de quebra e a reversão  $\rho_j$  remove outro ponto de quebra, e esses são distintos, a reversão  $\rho = \rho_i = \rho_j$  irá remover dois pontos de quebra simultaneamente.

Suponhamos que o intervalo  $I = [l + 1, i] \setminus [j, k - 1]$  seja não vazio. Se ele contém uma faixa decrescente então a reversão  $\rho_j$ , que não atua em  $I$  deixa uma faixa decrescente. Se ele contém uma faixa crescente, a reversão  $\rho_i$  atua sobre  $I$  e deixa na palavra uma faixa decrescente. Em ambos os casos, temos a hipótese contrariada.

Uma análise similar pode ser feita para o intervalo  $J = [j, k - 1] \setminus [l + 1, i]$ .

Disso segue que  $I$  e  $J$  são vazios. De onde,  $l + 1 = j$  e  $k - 1 = i$ .

E isso conclui a demonstração do lema. □

Descrevamos agora o procedimento.

**Procedimento**  $KS(\pi)$

**while**  $\pi$  contém um ponto de quebra faça

$\rho = Greedy(\pi)$

$\pi = \pi.\rho$

**endwhile**

**Procedimento**  $Greedy(\pi)$

**begin**

Retorne uma reversão que remova o maior número de pontos de quebra de  $\pi$  e que deixe na permutação ao menos uma faixa decrescente.

**end**

O Lema 2.3 implica que no Procedimento  $KS(\pi)$ , cada execução do algoritmo termina com uma reversão que remove dois pontos de quebra (exceto, possivelmente, na primeira execução). Logo, cada reversão, possivelmente feita para deixar  $\pi$  com uma faixa decrescente, pode ser amortizada por uma reversão que remove dois pontos de quebra. Sendo assim, em média, necessitamos de uma reversão para remover um ponto de quebra.

Levando em consideração o limite superior de  $b(\pi)$  e o limite inferior de  $\frac{b(\pi)}{2}$  reversões para se ordenar uma permutação, temos que o algoritmo  $KS(\pi)$  tem um raio de aproximação de 2.

O que irá se fazer é fornecer algoritmos que tenham um melhor desempenho.

Pelo Teorema 2.2 temos que para qualquer permutação  $\pi$  vale que  $d(\pi) \geq b(\pi) - c(\pi)$ . Consideremos ciclos de comprimento quatro em uma decomposição máxima de  $G(\pi)$ .

Então,

$$d(\pi) \geq b(\pi) - c_4(\pi) - (c(\pi) - c_4(\pi)) \quad (2.11)$$

Agora,  $c(\pi) - c_4(\pi)$  corresponde ao número de ciclos, em uma decomposição máxima, de comprimento no mínimo seis. Este número não excede a quantidade de ciclos de comprimento exatamente seis. Logo,  $c(\pi) - c_4(\pi) \leq \frac{2b(\pi) - 4c_4(\pi)}{6}$ .

Segue da Equação (2.11) que

$$d(\pi) \geq b(\pi) - c_4(\pi) - \frac{b(\pi) - 2c_4(\pi)}{3}$$

De onde,

$$d(\pi) \geq \frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi)$$

Nas seções seguintes serão fornecidos algoritmos que ordenam uma permutação  $\pi$  em no máximo  $b(\pi) - \epsilon c_4(\pi)$  passos, para algum  $\epsilon > 0$ . Conseqüentemente, o raio de aproximação desses algoritmos será de

$$\max_{0 \leq c_4(\pi) \leq \frac{b(\pi)}{2}} \left( \frac{b(\pi) - \epsilon c_4(\pi)}{\frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi)} \right)$$

Notemos que, se  $\epsilon > \frac{1}{2}$  temos que

$$\frac{b(\pi) - \epsilon c_4(\pi)}{\frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi)} \leq \frac{b(\pi) - \frac{1}{2}c_4(\pi)}{\frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi)} = \frac{3}{2}$$

Agora, se  $\epsilon \leq \frac{1}{2}$  então

$$1 \geq 2\epsilon$$

$$1 \geq 2\epsilon \left( \frac{2c_4(\pi) - b(\pi)}{2c_4(\pi) - b(\pi)} \right)$$

Para  $c_4(\pi) < \frac{b(\pi)}{2}$ , temos  $2c_4(\pi) - b(\pi) < 0$ .

Disso, segue que

$$2c_4(\pi) - b(\pi) \leq 2\epsilon(2c_4(\pi) - b(\pi))$$

$$3(b(\pi) - \epsilon c_4(\pi)) \leq 4b(\pi) - 2c_4(\pi) - 2b(\pi)\epsilon + c_4(\pi)\epsilon$$

De onde,

$$b(\pi) - \epsilon c_4(\pi) \leq (2 - \epsilon) \left( \frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi) \right)$$

$$\frac{b(\pi) - \epsilon c_4(\pi)}{\frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi)} \leq 2 - \epsilon$$

Portanto,

$$\max_{0 \leq c_4(\pi) \leq \frac{b(\pi)}{2}} \left( \frac{b(\pi) - \epsilon c_4(\pi)}{\frac{2}{3}b(\pi) - \frac{1}{3}c_4(\pi)} \right) \leq \begin{cases} 2 - \epsilon, & \epsilon \leq \frac{1}{2} \\ \frac{3}{2}, & \text{caso contrário.} \end{cases} \quad (2.12)$$

No capítulo que segue, irá se fornecer algoritmos que garantem soluções aproximadas para o problema da distância de reversão. Primeiramente, trataremos de permutações

---

com sinais e dar-se-á um algoritmo, que ordena estas permutações, com raio de aproximação de  $\frac{3}{2}$ . Depois, usar-se-á esse algoritmo para obter um procedimento que ordena permutações sem sinais com um raio de aproximação de  $\frac{9}{5}$ . Este último procedimento servirá de base para a construção de outro de raio de aproximação  $\frac{7}{4}$  para a ordenação de permutações sem sinal.

---

## Algoritmos para o problema de rearranjo via reversões

---

Neste capítulo iremos tratar dos algoritmos que aproximam a solução do problema de rearranjos de genomas via reversões. Para isso, consideraremos, inicialmente, permutações com sinal; irá se obter um algoritmo de raio de aproximação  $\frac{3}{2}$  para esse tipo de permutação. Usar-se-á esse algoritmo para conceber outros que fornecem soluções aproximadas para reversões sem sinal.

### 3.1 Permutações com sinal

Dado parte de um genoma, caso conheçamos a direção em que é feita a leitura dos genes, podemos representar esse fragmento por uma sequência numérica com sinais que representa a orientação em que são lidos os genes. Essa modelagem caracteriza o problema de rearranjo de genomas para permutações com sinal.

No entanto, é possível transformar permutações com sinal em permutações sem sinal. E dessa forma, aplicar os resultados que vimos anteriormente.

**Definição 3.1. (Transformação de permutação com sinal para permutação sem sinal)** Seja  $\pi$  uma permutação com sinal de  $n$  elementos. A *transformação* de  $\pi$  em uma permutação sem sinal  $\pi' \in S_{2n}$  é feita substituindo, simultaneamente, os elementos  $+i$  por  $2i - 1, 2i$  e os elementos  $-i$  por  $2i, 2i - 1$ .

**Observação 3.1.** A identidade com sinal,  $id_+ = (+1 + 2 + 3 \dots + n)$ , é levada na identidade sem sinal  $(1 2 3 \dots 2n)$  de  $S_{2n}$ .

Considerando a transformação de uma permutação  $\pi$  com sinal para uma permutação sem sinal  $\pi'$ , temos que o efeito de uma reversão sobre  $\pi$  pode ser representado por uma reversão em  $\pi'$ . Logo, qualquer limite inferior sobre  $\pi'$  é também limite inferior sobre  $\pi$ . E em particular, o Teorema 2.2 é válido.

Para conseguirmos um limitante superior, devemos realizar reversões a fim de eliminar o máximo de pontos de quebra. Essas reversões em permutações com sinal também podem ser transportadas para permutações sem sinal.

