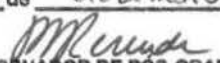


Este exemplar corresponde à redação final da
Tese/Dissertação devidamente corrigida e defendida
por: Maria Emilia Machado
Telles Walter
e aprovada pela Banca Examinadora.
Campinas, 20 de dezembro de 99

COORDENADOR DE PÓS-GRADUAÇÃO
CPG-IC

Algoritmos para Problemas
em Rearranjos de Genomas

Maria Emilia Machado Telles Walter

Tese de Doutorado

Algoritmos para Problemas em Rearranjos de Genomas

Maria Emília Machado Telles Walter¹

Outubro de 1999

Banca Examinadora:

- João Meidanis (Orientador)
- Katia Silva Guimarães
DI - UFPE
- José Augusto Ramos Soares
IME - USP
- Cid Carvalho de Souza
IC - UNICAMP
- João Carlos Setubal
IC - UNICAMP
- Yoshiko Wakabayashi (Suplente)
IME - USP
- Flávio Keidi Miyazawa (Suplente)
IC - UNICAMP

¹Suporte parcial da CAPES.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Walter, Maria Emília Machado Telles

W171a Algoritmos para problemas em rearranjos de genomas / Maria
Emília Machado Telles Walter -- Campinas, [S.P. :s.n.], 1999.

Orientador : João Meidanis

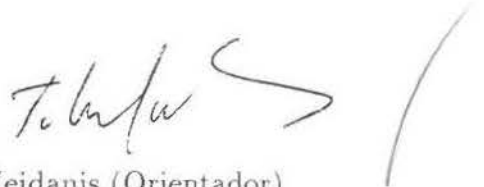
Tese (doutorado) - Universidade Estadual de Campinas, Instituto
de Computação.

1. Teoria da computação. 2. Algoritmos. 3. Evolução molecular.
4. Biologia molecular. I. Meidanis, João. II. Universidade Estadual de
Campinas. Instituto de Computação. III. Título.

Algoritmos para Problemas em Rearranjos de Genomas

Este exemplar corresponde à redação final da Tese devidamente corrigida e defendida por Maria Emília Machado Telles Walter e aprovada pela Banca Examinadora.

Campinas, 12 de dezembro de 1999.

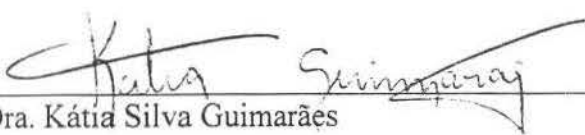
A handwritten signature in black ink, appearing to read 'João Meidanis', followed by a long horizontal stroke.

João Meidanis (Orientador)

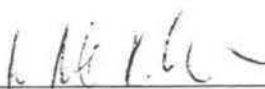
Tese apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Doutora em Ciência da Computação.

TERMO DE APROVAÇÃO

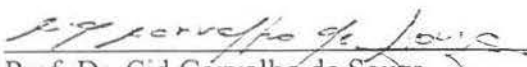
Tese defendida e aprovada em 25 de novembro de 1999, pela Banca Examinadora composta pelos Professores Doutores:



Prof. Dra. Kátia Silva Guimarães
DI - UFPe



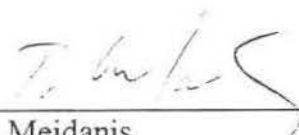
Prof. Dr. José Augusto Ramos Soares
IME - USP



Prof. Dr. Cid Carvalho de Souza
IC - UNICAMP



Prof. Dr. João Carlos Setubal
IC - UNICAMP



Prof. Dr. João Meidanis
IC - UNICAMP

© Maria Emília Machado Telles Walter, 1999.
Todos os direitos reservados.

Dedicatória

Ao João Luiz, meu marido, cúmplice companheiro pelos caminhos da vida,
A Maria Beatriz, João Francisco e Maria Luiza, meus filhos,
A Ernesto (in memoriam) e Maria de Lourdes, meus pais,
com todo o meu coração

Agradecimentos

Devo a muitas pessoas o fato de ter chegado até aqui. Com elas divido o esforço de ter alcançado esta meta, embora a responsabilidade de incorreções e imperfeições seja da minha inteira responsabilidade.

Primeiro, quero agradecer ao meu orientador, João Meidanis, por ter permitido que eu compartilhasse uma parte do seu trabalho, pela oportunidade de um intenso aprendizado, pelo exemplo de honestidade e dedicação profissional, pelo esmero e pela busca incessante do melhor de nós mesmos.

Ao meu amigo Zanoni Dias, que discutiu comigo grande parte dos assuntos desta tese, que me deu sempre tanto apoio, com suas sugestões, suas críticas e sobretudo com sua amizade e carinho.

Às minhas queridas Ana Monteiro e Anamaria Gomide, que me ajudaram com sua grande amizade, que me doaram sempre uma parte do seu tempo, que estiveram sempre ao meu lado.

Aos meus colegas de Doutorado, que partilharam comigo estes últimos anos de muito estudo e trabalho, Nalvo F. Almeida, José Roberto Monteiro e Bruno Schulze. A outros colegas do IC-Unicamp, pela sua ajuda em inúmeros momentos de dificuldades, Juliano, Helena, Mário, Marcus Vinicius, Júlio, Luis Mariano, Jerônimo, Delano, Guilherme Pimentel, Guilherme Albuquerque, Christiane, Islene, Luiz Arturo e à minha querida e especial Lin Tzy Li.

Aos Professores do IC-Unicamp, que tanto me ensinaram e ajudaram, principalmente a Cláudia M. Bauzer Medeiros, Ricardo Dahab, Ariadne M. B. R. Carvalho e Eliane Martins.

A Marco Antonio Casanova, da IBM, com quem tanto aprendi, pela sua confiança em meu trabalho.

Aos Professores da UnB, Tania Schmidt e Noraí R. Rocco do Departamento de Matemática e a Hilton V. Machado do Departamento de Estatística. Aos meus cole-

gas do Departamento de Ciência da Computação da UnB, que em mim depositaram a sua confiança, quando autorizaram meu afastamento para realizar este Doutorado, e sobretudo a Mário Dantas e Alba C. M. A. Mello, pelo seu apoio e incentivo.

A amigos especiais de Brasília, que me apoiaram com sua amizade e cuidados, Paulo Eduardo Mello, Maria Cristina Manfrinato, Gláucia Frazão e particularmente a Marta Vieira.

Finalmente, à minha família, que sempre me garantiu a tranquilidade dos seus cuidados e a quem confiei a guarda da minha casa, nas inúmeras vezes em que me ausentei. À minha querida Marisa, que acendeu para mim a chama de uma vela e me ajudou a reencontrar meu caminho. A Maria Inez, Nelson, Maria Tereza, Walmir, Bruno, Débora, Gil, Márcia, Carlinhos e Cristina, meus amados irmãos. A Olintho, Antônio, Elza e Auristela, pelo carinho e proteção. Aos meus queridos Geraldo e Beatriz, que tantas e tantas vezes me deram a sua casa, seu carinho, e acima de tudo, seu amor de pai e mãe. Ao Sr. Elias, que também me acolheu em sua casa, pelos seus cuidados, sua generosidade e seu otimismo.

A todos vocês, que fizeram parte desta etapa da minha vida, que foram e são tão importantes para mim, com muita emoção, minha sincera e profunda gratidão.

Prefácio

Desenvolver teorias para o grande volume de dados biológicos disponível na área de Biologia Molecular constitui-se num dos problemas mais desafiadores existentes atualmente. Laboratórios com tecnologia cada vez mais sofisticada fornecem detalhes precisos das estruturas moleculares estudadas. Assim, a área de Biologia Computacional tem por objetivo o estudo e aplicação de técnicas e ferramentas computacionais aos problemas de Biologia Molecular. Dentre os problemas pesquisados encontra-se o de evolução molecular, onde são estudados métodos de comparar genomas de organismos de espécies distintas, baseados em prováveis eventos que levaram a mutações. Estes métodos geram medidas de distância. Uma técnica de computar distância é comparar blocos, formados por um ou mais genes, de genomas de dois organismos. Esta tese situa-se na área de **Rearranjo de Genomas**, que visa resolver um problema combinatorial de encontrar uma série mínima de eventos de rearranjo que transforma um genoma em outro. De forma genérica, estudos de evolução baseados em rearranjos de genomas levam ao **problema da distância de rearranjo**, que é computar o número mínimo e encontrar a menor seqüência de eventos de rearranjo necessários para transformar um genoma em outro.

Especificamente, nesta tese estudamos dois eventos de rearranjo, *reversão* e *transposição*, os problemas de distância relacionados a eles e os diâmetros (maior distância entre dois cromossomos quaisquer) para os problemas de distância investigados. Inicialmente, apresentamos de forma cuidadosa uma equivalência entre os problemas de distância de reversão de cromossomos com sinais lineares e circulares, que implica imediatamente num algoritmo polinomial para o problema da distância de reversão de cromossomos circulares com sinais, baseado num algoritmo polinomial para o problema equivalente de cromossomos lineares. Mostramos também o diâmetro de reversão (a maior distância de reversão entre quaisquer dois cromossomos) de cromossomos lineares e circulares. A prova do diâmetro linear foi baseada no cálculo de distâncias de reversão entre determinadas permutações e a permutação identidade.

Em seguida apresentamos um algoritmo de aproximação e um algoritmo exato para

o problema da distância de transposição de cromossomos lineares sem sinais, baseados numa estrutura denominada *diagrama de pontos-de-quebra*. Embora este algoritmo tenha uma razão teórica elevada (2.25), quando comparado ao melhor algoritmo conhecido (1.5), mostrou uma razão bem melhor nos experimentos realizados, sugerindo que ele possa ser útil na prática. Mostramos também um limite inferior para o diâmetro de transposição (a maior distância de transposição entre quaisquer dois cromossomos) de cromossomos lineares. Esta prova foi baseada no cálculo da distância de transposição entre uma permutação e sua inversa.

Finalmente, apresentamos algoritmos de aproximação para o problema da distância de reversão e transposição de cromossomos lineares, sem e com sinais, que embora simples, levaram a um resultado bastante mais complexo, que foi o de estabelecer um limite inferior para o diâmetro de reversão e transposição. Esta prova utilizou o cálculo da distância de reversão e transposição entre uma particular permutação e a permutação identidade, cuja prova, por sua vez, foi baseada na análise de todas as possibilidades de aplicar reversões e transposições (operações) na permutação inicial, nos dois primeiros passos, em qualquer seqüência de operações que transforma a permutação inicial na permutação identidade.

Conteúdo

Dedicatória	v
Agradecimentos	vi
Prefácio	viii
1 Introdução	1
1.1 Modelos computacionais	6
1.1.1 Distância de reversão	8
1.1.2 Distância de transposição	9
1.1.3 Distância de transversão	12
1.1.4 Distância de translocação	12
1.1.5 Distância de reversão e transposição	14
1.1.6 Distância de reversão, transposição e transversão	15
1.1.7 Distância de reversão e translocação	15
1.1.8 Distância de transposição e translocação	17
1.2 Direções de pesquisa	18
1.3 Objetivos da tese	22
2 Estruturas e conceitos básicos	24
2.1 O grafo de pontos-de-quebra	24
2.2 O grafo de ciclos	30

3	Distância de reversão de cromossomos com sinais	34
3.1	Uma equivalência entre os problemas circular e linear	34
3.1.1	Uma formalização para o problema circular	38
3.1.2	Um isomorfismo entre os problemas circular e linear	44
3.2	O diâmetro de reversão	49
3.3	Distâncias de reversão de cromossomos circulares e lineares	55
3.4	Sumário e questões	63
4	Distância de transposição de cromossomos lineares sem sinais	65
4.1	Um algoritmo de aproximação	66
4.1.1	Definições	66
4.1.2	Apresentação do algoritmo	69
4.1.3	Resultados e análise dos experimentos	85
4.2	Um limite inferior para o diâmetro de transposição	89
4.2.1	Distância de transposição entre r_n e t_n	89
4.2.2	Um algoritmo para computar $d_t(r_n, t_n)$	103
4.3	Sumário e questões	105
5	Distância de reversão e transposição de cromossomos lineares	106
5.1	Definições	107
5.2	Algoritmos de aproximação	108
5.3	Um limite inferior para o diâmetro de reversão e transposição	115
5.4	Sumário e questões	141
6	Conclusões	142
A	Glossário de definições e notações	144
B	Referências bibliográficas comentadas	148
C	Glossário de termos de Biologia Molecular	159

Capítulo 1

Introdução

Pesquisas em **Biologia Molecular** estão basicamente voltadas para o entendimento da estrutura e função de proteínas e ácidos nucleicos. Existem dois tipos de ácidos nucleicos em organismos vivos, o RNA - ácido ribonucleico e o DNA - ácido desoxirribonucleico.

O **DNA** é uma molécula muito grande (macromolécula) formada por uma cadeia dupla de moléculas menores. Cada uma destas cadeias (chamadas de **fitas**) consiste de uma espinha dorsal com repetições das mesmas unidades básicas. Uma **unidade básica** é composta por açúcar (desoxirribose), fosfato e por uma das quatro bases: adenina (A), guanina (G), citosina (C) e timina (T). As duas fitas do DNA, que formam uma estrutura helicoidal, encontram-se ligadas por causa do pareamento da base A com a base T, e da C com a G (Figura 1.1). Os pares de bases (pb) fornecem uma unidade de comprimento para o DNA.

A estrutura química das ligações das bases impõe para cada uma das fitas uma **orientação**, que é indicada denominando-se as extremidades de cada fita por 5' e 3'. As duas fitas são orientadas de forma contrária, de modo que a extremidade 5' de uma corresponde à extremidade 3' da outra (Figura 1.2).

Denota-se o DNA por uma seqüência de letras, onde cada letra representa uma base. A cadeia dupla é representada colocando uma fita acima da outra (Figura 1.2).

Cada célula do organismo tem poucas cadeias de moléculas de DNA bastante longas. Estas moléculas podem ser lineares (quando as extremidades das fitas são livres) ou circulares (quando as duas extremidades encontram-se ligadas). Existem certos vírus constituídos apenas por DNA circular [27] (Figura 1.3).

A importância das moléculas de DNA é que nelas são codificadas todas as in-

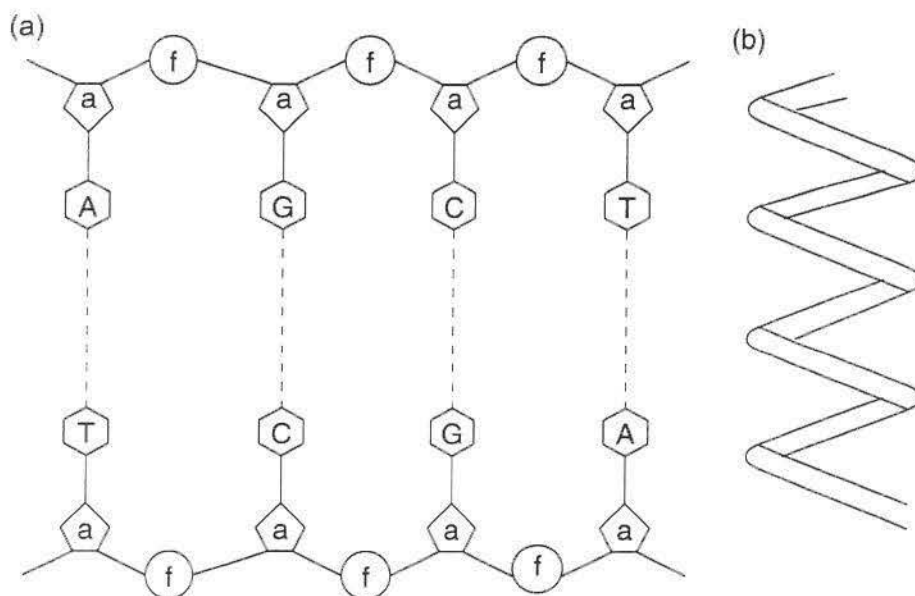


Figura 1.1: O esquema das ligações das bases e a estrutura espacial do DNA. (a) O pareamento das bases liga as duas fitas do DNA. (b) As duas fitas formam uma dupla hélice.

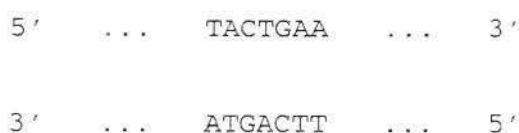


Figura 1.2: O DNA pode ser representado por uma sequência de letras, onde cada letra representa uma base. A orientação de cada fita é indicada por 5' e 3', e as duas fitas são orientadas de forma contrária. Observe ainda o pareamento das bases.

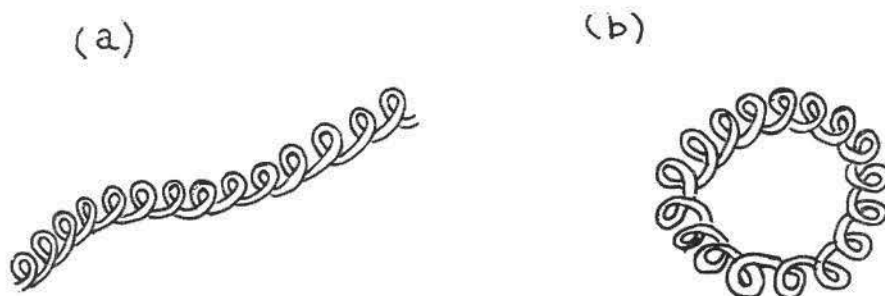


Figura 1.3: Exemplos das estruturas do DNA. (a) DNA representado como molécula linear. (b) DNA formando círculo.

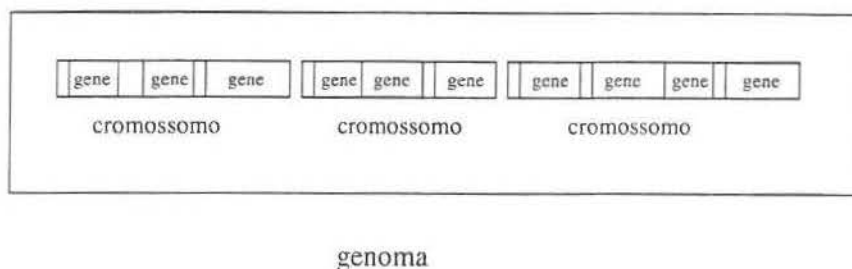


Figura 1.4: Esta figura mostra uma visão esquemática dos genes dentro dos cromossomos e dos cromossomos dentro do genoma.

formações necessárias para construir cada proteína ou RNA encontrado no organismo, garantindo assim a sobrevivência do indivíduo e a perpetuação da espécie.

Cada uma das moléculas de DNA é chamada de um **cromossomo**. Um fato importante sobre um cromossomo é que a cada tipo diferente de proteína num organismo usualmente corresponde um e apenas um trecho contíguo ao longo do cromossomo, chamado de **gene**. Comprimentos de genes variam, mas no caso de seres humanos, um gene tem aproximadamente 10.000 pb. Certos mecanismos celulares são capazes de reconhecer no DNA os pontos precisos onde um gene começa e onde termina.

Denomina-se de **genoma** a um conjunto completo de cromossomos dentro de uma célula. A Figura 1.4 mostra o esquema dos genes dentro dos cromossomos e dos cromossomos dentro do genoma. O número de cromossomos num genoma é característico da espécie. Por exemplo, cada célula humana tem 46 cromossomos, sendo que o tamanho do genoma é de aproximadamente 3×10^9 pb, enquanto a da bactéria *Escherichia coli* tem apenas um cromossomo, sendo o tamanho do genoma 5×10^6 pb.

A Ciência Moderna mostrou que todas as espécies de organismos vivos na Terra passaram por um lento processo de transformação através do tempo, ou seja, sofreram **evolução**. Portanto, uma área bastante interessante em Biologia é modelar processos evolucionários das espécies existentes atualmente. Isto é feito normalmente construindo-se árvores, chamadas de **árvores filogenéticas**, cujas folhas representam espécies atuais e cujos nós internos representam ancestrais hipotéticos. Uma característica importante destas árvores é a **distância** entre pares de nós, que fornece uma estimativa da distância evolucionária entre estes nós. Usar seqüências homólogas de genes para inferir distância evolucionária continua a ter grande interesse para os pesquisadores de Biologia Molecular, uma vez que proteínas e ácidos nucleicos também evoluem. Assim, novas técnicas de medir distâncias continuam a ser investigadas nesta área, visando a reconstrução das árvores filogenéticas.

Por outro lado, desenvolver teorias para o grande volume de dados disponível hoje na área de Biologia Molecular é um dos problemas mais desafiadores existentes atualmente. Laboratórios com tecnologia cada vez mais sofisticada fornecem detalhes precisos das estruturas moleculares estudadas. Esses dados necessitam de técnicas eficientes de manipulação, para que possam ser utilizados de forma mais efetiva pelos biólogos.

Neste contexto, surge uma nova área em Ciência da Computação. **Biologia Molecular Computacional** pode ser definida como a área que tem por objetivo o estudo e aplicação de técnicas e ferramentas computacionais aos problemas de Biologia Molecular. Esta área vem mostrando notável crescimento nos últimos anos [28, 41].

Em Biologia Molecular Computacional, dentre os problemas pesquisados, encontra-se o de evolução molecular. Mais especificamente, estudam-se métodos para *comparar genomas* de organismos de espécies distintas, baseados em prováveis eventos que levaram a mutações. Estes métodos geram medidas de distância, que podem ser empregadas para reconstrução de árvores filogenéticas, ou ainda para verificar o relacionamento funcional entre dois organismos. Assim, neste trabalho, de forma genérica, visamos estudar formas de computar distância entre genomas de organismos de espécies diferentes.

Em Biologia Molecular Computacional, para computar distância entre genomas, tradicionalmente utiliza-se a técnica de **alinhamento de seqüências**. De forma genérica, alinhar duas seqüências é encontrar uma correspondência entre bases similares. Para o alinhamento são utilizadas mutações pontuais nos genes tais como substituições, remoções e inserções de bases. A distância é computada associando custos a estas operações, e procurando pela composição menos cara dentre as que transformam uma seqüência na outra.

Entretanto, existem organismos (como os vírus da herpes) para os quais o alinhamento de seqüências produz uma similaridade tão ruim que não fornece nenhuma evidência não ambígua de que tenham tido uma origem evolucionária comum. Como resultado, os métodos clássicos de comparação de seqüências não são muito úteis para genomas com similaridade ruim, e utilizá-los pode levar até mesmo a contradições, pois genes diferentes podem originar árvores evolucionárias diferentes.

Em vista disso, os pesquisadores tentaram encontrar uma maneira diferente para computar distância, comparando porções maiores dos genomas, em lugar de comparar bases. Esta alternativa foi baseada em inúmeros trabalhos realizados em Biologia Molecular, destacando-se os trabalhos pioneiros de Nadeau e Taylor [31] e de Palmer e Herbon [32], que descobriram genomas de duas espécies, quase idênticos em seqüências de genes, mas muito diferentes na ordem destes genes (Figura 1.5). Estes trabalhos

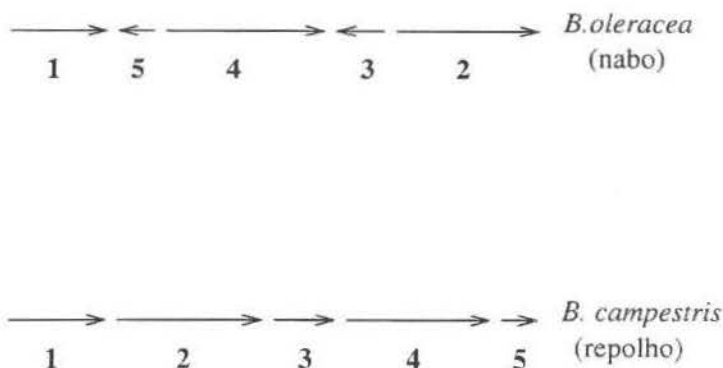


Figura 1.5: Os genomas de duas espécies de plantas, onde cada número denota um bloco de um ou mais genes, e as setas indicam as orientações relativas dos blocos de uma espécie em relação à outra.

provaram de forma convincente que rearranjo em porções grandes dos genomas é um modo comum de evolução molecular.

Todo o estudo realizado em **rearranjo de genomas** visa resolver um quebra-cabeça combinatorial para encontrar uma série mínima de rearranjos, ocorridos em porções grandes do genoma, que transforma um genoma no outro. Assim, em Biologia Molecular Computacional, estudos de evolução baseados em rearranjos levam ao **problema de distância de rearranjo**, que, de forma genérica, é computar o número mínimo e encontrar a menor sequência de eventos de rearranjo necessários para transformar um genoma em outro.

A razão para encontrar uma série mínima de eventos de rearranjo é justificada pela *hipótese da parsimônia*, na qual assume-se que a Natureza sempre encontra caminhos que necessitam de um mínimo de mudanças, e portanto, se desejarmos investigar como um organismo de uma espécie pode ter sido transformado num organismo de outra espécie (ou vice-versa), devemos tentar encontrar uma série mínima de eventos de rearranjo que possivelmente tenham realizado esta transformação.

Um enfoque computacional baseado em comparação da ordem de genes em vez da comparação tradicional das bases dos genes foi introduzido pioneiramente por Sankoff [35, 37, 36, 38]. Kececioğlu e Sankoff [25], e Pevzner e Waterman [34] fazem uma revisão de problemas combinatoriais motivados por rearranjo de genomas.