Sendo assim, o problema de ordenação por reversões para permutações com sinal pode ser resolvido por análise de permutações sem sinal através da transformação sugerida.

É de fácil observação que os vértices do grafo de pontos de quebra de uma permutação com sinal possuem grau dois (de cada vértice, rotulado por um par de números inteiros, sai apenas uma aresta preta e uma aresta cinza). Este fato garante que a decomposição em ciclos disjuntos do grafo de pontos de quebra de permutações com sinal é única. É esse o ponto chave para se obter um algoritmo exato para a solução do problema de rearranjo por reversões com sinal, como demonstrado por [12].

Nesta seção, no que se segue, sempre que citarmos grafo de pontos de quebra ou faixas de permutações com sinal, estaremos nos referindo a grafo de pontos de quebra ou faixas da permutação sem sinal associada.

**Definição 3.2. (Reversão sobre um ciclo)** Seja  $\pi$  uma permutação tal que os pares  $(\pi_{i-1}, \pi_i)$  e  $(\pi_j, \pi_{j+1})$  são pontos de quebra. Dizemos que  $\rho(i, j)$  é uma *reversão sobre um ciclo* se os pontos de quebra  $(\pi_{i-1}, \pi_i)$  e  $(\pi_j, \pi_{j+1})$  pertencem ao mesmo ciclo em  $G(\pi)$ .

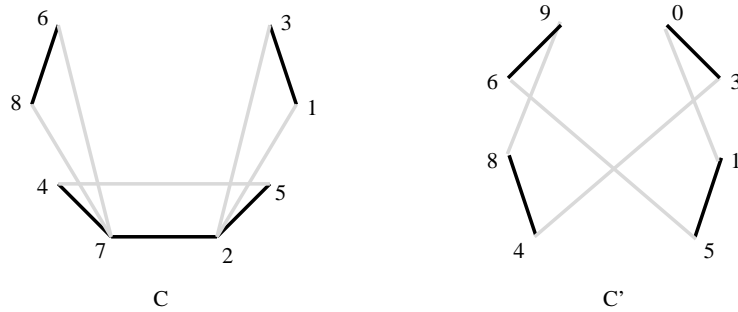
**Definição 3.3. (Ciclo orientado)** Um ciclo é dito *orientado* se existe uma reversão sobre ele que remove ao menos um ponto de quebra.

**Definição 3.4. (Ciclos cruzados)** Seja  $\pi$  uma permutação de  $n$  elementos e  $G(\pi)$  o grafo de pontos de quebra de  $\pi$ . Consideremos uma decomposição de  $G(\pi)$  em ciclos. Dizemos que, nessa decomposição, dois ciclos  $C_1$  e  $C_2$  se *cruzam* se existem arestas pretas  $(\pi_i, \pi_{i+1}), (\pi_j, \pi_{j+1})$  em  $C_1$  ( $C_2$ ),  $i < j$ , e uma aresta preta  $(\pi_k, \pi_{k+1})$  em  $C_2$  ( $C_1$ ) tal que  $i < k < j$ . Isto é, existem arestas pretas em  $C_1$  e em  $C_2$  que se intercalam no grafo  $G(\pi)$ .

**Exemplo 3.1.** No grafo da permutação  $\gamma_8 = (7 \ 5 \ 3 \ 1 \ 2 \ 4 \ 6 \ 8)$  destaquemos dois ciclos  $C$  e  $C'$ , mostrados na figura abaixo.

Observe que  $\rho = (3 \ 5)$  é uma reversão sobre o ciclo  $C$ , já que os pares  $(5, 2)$  e  $(2, 7)$  são pontos de quebra que pertencem a  $C$ .

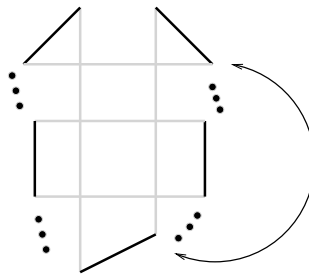
O ciclo  $C$  é orientado já que  $\rho$  remove um ponto de quebra. A saber, o ponto de quebra representado pela aresta com par de rótulos  $(7, 4)$ .



Note que os ciclos  $C$  e  $C'$  se cruzam, já que, por exemplo, o ponto de quebra  $(8, 4)$  em  $C'$  intercala com os pontos de quebra dados pelas arestas  $(6, 8)$  e  $(4, 7)$  em  $C$ .

Uma reversão pode transformar um ciclo não orientado em um ciclo orientado.

**Lema 3.1.** Qualquer 4-ciclo  $C$  que não é orientado tem um ciclo  $C'$  que o cruza. Além disso, existe uma reversão sobre  $C'$  que orientará  $C$ .



*Demonstração.* Um 4-ciclo  $C$  não orientado ocorre na palavra como  $\dots i j \dots j' i' \dots$  com  $i \sim i'$  e  $j \sim j'$ . Como a aresta representada pelos rótulos  $(j, j')$  é cinza, o conjunto de elementos no grafo de pontos de quebra entre os vértices de rótulos  $j$  e  $j'$  é não-vazio. Chamemos de  $S$  esse conjunto. Considere o maior e o menor, não necessariamente distintos, elementos em  $S$ . Para ao menos um desses elementos, digamos  $k$ , deve existir  $k' \sim k$  tal que  $k \notin S$ ,  $k' \neq j$  e  $k' \neq j'$ .

De fato, sem perda de generalidade, suponhamos que  $k$  seja máximo. Como estamos considerando permutações com sinal o vértice de rótulo  $k$  é transformado em um vértice de rótulos  $k_1 k_2$  com  $k_1 \sim k_2$ , tal que,  $k_1$  ou  $k_2$  é máximo entre os elementos que estão em  $S$  na permutação sem sinal. Como os vértices do grafo de pontos de quebra de uma permutação com sinal têm grau no máximo dois, e dos vértices  $j$  e  $j'$  já partem uma aresta preta e uma aresta cinza segue que  $k' \neq j$  e  $k' \neq j'$ . Conseqüentemente existem pontos de quebra  $(k, l)$  e  $(k', m)$  que são intercalados com os pontos de quebra  $(i, j)$  e  $(j', i')$ . O ciclo  $C'$  que contém esses pontos de quebra cruza o 4-ciclo  $C$ .

Observe que a reversão feita sobre o ciclo  $C'$  ao longo dos pontos de quebra  $(k, l)$  e  $(k', m)$  reverte os vértices de rótulos  $j$  e  $j'$  e tornará o ciclo  $C$  orientado.  $\square$



O Lema (3.1) motiva o algoritmo *SignedSort*, presente em [3], que ordena permutações com sinal.

**Procedimento**  $SignedSort(\pi)$

1. **enquanto**  $\pi$  contém um ponto de quebra faça
2.     **se**  $\pi$  não tem faixas decrescentes
3.         **se** em  $G(\pi)$  temos um 4-ciclo  $C$
4.             encontre um ciclo  $C'$  que cruza  $C$
5.             faça uma reversão sobre  $C'$  de maneira que o 4-ciclo  $C$  torne-se orientado
6.             faça uma reversão que remova dois pontos de quebra do 4-ciclo  $C$
7.     **caso contrário**
8.         faça uma reversão sobre um ciclo arbitrário
9.     **caso contrário**
10.          $\rho = Greedy(\pi)$
11.          $\pi = \pi\rho$
12. **finalizar enquanto**

**Lema 3.2.** Depois do passo (6) no procedimento *SignedSort* alguma faixa decrescente permanece na permutação.

*Demonstração.* Observemos que o passo (4) é executado quando todas as faixas são crescentes, já que este está dentro do laço que se inicia no passo (2). A reversão efetuada no passo (5), que age sobre uma faixa crescente, cria uma faixa decrescente. Para remover essa faixa decrescente, o passo (6) deveria ser executado sobre o mesmo intervalo em que foi realizada a reversão do passo (5). Isto não ocorre, pois o passo (6) remove dois pontos de quebra da permutação; se ele agisse sobre a faixa decrescente anterior ele apenas a transformaria na faixa crescente original. Logo, garantimos a existência de alguma faixa decrescente depois do passo (6).  $\square$

**Lema 3.3.** Se existe um 4-ciclo em  $G(\pi)$  no início de alguma execução do Procedimento *SignedSort* (exceto, possivelmente, a primeira), então existem no mínimo duas reversões que removem dois pontos de quebra nessa execução.