Existem duas técnicas em Biologia Molecular para obter dados sobre a ordem dos genes num genoma: **mapeamento físico** e **seqüenciamento**. Mapeamento físico fornece a ordem e a distância relativa dos genes. Seqüenciamento, além da ordem, fornece as orientações relativas dos genes. O seqüenciamento de genomas inteiros está se tornando cada vez mais comum, estando hoje disponíveis cerca de 25 geno-

mas completos de microorganismos, além de inúmeros genomas completos de vírus e organelas. No Brasil há dois projetos em andamento para seqüenciar totalmente genomas de bactérias. Contudo, para genomas muito grandes como os de plantas, ainda grande parte dos dados experimentais baseia-se em mapeamento físico, e portanto não contém as orientações dos genes. Notamos ainda que a disponibilidade destes dados gera novos problemas de comparação de genomas em diferentes áreas, como em DNA mitocondrial de plantas [32], DNA mitocondrial de animais [38], virologia [17], ou genética de *Drosophila* [43].

Na Seção 1.1 serão apresentados modelos computacionais que formalizam determinados eventos de rearranjo. Na Seção 1.2 apontamos possíveis direções de pesquisa na área de Rearranjo de Genomas, e fazemos comentários específicos em cada um dos problemas relacionados. Finalmente, na Seção 1.3 apresentamos os objetivos desta tese.

1.1 Modelos computacionais

Examinaremos agora modelos para formalizar determinados eventos de rearranjo de genomas. Os eventos que serão modelados são: **reversão**, que tem o efeito de inverter a ordem e a orientação dos genes na porção do genoma que sofreu a mutação, **transposição**, que move uma porção de uma região para outra dentro do genoma, **transversão**, que move os blocos de genes de um local para outro dentro do genoma, mas reverte a ordem e a orientação dos genes, e **translocação**, que troca porções entre dois cromossomos diferentes dentro do genoma (Figura 1.6). Na figura, cada número denota um bloco do cromossomo contendo um ou mais genes, e a seta denota o fato de que estes blocos têm uma orientação conhecida. O sentido da seta está associado às orientações relativas dos blocos de genes de uma espécie em relação à outra. Dois blocos de genes em genomas diferentes têm o mesmo número se contêm exatamente os mesmos genes.

Inicialmente, vamos modelar um cromossomo linear que, conforme visto na seção anterior, é uma molécula de DNA cujas pontas estão soltas.

Um cromossomo linear, possuindo n blocos de genes, será modelado por uma permutação $\pi = (\pi_1 \pi_2 \dots \pi_n)$, sendo cada π_i , $1 \leq i \leq n$, um bloco de genes, e todos eles distintos entre si. Se as orientações relativas dos blocos de genes não são conhecidas, cada π_i é um número inteiro sem sinais (Figura 1.7 (a)). Se estas orientações relativas são conhecidas, o cromossomo é modelado por uma permutação com sinais, idêntica ao caso anterior, exceto que cada π_i possui um sinal $+$ ou $-$, indicando a orientação

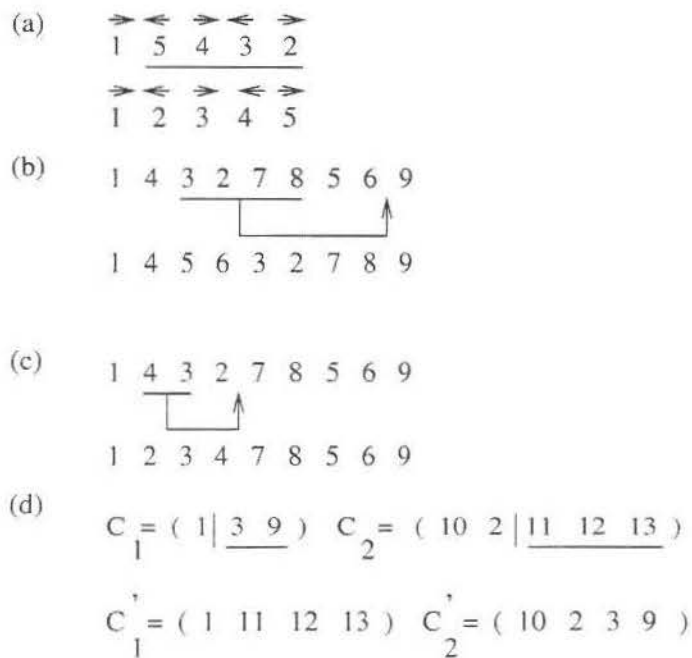


Figura 1.6: Exemplos dos eventos de rearranjo que serão modelados. (a) *Reversão* inverte a ordem e a orientação dos blocos de genes do cromossomo indicados pelo traço. (b) *Transposição* move os blocos de genes do cromossomo indicados pelo traço para o local indicado pela seta. (c) *Transversão* move os blocos de genes do cromossomo indicados pelo traço para o local indicado pela seta, mas invertendo a ordem e orientação dos blocos de genes. (d) *Translocação* troca os blocos de genes dos dois cromossomos indicados pela barra horizontal e pelo traço.

- (a) (7 1 2 4 5 3 6 8) cloroplasto de *Lobelia fervens*
- (b) (-4 -6 +1 +7 -2 -3 +5 +8) cromossomo X humano
- (c) ((1 3 9) (7 8 4 5 6) (10 2 11 12 13))
 outros genomas equivalentes:
 ((7 8 4 5 6) (1 3 9) (10 2 11 12 13))
 ((10 2 11 12 13) (7 8 4 5 6) (1 3 9))

Figura 1.7: Exemplos de representação de cromossomos e genomas, considerando apenas DNA linear. (a) O cromossomo linear sem sinais. (b) O cromossomo linear com sinais. (c) Como o genoma é um conjunto de cromossomos, não há uma ordem entre estes cromossomos.

do bloco (Figura 1.7 (b)). Observamos ainda que o sinal “+” representa a seta para a direita da Figura 1.6 (a), e o sinal “-” denota a seta para a esquerda.

O genoma, também visto na seção anterior, é um conjunto de cromossomos e será representado por um conjunto de permutações (Figura 1.7 (c)).

Estes diferentes eventos de rearranjo levam a diversos problemas combinatórios, cuja complexidade em alguns casos ainda não é conhecida. Apresentaremos oito modelos: *distância de reversão*, *distância de transposição*, *distância de transversão*, *distância de translocação*, *distância de reversão e transposição*, *distância de reversão, transposição e transversão*, *distância de reversão e translocação* e *distância de transposição e translocação*.

De forma genérica, os formalismos modelam eventos de rearranjo ocorrendo internamente a um único cromossomo (reversões, transposições e transversões) e ocorrendo entre diferentes cromossomos (translocações). Em cada uma das seções a seguir, inicialmente caracterizaremos cada evento de rearranjo e em seguida enunciaremos o problema relacionado a cada um destes eventos.

1.1.1 Distância de reversão

Rearranjos em DNA mitocondrial, de cloroplasto, viral e bacterial são basicamente constituídos por reversões [18]. A formulação combinatorial do problema de distância

de reversão é descrita em seguida.

Já sabemos que uma *reversão* é um evento de rearranjo de genomas que transforma um cromossomo em outro, “invertendo” a seqüência de blocos de genes de um determinado trecho do cromossomo.

Definimos uma **reversão** r do intervalo $[i, j]$, denotado por $r(i, j)$, atuando numa permutação linear com sinais $\pi = (\pi_1 \dots \pi_{i-1} \pi_i \pi_{i+1} \dots \pi_{j-1} \pi_j \pi_{j+1} \dots \pi_n)$ por

$$r(i, j) \cdot \pi = (\pi_1 \dots \pi_{i-1} \bar{\pi}_j \bar{\pi}_{j-1} \dots \bar{\pi}_{i+1} \bar{\pi}_i \pi_{j+1} \dots \pi_n)$$

onde $\bar{\pi}_k$ indica a inversão do sinal de π_k (Figura 1.8 (a)).

O problema da distância de reversão de cromossomos lineares, com orientações conhecidas, é formalizado como se segue. Dadas duas permutações lineares com sinais π e σ , modelando dois cromossomos lineares com orientações relativas conhecidas, o **problema da distância de reversão** de π e σ é encontrar uma série de reversões $r_1, r_2, \dots, r_\varrho$ tais que $r_\varrho \cdot r_{\varrho-1} \cdot \dots \cdot r_2 \cdot r_1 \cdot \pi = \sigma$ e ϱ é mínimo. Chamamos ϱ de **distância de reversão de π e σ** , denotada por $d_r(\pi, \sigma)$.

Quando não temos informações sobre a orientação dos blocos de genes, isto é, a permutação π não contém sinais, a reversão r inverterá apenas a ordem dos blocos de genes (Figura 1.8 (b)).

Podemos supor que σ é a permutação **identidade** $\iota = (+1 \ +2 \ \dots \ +n)$, e considerar o problema equivalente de transformar π em ι . O problema de **ordenar π por reversões** é encontrar a distância de reversão, $d_r(\pi, \iota)$.

Para permutações sem sinais, o enunciado do problema é inteiramente análogo, apenas lembrando que as reversões invertem apenas a ordem de $\pi_i \dots \pi_j$.

1.1.2 Distância de transposição

Análises dos genomas do vírus Epstein-Barr EBV e do vírus simples da Herpes HSV-1 revelaram que a evolução destes vírus envolvia um certo número de reversões e transposições de partes grandes dos genomas [4]. A formulação combinatorial para o problema de distância de transposição é descrita a seguir.

Uma *transposição* é um evento de rearranjo de genomas que transforma um cromossomo em outro, “cortando” uma certa porção do cromossomo e “colando” esta porção em um outro local no mesmo cromossomo.

Formalmente, uma transposição $t(i, j, k)$ é definida por três inteiros i, j e k tais que $1 \leq i < j \leq n + 1$, $1 \leq k \leq n + 1$, e $k \notin [i, j]$, da seguinte forma. Ela “corta” a

(a)

$$\pi = (+1 \ -5 \ +4 \ -3 \ \underline{+2})$$

$$\overset{1}{\pi} = r(5,5) . \ \pi = (+1 \ -5 \ \underline{+4} \ -3 \ -2)$$

$$\overset{2}{\pi} = r(3,3) . \ \overset{1}{\pi} = (+1 \ -5 \ \underline{-4 \ -3 \ -2})$$

$$\overset{3}{\pi} = r(2,5) . \ \overset{2}{\pi} = (+1 \ +2 \ +3 \ +4 \ +5) = \overset{1}{\pi}$$

(b)

$$\pi = (7 \ \underline{1 \ 2 \ 4 \ 5 \ 3 \ 6} \ 8)$$

$$\overset{1}{\pi} = r(2,7) . \ \pi = (7 \ \underline{6 \ 3 \ 5 \ 4 \ 2 \ 1} \ 8)$$

$$\overset{2}{\pi} = r(1,7) . \ \overset{1}{\pi} = (1 \ 2 \ \underline{4 \ 5 \ 3} \ 6 \ 7 \ 8)$$

$$\overset{3}{\pi} = r(3,5) . \ \overset{2}{\pi} = (1 \ 2 \ 3 \ \underline{5 \ 4} \ 6 \ 7 \ 8)$$

$$\overset{4}{\pi} = r(4,5) . \ \overset{3}{\pi} = (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8) = \overset{1}{\pi}$$

Figura 1.8: Exemplos de reversão atuando em permutação linear. O trecho sublinhado indica os elementos afetados pela reversão. (a) Reversão atuando em permutação com sinais. (b) Reversão atuando em permutação sem sinais.

$$\begin{aligned}
\pi &= (8 \ 5 \ 1 \ 4 \ 3 \ 2 \ 7 \ 6) \\
&\quad \underbrace{\hspace{1.5cm}} \uparrow \\
\pi^1 &= t(1,3,8) \cdot \pi = (1 \ 4 \ 3 \ 2 \ 7 \ 8 \ 5 \ 6) \\
&\quad \underbrace{\hspace{1.5cm}} \uparrow \\
\pi^2 &= t(3,7,9) \cdot \pi^1 = (1 \ 4 \ 5 \ 6 \ 3 \ 2 \ 7 \ 8) \\
&\quad \underbrace{\hspace{1.5cm}} \uparrow \\
\pi^3 &= t(2,5,7) \cdot \pi^2 = (1 \ 3 \ 2 \ 4 \ 5 \ 6 \ 7 \ 8) \\
&\quad \underbrace{\hspace{1.5cm}} \uparrow \\
\pi^4 &= t(2,3,4) \cdot \pi^3 = (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8) = \mathbf{1}
\end{aligned}$$

Figura 1.9: Exemplo de transposição atuando em permutação linear. Conhecer as orientações dos blocos de genes não altera a transposição, pois este evento envolve apenas mudanças de local no cromossomo.

porção entre as posições i e $j - 1$, incluindo os extremos, e “cola” estes elementos exatamente antes da posição k . Portanto, podemos escrever

$$\begin{aligned}
t(i, j, k) \cdot (\pi_1 \pi_2 \dots \pi_{i-1} \pi_i \dots \pi_{j-1} \pi_j \dots \pi_{k-1} \pi_k \dots \pi_n) = \\
(\pi_1 \pi_2 \dots \pi_{i-1} \pi_j \dots \pi_{k-1} \pi_i \dots \pi_{j-1} \pi_k \dots \pi_n),
\end{aligned}$$

se $i < j < k$, e

$$\begin{aligned}
t(i, j, k) \cdot (\pi_1 \pi_2 \dots \pi_{k-1} \pi_k \dots \pi_{i-1} \pi_i \dots \pi_{j-1} \pi_j \dots \pi_n) = \\
(\pi_1 \pi_2 \dots \pi_{k-1} \pi_i \dots \pi_{j-1} \pi_k \dots \pi_{i-1} \pi_j \dots \pi_n),
\end{aligned}$$

se $k < i < j$. Observe que $t(i, j, k) = t(j, k, i)$ quando $i < j < k$ (Figura 1.9).

O problema da distância de transposição de dois cromossomos lineares é formalizado como se segue. Dadas duas permutações π e σ , modelando dois cromossomos, o **problema da distância de transposição** de π e σ é encontrar uma série de transposições $t_1, t_2, \dots, t_\delta$ tais que $t_\delta \cdot t_{\delta-1} \cdot \dots \cdot t_2 \cdot t_1 \cdot \pi = \sigma$ e δ é mínimo. Chamamos δ de **distância de transposição** de π e σ , denotado por $d_t(\pi, \sigma)$.

Novamente, supondo que σ seja a permutação identidade $\mathbf{1} = (1 \ 2 \ \dots \ n)$, o problema de **ordenar π por transposições** é encontrar a distância de transposição, $d_t(\pi, \mathbf{1})$.

Observamos que no caso de transposições, o conhecimento das orientações dos blocos de genes não altera as definições anteriores, pois este evento de rearranjo envolve apenas mudanças de locais, não afetando portanto as orientações dos blocos de genes.

1.1.3 Distância de transversão

Podemos conceber também um evento que “corta” certos blocos de genes do cromossomo e “cola” estes blocos de genes em um outro local no mesmo cromossomo, mas com a ordem e as orientações dos genes invertidas. Chamaremos este evento de **transversão**.

Definimos uma transversão $\bar{l}(i, j, k)$ por três inteiros i, j e k tais que $1 \leq i < j \leq n+1$, $1 \leq k \leq n+1$, e $k \notin [i, j]$, da seguinte forma. Ela “corta” a porção entre as posições i e $j-1$, incluindo os extremos, e “cola” estes elementos exatamente antes da posição k , invertendo a ordem e a orientação dos genes. Portanto, podemos escrever

$$\begin{aligned} \bar{l}(i, j, k) \cdot (\pi_1 \pi_2 \dots \pi_{i-1} \pi_i \dots \pi_{j-1} \pi_j \dots \pi_{k-1} \pi_k \dots \pi_n) = \\ (\pi_1 \pi_2 \dots \pi_{i-1} \pi_j \dots \pi_{k-1} \bar{\pi}_{j-1} \dots \bar{\pi}_i \pi_k \dots \pi_n), \end{aligned}$$

se $i < j < k$, onde $\bar{\pi}_l$ indica a inversão do sinal de π_l , e

$$\begin{aligned} \bar{l}(i, j, k) \cdot (\pi_1 \pi_2 \dots \pi_{k-1} \pi_k \dots \pi_{i-1} \pi_i \dots \pi_{j-1} \pi_j \dots \pi_n) = \\ (\pi_1 \pi_2 \dots \pi_{k-1} \bar{\pi}_{j-1} \dots \bar{\pi}_i \pi_k \dots \pi_{i-1} \pi_j \dots \pi_n), \end{aligned}$$

se $k < i < j$.

O problema da distância de transversão de dois cromossomos lineares é formalizado como se segue. Dadas duas permutações π e σ , modelando dois cromossomos, o **problema da distância de transversão** de π e σ é encontrar uma série de transversões $\bar{l}_1, \bar{l}_2, \dots, \bar{l}_{\bar{\delta}}$ tais que $\bar{l}_{\bar{\delta}} \cdot \bar{l}_{\bar{\delta}-1} \dots \bar{l}_2 \cdot \bar{l}_1 \cdot \pi = \sigma$ e $\bar{\delta}$ é mínimo. Chamamos $\bar{\delta}$ de **distância de transversão de π e σ** , denotado por $d_{\bar{l}}(\pi, \sigma)$.

Novamente, assumindo que σ seja a permutação identidade $\iota = (1 \ 2 \ \dots \ n)$, o problema de **ordenar π por transversões** é encontrar a distância de transversão, $d_{\bar{l}}(\pi, \iota)$.

Para permutações sem sinais, o enunciado do problema é inteiramente análogo, apenas lembrando que as reversões apenas invertem a ordem de $\pi_i \dots \pi_{j-1}$.

1.1.4 Distância de translocação

Translocações podem ocorrer na formação dos cromossomos sexuais de um indivíduo da espécie humana, podendo resultar em indivíduos normais ou não [39]. A formulação combinatorial para o problema de rearranjo por translocações é apresentada em seguida.

$$\begin{aligned}
\Pi &= (\pi_1 = (1 \mid \underline{3 \ 9}) \quad \pi_2 = (7 \ 8 \ 4 \ 5 \ 6) \quad \pi_3 = (10 \ 2 \mid \underline{11 \ 12 \ 13})) \\
\Gamma &= (\gamma_1 = (1 \ 2 \ 3 \ 4 \ 5 \ 6) \quad \gamma_2 = (7 \ 8 \ 9) \quad \gamma_3 = (10 \ 11 \ 12 \ 13)) \\
\Pi^1 &= \text{Tr}(\pi_1, \pi_3, 2, 3) \cdot \Pi = ((1 \mid \underline{11 \ 12 \ 13}) (7 \ 8 \ 4 \ 5 \ 6) (10 \mid \underline{2 \ 3 \ 9})) \\
\Pi^2 &= \text{Tr}(\pi_1^1, \pi_3^1, 2, 2) \cdot \Pi^1 = ((1 \ 2 \ 3 \mid \underline{9}) (7 \ 8 \mid \underline{4 \ 5 \ 6}) (10 \ 11 \ 12 \ 13)) \\
\Pi^3 &= \text{Tr}(\pi_1^2, \pi_2^2, 2, 4) \cdot \Pi^2 = ((1 \ 2 \ 3 \ 4 \ 5 \ 6) (7 \ 8 \ 9) (10 \ 11 \ 12 \ 13)) = \Gamma
\end{aligned}$$

Figura 1.10: Exemplo de transformação de um genoma Π num genoma alvo Γ , por meio de translocações recíprocas entre dois cromossomos dos genomas.

De forma simplificada, para definir este modelo vamos supor que os dois genomas contêm o mesmo número N de cromossomos, e que cada bloco de genes aparece em cada genoma exatamente uma vez. Temos então:

$$\Pi = ((\pi_{11} \pi_{12} \dots \pi_{1m_1}), (\pi_{21} \pi_{22} \dots \pi_{2m_2}), \dots, (\pi_{N1} \pi_{N2} \dots \pi_{Nm_N}))$$

e

$$\Gamma = ((\gamma_{11} \gamma_{12} \dots \gamma_{1n_1}), (\gamma_{21} \gamma_{22} \dots \gamma_{2n_2}), \dots, (\gamma_{N1} \gamma_{N2} \dots \gamma_{Nn_N}))$$

Os genomas Π e Γ são ditos **idênticos** ($\Pi = \Gamma$) se e somente se é possível fazer uma correspondência entre os cromossomos de Π e Γ de modo que contenham os mesmos blocos de genes.

Para dois cromossomos $\pi = (\pi_1 \pi_2 \dots \pi_m)$ e $\gamma = (\gamma_1 \gamma_2 \dots \gamma_n)$ denotamos uma translocação atuando em π e γ como $\text{Tr}(\pi, \gamma, i, j)$, $1 < i \leq m$, $1 < j \leq n$, onde os cortes ocorrem em π entre π_{i-1} e π_i e em γ entre γ_{j-1} e γ_j . As porções $[\pi_i, \pi_m]$ de π e $[\gamma_j, \gamma_n]$ de γ são trocadas, transformando π e γ em dois novos cromossomos: $\pi^1 = (\pi_1 \dots \pi_{i-1} \gamma_j \dots \gamma_n)$ e $\gamma^1 = (\gamma_1 \dots \gamma_{j-1} \pi_i \dots \pi_m)$. Será modelado apenas o tipo mais comum de translocação, a **translocação recíproca**, onde cada uma das quatro porções $[\pi_1, \pi_{i-1}]$, $[\pi_i, \pi_m]$, $[\gamma_1, \gamma_{j-1}]$ e $[\gamma_j, \gamma_n]$ é não-vazia.

Para um genoma Π e uma translocação Tr atuando num par de cromossomos π e σ de Π , denotamos o genoma resultante como $\text{Tr}(\pi, \sigma, i, j) \cdot \Pi$. Se Tr for uma translocação recíproca então o número de cromossomos em Π e $\text{Tr}(\pi, \sigma, i, j) \cdot \Pi$ é o mesmo (Figura 1.10).

Dados os genomas Π e Γ , o **problema da distância de translocação** é encontrar

$$\begin{aligned}
\pi &= (-4 \ -6 \ +1 \ +7 \ +2 \ -3 \ +5 \ +8) \\
&\quad \underbrace{\hspace{1.5cm}} \quad \uparrow \\
\pi^1 &= t(1,3,5) \cdot \pi = (+1 \ +7 \ -4 \ -6 \ +2 \ -3 \ +5 \ +8) \\
\pi^2 &= r(3,6) \cdot \pi^1 = (+1 \ +7 \ +3 \ -2 \ +6 \ +4 \ +5 \ +8) \\
&\quad \underbrace{\hspace{1.5cm}} \quad \uparrow \\
\pi^3 &= t(2,4,6) \cdot \pi^2 = (+1 \ -2 \ +6 \ +7 \ +3 \ +4 \ +5 \ +8) \\
&\quad \underbrace{\hspace{1.5cm}} \quad \uparrow \\
\pi^4 &= t(3,5,8) \cdot \pi^3 = (+1 \ -2 \ +3 \ +4 \ +5 \ +6 \ +7 \ +8) \\
\pi^5 &= r(2,2) \cdot \pi^4 = (+1 \ +2 \ +3 \ +4 \ +5 \ +6 \ +7 \ +8)
\end{aligned}$$

Figura 1.11: Exemplo de transformação de um cromossomo em outro por meio de reversões e transposições.

uma série de translocações $Tr_1, Tr_2, \dots, Tr_\tau$ tal que $Tr_\tau \dots Tr_2 \cdot Tr_1 \cdot \Pi = \Gamma$ e τ é mínimo. Chamamos τ de **distância de translocação entre Π e Γ** , $d_{Tr}(\Pi, \Gamma)$.

Observamos que também no caso das translocações, o conhecimento das orientações dos blocos de genes não altera as definições anteriores, pois a mutação envolve apenas mudanças de locais, não afetando as orientações dos blocos de genes.

1.1.5 Distância de reversão e transposição

Podemos estudar a transformação de um genoma em outro por reversões e transposições (Figura 1.11). Utilizando as definições das seções anteriores, enunciamos dois problemas que modelam os eventos de reversão e transposição agindo em cromossomos lineares.

Dadas as permutações com sinais π e γ , o **problema da distância de reversão e transposição** é encontrar uma série de eventos $e_1, \dots, e_{\varrho\delta}$, sendo cada um destes eventos uma reversão ou transposição, tal que $e_{\varrho\delta} \cdot \dots \cdot e_1 \cdot \pi = \gamma$ e $\varrho\delta$ é mínimo. Chamamos $\varrho\delta$ de **distância de reversão/transposição entre π e γ** , $d_{rt}(\pi, \gamma)$.

Novamente, supondo que σ seja a permutação identidade $\iota = (+1 \ +2 \ \dots \ +n)$, o problema de **ordenar π por reversões e transposições** é encontrar a distância de reversão/transposição, $d_{rt}(\pi, \iota)$.

Para permutações sem sinais, o enunciado do problema é inteiramente análogo, lembrando que as reversões invertem apenas a ordem dos elementos, ficando as transposições definidas exatamente da mesma forma.

1.1.6 Distância de reversão, transposição e transversão

Podemos estudar a transformação de um genoma em outro por reversões, transposições e transversões. Utilizando as definições das seções anteriores, enunciamos dois problemas que modelam os eventos de reversão, transposição e transversão agindo em cromossomos lineares.

Dadas as permutações com sinais π e γ , o **problema da distância de reversão, transposição e transversão** é encontrar uma série de eventos $e_1, \dots, e_{\rho\delta\bar{\delta}}$, sendo cada um destes eventos uma reversão, uma transposição ou uma transversão, tal que $e_{\rho\delta\bar{\delta}} \cdot \dots \cdot e_1 \cdot \pi = \gamma$ e $\rho\delta\bar{\delta}$ é mínimo. Chamamos $\rho\delta\bar{\delta}$ de **distância de reversão/transposição/transversão** entre π e γ , $d_{rt\bar{t}}(\pi, \gamma)$.

Novamente, supondo que σ seja a permutação identidade $\iota = (+1 +2 \dots +n)$, o problema de **ordenar π por reversões, transposições e transversões** é encontrar a distância de reversão/transposição/transversão, $d_{rt\bar{t}}(\pi, \iota)$.

Para permutações sem sinais, o enunciado do problema é inteiramente análogo, lembrando apenas que as reversões e transversões invertem apenas a ordem dos elementos, ficando as transposições definidas exatamente da mesma forma.

1.1.7 Distância de reversão e translocação

Conforme pesquisas em Biologia Molecular, homens e camundongos possuem muitas semelhanças na organização geral de seus genomas [11, 31]. O mapeamento físico comparativo homem-camundongo começou há cerca de vinte anos atrás e atualmente existem mais de 1300 pares de genes homólogos mapeados entre estas espécies. Como resultado, os biólogos acreditam que genes relacionados entre homens e camundongos não estão caoticamente distribuídos pelos genomas mas formam “blocos conservados” (grupos sintênicos).

Dados fornecidos por mapeamento físico indicam que os genomas humanos e do camundongo combinam em aproximadamente 150 blocos de genes, mas encontram-se embaralhados em humanos quando comparados aos do camundongo. Rearranjos dos blocos acontecem muito raramente (cerca de uma vez em um milhão de anos). Torna-se possível portanto reconstruir um cenário de rearranjo para a evolução humana e a

do camundongo a partir de um ancestral comum.

Nadeau e Taylor [31] realizaram a primeira tentativa computacional para analisar eventos de rearranjo em genomas de mamíferos, utilizando exatamente cromossomos humanos e do camundongo. Estes pesquisadores estimaram que aproximadamente 178 ± 39 eventos de rearranjo aconteceram desde que a separação das linhagens levando aos homens e aos camundongos aconteceu há 80 milhões de anos atrás. Esta estimativa foi confirmada por Copeland e co-autores [11], com base num mapa de ligação genética de resolução muito melhor quando comparado àquele disponível há dez anos atrás. De acordo com estes trabalhos, sabe-se que translocações e reversões são eventos comuns de rearranjo em evolução de mamíferos.

Em seguida é apresentada a formulação combinatorial do problema que modela os eventos de reversão e translocação para cromossomos lineares.

Dados dois genomas Π e Γ com o mesmo número de cromossomos, queremos obter a menor sequência de reversões e translocações que transforma Π em Γ .

Seja um genoma $\Pi = (\pi(1), \dots, \pi(N))$ consistindo de N cromossomos e seja o k -ésimo cromossomo $\pi(k) = (\pi(k)_1 \dots \pi(k)_{n_k})$, onde n_k é o número de genes do k -ésimo cromossomo. Para simplificar a notação, consideraremos π e σ como $\pi(k)$ e $\pi(l)$ para $k \neq l$. Seja $\pi = (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_j \pi_{j+1} \dots \pi_n)$ um cromossomo com $1 \leq i \leq j \leq n$. Uma **reversão** $r(\pi, i, j)$ num cromossomo π rearranja os blocos de genes de π , transformando π em $(\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n)$. Sejam $\pi = (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_n)$ e $\sigma = (\sigma_1 \dots \sigma_{j-1} \sigma_j \dots \sigma_m)$ dois cromossomos. Uma **translocação** $Tr(\pi, \sigma, i, j)$, com $1 \leq i \leq n+1, 1 \leq j \leq m+1$, troca blocos de genes entre os dois cromossomos π e σ e os transforma em dois cromossomos diferentes $(\pi_1 \dots \pi_{i-1} \sigma_j \dots \sigma_m)$ e $(\sigma_1 \dots \sigma_{j-1} \pi_i \dots \pi_n)$ com $(i-1) + (m-j+1)$ e $(j-1) + (n-i+1)$ blocos de genes respectivamente. Denotamos $e \cdot \Pi$ como o genoma obtido de Π como resultado de um evento de rearranjo e , onde e é uma reversão ou uma translocação.

Distinguimos entre **reversões internas** que não envolvem os limites dos cromossomos (isto é as reversões $r(\pi, i, j)$ de um cromossomo com n genes com $1 < i < j < n$) e **reversões prefixadas** envolvendo limites dos cromossomos (isto é, $i = 1$ ou $j = n$). Notemos que uma translocação $Tr(\pi, \sigma, n+1, 1)$ concatena os cromossomos π e σ , resultando num cromossomo $\pi_1 \dots \pi_n \sigma_1 \dots \sigma_m$ e num cromossomo vazio $\emptyset$. Esta translocação especial levando a uma redução no número de cromossomos não-vazios é conhecida em Biologia Molecular como **fusão**. A translocação $\varrho(\pi, \emptyset, i, 1)$ para $1 < i \leq n$ “quebra” um cromossomo π em dois cromossomos $(\pi_1 \dots \pi_{i-1})$ e $(\pi_i \dots \pi_n)$. Esta translocação levando a um aumento no número de cromossomos não-vazios é chamada de **fissão**. Uma translocação é **recíproca** ou **interna** se não é fusão nem fissão. Fusões e fissões são comuns em evolução de mamíferos. Por exemplo, a única

$$\Pi = (\pi_1 = (-3 \ -2 \ -1 \ +4 \ +5 \ +6 \ +7 \ +11) \ \pi_2 = (\ +9 |) \ \pi_3 = (| \ +10 \ +8))$$

$$\Gamma = (\gamma_1 = (+1 \ +2 \ +3 \ +4) \ \gamma_2 = (+5 \ +6 \ +7 \ +8) \ \gamma_3 = (+9 \ +10 \ +11))$$

fusão:

$$\Pi \stackrel{1}{=} \text{Tr} (\pi_2, \pi_3, 2, 1) . \Pi = ((-3 \ -2 \ -1 \ +4 | +5 \ +6 \ +7 \ +11) (+9 \ +10 \ +8))$$

fissão:

$$\Pi \stackrel{2}{=} \text{Tr} (\pi_1, \phi, 5, 1) . \Pi \stackrel{1}{=} ((-3 \ -2 \ -1 \ +4) (+5 \ +6 \ +7 | +11) (+9 \ +10 | +8))$$

translocação recíproca:

$$\Pi \stackrel{3}{=} \text{Tr} (\pi_2^2, \pi_3^2, 4, 3) . \Pi \stackrel{2}{=} ((\underline{-3 \ -2 \ -1} \ +4) (+5 \ +6 \ +7 \ +8) (+9 \ +10 \ +11))$$

reversão:

$$\Pi \stackrel{4}{=} r (\pi_1^3, 1, 3) . \Pi \stackrel{3}{=} ((+1 \ +2 \ +3 \ +4) (+5 \ +6 \ +7 \ +8) (+9 \ +10 \ +11)) = \Gamma$$

Figura 1.12: Exemplo de transformação de um genoma em outro por meio de diferentes eventos de rearranjo.

diferença na organização geral dos genomas de homens e chimpanzés é a fusão dos cromossomos 12 e 13 do chimpanzé no cromossomo 2 do homem.

A Figura 1.12 ilustra quatro eventos de rearranjo transformando um genoma em outro.

Dados dois genomas Π e Γ , o **problema da distância de reversão e translocação** é encontrar uma série de eventos $e_1, \dots, e_{\varrho\tau}$, sendo cada um destes eventos uma reversão ou translocação, tal que $e_{\varrho\tau} \cdot \dots \cdot e_2 \cdot e_1 \cdot \Pi = \Gamma$ e $\varrho\tau$ é mínimo. Chamamos $\varrho\tau$ de **distância de reversão/translocação** entre Π e Γ , $d_{rT}(\Pi, \Gamma)$.

1.1.8 Distância de transposição e translocação

Podemos estudar a transformação de um genoma em outro por transposições e translocações. Utilizando as definições das seções anteriores, podemos enunciar o problema que modela os eventos de transposição e translocação para cromossomos lineares.

Dados dois genomas Π e Γ com o mesmo número de cromossomos, queremos obter a menor sequência de transposições e translocações que transforma Π em Γ .

Seja um genoma $\Pi = (\pi(1), \dots, \pi(N))$ consistindo de N cromossomos e seja o k -ésimo cromossomo $\pi(k) = (\pi(k)_1 \dots \pi(k)_{n_k})$, onde n_k é o número de genes do k -ésimo cromossomo. Para simplificar a notação, consideraremos π e σ como $\pi(k)$ e $\pi(l)$ para $k \neq l$. Seja $\pi = (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_j \pi_{j+1} \dots \pi_n)$ um cromossomo com $1 \leq i \leq j \leq n$. Uma **transposição** $t(\pi, i, j, k)$, com $1 \leq i < j < k \leq n+1$, troca blocos de genes de π , transformando π em $(\pi_1 \dots \pi_{i-1} \pi_j \dots \pi_{k-1} \pi_i \dots \pi_{j-1} \pi_k \dots \pi_n)$. Sejam $\pi = (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_n)$ e $\sigma = (\sigma_1 \dots \sigma_{j-1} \sigma_j \dots \sigma_m)$ dois cromossomos. Uma **translocação** $Tr(\pi, \sigma, i, j)$, com $1 \leq i \leq n+1, 1 \leq j \leq m+1$, troca blocos de genes entre os dois cromossomos π e σ e os transforma em dois cromossomos diferentes $(\pi_1 \dots \pi_{i-1} \sigma_j \dots \sigma_m)$ e $(\sigma_1 \dots \sigma_{j-1} \pi_i \dots \pi_n)$ com $(i-1)+(m-j+1)$ e $(j-1)+(n-i+1)$ blocos de genes respectivamente. Denotamos $e \cdot \Pi$ como o genoma obtido de Π como resultado de um evento de rearranjo e , onde e é uma transposição ou uma translocação.

Dados dois genomas Π e Γ , o **problema da distância de transposição e translocação** é encontrar uma série de eventos $e_1, \dots, e_{\delta\tau}$, sendo cada um destes eventos uma transposição ou translocação, tal que $e_{\delta\tau} \cdot \dots \cdot e_2 \cdot e_1 \cdot \Pi = \Gamma$ e $\delta\tau$ é mínimo. Chamamos $\delta\tau$ de **distância de transposição/translocação** entre Π e Γ , $d_{tTr}(\Pi, \Gamma)$.

1.2 Direções de pesquisa

As tabelas abaixo apresentam uma síntese dos resultados e algoritmos encontrados na literatura para resolver alguns dos problemas enunciados na seção anterior. Para cada um destes problemas, indicamos primeiro a sua complexidade, e em seguida uma lista dos trabalhos relacionados e comentários específicos. As posições da tabela que estão em **negrito** indicam que contribuições ao problema foram dadas nesta tese.

eventos	sem sinais	com sinais
reversão	NP-difícil (1)	P (2)
transposição	complexidade em aberto, casos particulares e algoritmos de aproximação (3)	complexidade em aberto (4)
transversão	complexidade em aberto (15)	complexidade em aberto (15)
reversão e transposição	complexidade em aberto, algoritmo de aproximação (5)	complexidade em aberto, aplicações e algoritmo de aproximação (6)
reversão, transposição e transversão	complexidade em aberto (15)	complexidade em aberto (7)

Tabela 1.1: Genomas com um cromossomo linear

eventos	sem sinais	com sinais
reversão	NP-difícil (8)	P (9)
transposição	complexidade em aberto, casos particulares (10)	complexidade em aberto (4)
transversão	complexidade em aberto (15)	complexidade em aberto (15)
transposição e reversão	complexidade em aberto (15)	complexidade em aberto (15)
reversão, transposição e transversão	complexidade em aberto (15)	complexidade em aberto (15)

Tabela 1.2: Genomas com um cromossomo circular

eventos	sem sinais	com sinais
translocação	complexidade em aberto, algoritmo de aproximação (11)	P (12)
reversão e translocação	complexidade em aberto, algoritmo de aproximação (13)	P (14)
transposição e translocação	complexidade em aberto (15)	complexidade em aberto (15)

Tabela 1.3: Genomas multi-cromossomais, onde cada cromossomo é linear

1. Sob o enfoque de teoria dos grupos, Even e Goldreich [12], e Jerrum [22] apresentaram resultados considerando que reversões geram o grupo simétrico S_n .

Admitindo reversões com certas restrições, citamos os trabalhos de Gates e Papadimitriou [13], e Aigner e West [1]. Para reversões sem restrições, Kececioğlu e Sankoff [25], Bafna e Pevzner [2], Hannenhalli e Pevzner [21] e Christie [9] desenvolveram algoritmos de aproximação para o problema. Kececioğlu e Sankoff [25] e Caprara, Lancia e Ng [7] desenvolveram algoritmos exatos branch-and-bound para este problema.

Caprara [6] apresentou a prova de que este problema é NP-difícil.

Métodos para reconstrução de árvores filogenéticas considerando evolução de organismos por reversões foram propostos por Sankoff [35], Bafna e Pevzner [3], e Sankoff e co-autores [38].

2. Bafna e Pevzner [2] apresentaram um algoritmo de aproximação para este problema. Hannenhalli e Pevzner [19] apresentaram o primeiro algoritmo polinomial para o problema com complexidade de tempo $O(n^4)$. Baseados neste algoritmo, Berman e Hannenhalli [5] apresentaram um algoritmo com complexidade de tempo $O(n^2\alpha(n))$, onde $\alpha(n)$ é a inversa da função de Ackerman. Kaplan e co-autores [23] propuseram um algoritmo com a menor complexidade de tempo até este momento $O(n^2)$, também com base na teoria desenvolvida por Hannenhalli e Pevzner [19], e indicaram a possibilidade de existir algoritmo mais eficiente para este problema.

Hannenhalli e co-autores [17] propuseram um método para reconstrução de árvores filogenéticas, utilizando a distância de reversão computada pelo algoritmo de Hannenhalli e Pevzner [19].

3. Aigner e West [1], e Jerrum [22] apresentaram variantes deste problema que podem ser resolvidas em tempo polinomial, respectivamente, transposição restrita à reinserção do primeiro elemento $r(1, 2, i)$, e transposição restrita a elementos adjacentes $r(i, i+1, i+2)$. Meidanis, Walter e Dias [30], e de forma independente Christie [10], calcularam a distância de transposição entre uma permutação e sua inversa.

Sem restringir a forma das transposições, Bafna e Pevzner [4] e Christie [10] apresentaram algoritmos de aproximação para o problema. Nesta tese, apresentamos um novo algoritmo de aproximação para o problema.

A complexidade deste problema ainda é desconhecida.

4. Para que o problema possa ser definido, os blocos de genes correspondentes nas duas permutações devem ter a mesma orientação. Se este for o caso, notamos que os algoritmos desenvolvidos para o problema da distância de transposição de permutações sem sinais aplicam-se a este problema sem modificações.

5. Nesta tese apresentamos um algoritmo de aproximação de razão 3 para este problema.
6. Hannenhalli e co-autores [17] propuseram um método para reconstrução de árvores filogenéticas para este problema, utilizando a distância de reversão computada pelo algoritmo de Hannenhalli e Pevzner [19], e a distância de transposição computada pelo algoritmo de Bafna e Pevzner [4].

Nesta tese, apresentamos um algoritmo de aproximação para este problema, notando-se que parte deste trabalho já foi publicado [40].

7. Gu, Peng e Sudborough [14] apresentaram um algoritmo de aproximação para este problema.
8. Para este problema, Watterson e co-autores [42] propuseram dois algoritmos: um de aproximação e outro baseado num modelo estocástico, e Sankoff e co-autores [37] apresentaram um modelo probabilístico.

Bafna e Pevzner [3], em trabalho já citado no item anterior, propuseram um método para reconstruir árvores filogenéticas considerando também permutações circulares.

A prova de Caprara [6], de que ordenação por reversões de permutações lineares sem sinais é NP-difícil, estende-se também para permutações circulares sem sinais.

9. Kececioğlu e Sankoff [26] apresentaram um algoritmo exato branch-and-bound para este problema, que levou a limites bastante precisos.

Nesta tese, mostramos que o problema da distância de reversão de cromossomo com sinais lineares e circulares são equivalentes. Portanto, utilizando um algoritmo polinomial para o problema linear, temos imediatamente um algoritmo polinomial para o problema circular. Parte deste estudo já foi publicado [29].

10. Jerrum [22] apresentou um algoritmo polinomial para transposição restrita a elementos adjacentes.
11. Para tentar obter uma solução aproximada para este problema, Hannenhalli [16] propôs uma maneira de utilizar o algoritmo polinomial para resolver o problema da distância de translocação de permutações lineares com sinais apresentado neste mesmo artigo. A complexidade deste problema ainda não é conhecida.
12. Para este problema, Hannenhalli [16] apresentou um algoritmo polinomial com complexidade de tempo $O(n^3)$, utilizando basicamente os conceitos desenvolvidos por Hannenhalli e Pevzner [19], para resolver o problema de ordenar por

reversões permutações lineares com sinais. Pode-se investigar se é possível encontrar algoritmo mais eficiente para este problema.

13. Kececioğlu e Ravi [24] apresentaram um algoritmo de aproximação para este problema, mas a sua complexidade ainda não é conhecida.
14. Kececioğlu e Ravi [24] apresentaram um algoritmo de aproximação para este problema. Hannenhalli e Pevzner [20] apresentaram um algoritmo polinomial com complexidade de tempo $O(n^4)$, que utiliza o algoritmo de ordenação por reversões de permutações lineares com sinais de Hannenhalli e Pevzner [19]. Pode-se investigar se é possível encontrar algoritmo mais eficiente para este problema.
15. Não encontramos referências de trabalhos desenvolvidos para este problema.

Encontramos ainda na literatura trabalhos na área de rearranjo de genomas, que definem outras operações diferentes daquelas encontradas na área de Biologia Molecular. Citamos o artigo de Christie [8], que estuda o problema da ordenação de permutações por uma operação chamada por ele de “troca de blocos”, que troca blocos não necessariamente adjacentes de elementos da permutação.

1.3 Objetivos da tese

Os objetivos gerais desta tese são propor algoritmos para os problemas de distância relacionados aos eventos de reversão e transposição agindo num único cromossomo, e computar os diâmetros (maior distância entre dois cromossomos quaisquer) para os problemas de distância investigados.

Para estes problemas de distância de rearranjo de genomas, nossa intenção é desenvolver algoritmos, que possam ter utilidade para os biólogos moleculares. Calcular o diâmetro visa obter a informação de, dado um valor de distância entre dois genomas, quanto estes dois organismos estão “próximos” em termos de evolução, no sentido de que quanto mais próximo o valor da distância estiver do valor do diâmetro menor o seu relacionamento em termos evolutivos.

Para isto, inicialmente, no Capítulo 2 apresentamos estruturas e conceitos básicos que vêm sendo utilizados na literatura que trata destes dois eventos e que serão utilizados nos outros capítulos.

No Capítulo 3 mostramos uma equivalência entre os problemas de distância de reversão de permutações circulares e de permutações lineares. Como consequência

temos um algoritmo polinomial para o caso circular, bastando apenas fornecer determinadas permutações de entrada para um algoritmo polinomial que resolve o problema linear. Além disto, mostramos o valor exato do diâmetro de reversão de permutações lineares e circulares.

No Capítulo 4 apresentamos um algoritmo de aproximação com razão 2.25 para o problema da distância de transposição de permutações lineares sem sinais, que embora tendo uma razão elevada em comparação ao melhor conhecido, que tem razão 1.5, é simples de implementar e mostrou uma razão bem mais baixa quando utilizado em experimentos, sugerindo que possa ser útil na prática. Ainda, mostramos um limite inferior para o diâmetro de transposição de permutações lineares sem sinais, calculando a distância de transposição entre uma permutação e sua inversa.

No Capítulo 5 mostramos algoritmos de aproximação para o problema da distância de reversão e transposição de permutações sem e com sinais. Mostramos também um limite inferior para o diâmetro de reversão e transposição de permutações lineares com sinais, por meio do cálculo da distância entre duas particulares permutações.

Finalmente, no Capítulo 6 apresentamos conclusões deste trabalho e sugerimos possíveis linhas de pesquisa.

Capítulo 2

Estruturas e conceitos básicos

Neste capítulo, visamos apresentar estruturas e conceitos básicos que vêm sendo empregados comumente na literatura referente aos eventos de reversão e transposição, e que serão utilizados nos próximos capítulos.

Inicialmente, na Seção 2.1 apresentamos uma estrutura denominada de *grafo de pontos-de-quebra*, criada por Bafna e Pevzner [2], para solucionar o problema da distância de reversão, e ainda determinados resultados baseados nela. Em seguida, na Seção 2.2, apresentamos outra estrutura, o *grafo de ciclos*, também criada por Bafna e Pevzner [4], e utilizada num algoritmo de aproximação para o problema da distância de transposição de permutação sem sinais, além de outros resultados baseados nesta estrutura.

2.1 O grafo de pontos-de-quebra

O *grafo de pontos-de-quebra* é uma estrutura criada por Bafna e Pevzner [2], e utilizada por eles em algoritmos de aproximação para os problemas de distância de reversão de permutações sem e com sinais. Esta mesma estrutura foi usada num algoritmo polinomial para o problema da distância de reversão de permutações com sinais, proposto por Hannenhalli e Pevzner [19].

O **grafo de pontos-de-quebra** de duas permutações com sinais contendo n blocos de genes $\pi = (\pi_1 \dots \pi_n)$ e $\sigma = (\sigma_1 \dots \sigma_n)$, denotado por $G(\pi, \sigma)$, é construído como se segue.

Cada um dos números inteiros com sinais π_i é representado por dois rótulos, $-\pi_i$ e $+\pi_i$, nesta ordem, se π_i tem sinal $+$, e $+\pi_i$ e $-\pi_i$, nesta ordem, se π_i tem sinal $-$.

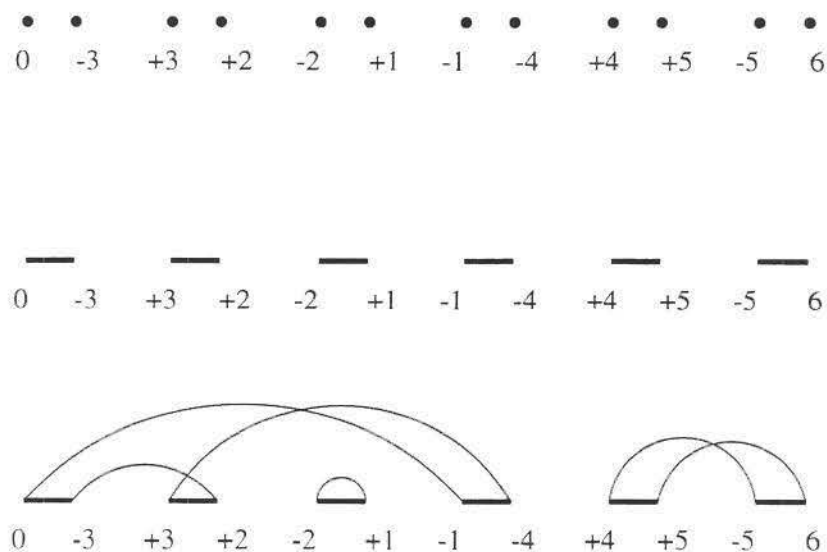


Figura 2.1: Exemplo da construção de um grafo de pontos-de-quebra para duas permutações com sinais $\pi = (+3 \ -2 \ -1 \ +4 \ -5)$ e $\sigma = (+1 \ +2 \ +3 \ +4 \ +5)$. Note que as arestas pretas estão em negrito e as arestas cinzas estão desenhadas com traço normal.

Cada um destes rótulos é um vértice do grafo. Além disso, adicionamos a π mais dois elementos, $\pi_0 = 0$ e $\pi_{n+1} = n + 1$, e acrescentamos ao grafo mais dois vértices, correspondentes a estes dois elementos, um à esquerda da sequência (rotulado por 0) e o outro à direita (rotulado por $n + 1$). Depois disso, criamos arestas *pretas* unindo dois rótulos de elementos adjacentes em π , e arestas *cinzas* unindo dois rótulos de elementos adjacentes em σ (Figura 2.1). Intuitivamente, as arestas pretas representam a situação de uma permutação em *relação* à outra tal como ela está, e que as arestas cinzas indicam a situação desejada.

Explicamos agora o motivo do nome desta estrutura. Primeiro, dadas duas permutações π e σ , um **ponto-de-quebra** de π com relação a σ indica um ponto em π que deverá ser necessariamente “quebrado” por uma reversão para que π possa ser

$$0 \bullet +3 \bullet -2 \quad -1 \bullet +4 \bullet -5 \bullet 6$$

Figura 2.2: Exemplo de pontos-de-quebra da permutação $\pi = (+3 \ -2 \ -1 \ +4 \ -5)$ com relação à $\sigma = (+1 \ +2 \ +3 \ +4 \ +5)$. O diagrama $G(\pi, \sigma)$ foi mostrado na figura anterior. Podemos verificar que os ciclos de tamanho maior do que 2 em $G(\pi, \sigma)$ indicam os pontos-de-quebra de π com relação a σ , e que o ciclo de tamanho 2 indica que não há pontos-de-quebra nos locais correspondentes de π relativamente à σ .

transformada em σ . Em termos do grafo de pontos-de-quebra $G(\pi, \sigma)$, a existência de um ciclo de tamanho maior do que 2 indica pontos-de-quebra em π com relação a σ . O ciclo de tamanho 2, formado por exatamente uma aresta preta e uma cinza, é uma indicação de que este não é um ponto-de-quebra em π com relação a σ (Figura 2.2). Neste ponto notamos que há diferentes definições de ponto-de-quebra, que dependem da permutação ter ou não sinais, e dos diferentes tipos de evento de rearranjo. Então, para facilitar a leitura, enunciamos estas definições nos capítulos em que são utilizadas.

O grafo de pontos-de-quebra é formado por uma coleção de ciclos. Cada um destes ciclos tem um número par de arestas, sendo metade delas arestas pretas e metade arestas cinzas. Denotamos por $c(\pi, \sigma)$ o número total de ciclos em $G(\pi, \sigma)$. Na Figura 2.1 temos $c(\pi, \sigma) = 3$.