*Demonstração.* Considerando que exista um 4-ciclo em  $G(\pi)$  no início de alguma execução do procedimento, essa começa com uma reversão que orienta o 4-ciclo seguida de uma reversão que remove dois pontos de quebra. Pelo Lema (3.2) alguma faixa decrescente permanece depois dessa última reversão. Agora, se a permutação tem alguma faixa decrescente podemos aplicar *Greedy*. Pelo Lema 2.3 podemos efetuar uma reversão que remove dois pontos de quebra.

E isso conclui a demonstração.  $\square$

**Teorema 3.1.** *O Procedimento SignedSort ordena a permutação com sinal  $\pi$  em no máximo  $b(\pi) - \frac{1}{2}c_4(\pi)$  reversões.*

*Demonstração.* O número de 4-ciclos em um grafo de pontos de quebra decresce somente se aplicarmos uma reversão em que dois pontos de quebra são eliminados. Agora, pelo lema anterior para duas reversões que removam dois pontos de quebra, possivelmente será necessário fazer uma reversão que oriente o 4-ciclo. Isto é, no máximo metade de duas reversões que removam dois pontos de quebra devem ser amortizadas por uma reversão que oriente o 4-ciclo. Consequentemente, uma permutação  $\pi$  com sinal pode ser ordenada em no máximo  $b(\pi) - \frac{1}{2}c_4(\pi)$  reversões.  $\square$

**Corolário 3.1.** O algoritmo *SignedSort* nos fornece um raio de aproximação de  $\frac{3}{2}$ .

*Demonstração.* Pelo Teorema anterior uma permutação com sinal é ordenada em  $b(\pi) - \frac{1}{2}c_4(\pi)$  reversões. A Equação (2.12) irá garantir um raio de aproximação de  $\frac{3}{2}$  para o algoritmo.  $\square$

## 3.2 Um algoritmo de aproximação para ordenação por reversões

Nesta seção usaremos o algoritmo *SignedSort* para propormos outro procedimento que ordena permutações sem sinais com aproximação de  $\frac{9}{5}$ .

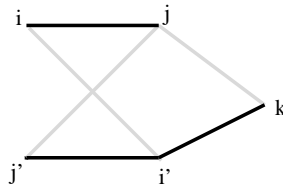
Considere o conjunto de 4-ciclos em um grafo de pontos de quebra de uma permutação  $\pi$ . Quaisquer dois 4-ciclos podem compartilhar no máximo duas arestas.

**Definição 3.5. (Ciclos 2-sobrepostos)** Dados dois 4-ciclos em um grafo de pontos de quebra de uma permutação  $\pi$ , diremos que eles são *2-sobrepostos* se compartilham duas arestas.

**Lema 3.4.** Se dois 4-ciclos de  $G(\pi)$  são 2-sobrepostos, então um deles é orientado.

*Demonstração.* Primeiramente, observemos que dois 4-ciclos que são 2-sobrepostos não podem compartilhar arestas de mesma cor; pois isso implicaria que eles seriam o mesmo ciclo.

Observe que a única maneira de dois 4-ciclos serem 2-sobrepostos é mostrada na figura abaixo.



O ciclo formado pelos vértices de rótulos  $i, j, i'$  e  $j'$  é orientado.  $\square$

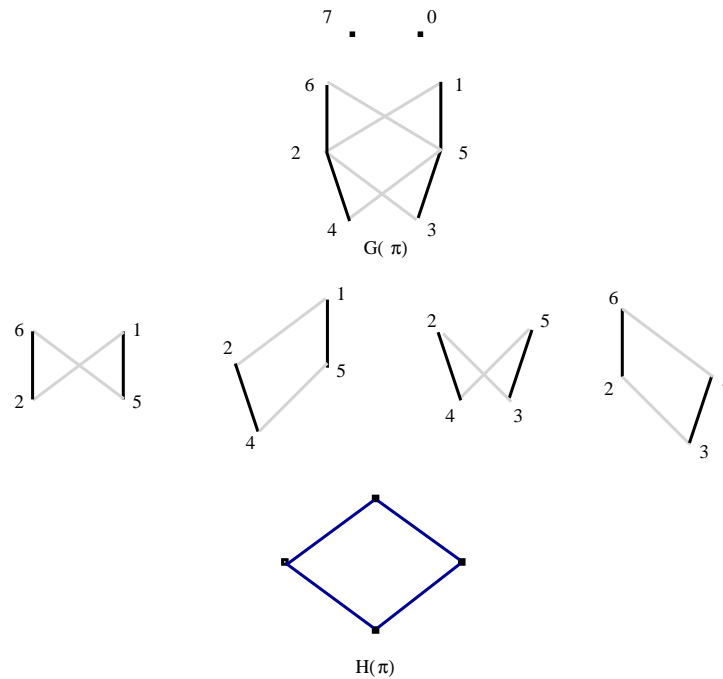


Figura 3.1: Grafo de pontos de quebra e de 4-ciclos da permutação  $\pi = (2\ 5)$

O lema anterior nos garante que se dois 4-ciclos de  $G(\pi)$  são 2-sobrepostos então é possível realizar uma reversão que remova dois pontos de quebra.

**Definição 3.6. (Grafo de 4-ciclos)** O *grafo de 4-ciclos*, denotemos por  $H(\pi)$ , de uma permutação  $\pi$  é definido como um grafo em que cada nó corresponde a um 4-ciclo em  $G(\pi)$ . Além disso, dois nós estão conectados por uma aresta se os correspondentes 4-ciclos compartilham ao menos uma aresta em  $G(\pi)$ .

**Definição 3.7. (Conjunto máximo independente)** Seja  $G(V, E)$  um grafo. Um subconjunto  $V'$  de  $V$  é dito independente se quaisquer dois vértices em  $V'$  não são adjacentes em  $G$ . Um conjunto  $V'$  independente é *máximo* se  $G$  não possui nenhum conjunto independente  $V''$  com  $|V''| > |V'|$ .

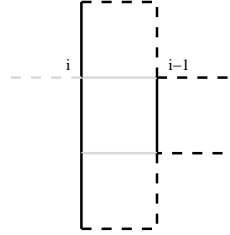
**Observação 3.2.** Um conjunto independente em  $H(\pi)$  corresponde a um conjunto de 4-ciclos em  $G(\pi)$  de arestas disjuntas. Isto é, corresponde a um conjunto de vértices em  $H(\pi)$  que não estão diretamente conectados por uma aresta.

**Exemplo 3.2.** Seja  $\pi = (2\ 5)$ . Na Figura 3.1 temos os grafos de pontos de quebra e de 4-ciclos,  $G(\pi)$  e  $H(\pi)$ , respectivamente, dessa permutação.

Denotemos por  $i(\pi)$  o comprimento de um conjunto máximo independente em um grafo de 4-ciclos  $H(\pi)$ .

**Definição 3.8. (Grafo fortemente  $d$ -limitado)** Dizemos que um grafo é *fortemente  $d$ -limitado* se o grau de qualquer um de seus vértices é limitado por  $d$  e o grau de ao menos um de seus vértices é estritamente menor que  $d$ .

**Lema 3.5.** Se  $G(\pi)$  não contém ciclos 2- sobrepostos, então o grafo de 4-ciclos  $H(\pi)$  é fortemente 4-limitado.



*Demonstração.* Os ciclos em  $G(\pi)$  compartilham, no máximo, uma aresta, já que  $G(\pi)$  não tem ciclos 2-sobrepostos. Logo, um 4-ciclo de  $G(\pi)$  pode ter no máximo quatro outros ciclos alternados compartilhando uma aresta com ele.