Note que $c(\sigma, \sigma) = n + 1$, de tal forma que os ciclos são formados por exatamente duas arestas paralelas entre o mesmo par de rótulos. Como o grafo tem $2n + 2$ rótulos, teremos $n + 1$ ciclos neste caso. Além disso, $\pi = \sigma$ é a única permutação para a qual $c(\pi, \sigma) = n + 1$. Portanto, podemos visualizar o processo de computar a distância de reversão entre duas permutações lineares com sinais como o processo de transformar os $c(\pi, \sigma)$ ciclos de $G(\pi, \sigma)$ em $n + 1$ ciclos, isto é, devemos criar $(n + 1) - c(\pi, \sigma)$ ciclos de tamanho 2, da forma mais rápida possível.

Neste ponto, devemos investigar como uma reversão aplicada em π afeta os ciclos em $G(\pi, \sigma)$. Observemos primeiro que uma reversão é caracterizada por dois pontos que “cortam” a permutação, os quais correspondem cada um a uma aresta preta no grafo. O resultado seguinte mostra como um grafo de pontos-de-quebra é afetado por uma reversão. A Figura 2.3 mostra exemplos.

Lema 2.1.1 *Dadas duas permutações π e σ , e $G(\pi, \sigma)$, temos que, para qualquer reversão r ,*

$$c(r \cdot \pi, \sigma) = c(\pi, \sigma) + x,$$

onde $x = -1, 0$ ou 1 .

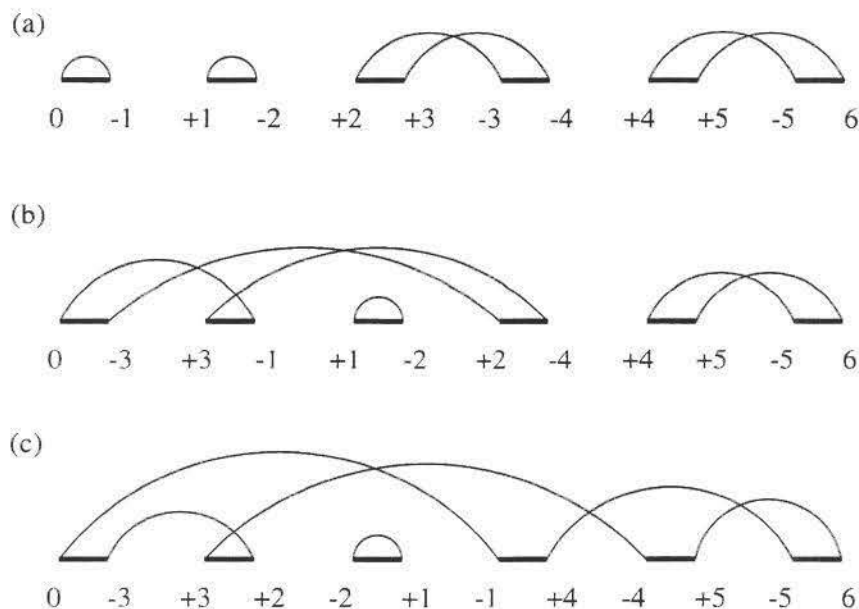


Figura 2.3: Exemplo de como uma reversão afeta os ciclos de $G(\pi, \sigma)$, sendo $\pi = (+3 \ -2 \ -1 \ +4 \ -5)$ e $\sigma = (+1 \ +2 \ +3 \ +4 \ +5)$. Notamos que o grafo $G(\pi, \sigma)$ está desenhado na primeira figura. (a) A reversão $r(1, 3)$ em π corresponde a aplicar a reversão no grafo na primeira e na quarta arestas pretas. A reversão aumenta o número de ciclos de 1. (b) A reversão $r(2, 3)$ em π corresponde a aplicar a reversão no grafo na segunda e quarta arestas pretas. Esta reversão não modifica o número de ciclos no grafo. (c) A reversão $r(4, 4)$ em π corresponde a aplicar a reversão no grafo na quarta e na quinta arestas pretas. A reversão diminui o número de ciclos de 1.

A consequência do resultado anterior é que qualquer reversão pode criar no máximo um ciclo. Estudando formas para criar ciclos no grafo de pontos-de-quebra, Hannenhalli e Pevzner [19] descobriram dois parâmetros que permitiram calcular exatamente a distância de reversão de permutações com sinais, e desenvolver um algoritmo polinomial para este problema. Passamos agora a explicar estes dois parâmetros, o *número de obstáculos* no grafo de pontos-de-quebra e um parâmetro que indica se o grafo é ou não uma *fortaleza*.

Para cada um dos ciclos de um grafo de pontos-de-quebra, podemos verificar se existe pelo menos uma reversão que pode ser aplicada no ciclo aumentando o número de ciclos de 1. Assim, definimos um ciclo c como sendo **ruim** quando, para qualquer reversão r agindo em duas arestas pretas de c , não aumentamos o número de ciclos, isto é, $c(\pi, \sigma) = c(r \cdot \pi, \sigma)$, para toda r aplicada em c . Caso contrário, definimos o ciclo como sendo **bom**.

Além disso, dizemos que dois ciclos se **entrelaçam** quando existem duas arestas

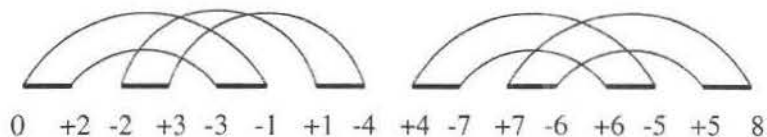


Figura 2.4: Exemplos de ciclos e componentes em $G(\pi, \sigma)$, para $\pi = (-2 \ -3 \ +1 \ +4 \ +7 \ +6 \ +5)$ e $\sigma = (+1 \ +2 \ +3 \ +4 \ +5 \ +6 \ +7)$. Como os dois primeiros ciclos são entrelaçados, eles formam uma componente, e de forma análoga, temos uma segunda componente contendo os dois últimos ciclos. Note que todos os ciclos são ruins, exceto o segundo, e portanto temos a primeira componente boa (contém um ciclo bom) e a segunda ruim (formada apenas por ciclos ruins).

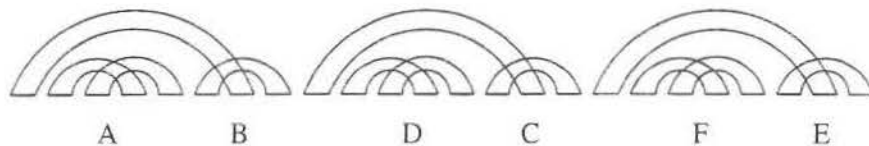


Figura 2.5: Exemplo de um grafo de pontos-de-quebra que não contém ciclos bons. Portanto, todas as componentes são ruins. (a) A componente B separa as componentes A e D . (b) As componentes A , D e F são obstáculos, enquanto que B , C e E são não-obstáculos. (c) O obstáculo A protege o não-obstáculo B , D protege C e F protege E . Portanto, A , D e F são super-obstáculos. (d) Como este grafo possui número ímpar (três) obstáculos e todos eles são super-obstáculos, então este grafo é uma fortaleza.

cinzas, uma de cada ciclo, que se cruzam.

Um conjunto de ciclos entrelaçados é denominado de uma **componente** de $G(\pi, \sigma)$. Chamamos uma componente de **ruim** quando todos os ciclos que a formam são ruins, e de componente **boa** quando pelo menos um dos ciclos que a formam é bom (Figura 2.4).

Uma componente ruim pode ainda ser classificada em *obstáculo* ou *não-obstáculo*, conceitos que passamos a explicar agora.

Uma componente A está **contida** numa outra componente B quando existe uma aresta cinza de um ciclo de B que contém inteiramente a componente A . Uma componente B **separa** duas componentes A e D se uma das duas está contida em B e a outra não está (Figura 2.5 (a)).

Dizemos ainda que uma componente ruim que não separa quaisquer duas outras componentes ruins é um **obstáculo**. Se uma componente ruim separa outras componentes então é chamada de **não-obstáculo** (Figura 2.5 (b)).

Nos nossos estudos, utilizaremos especificamente um tipo de obstáculo, que é uma

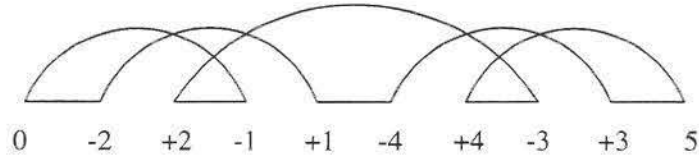


Figura 2.6: Exemplo de um grafo de pontos-de-quebra que contém um único ciclo ruim, que forma uma componente ruim, que é um obstáculo, mas não é uma fortaleza.

componente ruim composta por único ciclo ruim, que não se entrelaça nem contém qualquer outro ciclo (Figura 2.6). Como pode ser observado nos exemplos anteriores, estes não são os únicos tipos de obstáculos existentes num grafo de pontos-de-quebra, mas isto será suficiente para nossos objetivos. O número de obstáculos em um grafo $G(\pi, \sigma)$ será denotado por $h(\pi, \sigma)$.

Dizemos que um obstáculo A **protege** um não-obstáculo B quando a remoção de A torna B um obstáculo. Um obstáculo A é chamado de um **super-obstáculo** se ele protege um não-obstáculo B (Figura 2.5 (c)).

Um grafo $G(\pi, \sigma)$ é denominado de **fortaleza** se contém um número ímpar de obstáculos e todos eles são super-obstáculos (Figura 2.5 (d)).

Quando o grafo é uma fortaleza, necessita-se de uma reversão extra para calcular a distância de reversão. Portanto, Hannenhalli e Pevzner [19] definiram um parâmetro, denotado por $f(\pi, \sigma)$, cujo valor pode ser igual a 1 ou 0 apenas, dependendo do grafo de pontos-de-quebra ser, ou não, uma fortaleza, respectivamente. Para os nossos estudos, basta saber que numa fortaleza existe pelo menos um ciclo que não pertence a um obstáculo. Por exemplo, podemos observar que o grafo da Figura 2.6 não é uma fortaleza, pois é formado por apenas um ciclo que pertence a um obstáculo.

Utilizando como parâmetros o número de ciclos, o número de obstáculos e a indicação do grafo ser ou não uma fortaleza, Hannenhalli e Pevzner [19] apresentaram uma fórmula para computar a distância de reversão entre duas permutações com sinais π e σ :

$$d_r(\pi, \sigma) = (n + 1) - c(\pi, \sigma) + h(\pi, \sigma) + f(\pi, \sigma). \quad (2.1)$$

Podemos destacar os importantes trabalhos de Bafna e Pevzner, Hannenhalli e Pevzner e Kaplan, Shamir e Tarjan [2, 19, 23], ou o texto introdutório de Setubal e Meidanis [28] para uma explicação mais detalhada destes parâmetros.

Estes resultados serão utilizados no Capítulo 3.

$$\pi = (8 \ 5 \ 1 \ 4 \ 3 \ 2 \ 7 \ 6)$$

$$\sigma = (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8)$$

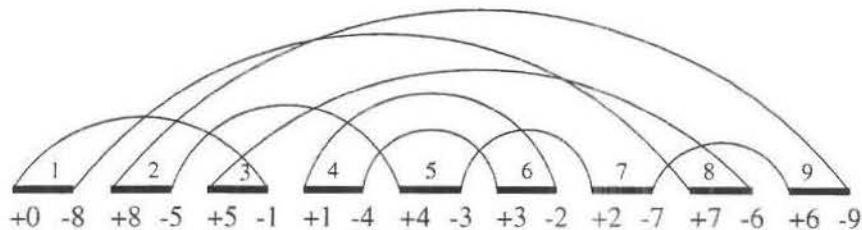


Figura 2.7: O grafo de ciclos para duas permutações, π e σ , como mostradas na figura. As arestas pretas estão representadas em negrito e as arestas cinza estão representadas por linhas normais.

2.2 O grafo de ciclos

Uma ferramenta poderosa para estudar a distância de transposição é o *grafo de ciclos* de duas permutações, proposto por Bafna e Pevzner [4], e utilizado em algoritmos de aproximação para este problema. Embora estes autores não o tenham descrito desta forma, podemos construí-lo de forma inteiramente análoga ao grafo de pontos-de-quebra descrito na seção anterior. Basta notar que, dadas duas permutações sem sinais π e σ , podemos tomar todos os elementos de π como sendo positivos, e portanto cada π_i vai ser associado aos rótulos $-\pi_i$ e $+\pi_i$.

O grafo de ciclos tem exatamente $n + 1$ arestas pretas e o mesmo número de arestas cinzas. Também como no caso das reversões, a idéia é que as arestas pretas indiquem a situação como ela está agora, e as arestas cinzas indiquem a situação desejada. Quando as arestas pretas tornam-se iguais às cinzas em todos os rótulos, temos $\pi = \sigma$ e $d_t(\pi, \sigma) = 0$. Então, nosso objetivo é aplicar transposições de tal forma que as arestas pretas tornem-se iguais às cinzas. A Figura 2.7 mostra o grafo correspondente a um par de permutações.

Bafna e Pevzner [4] apresentaram vários resultados interessantes baseados nesta estrutura, que passamos a apresentar.

Primeiro, como no caso das reversões, o grafo é composto por um certo número de ciclos, onde em cada ciclo alternam-se arestas pretas e arestas cinzas. O *comprimento* de um ciclo foi definido pelo número de arestas pretas que pertencem ao ciclo (que é o mesmo número de arestas cinzas). Um resultado importante segue.

Lema 2.2.1 *A soma dos comprimentos de todos os ciclos em qualquer grafo de ciclos é sempre $n + 1$.*

Além disso, uma transposição pode afetar o número de ciclos de uma forma muito específica, como mostra o seguinte lema. Denote por $c(\pi, \sigma)$ o número de ciclos no grafo de ciclos de π e σ .

Lema 2.2.2 *Para quaisquer permutações π e σ , e o grafo de ciclos correspondente, temos que, para qualquer transposição t ,*

$$c(t \cdot \pi, \sigma) = c(\pi, \sigma) + x,$$

onde $x = -2, 0$ ou 2 .

Uma transposição é chamada de -2 -movimento, de 0 -movimento ou de 2 -movimento, se x for $-2, 0$ ou 2 respectivamente, no lema anterior. Como $c(\sigma, \sigma) = n + 1$, o máximo possível, seria interessante termos tantos 2 -movimentos quanto possível.

De fato, um resultado mais específico pode ser verificado quanto ao efeito de uma transposição no grafo de ciclos. Denotemos por $c_{\text{impar}}(\pi, \sigma)$ o número de ciclos de paridade ímpar no grafo de ciclos de π e σ .

Lema 2.2.3 *Para quaisquer permutações π e σ e qualquer transposição t temos*

$$c_{\text{impar}}(t \cdot \pi, \sigma) = c_{\text{impar}}(\pi, \sigma) + x,$$

onde $x = -2, 0$ ou 2 .

Deste lema temos o seguinte limite inferior para a distância:

Teorema 2.2.4 *Para quaisquer permutações π e σ temos*

$$d(\pi, \sigma) \geq \frac{(n + 1) - c_{\text{impar}}(\pi, \sigma)}{2}$$

Bafna e Pevzner [4] mostraram ainda uma forma de representar um ciclo pelas suas arestas pretas. Primeiro, as arestas pretas do grafo de ciclos são numeradas associando um rótulo i à aresta preta (π_i, π_{i-1}) , com $1 \leq i \leq n + 1$, e assim as arestas pretas serão rotuladas de 1 até $n + 1$. Consideremos agora um ciclo c de tamanho k . Denotaremos c entre colchetes, tomando uma das arestas pretas como a primeira do ciclo, e as outras arestas pretas na ordem em que elas aparecem, seguindo as arestas cinzas de

$c, [i_1, \dots, i_k]$. Um ciclo c pode ser representado de k formas diferentes, dependendo da escolha da primeira aresta. Vamos escolher uma **representante canônica de um ciclo** c , tomando como aresta preta inicial i_1 a aresta mais à direita de c em π , isto é, $i_1 = \max_{1 \leq i \leq k} i_i$. No grafo de ciclos da Figura 2.7 temos três ciclos, com representantes canônicas $c_1 = [9, 7, 5, 2]$, $c_2 = [8, 1, 3]$ e $c_3 = [6, 4]$.

Escolhemos agora três arestas pretas x, y, z pertencendo ao mesmo ciclo c no grafo de ciclos. O ciclo c força uma ordem em x, y, z , e temos três possíveis representações desta ordem. Escolheremos como **representante canônica de uma tripla** (x, y, z) aquela começando na aresta preta mais à direita, isto é, na aresta preta rotulada por $\max(x, y, z)$. Uma tripla na ordem canônica é *não-orientada* se $x > y > z$ e *orientada* se $y < z < x$. No grafo de ciclos da Figura 2.7 temos as seguintes triplas não-orientadas: $(9, 7, 5)$, $(9, 7, 2)$ e $(7, 5, 2)$; e a tripla orientada $(8, 1, 3)$.

Dizemos ainda que um ciclo é **orientado** se ele admite um 2-movimento, e **não-orientado** se não existem 2-movimentos possíveis agindo nele. No grafo de ciclos da Figura 2.7 temos c_1 e c_3 não-orientados e c_2 orientado.

Finalmente, o Lema 2.2.2 sugere que qualquer algoritmo que visa computar a distância de transposição deve utilizar o máximo de 2-movimentos possível. Seguindo esta idéia, Bafna e Pevzner [4] mostraram uma forma de aumentar $c(\pi, \sigma)$ de pelo menos 2 em dois movimentos consecutivos, que passamos agora a descrever.

Uma transposição $t(i, j, k)$ **age num ciclo** c se as três arestas pretas i, j e k pertencem a c . O Lema 2.2.2 foi utilizado na prova do seguinte resultado.

Lema 2.2.5 *Se uma transposição t age num ciclo c e cria mais do que um novo ciclo no grafo de ciclos, então t é um 2-movimento. Se uma transposição t age em arestas pertencentes a ciclos diferentes então t é um 0-movimento.*

Tomando a representação canônica para ciclos, para $k > 1$, um ciclo $c = [i_1, \dots, i_k]$ é *não-orientado* se $i_1, \dots, i_k$ for uma sequência decrescente, e *orientado* em caso contrário.

Lema 2.2.6 *Se c for um ciclo orientado então existem 2-movimentos agindo em c . Se c for um ciclo não-orientado então não existem 2-movimentos agindo em c .*

Prova: Tome um ciclo orientado $c = [i_1, \dots, i_k]$, e $3 \leq v \leq k$ um índice tal que $i_v > i_{v-1}$. Uma transposição $t(i_{v-1}, i_v, i_1)$ agindo em c cria um 1-ciclo (na aresta preta (π_{i_v}, π_{i_v-2})). Então, pelo Lema 2.2.5, t é um 2-movimento. $\square$

Os Lemas 2.2.5 e 2.2.6 implicam o seguinte teorema.

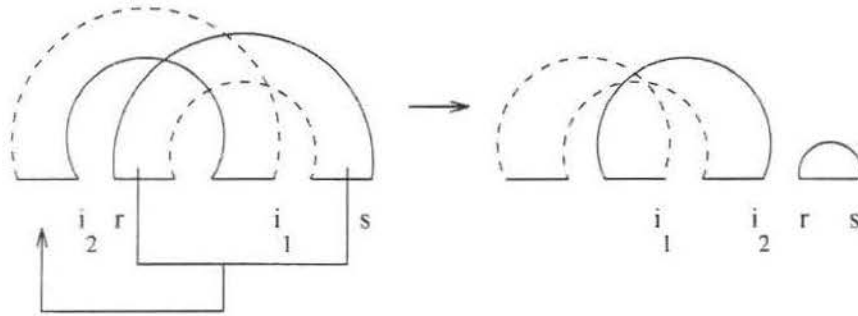


Figura 2.8: Um 0-movimento criando um ciclo orientado.

Teorema 2.2.7 *Para duas permutações π e σ existe um 2-movimento ou um 0-movimento seguido por um 2-movimento.*

Prova: Se $G(\pi, \sigma)$ tem um ciclo orientado então pelo Lema 2.2.6, é possível um 2-movimento.

Caso contrário, tomar um ciclo não-orientado $c = [i_1, \dots, i_k]$, e seja r a posição do elemento maximal de π com relação a σ no intervalo $[i_2, i_1 - 1]$. Seja s a posição de $\pi_r + 1$ em π com relação a σ , e note que $s \notin [i_2, i_1]$. Se $s > i_1$, a transposição $t(r + 1, s, i_2)$ age em arestas de dois ciclos diferentes, e portanto pelo Lema 2.2.5, t é um 0-movimento (Figura 2.8). Mas agora os dois ciclos criados por t são um 1-ciclo composto pela aresta preta (s, r) e um k -ciclo orientado, com $k \geq 3$. Um argumento análogo pode ser usado quando $s < i_2$. Neste caso use a transposição $t(s, i_2, r + 1)$.

□

Estes resultados serão utilizados nos Capítulos 4 e 5.

Capítulo 3

Distância de reversão de cromossomos com sinais

Neste capítulo estudamos o problema da distância de reversão de permutações com sinais. Na seção 3.1 mostramos uma equivalência entre os problemas da distância de reversão de permutações com sinais, lineares e circulares. Embora esta analogia seja razoavelmente simples, não consta da literatura da área, pelo menos na que conhecemos. Com base nesta equivalência, mostramos que um algoritmo polinomial para o caso circular pode utilizar qualquer algoritmo polinomial para computar a distância de reversão de permutações com sinais, bastando para isto fornecer determinadas permutações como entrada para o algoritmo escolhido. Na seção 3.2 calculamos o diâmetro de reversão de permutações com sinais, lineares e circulares. Na seção 3.3, inicialmente apresentamos alguns resultados relacionando a distância de reversão de permutações com sinais lineares e circulares, de mesmo tamanho. Em seguida, justificamos um procedimento comum em artigos onde há cálculos de distância de reversão de cromossomos lineares, que é o de fixar a priori uma das pontas das moléculas de DNA comum aos dois cromossomos. Finalmente, na seção 3.4, apresentamos um sumário do capítulo, além de algumas questões que surgiram a partir destes estudos.

3.1 Uma equivalência entre os problemas circular e linear

Um cromossomo circular, conforme visto no Capítulo 1, é uma molécula de DNA circular. A Figura 3.1 mostra exemplos de cromossomos circulares com sinais de duas espécies de plantas, onde cada número representa um bloco de genes, e a seta indica

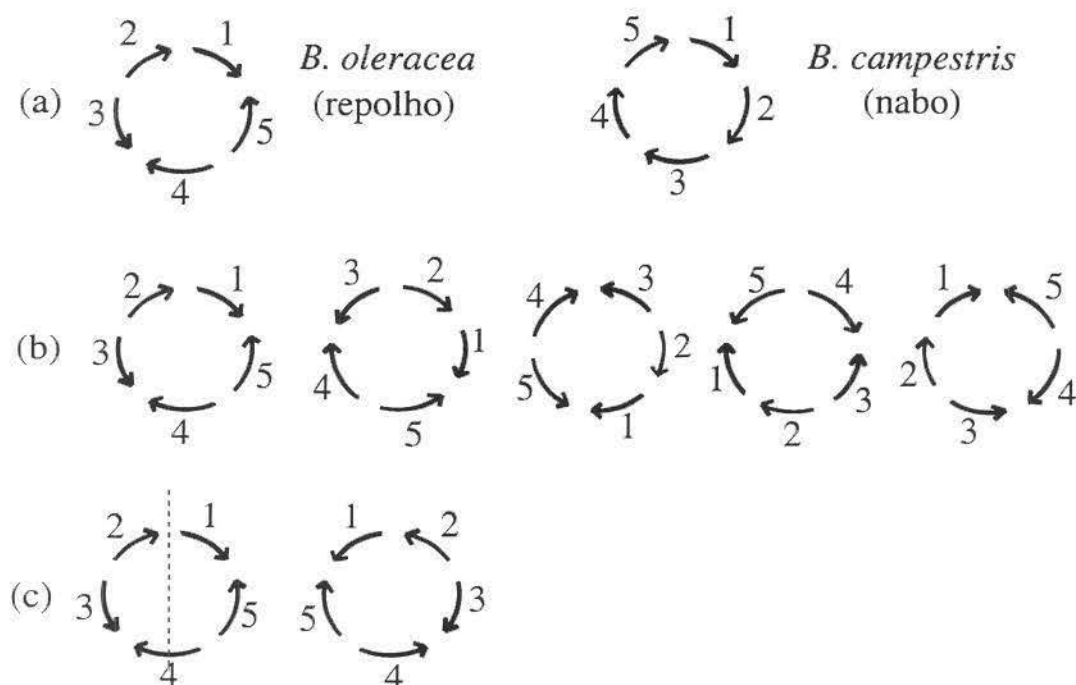


Figura 3.1: Exemplos de cromossomos circulares de duas espécies de plantas. (a) As setas indicam as orientações dos blocos de genes de uma espécie em relação à outra. (b) Diferentes representações do mesmo cromossomo circular. (c) As duas formas de visualizar os blocos de genes de um cromossomo circular, onde uma é obtida da outra por reflexão relativa ao eixo mostrado na figura. Estas duas formas são consideradas equivalentes.

a orientação de um bloco de genes de uma espécie em relação à outra.

Num cromossomo circular, uma reversão é definida fixando dois pontos de corte neste cromossomo, e revertendo a ordem dos genes em uma das regiões delimitadas por estes dois pontos (Figura 3.2).

De forma genérica, o *problema da distância de reversão de cromossomos circulares com sinais* é formulado da seguinte forma. Dados dois cromossomos circulares com sinais A e B , devemos encontrar a menor série de reversões que transforma A em B . A Figura 3.3 mostra um exemplo de um cromossomo circular transformado em outro com o menor número de reversões possível.

Nesta seção, apresentamos uma equivalência entre reversões circulares agindo em cromossomos circulares com sinais e reversões lineares agindo em cromossomos lineares com sinais. Para isto, inicialmente formalizamos um cromossomo circular por uma classe de equivalência, definimos reversões circulares agindo neste cromossomo

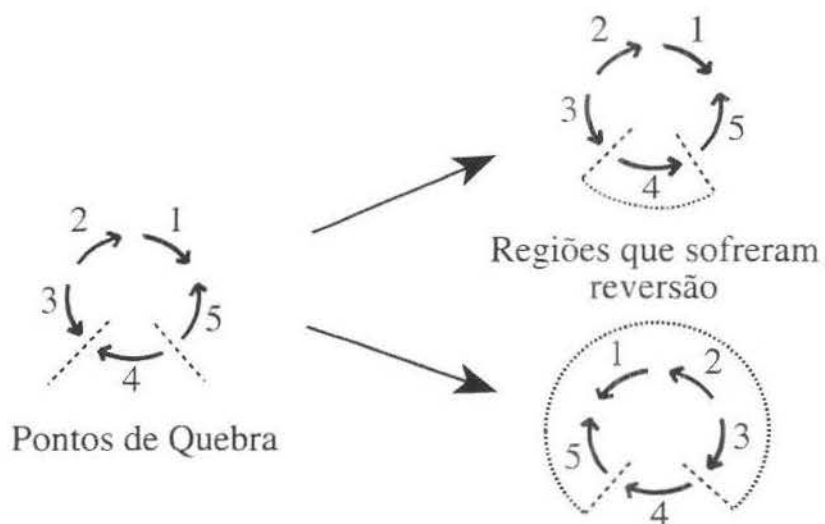


Figura 3.2: Existem duas possibilidades para aplicar uma reversão num cromossomo circular, dados os dois pontos de corte.

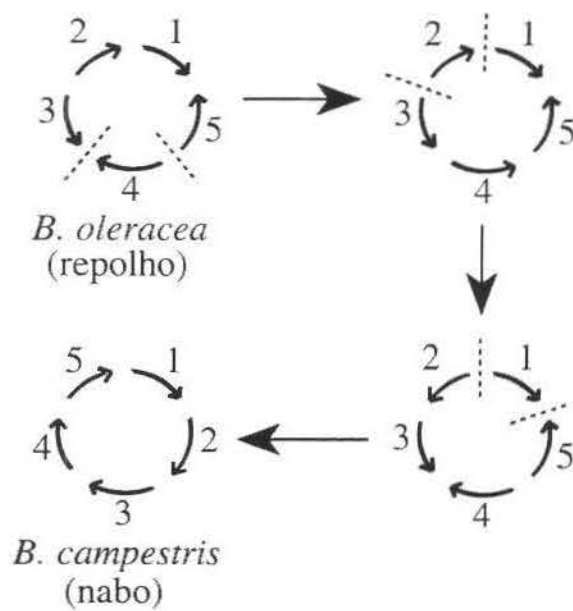


Figura 3.3: Exemplo de uma série mínima de reversões que transforma *B. oleracea* (repolho) em *B. campestris* (nabo).

circular, e definimos precisamente o problema. Utilizando estas definições, caracterizamos um isomorfismo entre reversões circulares agindo em cromossomos circulares e reversões lineares atuando em cromossomos lineares. Este isomorfismo permitiu computar a distância de reversão de cromossomos circulares por meio do cálculo da distância de reversão de cromossomos lineares com um bloco de genes a menos. Como consequência, obtivemos um algoritmo polinomial para o problema da distância de reversão de cromossomos circulares com sinais.

3.1.1 Uma formalização para o problema circular

Nesta seção inicialmente apresentamos formalismos para cromossomos circulares com sinais e reversões circulares, e enunciamos o problema da distância de reversão de cromossomos circulares com sinais.

Cromossomos Circulares

Apresentamos agora uma formalização de um cromossomo circular com sinais por uma classe de equivalência.

Intuitivamente, um cromossomo circular é um arranjo circular de blocos de genes com sinais (Figura 3.1). Um *bloco* de genes de um cromossomo será modelado por um inteiro com sinal. O sinal “+” indica uma seta na direção dos ponteiros do relógio na Figura 3.1, e o sinal “-” indica uma seta na direção contrária aos ponteiros do relógio. Dado um *bloco inicial* de genes, podemos representar um cromossomo circular por uma permutação como se segue. Seguimos ao longo do cromossomo, na direção dos ponteiros do relógio, começando no bloco inicial, e escrevendo os inteiros com sinais correspondentes aos blocos encontrados. Então, $(\pi_1 \pi_2 \dots \pi_n)$ denotará o cromossomo circular, com n blocos de genes, onde cada π_i tem o sinal “+” ou “-”, e π_1 é o bloco inicial de genes. Como um exemplo, o cromossomo de *B. oleracea* da Figura 3.1(a) pode ser representado pela permutação $(+1 - 5 + 4 - 3 + 2)$, se escolhermos +1 como o bloco inicial de genes.

Podemos escolher cada um dos blocos de genes como sendo o primeiro, e então podemos ter várias permutações diferentes representando o mesmo cromossomo. No entanto, todas as permutações são consideradas equivalentes (Figura 3.4). Além disso, duas permutações onde uma é obtida a partir da outra por reflexão são consideradas equivalentes, isto é, $(\pi_1 \pi_2 \dots \pi_n)$ e $(\bar{\pi}_n \bar{\pi}_{n-1} \dots \bar{\pi}_2 \bar{\pi}_1)$, onde $\bar{\pi}_j$, com $1 \leq j \leq n$, é o elemento π_j com o sinal invertido, são permutações consideradas equivalentes (Figura 3.5).

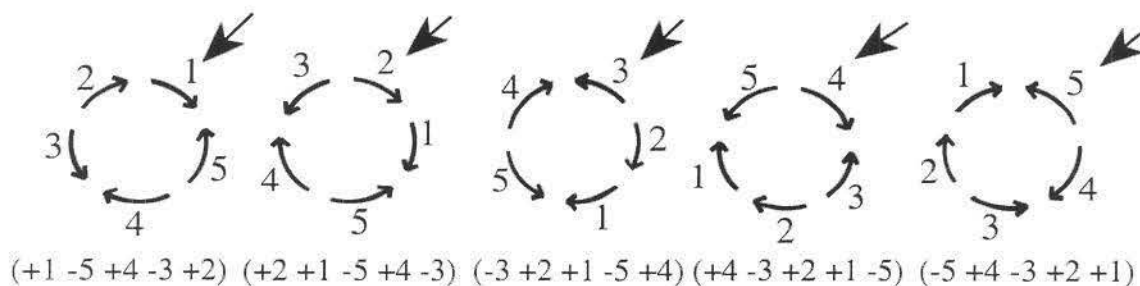


Figura 3.4: Num cromossomo circular podemos escolher cada um dos blocos de genes como sendo o primeiro. Então, todas estas permutações são consideradas equivalentes, e representam o cromossomo circular de *B. oleracea* relativamente a *B. campestris*, mostrados na Figura 3.1 (a).

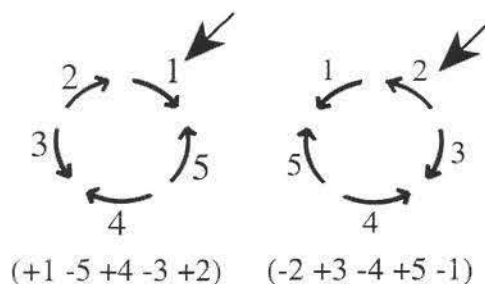


Figura 3.5: Num cromossomo circular duas permutações onde uma é obtida da outra por reflexão são consideradas equivalentes. O cromossomo circular representado é o de *B. oleracea* relativamente a *B. campestris*, mostrados na Figura 3.1 (a).

Desta forma, uma permutação de tamanho n , modelando um cromossomo circular com n blocos de genes, é uma representante de uma classe de equivalência, no conjunto de todas as permutações de tamanho n . Abaixo definimos as operações de *rotação* e de *reflexão*, que formalizarão as duas possibilidades descritas acima. A partir destas operações podemos definir uma relação de equivalência entre duas permutações, e também a classe de equivalência que representará o cromossomo circular.

Chamaremos de S_n o conjunto de todas as possíveis permutações com sinais, onde cada permutação tem tamanho n . Observemos que $|S_n| = 2^n n!$. Tomemos agora $\pi = (\pi_1 \pi_2 \dots \pi_n)$, uma permutação em S_n . Definiremos dois tipos de operações agindo em π como se segue:

- *Rotações*. Denotaremos por $\text{rot}(\pi)$ a rotação básica que move elementos da permutação π uma posição para a esquerda:

$$\text{rot}(\pi) = (\pi_2 \pi_3 \dots \pi_n \pi_1).$$

Definimos rot^i para cada $\pi \in Z$ da forma usual: rot^i é a composição de rot i vezes para $i > 0$ e rot^{n-i} é a inversa de rot^i . Além disso, rot^0 é a identidade. Temos as seguintes relações importantes:

$\text{rot}^n = \text{rot}^0$, ou mais genericamente, $\text{rot}^i = \text{rot}^j$ se $i \equiv j \pmod{n}$ para todo $i, j \in Z$.

$\text{rot}^i \text{rot}^j = \text{rot}^{i+j}$ para todo $i, j \in Z$.

As operações rot^i são chamadas de *rotações*.

- *Reflexões*. Denotaremos por $\text{refl}(\pi)$ a reflexão que inverte a ordem de todos os elementos de uma permutação π e também os seus sinais. Então,

$$\text{refl}(\pi) = (\bar{\pi}_n \bar{\pi}_{n-1} \dots \bar{\pi}_2 \bar{\pi}_1).$$

Definimos refl^i para cada $\pi \in Z$ da forma usual: refl^i é a composição de refl i vezes para $i > 0$ e refl^{i-1} é a inversa de refl^i . Notamos também que $\text{refl}^{-1} = \text{refl}^i$. Além disso, refl^0 é a identidade. Temos as seguintes relações importantes:

$\text{refl}^2 = \text{refl}^0$, ou mais genericamente, $\text{refl}^i = \text{refl}^j$ se $i \equiv j \pmod{2}$ para todo $i, j \in Z$.

$\text{refl}^i \text{refl}^j = \text{refl}^{i+j}$ para todo $i, j \in Z$.

Podemos aplicar rot e refl a uma permutação, usando as definições acima. Então,

$$\text{rot refl}(\pi) = \text{rot}(\text{refl}(\pi)) = \text{rot}(\bar{\pi}_n \bar{\pi}_{n-1} \dots \bar{\pi}_2 \bar{\pi}_1) =$$

$$(\bar{\pi}_{n-1} \bar{\pi}_{n-2} \dots \bar{\pi}_2 \bar{\pi}_1 \bar{\pi}_n).$$

Temos a seguinte relação:

$$\text{rot refl} = \text{refl rot}^{-1}. \quad (3.1)$$

Genericamente, as operações refl rot^i são chamadas de *reflexões*. Cada reflexão é igual à sua própria inversa.

Definiremos em seguida uma relação de equivalência entre duas permutações π e σ .

Definição 3.1.1 *Dadas duas permutações π e σ , definimos*

$$\pi \sim \sigma$$

se e somente se existem $i, j \in \mathbb{Z}$ tais que $\sigma = \text{rot}^i \text{refl}^j \pi$.

A relação acima é de equivalência. A prova do resultado é simples. A Equação 3.1 pode ser usada nesta prova.

Desta relação de equivalência, podemos definir a classe de equivalência da permutação π , denotada por $[\pi]$, que representa um cromossomo circular com sinais, como se segue

$$[\pi] = \{\sigma \in S_n | \pi \sim \sigma\}$$

Esta formalização é interessante do ponto de vista biológico, porque ela não fixa qualquer elemento da permutação, e então cada um dos blocos de genes pode ser o primeiro, bastando aplicar rotação. Além disso, duas permutações onde uma é obtida da outra por reflexão podem ser produzidas aplicando o operador refl .

Reversões Circulares

Denominaremos reversão agindo num cromossomo circular por **reversão circular**. Conforme notamos anteriormente, existem duas possibilidades para uma reversão agir num cromossomo circular, dados os dois pontos onde os cortes ocorreram (Figura 3.2). Modelamos agora como uma reversão circular agirá numa classe de equivalência A que representa um cromossomo circular.

Primeiro, escolhemos uma permutação $\pi = (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_j \pi_{j+1} \dots \pi_n)$ na classe A . Em seguida, vamos supor que os dois cortes da reversão circular tenham ocorrido em π entre $i \ominus 1, i$ e $j, j \oplus 1$, onde $1 \leq i \leq j \leq n$. Aqui $\ominus$ e $\oplus$ são as operações usuais

de subtração e adição, respectivamente, exceto que tomamos o resultado módulo n e escolhemos n no lugar de 0 como representante da classe dos múltiplos de n . Vamos supor que estes cortes sejam distintos, e então $i \neq (j \oplus 1)$.

Além disso podemos sempre escolher i e j tais que $i \leq j$. Se $i > j$ devemos escolher outros índices i' e j' que indiquem os mesmos pontos de corte e tais que $i' \leq j'$. Mas para isto, basta tomar $i' = j \oplus 1$ e $j' = i \oplus 1$, o que é sempre possível, pois por hipótese $i \neq (j \oplus 1)$.

Temos então o seguinte lema.

Lema 3.1.2 *Dada uma permutação $\pi = (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_j \pi_{j+1} \dots \pi_n)$ de uma classe de equivalência A que modela um cromossomo circular, e dois pontos de corte em π indicados por inteiros i e j , com $1 \leq i \leq j \leq n$ e $i \neq (j \oplus 1)$, tais que estes cortes ocorrem entre $i \oplus 1, i$, e $j, j \oplus 1$, as permutações resultantes das duas possíveis formas de reverter π com estes dois cortes pertencem à mesma classe de equivalência.*

Prova:

Denotaremos por P e Q as duas possíveis formas de reverter o cromossomo circular modelado por A (Figura 3.6). Tomando π , e aplicando a reversão circular P em π , com i e j nas condições do lema, temos

$$P \cdot (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_j \pi_{j+1} \dots \pi_n) = (\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n)$$

Aplicando a reversão circular Q em π , com i e j nas condições do lema, temos

$$Q \cdot (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_j \pi_{j+1} \dots \pi_n) = (\bar{\pi}_1 \bar{\pi}_n \dots \bar{\pi}_{j+1} \pi_i \dots \pi_j \bar{\pi}_{i-1} \dots \bar{\pi}_2)$$

Mas aplicando rot^{-1} e refl em $(\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n)$ temos

$$\text{rot}^{-1}(\text{refl}(\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n)) = (\bar{\pi}_1 \bar{\pi}_n \dots \bar{\pi}_{j+1} \pi_i \dots \pi_j \bar{\pi}_{i-1} \dots \bar{\pi}_2)$$

Então, tomando

$$\sigma_1 = (\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n) = P \cdot \pi$$

e

$$\sigma_2 = (\bar{\pi}_1 \bar{\pi}_n \dots \bar{\pi}_{j+1} \pi_i \dots \pi_j \bar{\pi}_{i-1} \dots \bar{\pi}_2) = Q \cdot \pi$$

temos que $\sigma_2 = \text{rot}^{-1}(\text{refl}(\sigma_1))$, implicando em $\sigma_1 \sim \sigma_2$, ou seja, σ_1 e σ_2 pertencem à mesma classe de equivalência. Temos assim que $[P \cdot \pi] = [\sigma_1] = [\sigma_2] = [Q \cdot \pi]$.

□

Definimos agora reversão circular.

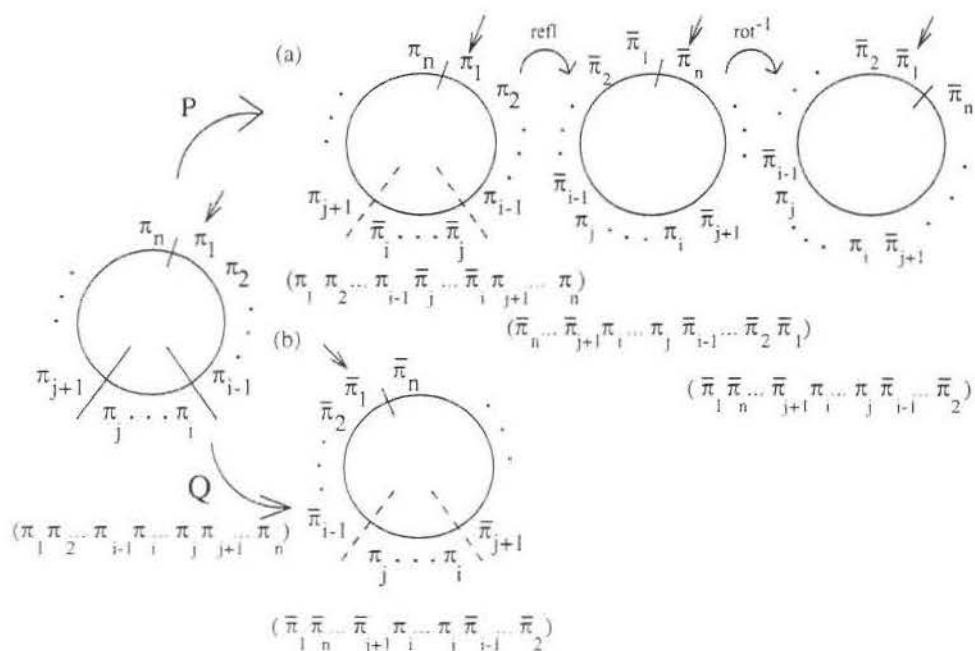


Figura 3.6: Os dois cromossomos circulares resultantes das duas formas de aplicar reversão circular são representados por duas permutações que pertencem à mesma classe de equivalência. Note que a seta, antes da reversão circular, indica o primeiro bloco da permutação escolhida da classe de equivalência que representa o cromossomo circular. A porção do cromossomo sofrendo a reversão circular pode incluir ou não a seta. As permutações resultantes das duas reversões circulares são mostradas. Podemos verificar que aplicar reflexão e rotação numa das permutações resultantes da reversão permite obter a outra permutação. Mostramos também as permutações resultantes de cada operação aplicada. (a) Neste caso, a reversão não inclui a seta. (b) Neste caso, a reversão inclui a seta.

Definição 3.1.3 Dada uma classe de equivalência A modelando um cromossomo circular, definimos uma **reversão circular** P agindo em A , denotada por $P \cdot A$, como sendo $[P \cdot \pi]$, $\pi \in A$, onde P inverte o mesmo bloco de elementos, qualquer que seja $\pi \in A$.

Agora podemos enunciar o problema de encontrar o menor número de reversões circulares agindo em cromossomos circulares com sinais.

Dadas duas classes de equivalência A e B , representando dois cromossomos circulares com sinais, o **problema da distância de reversão de cromossomos circulares com sinais** é encontrar uma série de reversões circulares $P_1, P_2, \dots, P_\varrho$ tais que $P_\varrho \cdot P_{\varrho-1} \cdot \dots \cdot P_2 \cdot P_1 \cdot A = B$ e ϱ é mínimo. Chamamos ϱ de **distância de reversão circular** de A e B , denotada por $d_r^c(A, B)$.

3.1.2 Um isomorfismo entre os problemas circular e linear

Como dito no início desta seção, definiremos reversão circular em termos de reversão linear, o que permitirá estabelecer um isomorfismo entre reversões circulares atuando em cromossomos circulares e reversões lineares agindo em cromossomos lineares. Deste isomorfismo deduzimos imediatamente um algoritmo polinomial para o problema da distância de reversão de cromossomos circulares com sinais, baseado num algoritmo polinomial que soluciona o problema equivalente para cromossomos lineares.

Uma reversão linear $r(i, j)$, com $1 \leq i \leq j \leq n$, como descrito no Capítulo 1, age na permutação $\pi = (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_j \pi_{j+1} \dots \pi_n)$, de tamanho n , da seguinte forma:

$$r(i, j) \cdot \pi = (\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n).$$

A definição de reversão circular P agindo numa classe de equivalência A induz imediatamente à utilização da reversão linear. Uma primeira intuição seria escolher em A uma permutação π , escolher os pontos de corte i e j , e utilizar estes mesmos índices numa reversão linear atuando em π , assim:

$$P \cdot A = [P \cdot \pi] = [r(i, j) \cdot \pi]$$

Entretanto esta definição não funciona corretamente, porque a mesma reversão linear aplicada a diferentes permutações π dentro de uma classe de equivalência A leva a classes de equivalência distintas. Para resolver este problema, não será permitida uma escolha aleatória de uma permutação em A na qual as reversões circular e linear atuarão.

Definimos uma *representante canônica* de A , denotada por $\text{can}(A)$, com as características de ter o bloco 1 fixado como sendo o primeiro elemento da permutação, e tendo a orientação + (Figura 3.7). Note que cada classe de equivalência tem uma única representante canônica.

$$\begin{aligned}
 [\pi] = \{ & (+1 -5 +4 -3 +2) (-5 +4 -3 +2 +1) (+4 -3 +2 +1 -5) \\
 & (-3 +2 +1 -5 +4) (+2 +1 -5 +4 -3) \\
 & (-2 +3 -4 +5 -1) (+3 -4 +5 -1 -2) (-4 +5 -1 -2 +3) \\
 & (+5 -1 -2 +3 -4) (-1 -2 +3 -4 +5) \} \\
 \text{can}([\pi]) = & (+1 -5 +4 -3 +2)
 \end{aligned}$$

Figura 3.7: Exemplo de uma classe de equivalência $A = [\pi]$ e da sua representante canônica $\text{can}(A)$.

Definimos agora reversão circular em termos de reversão linear, ambas atuando apenas na representante canônica.

Definição 3.1.4 Dadas uma classe de equivalência A modelando um cromossomo circular e uma reversão linear $r(i, j)$ com $1 \leq i \leq j \leq n$, exceto $(i, j) = (1, n)$, definimos uma **reversão circular** P atuando em A por

$$P \cdot A = [P \cdot \text{can}(A)] = [r(i, j) \cdot \text{can}(A)]$$

O próximo teorema mostra que, além da reversão circular P poder ser definida por uma reversão linear $r(i, j)$ atuando apenas na representante canônica, os índices i e j podem ser escolhidos no intervalo de 2 até n .

Teorema 3.1.5 Dada uma classe de equivalência A modelando um cromossomo circular, para qualquer reversão circular P atuando em A , existem inteiros i e j com $2 \leq i \leq j \leq n$ tais que

$$P \cdot A = [P \cdot \text{can}(A)] = [r(i, j) \cdot \text{can}(A)]$$

Prova:

Tomemos inicialmente a representante canônica de A , $can(A)$.

Existem duas possíveis formas de uma reversão circular agir em $can(A)$, mas ambas produzem permutações que pertencem à mesma classe de equivalência (Lema 3.1.2). Como podemos escolher qualquer uma das duas formas tomaremos a forma que não inclui $\pi_1 = +1$. Desta forma $P \cdot can(A)$ produzirá uma permutação que é também uma representante canônica. Em outras palavras, a representante canônica da classe de equivalência A que modela o cromossomo circular *antes* da reversão é transformada numa outra representante canônica da classe de equivalência que modela o cromossomo circular obtido *após* a reversão. Neste caso, a reversão circular age em $can(A)$ exatamente da mesma forma que agiria a reversão linear $r(i, j)$, com $i \geq 2$. Então,

$$r(i, j) \cdot (+1 \pi_2 \dots \pi_i \dots \pi_j \dots \pi_n) = (+1 \pi_2 \dots \bar{\pi}_j \dots \bar{\pi}_i \dots \pi_n),$$

com $2 \leq i \leq j \leq n$. Este resultado decorre da definição de reversão linear. Como a permutação da direita é canônica temos

$$r(i, j) \cdot can(A) = can(P \cdot A).$$

de onde temos

$$[r(i, j) \cdot can(A)] = [can(P \cdot A)] = [P \cdot can(A)] = P \cdot A.$$

□

Mostraremos agora que existe um isomorfismo entre reversões circulares agindo em cromossomos circulares e reversões lineares agindo em cromossomos lineares. Para isto, inicialmente definiremos duas bijeções. Tomemos S_n como sendo o conjunto de todas as permutações lineares com sinais de n elementos. Seja R_n o conjunto de todas as reversões lineares agindo em permutações lineares com n elementos. Sejam S_n^c o conjunto de todas as classes de equivalência das permutações lineares com sinais de comprimento n e R_n^c o conjunto de todas as reversões circulares agindo nestas classes de equivalência. Definimos

$$\varphi : S_n^c \longrightarrow S_{n-1}$$

de tal forma que

$$\varphi(A) = \text{tomar } can(A), \text{ remover } +1, \text{ subtrair } 1 \text{ dos valores absolutos}$$

dos outros elementos, conservando o sinal

Definimos também

$$\theta : R_n^c \longrightarrow R_{n-1}$$

de tal forma que

$$\theta(P) = r(i-1, j-1),$$

onde $P \cdot A = [P \cdot \text{can}(A)] = [r(i, j) \cdot \text{can}(A)]$, com $2 \leq i \leq j \leq n$.

Enunciamos agora o resultado.