Considere um 4-ciclo  $C$  que contém o vértice com rótulo  $i$  máximo dentre os outros 4-ciclos em uma componente conexa de  $H(\pi)$ . Pela maximalidade de  $i$ , o vértice com rótulo  $i + 1$  não pode pertencer a qualquer 4-ciclo nesta componente. Disso segue que  $C$  pode compartilhar uma aresta com no máximo três 4 – *ciclos* dessa componente. E portanto,  $H(\pi)$  é fortemente 4-limitado.  $\square$

**Lema 3.6.** Em um grafo  $G(V, E)$  fortemente  $d$ -limitado, um conjunto independente de comprimento no mínimo  $\frac{|V|}{d}$  pode ser computado em tempo linear.

*Demonstração.* No grafo  $G(V, E)$  escolha um vértice de grau menor que  $d$  e o adicione ao conjunto independente. Remova todos os vértices que estão ligados a esse. Repita esse processo até que não tenhamos mais vértices com grau menor que  $d$ . Como não mais que  $d$  vértices são removidos em cada passo do processo, já que os vértices escolhidos têm grau menor que  $d$ , o comprimento do conjunto independente é no mínimo  $\frac{|V|}{d}$ .  $\square$

**Definição 3.9. (Cobertura mínima)** Em um grafo  $G(V, E)$  dizemos que um conjunto  $K$  é uma *cobertura de  $G(V, E)$*  se qualquer aresta  $e = (u, v) \in E$  tem ao menos um vértice,  $u$  ou  $v$ , em  $K$ . Uma cobertura  $K$  é dita *mínima* se  $G(V, E)$  não tem uma cobertura  $K'$  tal que  $|K'| < |K|$ .

**Proposição 3.1.** Um conjunto  $V' \subset V$  é um conjunto independente em um grafo  $G(V, E)$  se e somente se  $V \setminus V'$  é uma cobertura de  $G(V, E)$ .

*Demonstração.* Seja  $V'$  um conjunto independente de  $G(V, E)$ . Consideremos uma aresta  $e = (u, v) \in E$ . Se  $u \in V'$  então  $v \notin V'$  o que implica que  $v \in V \setminus V'$ . Por outro lado, se  $v \in V'$  então  $u \notin V'$  o que nos dá  $u \in V \setminus V'$ .

Logo, qualquer aresta de  $G(V, E)$  possui ao menos um vértice no conjunto  $V \setminus V'$ . Disso segue que  $V \setminus V'$  é uma cobertura de  $G(V, E)$ .

Reciprocamente, seja  $V \setminus V'$  cobertura de  $G(V, E)$ . Suponhamos, por contradição, que  $V'$  não seja um conjunto independente. Então existem vértices  $u$  e  $v$  em  $V'$  tais que  $e = (u, v) \in E$ . Isso contraria o fato de  $V \setminus V'$  ser cobertura de  $G(V, E)$ .

Isso finaliza a demonstração.  $\square$

**Corolário 3.2.** Um conjunto  $V'$  é máximo independente de  $G(V, E)$  se e somente se  $V \setminus V'$  é cobertura mínima de  $G(V, E)$ .

*Demonstração.* Seja  $V'$  um conjunto máximo de  $G(V, E)$ . Pela proposição anterior,  $V \setminus V'$  é uma cobertura de  $G(V, E)$ . Suponhamos, por contradição, que exista uma cobertura mínima  $K$  de  $G(V, E)$  tal que  $K \not\equiv V \setminus V'$ . Sendo assim, vale que

$$|V| - |V'| = |V \setminus V'| > |K| \quad (3.1)$$

Como  $K$  é cobertura, temos que  $V \setminus K$  é conjunto independente de  $G(V, E)$ . Além disso,

$$|V| - |K| = |V \setminus K|$$

Logo pela Equação (3.1),  $|V \setminus K| > |V|$ ; isso contraria o fato de  $V'$  ser um conjunto máximo independente.

A recíproca é demonstrada com um argumento análogo.  $\square$

**Lema 3.7.** Em um grafo  $G(V, E)$  fortemente  $d$ -limitado, uma aproximação de  $\frac{2}{(d+1)}$  a um conjunto máximo independente é computada em tempo linear.

*Demonstração.* Para algum inteiro positivo  $a$ , seja  $an$  o comprimento do conjunto máximo independente em  $G(V, E)$ , onde  $|V| = n$ . Consequentemente, a cobertura mínima de vértices para  $G(V, E)$  tem comprimento  $(1 - a)n$ .

Em [4] é dito que em [9] podemos encontrar uma cobertura de vértices  $V'$  de comprimento no máximo  $\min(2(1 - a)n, n)$  em tempo linear. Sendo assim,  $I_1 = V \setminus V'$  é um conjunto independente cujo comprimento é no mínimo  $\max((2a - 1)n, 0)$ .

Podemos obter ainda, pelo Lema 3.6 um conjunto independente  $I_2$  de comprimento no mínimo  $\frac{n}{d}$ .

Selecionamos o maior desses conjuntos.

Devemos então, analisar os casos:

- $\max((2a - 1)n, 0) = 0$ ;

Neste caso, o conjunto máximo independente a ser escolhido deve ser  $I_2$ . Sua aproximação é dada por  $\frac{n}{an} = \frac{1}{ad}$

Como  $\max((2a - 1)n, 0) = 0$  então,  $2a - 1 \leq 0$ . De onde,  $a \leq \frac{1}{2}$ .

Dessa forma,  $\frac{1}{ad} \geq \frac{2}{d} > \frac{2}{d+1}$ .

- $\max((2a - 1)n, 0) = (2a - 1)n$ .

Neste caso, devemos escolher entre o maior dos conjuntos  $I_1$  e  $I_2$ . A aproximação será dada por  $\max(\frac{1}{da}, \frac{2a-1}{a})$ .

i)  $\max(\frac{1}{da}, \frac{2a-1}{a}) = \frac{1}{da}$ ;

Se isso ocorre então

$$\frac{1}{da} \geq \frac{2a-1}{a}$$

De onde,  $a \leq \frac{d+1}{2d}$ .

Sendo assim,

$$\frac{1}{da} \geq \frac{2d}{d+1} \frac{1}{d} = \frac{2}{d+1}$$

ii)  $\max(\frac{1}{da}, \frac{2a-1}{a}) = \frac{2a-1}{a}$ ;

Disso, segue que

$$\frac{2a-1}{a} \geq \frac{1}{da}$$

Logo,  $a \geq \frac{1+d}{2d}$ .

Assim,

$$\frac{2a-1}{a} = 2 - \frac{1}{a} \geq 2 - \frac{2d}{d+1} = \frac{2}{d+1}$$

iii)  $\frac{1}{da} = \frac{2a-1}{a}$ .

Neste caso,  $a = \frac{d+1}{2d}$  e o desempenho é exatamente  $\frac{2}{d+1}$ .

□

Os lemas anteriores levam à construção de um algoritmo que mostraremos ter um raio de aproximação de  $\frac{9}{5}$ . A seguir, está a descrição do procedimento presente em [4].

**Procedimento *ReversalSort*( $\pi$ )**

1. Comece realizando, na permutação  $\pi$ , reversões que removam dois pontos de quebra até que não restem em  $G(\pi)$  ciclos 2-sobrepostos. Denote por  $\sigma$  a permutação resultante.
2. Use a aproximação a um conjunto independente em  $H(\sigma)$  para encontrar no grafo de pontos de quebra  $G(\sigma)$  um conjunto de 4-ciclos que não compartilham arestas de comprimento no mínimo  $\frac{2}{5}c_4(\sigma)$ . Encontre uma decomposição cíclica arbitrária das arestas que restaram.
3. Decomponha os vértices de grau 4 em  $G(\sigma)$  de acordo com a decomposição cíclica encontrada no passo (2). Em termos de faixas, substitua elementos únicos por faixas, orientadas apropriadamente, resultando em uma permutação com sinal  $\sigma'$ .
4. Use  $SignedSort(\sigma')$  para ordenar  $\sigma'$ . Essa ordenação irá nos fornecer uma ordenação para  $\sigma$ .

**Teorema 3.2.** *O algoritmo  $ReversalSort$  fornece um raio de aproximação de  $\frac{9}{5}$ .*

*Demonstração.* Nossa estratégia será mostrar que  $d(\pi)$  é no máximo  $b(\pi) - \frac{1}{5}c_4(\pi)$ . Disso e da Equação (2.12) seguirá o resultado.

No passo (4) podemos ordenar a permutação  $\sigma'$  pelo Procedimento  $SignedSort(\sigma')$ . Logo, um limite superior para o número de reversões em  $SignedSort(\sigma')$  será um limite superior para o número de reversões no passo (4) de  $ReversalSort$ . Suponhamos que o número de 4-ciclos encontrados no passo (2) seja de  $c'_4(\sigma)$ . Pelo Teorema 3.1 segue que  $d(\sigma) \leq b(\sigma) - \frac{1}{2}c'_4(\sigma)$  para o número de reversões no passo (4) do Procedimento  $ReversalSort$ .

Agora, consideremos  $\rho_1, \rho_2, \dots, \rho_x$  a sequência de reversões que transforma  $\pi$  em  $\sigma$  no passo (1). Logo, o Algoritmo  $ReversalSort$  não requer mais que  $x + b(\sigma) - \frac{1}{2}c'_4(\sigma)$  para ordenar a permutação  $\pi$ . Mas, observe que, se são necessárias  $x$  reversões que removem dois pontos de quebra para se transformar  $\pi$  em  $\sigma$  temos que  $b(\pi) - b(\sigma) = 2x$ . De onde temos que  $d(\pi) \leq b(\pi) - x - \frac{1}{2}c'_4(\sigma)$ .

Os Lemas 3.5 e 3.7 nos garantem que  $c'_4(\sigma) \geq \frac{2}{5}i(\sigma)$ , onde  $i(\sigma)$  é o comprimento do conjunto máximo independente em  $H(\sigma)$ .

Já que existem no máximo 4-ciclos que não compartilham arestas com um outro 4-ciclo, cada reversão feita no passo (1) destrói no máximo  $i(\pi)$  vértices de um conjunto máximo independente em  $H(\pi)$ . Disso segue que  $i(\pi) - i(\sigma) \leq 4x$ .

Consequentemente,

$$d(\pi) \leq b(\pi) - x - \frac{1}{5}i(\sigma)$$

$$d(\pi) \leq b(\pi) - x - \frac{1}{5}(i(\pi) - 4x)$$

Recordemos que  $c_4(\pi)$  corresponde ao número de 4-ciclos em  $G(\pi)$  em uma decomposição cíclica máxima, ou seja,  $c_4(\pi)$  corresponde ao número de 4-ciclos que não compartilham arestas. Logo,  $c_4(\pi) \leq i(\pi)$ .

De onde,

$$d(\pi) \leq b(\pi) - \frac{1}{5}x - \frac{1}{5}c_4(\pi)$$

$$d(\pi) \leq b(\pi) - \frac{1}{5}c_4(\pi)$$

E isso demonstra o teorema. □

### 3.3 Uma melhor aproximação para ordenação por reversões

Nesta seção, nosso objetivo será modificar o Algoritmo *ReversalSort* a fim de melhorar seu raio de aproximação. A partir do Procedimento  $Transform(\pi)$ , que será descrito a seguir, transformaremos a permutação  $\pi$  dada em uma permutação  $\sigma$ , cujo grafo de 4- ciclos,  $H(\sigma)$ , terá uma propriedade especial. Isso possibilitará a construção de um procedimento, *ImprovedSort*, que garantirá um número máximo de  $b(\pi) - \frac{1}{4}c_4(\pi)$  reversões para se ordenar a permutação  $\pi$ .

Dadas permutações  $\pi$  e  $\nu$ , como já argumentamos, o problema de transformar  $\pi$  em  $\nu$  equivale ao problema de transformar  $\nu^{-1}\pi$  na permutação identidade. Denotemos por  $G(\nu^{-1}\pi)$ ,  $b(\nu^{-1}\pi) \equiv b(\pi, \nu)$  e  $i(\nu^{-1}\pi)$  o grafo de pontos de quebra, o número de pontos de quebra e o número de ciclos máximos independentes em  $H(\nu^{-1}\pi)$  da permutação  $\nu^{-1}\pi$ .

Observe que, se uma permutação  $\pi$  tem  $k$  pontos de quebra então  $k$  elementos de  $\pi_{str}$  estão desordenados se tomamos  $id_{str}$  como parâmetro. Se o elemento de rótulo  $i$  ocupa a posição  $j$  então o elemento  $j$  ocupa alguma posição  $k \neq j$ . Logo, a permutação  $\pi$  pode ser escrita como produto de ciclos disjuntos. Sendo assim, a permutação  $\pi^{-1}$  terá a mesma estrutura cíclica de  $\pi$  mas com a ordem dos elementos invertidas, e consequentemente também terá  $k$  pontos de quebra.

Pela observação feita acima  $b(\pi, \nu) = b(\nu, \pi)$ .

Feitas essas observações estamos aptos a começar o plano traçado no início da seção. Descrevamos o procedimento  $Transform(\pi)$ .

**Procedimento**  $Transform(\pi)$

**início**

$$\bar{\pi} = \pi$$

$$\bar{\nu} = id$$

**enquanto**  $\bar{\nu}^{-1}\bar{\pi}$  ou  $\bar{\pi}^{-1}\bar{\nu}$  tem uma reversão  $\rho$  que remove dois pontos de quebra



**if**  $\rho$  é uma reversão que remove dois pontos de quebra em  $\bar{\nu}^{-1}\bar{\pi}$   
 $\bar{\pi} = \bar{\pi}\rho$   
**caso contrário**  $\bar{\nu} = \bar{\nu}\rho$   
**finalizar enquanto**  
**retorne**  $\bar{\nu}^{-1}\bar{\pi}$   
**fim**

**Definição 3.10. (Grafo bipartido)** Um grafo  $G(V, E)$  é dito *bipartido* se seus vértices podem ser divididos em dois conjuntos disjuntos nos quais não há arestas entre os vértices de um mesmo conjunto.

Consideremos a permutação  $\sigma = \bar{\nu}^{-1}\bar{\pi}$ , dada pelo Procedimento *Transform*.

**Lema 3.8.** O grafo  $H(\sigma)$  é bipartido.

*Demonstração.* Considere um 4-ciclo em  $G(\sigma)$  formado pelos vértices cujos rótulos são  $\sigma_i, \sigma_{i+1}, \sigma_j$  e  $\sigma_{j+1}$ , como  $i + 1 < j$ . A permutação  $\sigma$  é obtida a partir do Procedimento *Transform*( $\pi$ ); pela análise do procedimento, não é possível realizar reversão que remova dois pontos de quebra em  $\sigma$ . Logo, qualquer 4-ciclo em  $G(\sigma)$  é não orientado. Conseqüentemente, temos  $\sigma_i \sim \sigma_{j+1}$  e  $\sigma_j \sim \sigma_{i+1}$ .

Afirmamos que os vértices  $\sigma_i, \sigma_j, \sigma_{i+1}$  e  $\sigma_{j+1}$  formam um 4-ciclo em  $G(\sigma^{-1})$ .

É importante ressaltar que a representação de permutações como ciclos é importante na demonstração a seguir, para visualizarmos com clareza como se dá a palavra em  $\sigma^{-1}$ .

De fato, observe que em  $G(\sigma^{-1})$  os vértices  $\sigma_i, \sigma_j, \sigma_{i+1}$  e  $\sigma_{j+1}$  têm rótulos  $i, i + 1, j$  e  $j + 1$ , respectivamente. Agora,  $\sigma_i \sim \sigma_{j+1}$  e  $\sigma_j \sim \sigma_{i+1}$ ; então os elementos  $i$  e  $j + 1$ , assim como  $j$  e  $i + 1$  são vizinhos na palavra associada à  $\sigma^{-1}$ . Portanto, em  $G(\sigma^{-1})$ , os vértices  $\sigma_i, \sigma_j, \sigma_{i+1}$  e  $\sigma_{j+1}$  formam um 4-ciclo.