Teorema 3.1.6 *Dadas duas bijeções φ e θ como definidas acima, então temos*

$$\varphi(P \cdot A) = \theta(P) \cdot \varphi(A)$$

Prova:

Inicialmente temos

$$\varphi(P \cdot A) = \text{tomar } \text{can}(P \cdot A), \text{ remover } +1, \text{ subtrair } 1 \text{ dos valores absolutos}$$

dos outros elementos, conservando o sinal

Seja (Teorema 3.1.5) $P \cdot A = [r(i, j) \cdot \text{can}(A)]$, com $2 \leq i \leq j \leq n$, e $A = [\pi]$, onde $\pi = \text{can}(A)$, isto é, $\pi = (\pi_1 \ \pi_2 \ \dots \ \pi_n)$ com $\pi_1 = +1$. Mas (Definição 3.1.4) $\text{can}(P \cdot A) = \text{can}([r(i, j) \cdot \text{can}(A)]) = \text{can}([r(i, j) \cdot \pi]) = r(i, j) \cdot \pi$, com $2 \leq i \leq j \leq n$. Então

$$\varphi(P \cdot A) = \text{tomar } r(i, j) \cdot \pi, \text{ remover } +1, \text{ subtrair } 1 \text{ dos valores absolutos}$$

dos outros elementos, conservando o sinal

Por outro lado, como $\text{can}(A) = \pi$, então temos

$$\varphi(A) = \text{tomar } \pi, \text{ remover } +1, \text{ subtrair } 1 \text{ dos valores absolutos}$$

dos outros elementos, conservando o sinal

e

$$\theta(P) = r(i-1, j-1)$$

Então o resultado segue porque $\theta(P) = r(i-1, j-1)$ age em $\varphi(A)$ exatamente nos mesmos elementos que $r(i, j)$ atua em π . Além disso, $\varphi(P \cdot A)$ gera uma permutação idêntica a $\theta(P)$ agindo em $\varphi(A)$.

□

Note que $|S_n^c| = |S_{n-1}| = 2^{n-1}(n-1)!$ e $|R_n^c| = |R_{n-1}| = (n-1)n/2$.

Do Teorema 3.1.6 decorre imediatamente o próximo resultado.

Corolário 3.1.7 *Dadas duas classes de equivalência A e B modelando dois cromossomos circulares, e a bijeção φ definida acima,*

$$d_r^c(A, B) = d_r(\varphi(A), \varphi(B))$$

Do Corolário 3.1.7 podemos deduzir um algoritmo para o problema da distância de reversão de cromossomos circulares com sinais. Basicamente, ele consiste em executar qualquer algoritmo que resolve o problema equivalente de cromossomos lineares com sinais, tomando como entrada duas permutações, obtidas aplicando a bijeção φ nas duas classes de equivalência que representam os cromossomos circulares.

Tomemos as duas permutações de entrada α e β , onde α e β são as permutações das duas classes que representam os cromossomos circulares. As representantes canônicas são obtidas pesquisando as duas permutações α e β para encontrar a posição k do bloco 1. Se ele tiver sinal + apenas aplicamos rot^{k-1} , e se ele tiver sinal - aplicamos rot^{k-n} seguido de refl.

Em particular, se utilizarmos o algoritmo de Kaplan, Shamir e Tarjan [23], a complexidade do algoritmo é $O(n^2)$ (encontrar as representantes canônicas e aplicar φ custam $O(n)$ e o algoritmo de Kaplan, Shamir e Tarjan [23] tem complexidade $O(n^2)$), onde n é o número de blocos de genes dos cromossomos circulares.

As reversões circulares são obtidas aplicando a inversa de θ em cada passo do algoritmo para cromossomos lineares. Este passo não afeta a complexidade do algoritmo acima pois custa $O(1)$.

3.2 O diâmetro de reversão

O **diâmetro de reversão circular**, denotado por $D_r^c(n)$, das classes de equivalência em S_n^c , com respeito à distância de reversão circular, é a máxima distância de reversão entre duas classes de equivalência das permutações de comprimento n . Analogamente, o **diâmetro de reversão linear**, denotado por $D_r(n)$, das permutações de n elementos do conjunto S_n , com respeito à distância de reversão linear, é a distância máxima entre duas permutações.

Nesta seção, mostramos que o diâmetro de reversão para cromossomos com sinais, lineares e circulares, são respectivamente $n + 1$ e n (exceto em poucos casos). Isto corrige um resultado de Kececioglu e Sankoff [26] que estabelece que $n - 2 \leq D_r(n) \leq n - 1$.

Primeiro, vamos provar o resultado sobre o diâmetro de reversão linear, $D_r(n) = n + 1$, exceto em dois casos. Para isto, inicialmente computamos a distância de reversão entre certas permutações e a permutação identidade, para cada n . Estas distâncias serão utilizadas como limite inferior para $D_r(n)$. Estes resultados estão enunciados nos lemas seguintes.

Nesta seção, todos os elementos das permutações têm sinais positivos (“+”), portanto, para não sobrecarregar a notação, omitiremos estes sinais. Além disto, nos grafos de pontos-de-quebra, os dois rótulos $-\pi_i$ e $+\pi_i$, associados a cada elemento π_i , também serão denotados apenas por π_i .

Lema 3.2.1 *Dadas as permutações*

$$\pi_n = (2 \ 1 \ 4 \ 3 \ 6 \ 5 \ \dots \ (n-4) \ (n-5) \ (n-2) \ (n-3) \ n \ (n-1))$$

e

$$\sigma_n = \iota_n = (1 \ 2 \ 3 \ 4 \ \dots \ (n-1) \ n),$$

para n par, $n \geq 2$, então $d(\pi_n, \sigma_n) = n + 1$.

Prova: Primeiro, construímos o grafo de pontos-de-quebra $G(\pi_n, \iota_n)$ para π_n com relação a ι_n , com n nas condições do lema. Este grafo é formado por exatamente um ciclo, de tamanho $n + 1$, envolvendo todos os rótulos. Conforme visto na Seção 2.1, este é um ciclo ruim e portanto um obstáculo (Figura 3.8).

Neste caso, usando a Fórmula 2.1 de Hannenhalli e Pevzner, também enunciada na Seção 2.1, e do grafo de pontos-de-quebra $G(\pi_n, \sigma_n)$, temos

$$d(\pi_n, \sigma_n) = (n + 1) - 1 + 1 + 0 = n + 1$$

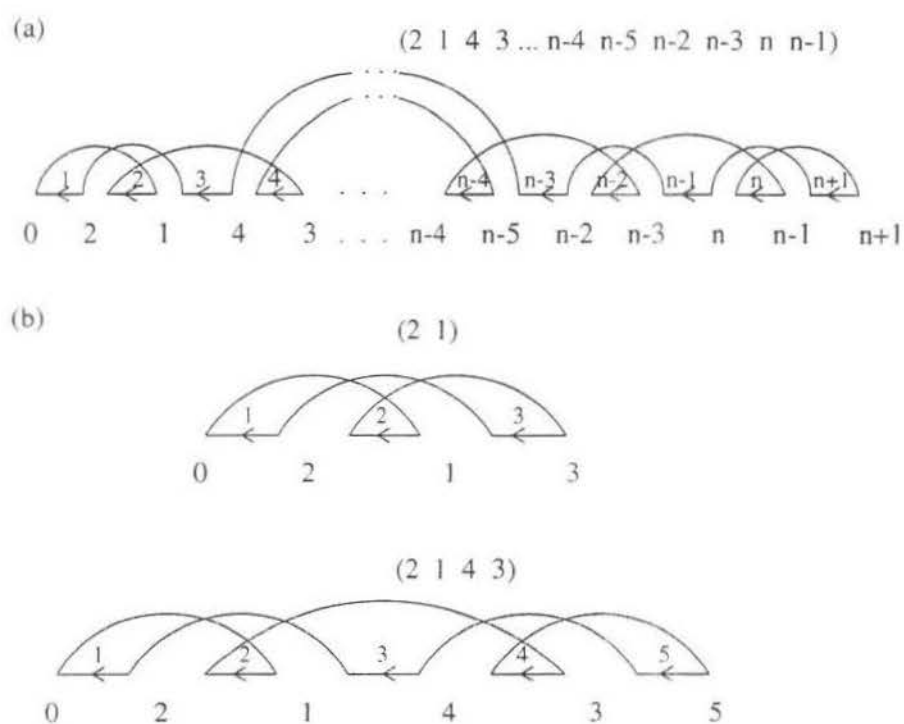


Figura 3.8: (a) O grafo de pontos-de-quebra para π_n e $\sigma_n = \iota_n$, n par, $n \geq 2$. (b) Exemplos para $n = 2$ e 4 com relação a ι_n .

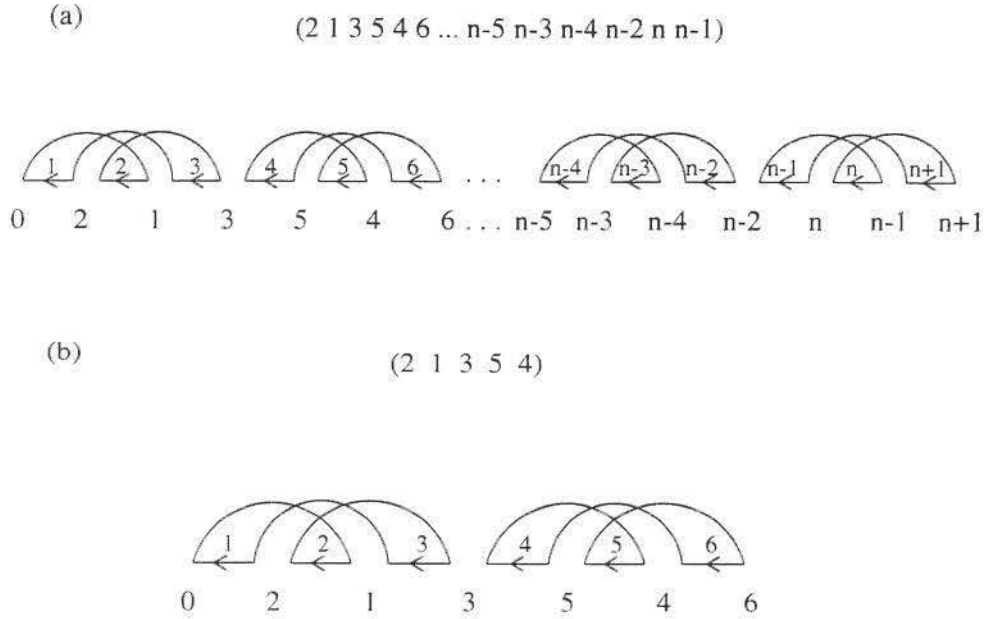


Figura 3.9: (a) O grafo de pontos-de-quebra para π_n e $\sigma_n = \iota_n$, n ímpar, $n \geq 5$, $(n+1) \bmod 3 = 0$. (b) Exemplo para $n = 5$ com relação a ι_5 .

□

Nos lemas seguintes calculamos distâncias de reversão entre diversas permutações e a permutação identidade, para n ímpar.

Lema 3.2.2 *Dadas as permutações*

$$\pi_n = (2\ 1\ 3\ 5\ 4\ 6\ 8\ 7\ 9\ \dots\ (n-6)\ (n-7)\ (n-5)\ (n-3)\ (n-4)\ (n-2)\ n\ (n-1))$$

e

$$\sigma_n = \iota_n = (1\ 2\ 3\ 4\ \dots\ (n-1)\ n),$$

para n ímpar, $n \geq 5$, e $(n+1) \bmod 3 = 0$, então $d(\pi_n, \sigma_n) = n+1$.

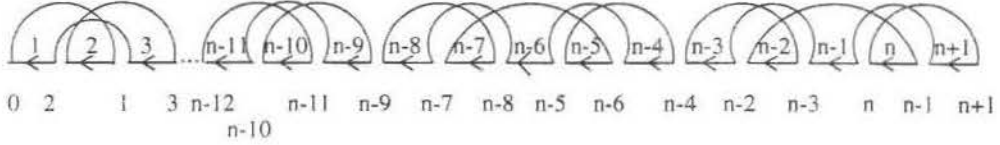
Prova: O grafo de pontos-de-quebra para π_n com relação a σ_n , com n nas condições da hipótese, é formado por exatamente $(n+1)/3$ ciclos de tamanho 3, com $n \geq 5$, construídos um ao lado do outro. Estes são ciclos ruins e portanto obstáculos (Figura 3.9).

Neste caso, usando a Fórmula 2.1, e do grafo de pontos-de-quebra $G(\pi_n, \iota_n)$, temos

$$d(\pi_n, \iota_n) = (n+1) - (n+1)/3 + (n+1)/3 + 0 = n+1$$

□

(a) $(2\ 1\ 3\ \dots\ n-12\ n-10\ n-11\ n-9\ n-7\ n-8\ n-5\ n-6\ n-4\ n-2\ n-3\ n\ n-1)$



(b) $(2\ 1\ 4\ 3\ 5\ 7\ 6\ 9\ 8)$

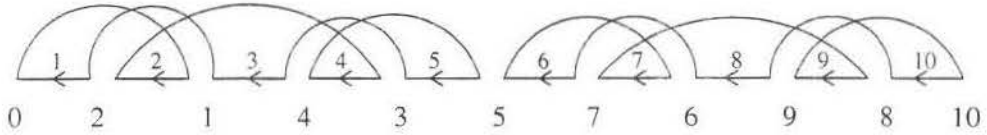


Figura 3.10: (a) O grafo de pontos-de-quebra para π_n e $\sigma_n = \iota_n$, n ímpar, $n \geq 9$, $(n+1) \bmod 3 = 1$. (b) Exemplo para $n = 9$ com relação a ι_9 .

Lema 3.2.3 *Dadas as permutações*

$$\begin{aligned} \pi_n = & (2\ 1\ 3\ 5\ 4\ 6\ \dots\ (n-15)\ (n-13)\ (n-14)\ (n-12)\ (n-10)\ (n-11) \\ & (n-9)\ (n-7)\ (n-8)\ (n-5)\ (n-6)\ (n-4)\ (n-2)\ (n-3)\ n\ (n-1)) \end{aligned}$$

e

$$\sigma_n = \iota_n = (1\ 2\ 3\ 4\ \dots\ (n-1)\ n),$$

para n ímpar, $n \geq 9$, $(n+1) \bmod 3 = 1$, então $d(\pi_n, \sigma_n) = n+1$.

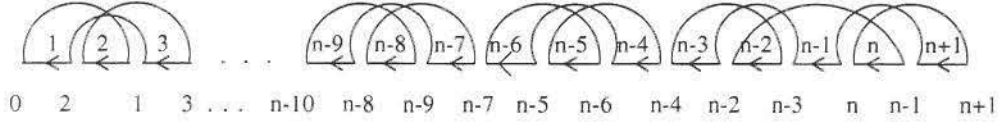
Prova: O grafo de pontos-de-quebra para π_n com respeito a ι_n , com n nas condições da hipótese, é formado por exatamente $(n-9)/3$ ciclos de tamanho 3, e dois ciclos de tamanho 5, com $n \geq 9$, construídos um ao lado do outro. Estes são ciclos ruins e portanto obstáculos (Figura 3.10).

Neste caso, usando a Fórmula 2.1, e do grafo de pontos-de-quebra $G(\pi_n, \iota_n)$, temos

$$d(\pi_n, \iota_n) = (n+1) - ((n-9)/3 + 2) + ((n-9)/3 + 2) + 0 = n+1$$

□

(a) $(2\ 1\ 3\ 5\ 4\ 6\ \dots\ n-10\ n-8\ n-9\ n-7\ n-5\ n-6\ n-4\ n-2\ n-3\ n\ n-1)$



(b)

$(2\ 1\ 3\ 5\ 4\ 7\ 6)$

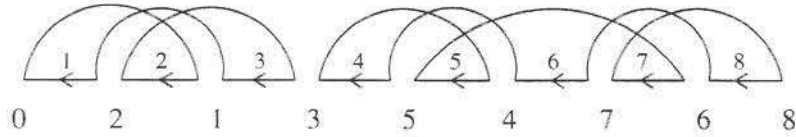


Figura 3.11: (a) O grafo de pontos-de-quebra para π_n e $\sigma_n = \iota_n$, n ímpar, $n \geq 7$, $(n+1) \bmod 3 = 2$. (b) Exemplo para $n = 7$ com relação a ι_7 .

Lema 3.2.4 *Dadas as permutações*

$$\pi_n = (2\ 1\ 3\ 5\ 4\ 6\ \dots\ (n-11)\ (n-12)\ (n-10)\ (n-8)\ (n-9)\ (n-7)\ (n-5)\ (n-6)\ (n-4)\ (n-2)\ (n-3)\ n\ (n-1))$$

e

$$\sigma_n = \iota_n = (1\ 2\ 3\ 4\ \dots\ (n-1)\ n),$$

para n ímpar, $n \geq 7$, $(n+1) \bmod 3 = 2$, então $d(\pi_n, \sigma_n) = n+1$.

Prova: O grafo de pontos-de-quebra para π_n com respeito a ι_n , com n nas condições da hipótese, é formado por exatamente $(n-4)/3$ ciclos de tamanho 3, e um ciclo de tamanho 5, com $n \geq 5$, construídos um ao lado do outro. Estes são ciclos ruins e portanto obstáculos (Figura 3.11).

Neste caso, usando a Fórmula 2.1, e do grafo de pontos-de-quebra $G(\pi_n, \iota_n)$, então temos

$$d(\pi_n, \iota_n) = (n+1) - ((n-4)/3 + 1) + ((n-4)/3 + 1) + 0 = n+1$$

□

Utilizando os lemas anteriores, provamos agora o diâmetro de reversão linear.

Teorema 3.2.5 *O diâmetro de reversão de cromossomos lineares com sinais é*

$$D_r(n) = \max_{\substack{\pi \in S_n \\ \sigma \in S_n}} \{d_r(\pi, \sigma)\} = \begin{cases} n & \text{se } n = 1 \text{ ou } n = 3 \\ n + 1 & \text{caso contrário} \end{cases}$$

Prova:

Para n par, $n \geq 2$, o Lema 3.2.1 apresenta uma família de permutações π_n , com n nestas condições, tais que $d_r(\pi_n, \iota_n) = n + 1$. Portanto, no caso de n par, temos $D_r(n) \geq n + 1$.

Analisemos agora o caso em que n é ímpar.

Quando $n = 1$ temos apenas duas permutações, e a distância entre elas é exatamente 1, e portanto $D_r(1) = 1$.

Quando $n = 3$ temos, decorrente de um teorema de Kececioglu e Sankoff [26], um algoritmo guloso que ordena qualquer permutação π com pelo menos um elemento negativo em no máximo $n - 1$ passos. Então, apenas as permutações com todos os elementos positivos são candidatos a possuir distância de reversão igual a $n + 1$. Usando este fato e construindo os grafos de pontos-de-quebra para todas as possíveis permutações com todos os seus elementos positivos para $n = 3$, concluímos que $d_r(\pi_3, \iota_3) \leq 3$. Por outro lado, $\pi_3 = (3 \ 2 \ 1)$ e $\iota_3 = (1 \ 2 \ 3)$ têm $d_r(\pi_3, \iota_3) = 3$, e portanto $D_r(3) = 3$.

Quando $n \geq 5$, vamos considerar os possíveis valores de $(n + 1) \bmod 3$. Temos três casos:

- $(n + 1) \bmod 3 = 0$: Do Lema 3.2.2, temos uma família de permutações π_n , com n nestas condições, para as quais $d_r(\pi_n, \iota_n) = n + 1$.
- $(n + 1) \bmod 3 = 1$: Do Lema 3.2.3, temos uma família de permutações π_n , com n nestas condições, para as quais $d_r(\pi_n, \iota_n) = n + 1$. Notamos que a restrição $n \geq 9$ do Lema 3.2.3 não elimina nenhum n para o qual $(n + 1) \bmod 3 = 1$, pois $n = 9$ é o primeiro número ímpar satisfazendo esta condição.
- $(n + 1) \bmod 3 = 2$: Do Lema 3.2.4, temos uma família de permutações π_n , com n nestas condições, para as quais $d_r(\pi_n, \iota_n) = n + 1$. Notamos que a restrição $n \geq 7$ do Lema 3.2.4 também não elimina nenhum n para o qual $(n + 1) \bmod 3 = 2$, pois $n = 7$ é o primeiro número ímpar satisfazendo esta condição.

Portanto, quando $n \geq 5$, mostramos que $D_r(n) \geq n + 1$.

Então, $D_r(1) = 1$ e $D_r(3) = 3$, enquanto que para $n = 2$ e para todo $n \geq 4$, temos um limite inferior de $n + 1$ para $D_r(n)$.

Para completar a prova, vamos mostrar que $D_r(n) < n + 2$, para todo n .

Da Fórmula 2.1 temos que

$$d_r(\pi_n, \iota_n) = (n + 1) - c(\pi_n, \iota_n) + h(\pi_n, \iota_n) + f(\pi_n, \iota_n)$$

para todo n .

Primeiro, observamos que $h(\pi_n, \iota_n) \leq c(\pi_n, \iota_n)$, pela definição de $h(\pi_n, \iota_n)$. Portanto, se $h(\pi_n, \iota_n) = c(\pi_n, \iota_n)$, então temos $d(\pi_n, \iota_n) \leq (n + 1) + 1$, isto é, $d(\pi_n, \iota_n) \leq n + 2$. Mas se $f(\pi_n, \iota_n) = 1$, então necessariamente $h(\pi_n, \iota_n) < c(\pi_n, \iota_n)$, e portanto $d(\pi_n, \iota_n) < n + 2$.

Isto prova o caso linear.

□

Das bijeções definidas na seção anterior, temos o próximo resultado.

Lema 3.2.6 *O diâmetro de reversão dos cromossomos circulares é*

$$D_r^c(n) = D_r(n - 1)$$

Deste lema, decorre o próximo teorema, que mostra o diâmetro de reversão circular das classes de equivalência em S_n^c .

Teorema 3.2.7 *O diâmetro de reversão de cromossomos circulares com sinais circulares é*

$$D_r^c(n) = \max_{\substack{A \in S_n^c \\ B \in S_n^c}} \{d_r^c(A, B)\} = \begin{cases} n - 1 & \text{se } n = 1, n = 2 \text{ ou } n = 4 \\ n & \text{caso contrário} \end{cases}$$

3.3 Distâncias de reversão de cromossomos circulares e lineares

Na Seção 3.1 mostramos que existe uma distância preservando a correspondência entre cromossomos circulares e cromossomos lineares de tamanho uma unidade menor.

Nesta seção estudamos como se relacionam cromossomos circulares e lineares do mesmo tamanho.

Mostraremos no Teorema 3.3.4 que $d_r^c([\pi], [\sigma]) \leq d_r(\pi, \sigma)$, para quaisquer π e σ . Antes de provar este resultado, precisamos de três lemas técnicos.

Lema 3.3.1 *Dadas duas permutações lineares σ e π , tais que $\sigma = op(\pi)$ onde $op = \text{rot}$ ou refl , então para toda reversão r existe uma reversão r' tal que $r \cdot \pi \sim r' \cdot \sigma$.*

Prova:

Seja $\pi = (\pi_1 \dots \pi_n)$.

Temos duas possibilidades para op .

- $op = \text{rot}$: Tomemos $\sigma = \text{rot}(\pi) = (\pi_2 \pi_3 \dots \pi_n \pi_1)$ e $r = r(i, j)$ de tal forma que

$$r(i, j) \cdot \pi = (\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n)$$

Temos três casos.

1. $i = 1, j = n$:

$$r(1, n) \cdot \pi = (\bar{\pi}_n \dots \bar{\pi}_1)$$

Neste caso $r = r(1, n) = \text{refl}$. Tomar $r' = \text{refl}$ também. Obteremos

$$r \cdot \pi = \text{refl}(\pi) \sim \pi \sim \sigma \sim \text{refl}(\sigma) = r' \cdot \sigma$$

2. $i = 1, j < n$: Neste caso

$$r(1, j) \cdot \pi = (\bar{\pi}_j \dots \bar{\pi}_1 \pi_{j+1} \dots \pi_n).$$

Então:

$$r(j, n-1) \cdot \sigma = (\pi_2 \pi_3 \dots \pi_{j-1} \pi_j \bar{\pi}_n \dots \bar{\pi}_{j+1} \pi_1)$$

$$\text{refl}(r(j, n-1) \cdot \sigma) = (\bar{\pi}_1 \pi_{j+1} \dots \pi_n \bar{\pi}_j \dots \bar{\pi}_3 \bar{\pi}_2)$$

$$\text{rot}^{n-j+1}(\text{refl}(r(j, n-1) \cdot \sigma)) = (\bar{\pi}_j \dots \bar{\pi}_3 \bar{\pi}_2 \bar{\pi}_1 \pi_{j+1} \dots \pi_n)$$

$$r(1, j) \cdot \pi = \text{rot}^{n-j+1}(\text{refl}(r(j, n-1) \cdot \sigma))$$

Então, tomando $r' = r(j, n-1)$ temos

$$r \cdot \pi \sim r' \cdot \sigma.$$

3. $i > 1, j \leq n$: Neste caso

$$r(i, j) \cdot \pi = (\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n).$$

Então:

$$r(i-1, j-1) \cdot \sigma = (\pi_2 \pi_3 \dots \bar{\pi}_j \dots \bar{\pi}_i \dots \pi_n \pi_1)$$

$$\text{rot}^{-1}(r(i-1, j-1) \cdot \sigma) = (\pi_1 \pi_2 \pi_3 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \dots \pi_n)$$

$$r(i, j) \cdot \pi = \text{rot}^{-1}(r(i-1, j-1) \cdot \sigma)$$

Portanto, tomando $r' = r(i-1, j-1)$ temos

$$r \cdot \pi \sim r' \cdot \sigma.$$

- $op = \text{refl}$: Tomemos $\sigma = \text{refl}(\pi) = (\bar{\pi}_n \dots \bar{\pi}_1)$ e $r = r(i, j)$ de tal forma que

$$r(i, j) \cdot \pi = (\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n)$$

Então:

$$r(n+1-j, n+1-i) \cdot \sigma = (\bar{\pi}_n \dots \bar{\pi}_{j+1} \pi_i \dots \pi_j \bar{\pi}_{i-1} \dots \bar{\pi}_1)$$

$$\text{refl}(r(n+1-j, n+1-i) \cdot \sigma) = (\pi_1 \dots \pi_{i-1} \bar{\pi}_j \dots \bar{\pi}_i \pi_{j+1} \dots \pi_n)$$

$$r(i, j) \cdot \pi = \text{refl}(r(n+1-j, n+1-i) \cdot \sigma)$$

Então, tomando $r' = r(n+1-j, n+1-i)$, temos

$$r \cdot \pi \sim r' \cdot \sigma.$$

□

Lema 3.3.2 Dadas duas permutações lineares π e σ , tais que $\pi \sim \sigma$ então para toda reversão r existe uma reversão r' tal que $r \cdot \pi \sim r' \cdot \sigma$.