Sem perda de generalidade, vamos considerar que o par  $(\sigma_i, \sigma_{j+1})$  antecede o par  $(\sigma_{i+1}, \sigma_j)$  na palavra. Se  $\sigma_j - \sigma_{i+1} = \sigma_{j+1} - \sigma_i$  então a palavra associada à  $\sigma^{-1}$  é da forma  $\dots i j + 1 \dots i + 1 j \dots$  ou  $\dots j + 1 i \dots j i + 1 \dots$ ; em ambos os casos os vértices de rótulos  $i, j + 1, i + 1, j$  formam um 4-ciclo orientado em  $G(\sigma^{-1})$ .

Por outro lado, pelo Procedimento *Transform*( $\pi$ ) não é possível realizar em  $\sigma^{-1}$  uma reversão que remova dois pontos de quebra. Logo,  $G(\sigma^{-1})$  não pode conter 4 – ciclos orientados. E disso segue que  $\sigma_j - \sigma_{i+1} = -(\sigma_{j+1} - \sigma_i)$ .

Sendo assim, duas possibilidades existem:

i)  $\sigma_j > \sigma_{i+1}$ ;

Neste caso  $\sigma_{j+1} < \sigma_i$  e as palavras associadas a  $\sigma$  e  $\sigma^{-1}$  são respectivamente da forma,  $\dots \sigma_{j+1} \sigma_i \dots \sigma_{i+1} \sigma_j \dots$  e  $\dots j + 1 i \dots i + 1 j \dots$

ii)  $\sigma_{i+1} > \sigma_j$ .

Então, temos que  $\sigma_i < \sigma_{j+1}$ . As palavras associadas a  $\sigma$  e  $\sigma^{-1}$  são da forma,  $\dots\sigma_i\sigma_{j+1}\dots\sigma_j\sigma_{i+1}\dots$  e  $\dots i\tilde{j} + 1 \dots \tilde{j}i + 1 \dots$ , respectivamente.

Para interpretarmos graficamente, direcionemos arestas pretas dos rótulos, não consecutivos,  $\sigma_k$  a  $\sigma_{k+1}$  e arestas cinzas dos rótulos  $k$  a  $k + 1$ , para algum  $k$ .

Somente dois tipos de 4 - *ciclos* são possíveis em  $G(\sigma)$  e  $G(\sigma^{-1})$ . A saber, os construídos a partir de  $(i)$  e  $(ii)$ . Denotemos os 4 - *ciclos* de  $(i)$  por  $(+)$  e os de  $(ii)$  por  $(-)$ . Segundo a interpretação gráfica os 4 - *ciclos* em  $G(\sigma)$  têm a mesma estrutura que os de  $G(\sigma^{-1})$ ; mudando apenas as cores das arestas (arestas pretas em  $G(\sigma)$  são cinzas em  $G(\sigma^{-1})$  e vice-versa).

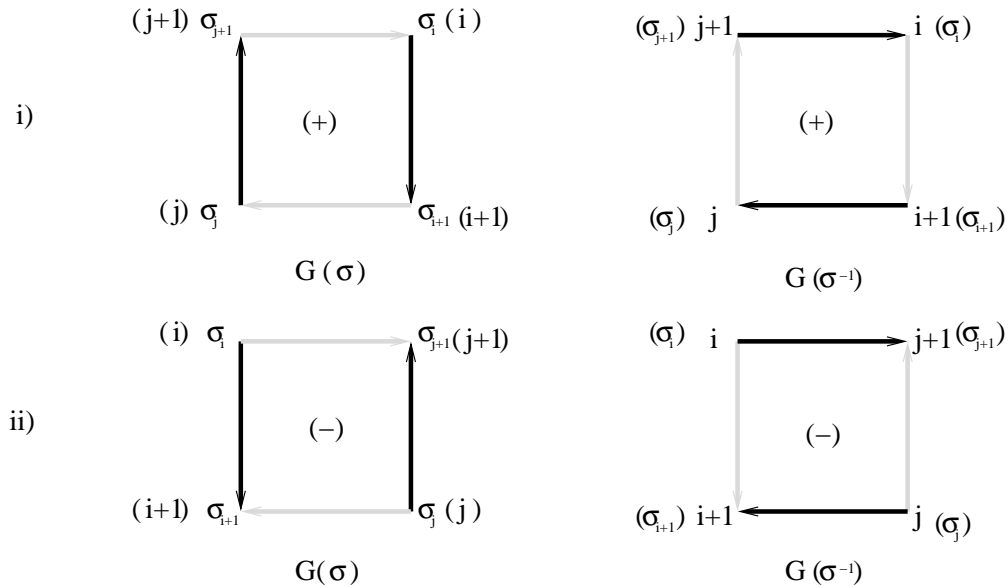


Figura 3.2: Representação gráfica de  $(i)$  e  $(ii)$ , respectivamente

Notemos ainda que, todos os ciclos que compartilham uma aresta com um ciclo do tipo  $(+)$  deve ser do tipo  $(-)$  e vice-versa.

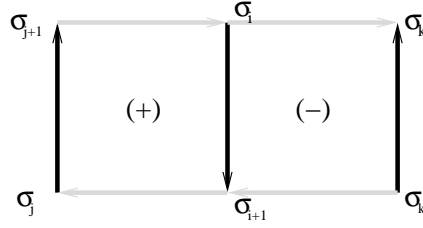
De fato, tomemos um 4-ciclo do tipo  $(+)$ , como em  $(i)$ , de vértices  $\sigma_{j+1}\sigma_i, \sigma_{i+1}$  e  $\sigma_j$ .

Do vértice de rótulo  $\sigma_i$  pode sair uma aresta preta ou uma aresta cinza para um vértice de rótulo.

A aresta cinza direciona-se de  $\sigma_i$  para um vértice de rótulo  $\sigma_k$ . Isto se deve ao fato de que  $\sigma_i < \sigma_k$ , já que  $\sigma_{j+1} < \sigma_i$ .

Do vértice de rótulo  $\sigma_{i+1}$  podemos ter uma aresta preta e uma aresta cinza. Digamos que a aresta cinza, esteja ligada a um vértice de rótulo  $\sigma_{k'}$ . A direção desta aresta é de  $\sigma_{k'}$  para  $\sigma_{i+1}$ ; isto porque  $\sigma_{i+1} < \sigma_j$ , logo  $\sigma_{k'} < \sigma_{i+1}$ .

Como existem apenas 4 - *ciclos* do tipo  $(+)$  ou  $(-)$ , o ciclo de vértices  $\sigma_i, \sigma_k, \sigma_{k'}$  e  $\sigma_{i+1}$  é do tipo  $(-)$ .



A análise dos outros vértices é análoga. Assim, como a análise para o caso (ii). Isto implica que o grafo  $H(\sigma)$  é bipartido. □

**Lema 3.9.** A permutação  $\sigma$  pode ser ordenada em  $b(\sigma) - \frac{1}{2}i(\sigma)$  reversões.

*Demonstração.*  $H(\sigma)$  é bipartido. Em [4] é dito que em [13] se pode encontrar um conjunto máximo independente em  $H(\sigma)$  em  $O(n^{\frac{3}{2}})$ . Logo, podemos encontrar uma decomposição de  $G(\sigma)$  em ciclos e que o número de 4-ciclos é  $i(\sigma)$ . O Teorema (3.1) nos garante que podemos ordenar a permutação  $\sigma$  em no máximo  $b(\sigma) - \frac{1}{2}c_4(\sigma)$ . E disso segue o resultado. □

Sejam  $\rho_1\rho_2\dots\rho_x$  a sequência de reversões em  $Transform(\pi)$  que transforma  $\pi$  em  $\bar{\pi}$  e  $\varphi_1\varphi_2\dots\varphi_y$  a sequência que transforma  $id$  em  $\bar{\nu}$ .

**Lema 3.10.**  $b(\pi) - b(\sigma) = 2(x + y)$  e  $i(\pi) - i(\sigma) \leq 4(x + y)$ .

*Demonstração.* Em cada execução do Procedimento  $Transform$   $\bar{\pi}$  e  $\bar{\nu}$  são transformados em permutações  $\bar{\pi}'$  e  $\bar{\nu}'$  de duas maneiras.

i)  $\bar{\pi}' = \bar{\pi}.\rho$  e  $\bar{\nu}' = \bar{\nu}$ ;

Neste caso,  $\rho$  é uma reversão que remove dois pontos de quebra na permutação  $\bar{\pi}^{-1}$ .

Disso temos que

$$2 = b(\bar{\pi}^{-1}\bar{\nu}) - b(\bar{\pi}'^{-1}\bar{\nu}') = b(\bar{\nu}, \bar{\pi}) - b(\bar{\nu}', \bar{\pi}') = b(\bar{\pi}, \bar{\nu}) - b(\bar{\pi}', \bar{\nu}')$$

ii)  $\bar{\pi}' = \bar{\pi}$  e  $\bar{\nu}' = \bar{\nu}.\rho$ .

Temos que  $\rho$  é uma reversão que remove dois pontos de quebra em  $\bar{\nu}^{-1}\bar{\pi}$ .

Logo,

$$2 = b(\bar{\nu}^{-1}\bar{\pi}) - b(\bar{\nu}'^{-1}\bar{\pi}') = b(\bar{\pi}, \bar{\nu}) - b(\bar{\pi}', \bar{\nu}')$$

Já que  $(x + y)$  iterações do Procedimento são realizadas para se transformar  $\pi$  em  $\sigma$ ; segue que  $b(\pi) - b(\sigma) = 2(x + y)$ .

Agora, cada reversão que remove dois pontos de quebra pode destruir no máximo quatro  $4 - \text{ciclos}$  que não compartilham arestas. Portanto,  $i(\bar{\pi}, \bar{\nu}) - i(\pi', \nu') \leq 4$ .

No procedimento *Transform* cada reversão feita pertence à sequência  $\rho_1\rho_2 \dots \rho_x$  ou  $\varphi_1\varphi_2 \dots \varphi_y$ . Disso, temos que  $i(\pi) - i(\sigma) \leq 4(x + y)$ . □

A partir dos resultados anteriores e do uso do Procedimento *Transform*( $\pi$ ) é possível construir um algoritmo que melhore o raio de aproximação do algoritmo *ReversalSort*( $\pi$ ) dado na seção anterior.

#### Procedimento *ImprovedSort*( $\pi$ )

1. Use o procedimento *Transform*( $\pi$ ) para encontrar duas sequências de reversões  $\rho_1\rho_2 \dots \rho_x$  e  $\varphi_1\varphi_2 \dots \varphi_y$  tais que  $\pi\rho_1\rho_2 \dots \rho_x = \bar{\pi}$  e  $\text{id}\varphi_1\varphi_2 \dots \varphi_y = \bar{\nu}$ . Considere  $\sigma = \bar{\nu}^{-1}$ .
2. Resolva o problema do conjunto máximo independente em um grafo bipartido  $H(\sigma)$  a fim de encontrar um conjunto máximo independente em  $G(\sigma)$  de  $4 - \text{ciclos}$  que não compartilham arestas. Com as arestas restantes, faça uma decomposição cíclica arbitrária.
3. Decomponha os vértices de grau 4 em  $G(\sigma)$  de acordo com a decomposição cíclica encontrada no passo (2). Em termos de faixas, substitua elementos únicos por faixas, orientadas apropriadamente, resultando em uma permutação com sinal  $\sigma'$ .
4. Use *SignedSort*( $\sigma'$ ) para encontrar uma sequência de reversões que ordenam  $\sigma'$ . Esta sequência fornece uma outra  $\theta_1\theta_2 \dots \theta_z$  que ordena  $\sigma$ .
5. Aplique a sequência de reversões  $\rho_1\rho_2 \dots \rho_x\theta_1\theta_2 \dots \theta_z\varphi_y\varphi_{y-1} \dots \varphi_1$  para ordenar  $\pi$ .

O teorema a seguir diz respeito ao desempenho do algoritmo *ImprovedSort*( $\pi$ ).

**Teorema 3.3.** *O algoritmo *ImprovedSort*( $\pi$ ) ordena a permutação  $\pi$  em  $b(\pi) - \frac{1}{4}c_4(\pi)$  passos.*

*Demonstração.* Observe que do passo (5) do Algoritmo *ImprovedSort*, temos que

$$d(\pi) \leq x + y + z$$

Pelo Lema (3.9) vale que

$$d(\pi) \leq x + y + b(\sigma) - \frac{1}{2}i(\sigma)$$

Usando o Lema (3.10) segue que

$$d(\pi) \leq b(\pi) - 2x - 2y + x + y - \frac{1}{2}i(\sigma)$$

De onde,

$$d(\pi) \leq b(\pi) - (x + y) - \frac{1}{2}i(\sigma)$$

Ainda pelo Lema (3.10)

$$d(\pi) \leq b(\pi) - \frac{(i(\pi) - i(\sigma))}{4} - \frac{1}{2}i(\sigma)$$

Logo,

$$d(\pi) \leq b(\pi) - \frac{i(\pi)}{4} - \frac{i(\sigma)}{4}$$

$$d(\pi) \leq b(\pi) - \frac{1}{4}i(\pi)$$

E finalmente, pelo fato de  $c_4(\pi) \geq i(\pi)$

$$d(\pi) \leq b(\pi) - \frac{1}{4}c_4(\pi)$$

□

**Corolário 3.3.** O algoritmo *ImprovedSort* provê um raio de aproximação de  $\frac{7}{4}$ .

*Demonstração.* Pelo Teorema (3.3) e Equação (2.12) segue o resultado.

□

---

### Conclusões e Trabalhos Futuros

---

Neste trabalho apresentamos uma abordagem na qual se discrimina com precisão as propriedades de soluções para o problema de rearranjo de genomas com transformações restritas à operação de reversão. Essa diferenciação permite escolher a representação mais adequada de permutações, a fim de tornar as provas dos resultados mais claras e precisas. Nosso principal cuidado foi distinguir entre as notações utilizadas para representar permutações: ciclos e palavras. Baseada nessa discriminação é possível analisar com precisão os resultados presentes no trabalho seminal de Bafna e Pevzner [3].

Nesta seção, vamos analisar alguns pontos do artigo com o intuito de justificar os ganhos obtidos com a abordagem proposta.

Como já observamos, podemos modelar o sequenciamento genético por sequências numéricas. Associar essas sequências às permutações é interessante visto que podemos nos valer das propriedades matemáticas que estas possuem.

Na apresentação do problema de rearranjo de genomas por reversões, por exemplo, dadas duas sequências, que representam o genoma de dois organismos, queremos encontrar o menor número de reversões que transformem uma sequência em outra. Identificando a essas sequências permutações ( visto que são funções ), temos bem-definidas as noções de inversa, composta e função identidade. Este é um fato importante, já que é essa associação que nos permite fazer as operações algébricas que foram realizadas ao longo deste trabalho e que também são consideradas em outros textos.

Outra questão que deve ser ressaltada é a permutação de Gollan. No artigo de

Bafna e Pevzner, define-se permutação de Gollan em notação cíclica como:

$$\gamma_n = \begin{cases} (1, 3, 5, \dots, n-1, n, \dots, 6, 4, 2), & n \text{ par} \\ (1, 3, 5, \dots, n, n-1, \dots, 6, 4, 2), & n \text{ ímpar} \end{cases}$$

Observe que esta notação foge à que é usual para permutações.

Em um ponto mais avançado do texto dá-se o exemplo de que  $\gamma_6 = 3\ 1\ 5\ 2\ 6\ 4$ . Porém, em momento algum define-se de que maneira a permutação como notação cíclica irá nos fornecer essa representação como palavra. Nesse ponto vimos a necessidade de se formalizar a idéia de como podemos representar uma palavra em notação cíclica e, reciprocamente, se nos é dada uma permutação em notação cíclica, como extrair a palavra que se associa a essa permutação.

Observe que na demonstração do Teorema 2.4, principalmente quando se provam as equações em (2.3), o formalismo proposto tem significância relevante. De fato, por exemplo, para mostrar que  $A \cap C = \emptyset$ , primeiramente usa-se as propriedades de função que têm as permutações para garantir que  $\gamma_{n-1} = (\gamma_{n-1})^+ \rho'_\pi \rho_\pi$ . Posteriormente, tomamos a permutação de Gollan em notação cíclica para constuirmos as palavras  $(\gamma_{n-1})_{str} n$  e  $1(\gamma_{n-1})^+_{str}$ . Neste ponto, observando como estão dispostos os elementos nestas palavras, fica fácil identificar as permutações  $\rho'_\pi$  e  $\rho_\pi$  necessárias para obtermos o resultado desejado.