Prova:

Tomemos $\sigma = op_v(op_{v-1}(\dots(op_1(\pi))\dots))$, onde $v \geq 0$, e $op_i = \text{rot}$ ou refl , para $1 \leq i \leq v$.

Esta prova sera feita por indução em v .

- $v = 0$: apenas tomamos $r' = r$.

- $v > 0$: Tomemos

$$\sigma' = op_{v-1}(\dots op_1(\pi)\dots)$$

Dada r , gostaríamos de obter r' tal que

$$r \cdot \pi \sim r' \cdot \sigma$$

pela hipótese de indução, temos

$$r \cdot \pi \sim r'' \cdot \sigma'$$

Mas, $\sigma \sim \sigma'$ (pois $\sigma' \sim \pi$), de tal forma que $\sigma = op_v(\sigma')$ e então, usando o Lema 3.3.1, existe uma r' tal que

$$r'' \cdot \sigma' \sim r' \cdot \sigma$$

Então,

$$r \cdot \pi \sim r' \cdot \sigma$$

□

Lema 3.3.3 *Dada uma permutação linear π e uma reversão linear r , então existe uma reversão circular P ou uma transformação identidade $P = I$, tal que*

$$[r \cdot \pi] = P \cdot [\pi].$$

Prova:

Seja σ a representante canônica de $[\pi]$:

$$\pi \sim \sigma = \text{can}([\pi])$$

O Lema 3.3.2 nos diz que dadas as permutações π e σ tais que $\pi \sim \sigma$, então para toda r existe uma r' tal que

$$r \cdot \pi \sim r' \cdot \sigma,$$

e portanto

$$[r \cdot \pi] = [r' \cdot \sigma]$$

Mas $r' = r(i, j)$ com $1 \leq i \leq j \leq n$. Então temos dois casos:

1. $i = j \oplus 1$. Então, $r' = \text{refl}$ e

$$[r' \cdot \sigma] = [\text{refl}(\sigma)] = [\sigma] = [\pi],$$

e então existe a transformação identidade $P = I$.

2. $i \neq j \oplus 1$. Pelo fato de σ ser a representante canônica de $[\pi]$, $\sigma = \text{can}([\pi])$, então temos (Definição 3.1.4)

$$[r(i, j) \cdot \sigma] = [r(i, j) \cdot \text{can}([\pi])] = [P \cdot \text{can}([\pi])] = P \cdot [\pi]$$

e portanto existe a reversão circular P .

□

Neste ponto mostramos que existem menos reversões no caso circular do que no caso linear, quando ambos os cromossomos tem o mesmo tamanho.

Teorema 3.3.4 *Dadas quaisquer permutações π e σ com o mesmo tamanho,*

$$d_r^c([\pi], [\sigma]) \leq d_r(\pi, \sigma)$$

Prova: Tomemos $\varrho = d_r(\pi, \sigma)$. Então,

$$r_\varrho \cdot r_{\varrho-1} \cdot \dots \cdot r_1 \cdot \pi = \sigma$$

$$[r_\varrho \cdot r_{\varrho-1} \cdot \dots \cdot r_1 \cdot \pi] = [\sigma]$$

$$P'_\varrho \cdot P'_{\varrho-1} \cdot \dots \cdot P'_1 \cdot [\pi] = [\sigma] \text{ (Lema 3.3.3)}$$

onde P'_i é uma reversão circular ou a transformação identidade. Então,

$$d_r^c([\pi], [\sigma]) \leq \varrho = d_r(\pi, \sigma)$$

□

Observamos ainda que há casos em que $d_r(\pi, \sigma)$ e $d_r^c([\pi], [\sigma])$ são diferentes. Por exemplo, consideremos as permutações $\pi = (-2 +3 +1)$ e $\sigma = (+1 +2 +3)$. Temos $d_r^c([\pi], [\sigma]) = 1$ porque $d_r^c([\pi], [\sigma]) = d_r(\text{can}([\pi]), \text{can}([\sigma])) = 1$, onde $\text{can}([\pi]) = (+1 -2 +3)$. Mas $d_r(\pi, \sigma) = 3$. Para fazer esta computação, é suficiente construir o grafo de pontos-de-quebra de π em relação a σ , e usar a Fórmula 2.1 de Hannenhalli e Pevzner [19].

O teorema seguinte também resolve o problema da distância de reversão para cromossomos circulares com sinais.

Teorema 3.3.5 *Dados dois cromossomos circulares do mesmo tamanho representados pelas classes de equivalência A e B então*

$$d_r^c(A, B) = d_r(\text{can}(A), \text{can}(B))$$

Prova:

Inicialmente mostraremos que

$$d_r^c(A, B) \leq d_r(\text{can}(A), \text{can}(B))$$

Do Teorema 3.3.4 sabemos que $d_r^c([\pi], [\sigma]) \leq d_r(\pi, \sigma)$. Em particular, tomando $\pi = \text{can}(A)$ e $\sigma = \text{can}(B)$, temos imediatamente o resultado.

Em seguida, mostraremos que

$$d_r^c(A, B) \geq d_r(\text{can}(A), \text{can}(B))$$

Para resolver o problema da distância de reversão de cromossomos circulares com sinais, usamos reversões circulares no intervalo $[2, n]$, que age sempre na permutação que é representante canônica. Consideremos então o cromossomo linear $\text{can}(A)$. Inicialmente, $\pi_1 = +1$ já está na sua posição correta, e isto não é modificado durante o processo. Estas mesmas reversões circulares fornecem uma série de reversões também para o caso linear.

□

Do Teorema 3.3.5 podemos deduzir um outro algoritmo para o problema da distância de reversão de cromossomos circulares com sinais. Este algoritmo consiste em utilizar o algoritmo que resolve o problema da distância de reversão de cromossomos lineares com sinais, fornecendo como entrada as representantes canônicas das classes de equivalência das permutações que representam os cromossomos circulares.

Em particular, se escolhermos o algoritmo de Kaplan, Shamir e Tarjan [23], a complexidade do algoritmo será $O(n^2)$: encontrar as representantes canônicas custa $O(n)$, como visto no final da Seção 3.1.2, e o algoritmo de Kaplan, Shamir e Tarjan [23] tem complexidade $O(n^2)$, onde n é o número de blocos de genes dos cromossomos circulares.

Dos resultados acima, pode ser demonstrado que o Corolário 3.1.7 e o Teorema 3.3.5 são equivalentes, no seguinte sentido:

Teorema 3.3.6 *Dadas duas classes de equivalência, A e B , modelando dois cromossomos circulares de mesmo tamanho, e a bijeção φ definida anteriormente, então*

$$d_r(\varphi(A), \varphi(B)) = d_r(\text{can}(A), \text{can}(B))$$

Além disso, do Teorema 3.3.5 e da fórmula de distância de reversão de cromossomos lineares de Hannenhalli e Pevzner [19], podemos deduzir uma fórmula para a distância de reversão de cromossomos circulares.

Lema 3.3.7 *Dadas duas classes de equivalência A e B modelando dois cromossomos circulares com sinais de mesmo tamanho, temos que*

$$d_r^c(A, B) = d_r(\text{can}(A), \text{can}(B)) = \\ (n + 1) - c(\text{can}(A), \text{can}(B)) + h(\text{can}(A), \text{can}(B)) + f(\text{can}(A), \text{can}(B))$$

No teorema seguinte provamos que as representantes canônicas, das classes de equivalência modelando os cromossomos circulares, fornecem uma distância mínima entre todas as permutações que pertencem às duas classes.

Teorema 3.3.8 *Dadas duas classes de equivalência de permutações de mesmo tamanho quaisquer A e B , modelando cromossomos circulares, temos*

$$d_r(\text{can}(A), \text{can}(B)) = \min_{\substack{\pi \in A \\ \sigma \in B}} \{d_r(\pi, \sigma)\}.$$

Prova:

Observemos inicialmente que $d_r^c(A, B) = d_r(\text{can}(A), \text{can}(B))$ (Teorema 3.3.5). Além disso, temos cada valor $d_r(\pi, \sigma)$ maior ou igual a $d_r^c([\pi], [\sigma])$ (Teorema 3.3.4).

□

Uma questão surge aqui. Quais são as permutações, das duas classes de equivalência modelando os cromossomos circulares, que levam a uma distância de reversão mínima?

Nossos resultados mostraram que as representantes canônicas das classes certamente levam. Mas elas não são as únicas. Temos um exemplo de Palmer e co-autores [33], onde as permutações não possuem as características das nossas representantes canônicas, mas levam a uma distância mínima. As permutações são $(-8 \ -7 \ -6 \ -5 \ -4 \ -3 \ -2 \ -1 \ -11 \ -10 \ -9C \ -9B \ -9A)$ e $(-4 \ +3 \ -2 \ +8 \ +7 \ -1 \ -5 \ -6 \ -11 \ +10 \ +9A \ -9B \ +9C)$.

Se denominarmos de *representantes ótimas* das duas classes de equivalência modelando os cromossomos circulares, duas permutações, uma de cada classe, que levam a uma distância de reversão mínima, seria interessante poder caracterizar este conjunto de representantes ótimas.

Passamos agora a mostrar resultados que explicam uma prática muito freqüente em artigos onde se estudam formas de calcular a distância de reversão de cromossomos lineares, que passamos a descrever.

Dado um cromossomo linear, há essencialmente duas maneiras de escrevê-lo como uma permutação, sendo uma a reflexão da outra. Isto é devido ao fato de que moléculas de DNA livres não possuem uma extremidade distinguível em geral, de modo que a leitura pode ser feita começando em qualquer uma das pontas. Assim sendo, ao comparar dois cromossomos, deveríamos na realidade considerar ambas as possibilidades para cada um deles. Isto implicaria na necessidade de calcular quatro distâncias de reversão, e tomar o mínimo entre elas como sendo a distância, isto é, dadas duas permutações π e σ modelando dois cromossomos lineares,

$$d_r(\pi, \sigma) = \min\{d_r(\pi, \sigma), d_r(\text{refl}(\pi), \sigma), d_r(\pi, \text{refl}(\sigma)), d_r(\text{refl}(\pi), \text{refl}(\sigma))\} \quad (3.2)$$

Contudo, em artigos, é comum a prática de comparar dois cromossomos tomando ambos com o mesmo bloco de genes numa das extremidades, se possível, ou, em termos de permutações, tomar as duas permutações contendo um dos elementos extremos comuns. Os resultados seguintes justificam isto mostrando que, neste caso, estas permutações escolhidas conduzem ao mínimo entre as quatro possibilidades existentes.

No teorema a seguir são reduzidas para duas as possibilidades de calcular a menor distância de reversão, formulada na Equação 3.2. Sua demonstração é simples e será omitida.

Teorema 3.3.9 *Dadas duas permutações π e σ que modelam dois cromossomos lineares, então*

$$d_r(\pi, \sigma) = d_r(\text{refl}(\pi), \text{refl}(\sigma))$$

e

$$d_r(\text{refl}(\pi), \sigma) = d_r(\pi, \text{refl}(\sigma))$$

O teorema abaixo comprova a correção do procedimento para calcular a distância de reversão, adotado na prática, de escolher permutações que comecem ou terminem com a mesma sequência de genes e com a mesma orientação.

Teorema 3.3.10 *Dadas duas permutações $\pi = (\pi_1 \dots \pi_n)$ e $\sigma = (\sigma_1 \dots \sigma_n)$ tais que $\pi_1 = \sigma_1$ ou $\pi_n = \sigma_n$, então*

$$d_r(\pi, \sigma) \leq d_r(\pi, \text{refl}(\sigma))$$

Então, considerando o Teorema 3.3.9, e se tivermos representantes com as características do Teorema 3.3.10, temos que os resultados encontrados nos artigos calculando apenas $d_r(\pi, \sigma)$ fornecem realmente a mínima distância de reversão entre as quatro possibilidades existentes.

Por fim, notamos que o mesmo exemplo de Palmer e co-autores [33], citado anteriormente, fornece um caso em que as duas permutações não têm $\pi_1 = \sigma_1$ nem $\pi_n = \sigma_n$, mas levam à uma distância de reversão mínima.

3.4 Sumário e questões

Neste capítulo, nosso objetivo foi o de iniciar um estudo sistemático da teoria do problema da distância de reversão de cromossomos circulares com sinais. Com esta perspectiva, contribuímos em certos pontos, descritos em seguida.

Inicialmente formalizamos um cromossomo circular por meio de uma classe de equivalência. Isto é interessante pois esta classe inclui as diferentes formas de visualizar um cromossomo circular com sinais, obtidas por rotações e reflexões. Definimos também reversões circulares usando definições conhecidas de reversões lineares. Utilizando estas definições, mostramos que há um isomorfismo entre reversões circulares agindo em cromossomos circulares e reversões lineares agindo em cromossomos lineares. Este isomorfismo permitiu resolver o problema da distância de reversão de cromossomos circulares com sinais utilizando algoritmos polinomiais que resolvem o problema da distância de reversão de cromossomos lineares com sinais, fornecendo como entrada determinadas permutações, escolhidas nas classes de equivalência que modelam os cromossomos circulares.

Em seguida, determinamos o diâmetro de reversão com sinais para cromossomos lineares ($D_r(n) = n + 1$) e circulares ($D_r^c(n) = n$), corrigindo um resultado de Kececioğlu e Sankoff [26] do diâmetro de reversão linear $D_r(n)$.

Além disso, apresentamos alguns resultados relacionando cromossomos circulares e lineares do mesmo tamanho, e justificamos a prática, comum nos artigos encontrados na literatura, de calcular apenas uma distância de reversão $d_r(\pi, \sigma)$, fixando a priori uma das pontas das moléculas de DNA, e não calculando as outras três possibilidades, obtidas invertendo os genes dos cromossomos. Mostramos que quando uma das duas pontas dos dois cromossomos é idêntica, o que em termos do formalismo significa considerar permutações com extremidade comum, basta calcular apenas $d_r(\pi, \sigma)$.

Para terminar este capítulo, colocamos questões que surgiram destes estudos. Primeiro, gostaríamos de investigar quais permutações das classes de equivalência

levam à distância de reversão mínima, isto é, gostaríamos de saber como caracterizar o conjunto das representantes ótimas. Segundo, seria interessante do ponto de vista prático caracterizar precisamente em que condições temos $d_r(\pi, \sigma) \leq d_r(\pi, \text{refl}(\sigma))$. Neste trabalho, fornecemos condições suficientes para que isto ocorra. Contudo, um exemplo apresentado no texto não seguia este padrão, mas também levava a uma distância de reversão mínima.

Uma outra direção de pesquisa seria investigar o problema da distância de reversão de cromossomos circulares visando descobrir um algoritmo polinomial que o resolva. Pelo isomorfismo apresentado solucionaríamos também o caso linear. O objetivo seria o de tentar desenvolver um algoritmo mais simples do que aquele proposto por Hannenhalli e Pevzner [19], ou ainda mais simples do que o algoritmo de Kaplan, Shamir e Tarjan [23], que tem a menor complexidade conhecida ($O(n^2)$, onde n é o tamanho das permutações), mas baseia-se em parte em conceitos apresentados por Hannenhalli e Pevzner. Um ponto de partida seria estudar diretamente a própria definição de reversão circular, que referencia os elementos de uma permutação da classe de equivalência, e não os índices associados a estes elementos, como no caso da reversão linear.

Capítulo 4

Distância de transposição de cromossomos lineares sem sinais

Neste capítulo estudaremos o problema da distância de transposição de duas permutações lineares sem sinais.

Bafna e Pevzner [4] estudaram este problema, apresentando vários algoritmos de aproximação, sendo de 1.5 o algoritmo de menor razão, e deixando diversas questões em aberto. Entre estas questões destacamos a complexidade do problema, o diâmetro de transposição e a distância de transposição entre uma permutação e sua inversa. Guyer, Heath e Vergara [15] também estudaram este problema e implementaram vários algoritmos para computar a distância de transposição entre dois cromossomos lineares sem sinais, baseados nos conceitos de subsequências e execuções.

Na primeira seção deste capítulo apresentamos um algoritmo de aproximação de razão 2.25 e um algoritmo exato para o problema da distância de transposição, mostrando os resultados de experimentos realizados com ambos os algoritmos. Embora a razão deste algoritmo seja elevada quando comparada com o melhor algoritmo conhecido (de razão 1.5 citado acima), os experimentos mostraram que esta razão é menor, sugerindo que o algoritmo possa ter utilidade na prática.

Na segunda seção, apresentamos um limite inferior para o diâmetro de transposição, mostrando que a distância de transposição entre uma permutação e sua inversa (sem complementação) é $\lfloor n/2 \rfloor + 1$, onde n é o tamanho das permutações. Notamos que este resultado foi encontrado independentemente por Christie [10]. Além disso, apresentamos um algoritmo para computar uma série ótima de transposições para este caso particular. Bafna e Pevzner [3] mostraram um limite superior $\frac{3}{4}n$ para este diâmetro.

Finalmente, na última seção, apresentamos o sumário do capítulo e colocamos questões que surgiram a partir destes estudos.

4.1 Um algoritmo de aproximação

Inicialmente, na Seção 4.1.1 apresentamos algumas definições e resultados que serão usados nas próximas seções. Na Seção 4.1.2 apresentamos um algoritmo de aproximação com razão 2.25 para computar a distância de transposição entre duas permutações, baseado numa estrutura denominada *diagrama de pontos-de-quebra*. Para verificar o desempenho deste algoritmo, elaboramos um algoritmo exato, também baseado no diagrama de pontos-de-quebra. Assim, na Seção 4.1.3, apresentamos resultados de experimentos, que exibem um fator de aproximação melhor, sugerindo que ele possa ser comparado aos melhores algoritmos conhecidos. Além disso, esta estrutura é mais fácil de ser implementada do que o grafo de ciclos de Bafna e Pevzner [4].

4.1.1 Definições

Podemos verificar facilmente que computar a distância de transposição entre duas permutações é equivalente ao **problema da ordenação por transposições**, que é encontrar a distância de transposição entre uma permutação qualquer e a permutação identidade ι , $d_t(\pi, \iota)$. Nesta seção estudaremos portanto o problema da ordenação por transposições.

Antes de apresentar as próximas definições, estendemos a permutação original π , de n elementos, com mais dois elementos $\pi_0 = 0$ e $\pi_{n+1} = (n + 1)$, e chamaremos esta permutação estendida simplesmente de permutação.

Dada uma permutação π , escrevemos $\pi_i \prec \pi_j$ se e somente se $\pi_j - \pi_i = 1$ e $\pi_i \neq \pi_j$ em caso contrário. Um **ponto-de-quebra** é um par de elementos adjacentes π_i, π_{i+1} , com $0 \leq i \leq n$, tais que $\pi_i \neq \pi_{i+1}$. O número de pontos-de-quebra de π em relação à identidade é denotado por $b(\pi, \iota)$. Quando aplicamos uma transposição t , a variação no número de pontos-de-quebra é denotado por $\Delta b(\pi, \iota) = b(t \cdot \pi, \iota) - b(\pi, \iota)$. Uma **faixa** é uma série maximal de elementos consecutivos $\pi_i \pi_{i+1} \dots \pi_j$ sem pontos-de-quebra. Finalmente, dadas duas faixas s_1 e s_2 , escrevemos $s_1 \prec s_2$ se o último elemento π_l de s_1 e o primeiro elemento π_m de s_2 são tais que $\pi_l \prec \pi_m$.

A observação de que podemos remover no máximo três pontos-de-quebra numa transposição leva imediatamente ao próximo lema.

Teorema 4.1.1 *Dada uma permutação π , então*

$$\frac{b(\pi, \iota)}{3} \leq d_t(\pi, \iota)$$

Definimos agora uma estrutura denominada **diagrama de pontos-de-quebra** de uma permutação qualquer π com relação à permutação ι , denotada por $D(\pi, \iota)$, da seguinte forma:

- definimos um conjunto de nós, $V = \{1, 2, \dots, b(\pi, \iota)\}$, um para cada ponto-de-quebra de π com relação a ι , na ordem em que eles aparecem, onde $i = s_{i-1} \cdot s_i$, s_i é uma faixa, com $1 \leq i \leq b(\pi, \iota)$, e o último elemento π_l de s_{i-1} e o primeiro elemento π_m de s_i são tais que $\pi_l \not\prec \pi_m$, claramente, senão s_{i-1} e s_i seriam uma única faixa.
- definimos três tipos de arestas, $\rightarrow$, $\Rightarrow$ e $->$, da seguinte forma: dados dois nós $i = s_{i-1} \cdot s_i$ e $j = s_{j-1} \cdot s_j$,
 - $i \rightarrow j$ quando $i < j$, $s_{i-1} \prec s_j$ e $n+1 \notin s_j$.
 - $i \Rightarrow j$ quando $i < j$, $s_{i-1} \prec s_j$ e $0 \notin s_{i-1}$.
 - $i -> j$ quando $i < j$ e $s_{j-1} \prec s_i$.

Por exemplo, dada a permutação $\pi = (0 \ 5 \ 3 \ 1 \ 4 \ 2 \ 6)$, temos os nós $1 = 0.5$, $2 = 5.3$, $3 = 3.1$, $4 = 1.4$, $5 = 4.2$ e $6 = 2.6$, e as arestas $1 \rightarrow 3$, $1 -> 5$, $2 \Rightarrow 6$, $2 -> 6$, $3 \rightarrow 4$, $3 \Rightarrow 4$, $4 \rightarrow 5$ e $4 \Rightarrow 5$ (Figura 4.1). Observe que as arestas $\rightarrow$ e $\Rightarrow$ coincidem, exceto quando saem do primeiro nó ou entram no último nó. Do primeiro nó apenas $\rightarrow$ pode sair, enquanto que somente $\Rightarrow$ pode entrar no último nó.

Denotaremos por $T(i, j, k)$ uma transposição aplicada aos vértices i, j e k do diagrama de pontos-de-quebra. Isto implica na necessidade de obter índices i', j' e k' tais que $t(i', j', k')$ seja aplicada na permutação nos pontos correspondentes a i, j e k .

Podemos observar que, dados três nós i, j e k num diagrama, a transposição $T(i, j, k)$ (Figura 4.2):

- elimina três pontos-de-quebra, quando $i \rightarrow j$, $j \Rightarrow k$ e $i -> k$,
- elimina dois pontos-de-quebra, quando $i \rightarrow j$ e $j \Rightarrow k$ mas $i \not-> k$, ou $i \rightarrow j$ e $i -> k$, mas $j \not\Rightarrow k$, ou ainda quando $i -> k$ e $j \Rightarrow k$ mas $i \not\rightarrow j$,
- elimina um ponto-de-quebra, quando existe apenas uma das três arestas entre estes três nós, e não existe nenhuma das duas outras arestas,

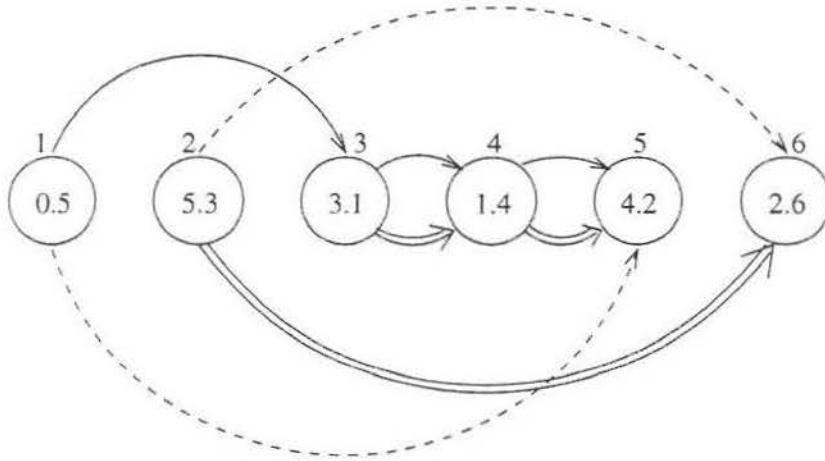


Figura 4.1: O diagrama de pontos-de-quebra $D(\pi, \iota)$, para $\pi = (0\ 5\ 3\ 1\ 4\ 2\ 6)$.

- não elimina nenhum ponto-de-quebra, quando não existe nenhuma aresta entre quaisquer dois nós de i, j e k .

Dizemos que uma transposição t é uma x -transposição, $x \in \{-3, -2, -1, 0, 1, 2, 3\}$, se $\Delta b(\pi, \iota) = x$.

Lema 4.1.2 Dada a permutação π , se $V \neq \emptyset$ em $D(\pi, \iota)$, então $|V| \geq 3$.

Lema 4.1.3 Dada a permutação π , se $|V| = r \geq 4$ em $D(\pi, \iota)$, então existem pelo menos quatro arestas $1 \rightarrow j$, $i \Rightarrow r$, $1 \rightarrow l$ e $m \rightarrow r$, com $2 \leq i, j, l, m \leq r-1$, i, j, l, m não necessariamente distintos.

Do Lema 4.1.3 obtemos o seguinte resultado.

Lema 4.1.4 Dada a permutação π , se $|V| \geq 4$ em $D(\pi, \iota)$, então sempre existe uma -1 -transposição.

Do Lema 4.1.4 temos um limite superior para $d_t(\pi, \iota)$.