Como última observação, consideremos o Lema 3.8, que garante que o grafo  $H(\sigma)$  é bipartido. Em [4] na demonstração do lema, considera-se um 4-ciclo de vértices  $\sigma_i, \sigma_{i+1}, \sigma_j$  e  $\sigma_{j+1}$  tal que  $\sigma_i \sim \sigma_{j+1}$  e  $\sigma_j \sim \sigma_{i+1}$ . Afirma-se então que os vértices  $\sigma_i, \sigma_{i+1}, \sigma_j$  e  $\sigma_{j+1}$  formam um 4-ciclo em  $G(\sigma^{-1})$ . Note que não é feita uma distinção entre rótulos e vértices de  $G(\sigma)$ . Pelo que é apresentado ao longo do texto,  $\sigma_i, \sigma_{i+1}, \sigma_j$  e  $\sigma_{j+1}$  são tomados como os rótulos do grafo  $G(\sigma)$  na nossa representação. Então, dizer que em  $G(\sigma^{-1})$  existe um 4-ciclo de vértices  $\sigma_i, \sigma_{i+1}, \sigma_j$  e  $\sigma_{j+1}$  nos leva a pensar que existe um 4-ciclo com esses rótulos em  $G(\sigma^{-1})$ , o que não ocorre. Portanto, a noção de rótulos e vértices em um grafo de pontos de quebra é importante, para que não se torne dúbida a apresentação dos resultados. Neste exemplo, a representação das permutações como palavras nos permite ter a visão de como extrair os 4-ciclos de  $G(\sigma)$  e  $G(\sigma^{-1})$ . Por outro lado, é a representação como ciclos que facilita a construção de  $(\sigma^{-1})_{str}$ . Com efeito, ciclicamente,  $\sigma$  deve ser da forma,  $\sigma = (\dots \sigma_i\ i \dots \sigma_{j+1}\ j+1 \dots \sigma_j\ j \dots \sigma_{i+1}\ i+1 \dots)$ ; disso segue claramente que  $\sigma^{-1} = (\dots i+1\ \sigma_{i+1} \dots j\ \sigma_j \dots j+1\ \sigma_{j+1} \dots i\ \sigma_i \dots)$ . Com a informação de que  $\sigma_i \sim \sigma_{j+1}$  e  $\sigma_j \sim \sigma_{i+1}$  se pode observar que em  $G(\sigma^{-1})$  temos um 4-ciclo de rótulos  $i, i+1, j$  e  $j+1$ .

Como já dissemos, na literatura, existem trabalhos mais recentes que apresentam algoritmos de melhor raio de aproximação para distância de reversão (por exemplo, Hannenhalli e Pevzner [12] e Caprara [7] mostraram, respectivamente, que o prob-

lema de rearranjo por reversões para permutações com sinais é resolvido em tempo polinomial e que o problema para reversões sem sinais é *NP*-difícil) e que nos proporcionam informações sobre a complexidade desse problema e também de problemas de rearranjos em que outras operações são consideradas. Porém, nosso maior cuidado foi em discriminar a notação dos objetos matemáticos que modelam os genes, a fim de obter um embasamento teórico sólido para, em particular, o problema de rearranjo por reversões apresentado por Bafna e Pevzner.

Neste trabalho, utilizamos a representação de permutações como palavras e ciclos, de acordo com o que nos era conveniente, a fim de precisar e deixar clara a prova de resultados presentes na literatura sobre rearranjo de genomas por reversões. Como trabalho futuro podemos tentar formalizar o problema de rearranjo de genomas por meio apenas de representações algébricas, ou seja permutações como ciclos. Tendo em mãos essa abordagem algébrica, poderíamos tentar identificar o conjunto das operações de rearranjo (reversões, por exemplo) com algum elemento já conhecido da Teoria de Grupos (talvez Grupos Quocientes ou outro elemento). Isto permitiria o uso de programas matemáticos, como o GAP, com o objetivo de computar quais e quantas são as operações necessárias para se transformar um genoma em outro, além de talvez nos possibilitar a obtenção de relações que forneçam dados importantes na solução do problema de rearranjo.



---

## Referências Bibliográficas

---

- [1] Max A. Alekseyev and Pavel A. Pevzner. Breakpoint Graphs and Ancestral Genome Reconstructions. *Genome Research*, 19(5):943–957, 2009.
- [2] David A. Bader, Bernard M. E. Moret, and Mi Yan. A linear-time algorithm for computing inversion distance between signed permutations with an experimental study. 2125:365–376, 2001.
- [3] Vineet Bafna and Pavel A. Pevzner. Genome Rearrangements and Sorting by Reversals. In *34<sup>th</sup> Annual Symposium on Foundations of Computer Science FOCS*, pages 148–157. IEEE, 1993.
- [4] Vineet Bafna and Pavel A. Pevzner. Genome Rearrangements and Sorting by Reversals. *SIAM Journal on Computing*, 25(2):272–289, 1996.
- [5] Piotr Berman, Sridhar Hannenhalli, and Marek Karpinski. 1.375-Approximation Algorithm for Sorting by Reversals. In Rolf H. Möhring and Rajeev Raman, editors, *Algorithms - ESA 2002, 10th Annual European Symposium ESA*, volume 2461 of *Lecture Notes in Computer Science*, pages 200–210. Springer, 2002.
- [6] Hans-Joachim Böckenhauer and Dirk Bongartz. *Algorithmic Aspects of Bioinformatics*. Natural Computing Series. Springer-Verlag, 2007.
- [7] Alberto Caprara. Sorting by reversals is difficult. In *RECOMB*, pages 75–83, 1997.
- [8] David A. Christie. Sorting Permutations by Block-Interchanges. *Inf. Process. Lett.*, 60(4):165–169, 1996.

- 
- [9] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*. MIT Press, Cambridge, 4<sup>th</sup> edition, 1990.
- [10] Isaac Elias and Tzvika Hartman. A 1.375-Approximation Algorithm for Sorting by Transpositions. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 3(4):369–379, 2006.
- [11] Stephen Jay Gould. *The Structure of Evolutionary Theory*. Harvard Press, Cambridge, 2002.
- [12] Sridhar Hannenhalli and Pavel A. Pevzner. Transforming cabbage into turnip: Polynomial algorithm for sorting signed permutations by reversals. *J. ACM*, 46(1):1–27, 1999.
- [13] John E. Hopcroft and Richard M. Karp. An  $n^{5/2}$  algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.*, 2(4):225–231, 1973.
- [14] Neil C. Jones and Pavel A. Pevzner. *An Introduction to Bioinformatics Algorithms*. MIT Press, Cambridge, 2004.
- [15] John D. Kececioglu and David Sankoff. Exact and Approximation Algorithms for Sorting by Reversals, with Application to Genome Rearrangement. *Algorithmica*, 13(1/2):180–210, 1995.
- [16] Ying Chih Lin, Chin Lung Lu, Ying-Chuan Liu, and Chuan Yi Tang. SPRING: a tool for the analysis of genome rearrangement using reversals and block-interchanges. *Nucleic Acids Research*, 34(Web-Server-Issue):696–699, 2006.
- [17] Cleber V. G. Mira and João Meidanis. Sorting by Block-Interchanges and Signed Reversals. In 4<sup>th</sup> *International Conference on Information Technology: New Generations ITNG*, pages 670–676. IEEE Computer Society, 2007.
- [18] João Meidanis and Pedro C. Feijão. A Survey on Genome Rearrangement Problems and Gene Order Based Phylogenies. Technical report, Instituto de Computação UNICAMP, dezembro 2008.
- [19] G.A. Watterson and Warren J. Ewens. The Chromosome Inversion Problem. *Journal of Theoretical Biology*, 99:1–7, 1982.