Teorema 4.1.5 Dada a permutação π , então

$$d_t(\pi, \iota) \leq b(\pi, \iota)$$

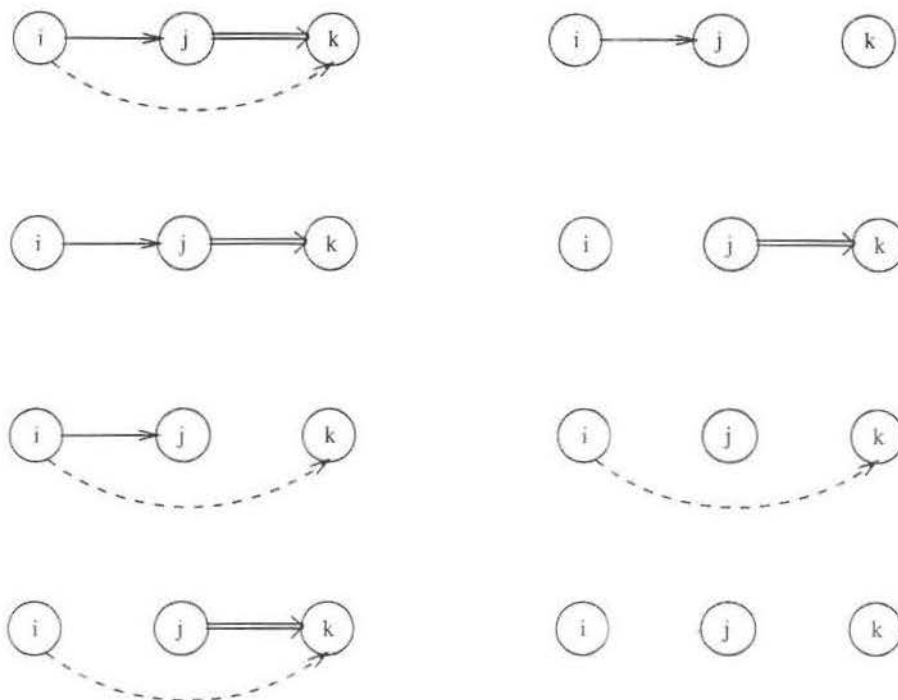


Figura 4.2: As várias formas de eliminar três, dois, um e nenhum ponto-de-quebra em $D(\pi, \iota)$, para uma permutação π .

4.1.2 Apresentação do algoritmo

Observamos inicialmente que, em termos do diagrama de pontos-de-quebra, ordenar por transposições é transformar V num conjunto vazio utilizando o menor número de transposições possível.

Então, dos Teoremas 4.1.1 e 4.1.5, e do Lema 4.1.3, temos imediatamente um 3-algoritmo de aproximação para o problema da ordenação por transposições.

Mas, nesta seção, apresentamos um algoritmo de aproximação de razão 2.25, baseado no diagrama de pontos-de-quebra.

Podemos verificar que num diagrama com $|V| = 3$ temos necessariamente uma -3 -transposição, enquanto que num diagrama com $|V| = 4$ temos necessariamente uma -1 -transposição, seguida por uma -3 -transposição (Figura 4.3).

Podemos verificar ainda que quando $|V| = 5$ temos uma única permutação, $\pi = (4\ 3\ 2\ 1)$, que gera um diagrama onde necessariamente temos duas -1 -transposições, seguidas por uma -3 -transposição, isto é, $d_t(\pi, \iota) = 3$. Todas as outras permutações geram diagramas onde temos uma -2 -transposição seguida por uma -3 -transposição, isto é, $d_t(\pi, \iota) = 2$. Estas distâncias foram computadas pelo programa descrito na

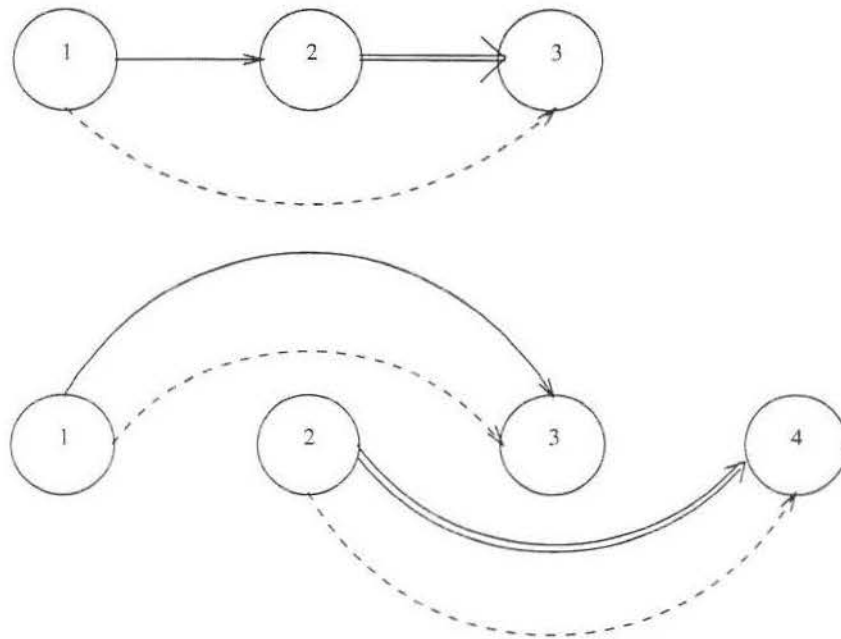


Figura 4.3: Os diagramas de pontos-de-quebra para $|V| = 3$ e $|V| = 4$. Observe que, em cada um destes casos, temos apenas uma forma geral de diagrama.

Seção 4.1.3.

O Teorema 4.1.10 e o Corolário 4.1.11 mostram que, quando $|V| \geq 6$, podemos remover no mínimo quatro pontos-de-quebra em três passos.

Mas antes de provarmos estes dois resultados, mostraremos alguns lemas técnicos. Notamos neste ponto que quando existe uma aresta $\rightarrow$ entre dois nós $i = s_{i-1} \cdot s_i$ e $k = s_{k-1} \cdot s_k$, certamente existe um nó $j = s_{j-1} \cdot s_j$ entre i e k , senão $s_i = s_{k-1}$.

Lema 4.1.6 *Dada uma permutação π tal que $D(\pi, \iota)$:*

- possui $|V| = r \geq 6$,
- não contém -2 -transposição nem -3 -transposição,
- possui nós $(0.a)$, $(b.1)$, $(r-1.c)$ e $(d.r)$ e arestas $(0.a) \rightarrow (b.1)$, $(r-1.c) \Rightarrow (d.r)$ tais que o nó $(r-1.c)$ está no intervalo entre $(0.a)$ e $(b.1)$,

e existem nó $(e.f)$ e aresta $(0.a) \rightarrow (e.f)$ com $(e.f)$ no intervalo entre $(0.a)$ e $(r-1.c)$ então é possível remover pelo menos quatro nós em três passos.

Prova:

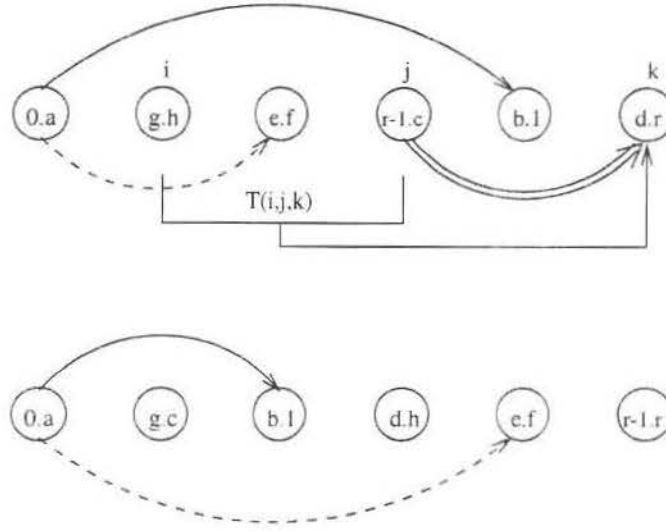


Figura 4.4: A -1 -transposição $T(i, j, k)$ indicada na figura gera um novo diagrama contendo uma -2 -transposição.

Nas condições do lema, temos necessariamente um nó $i = (g.h)$ entre $(0.a)$ e $(e.f)$. Tomando $j = (r-1.c)$ e $k = (d.r)$, aplicando a -1 -transposição $T(i, j, k)$ obtemos um diagrama com $|V| = r-1$ (o nó $(r-1.r)$ é eliminado), onde temos uma -2 -transposição, $i = (0.a)$, $j = (b.l)$ e $k = (e.f)$ (Figura 4.4).

□

Lema 4.1.7 Dada uma permutação π tal que $D(\pi, \iota)$:

- possui $|V| = r \geq 6$,
- não contém -2 -transposição nem -3 -transposição,
- possui nós $(0.a)$, $(b.l)$, $(r-1.c)$ e $(d.r)$ e arestas $(0.a) \rightarrow (b.l)$, $(r-1.c) \Rightarrow (d.r)$ tais que o nó $(r-1.c)$ está no intervalo entre $(0.a)$ e $(b.l)$,

e existem:

- nó $(e.f)$ e aresta $(0.a) \rightarrow (e.f)$ com $(e.f)$ no intervalo entre $(r-1.c)$ e $(b.l)$
- nó $(g.h)$ e aresta $(g.h) \rightarrow (d.r)$ com $(g.h)$ entre $(e.f)$ e $(b.l)$,

então é possível remover pelo menos quatro nós em três passos.

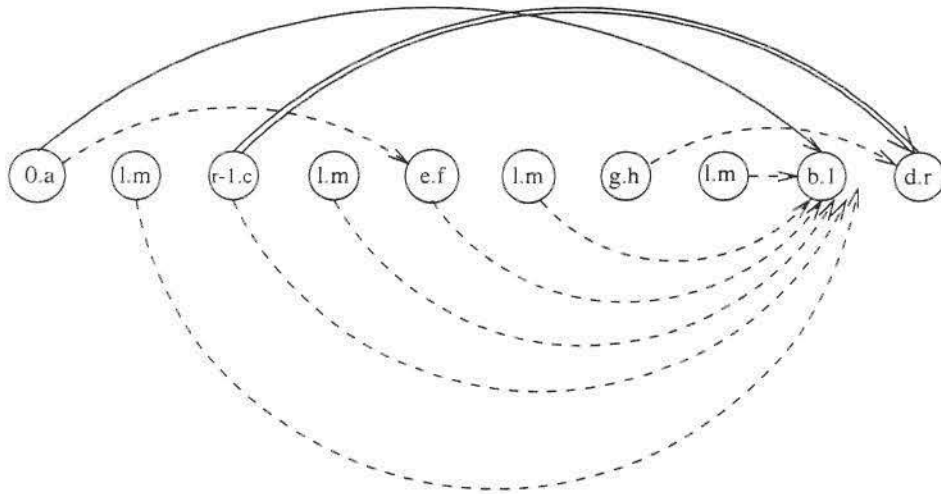


Figura 4.5: As formas gerais dos diagramas de pontos-de-quebra com $|V| \geq 6$, para as seis possibilidades da aresta $(l.m) \rightarrow (b.l)$, quando $(g.h)$ está no intervalo entre $(e.f)$ e $(b.l)$.

Prova:

Nas condições do lema, temos seis possibilidades para a aresta $(l.m) \rightarrow (b.l)$: $(l.m)$ entre $(g.h)$ e $(b.l)$, $(l.m)$ entre $(e.f)$ e $(g.h)$, $(l.m) = (e.f)$, $(l.m)$ entre $(r-1.c)$ e $(e.f)$, $(l.m) = (r-1.c)$ e finalmente $(l.m)$ entre $(0.a)$ e $(r-1.c)$ (Figura 4.5):

- $(l.m)$ entre $(g.h)$ e $(b.l)$: temos necessariamente um nó $(s.t)$ entre $(l.m)$ e $(b.l)$. Portanto, tomando $i = (g.h)$, $j = (s.t)$ e $k = (d.r)$, a -1 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r - 1$ (o nó $(d.h)$ é eliminado), contendo uma -2 -transposição, $i = (0.a)$, $j = (b.l)$ e $k = (l.m)$ (Figura 4.6).
- $(l.m)$ entre $(e.f)$ e $(g.h)$: tomando $i = (r-1.c)$, $j = (g.h)$ e $k = (d.r)$, a 0 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r$, contendo uma -2 -transposição, $i = (0.a)$, $j = (b.l)$ e $k = (l.m)$. Esta -2 -transposição gera um diagrama com $|V| = r - 2$ nós (os nós $(0.l)$ e $(b.m)$ são eliminados), contendo ainda uma outra -2 -transposição, $i = (d.c)$, $j = (r-1.h)$ e $k = (g.r)$ (Figura 4.7).
- $(l.m) = (e.f)$: tomando $i = (r-1.c)$, $j = (g.h)$ e $k = (d.r)$, a 0 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r$, contendo uma -3 -transposição, $i = (0.a)$, $j = (b.l)$ e $k = (e.f)$ (Figura 4.8).
- $(l.m)$ entre $(r-1.c)$ e $(e.f)$: tomando $i = (0.a)$, $j = (g.h)$ e $k = (d.r)$, a 0 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r$, contendo uma -2 -

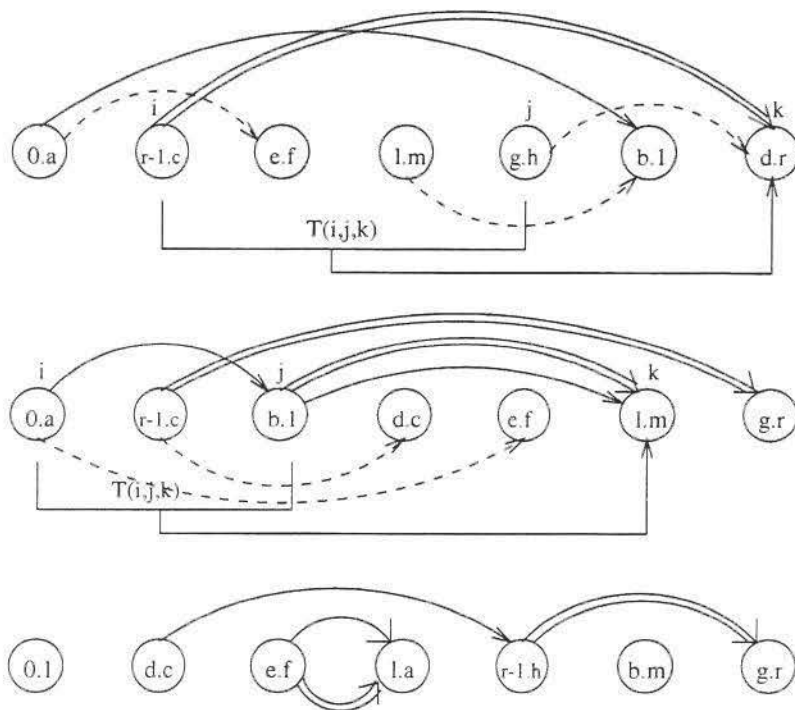


Figura 4.7: A primeira 0-transposição indicada na figura gera um diagrama contendo uma -2-transposição, e esta transposição gera um novo diagrama contendo uma outra -2-transposição.

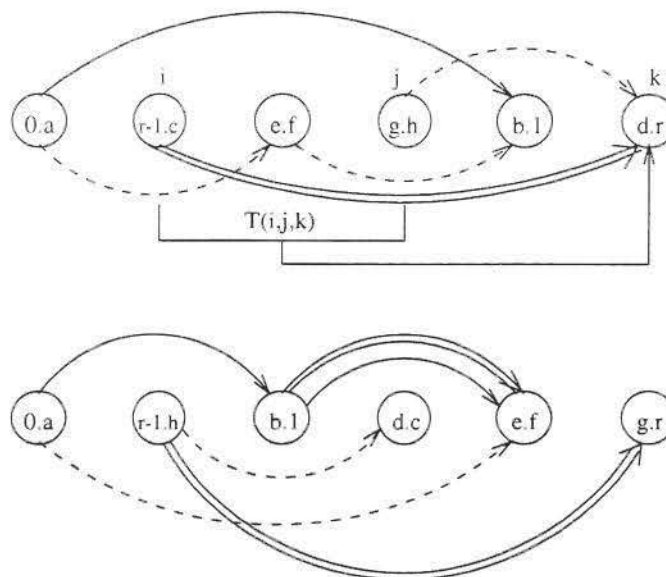


Figura 4.8: A 0-transposição $T(i,j,k)$ indicada na figura gera um novo diagrama contendo uma -3-transposição.

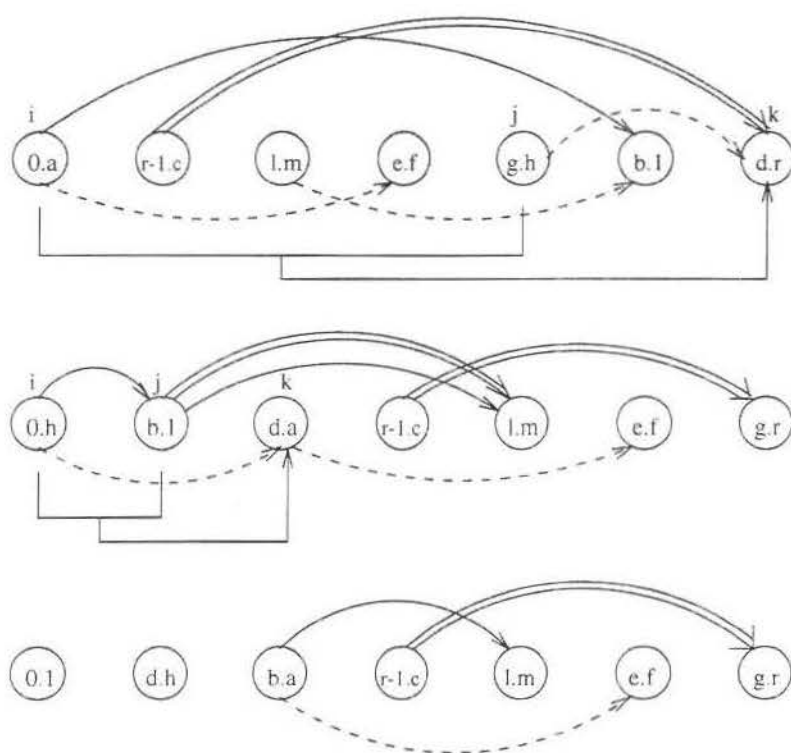


Figura 4.9: A primeira 0-transposição indicada na figura gera um diagrama contendo uma -2 -transposição, e esta transposição gera um novo diagrama contendo uma outra -2 -transposição.

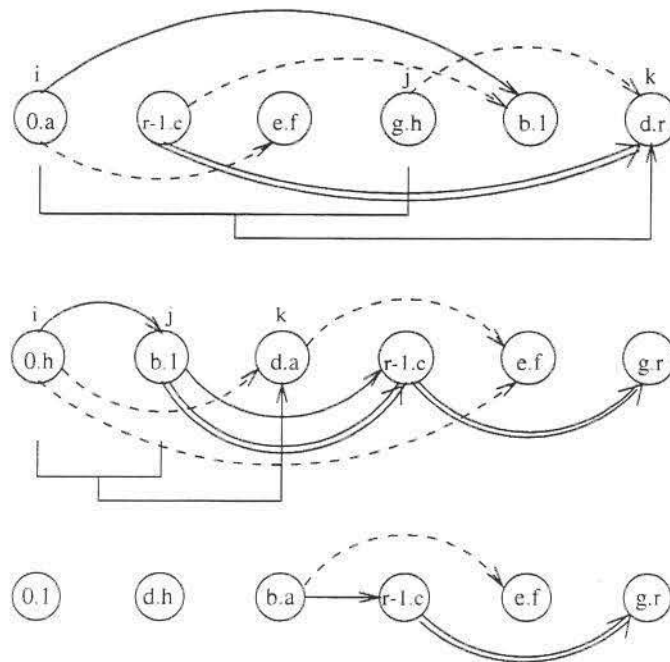


Figura 4.10: A primeira 0-transposição indicada na figura gera um diagrama contendo uma -2 -transposição, e esta transposição gera um novo diagrama contendo uma outra -2 -transposição.

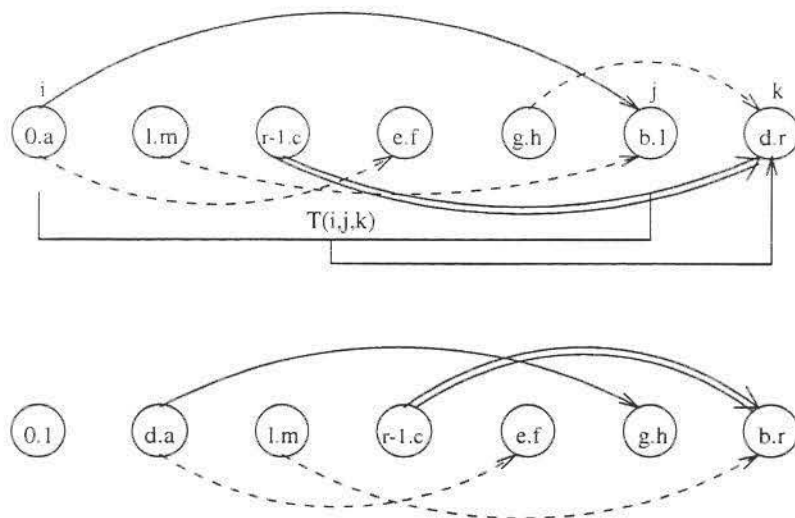


Figura 4.11: A -1 -transposição $T(i, j, k)$ indicada na figura gera um novo diagrama contendo uma -2 -transposição.

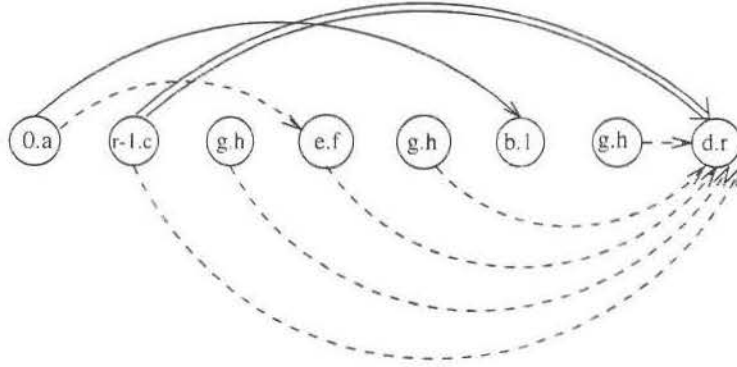


Figura 4.12: As formas gerais dos diagramas de pontos-de-quebra com $|V| \geq 6$, para as cinco possibilidades da aresta $(g.h) \rightarrow (d.r)$.

- não contém -2 -transposição nem -3 -transposição,
- possui nós $(0.a)$, $(b.l)$, $(r-1.c)$ e $(d.r)$ e arestas $(0.a) \rightarrow (b.l)$, $(r-1.c) \Rightarrow (d.r)$ tais que o nó $(r-1.c)$ está no intervalo entre $(0.a)$ e $(b.l)$,

e existem nó $(e.f)$ e aresta $(0.a) \rightarrow (e.f)$ com $(e.f)$ no intervalo entre $(r-1.c)$ e $(b.l)$ então é possível remover pelo menos quatro nós em três passos.

Prova:

Nas condições do lema, temos cinco possibilidades para $(g.h) \rightarrow (d.r)$: $(g.h) = (r-1.c)$, $(g.h)$ entre $(r-1.c)$ e $(e.f)$, $(g.h) = (e.f)$, $(g.h)$ entre $(e.f)$ e $(b.l)$, e finalmente $(g.h)$ entre $(b.l)$ e $(d.r)$ (Figura 4.12):

- $(g.h) = (r-1.c)$: tomando $i = (0.a)$, $j = (r-1.c)$ e $k = (d.r)$, a -1 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r-1$ (o nó $(r-1.r)$ é eliminado), onde temos uma -2 -transposição, $i = (0.c)$, $j = (b.l)$ e $k = (d.a)$ (Figura 4.13).
- $(g.h)$ entre $(r-1.c)$ e $(e.f)$: tomando $i = (0.a)$, $j = (r-1.c)$ e $k = (d.r)$, a -1 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r-1$ (o nó $(r-1.r)$ é eliminado), onde temos uma -2 -transposição, $i = (g.h)$, $j = (e.f)$ e $k = (d.a)$ (Figura 4.14).
- $(g.h) = (e.f)$: tomando $i = (0.a)$, $j = (b.l)$ e $k = (d.r)$, a -1 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r-1$ (o nó (0.1) é eliminado). Em seguida, tomando $i = (d.a)$, $j = (r-1.c)$ e $k = (e.f)$, aplicamos a -1 -transposição (o nó $(e.a)$ é eliminado), geramos um diagrama onde temos uma -2 -transposição, $i = (d.c)$, $j = (r-1.f)$ e $k = (b.r)$ (Figura 4.15).

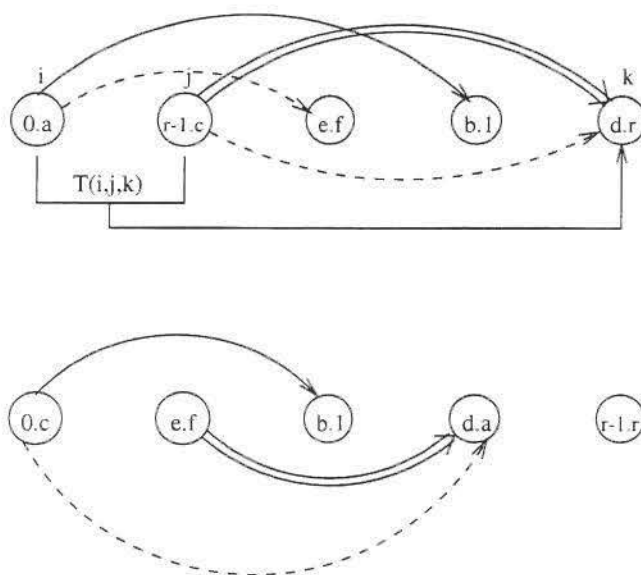


Figura 4.13: A -1 -transposição $T(i, j, k)$ indicada na figura gera um novo diagrama contendo uma -2 -transposição.

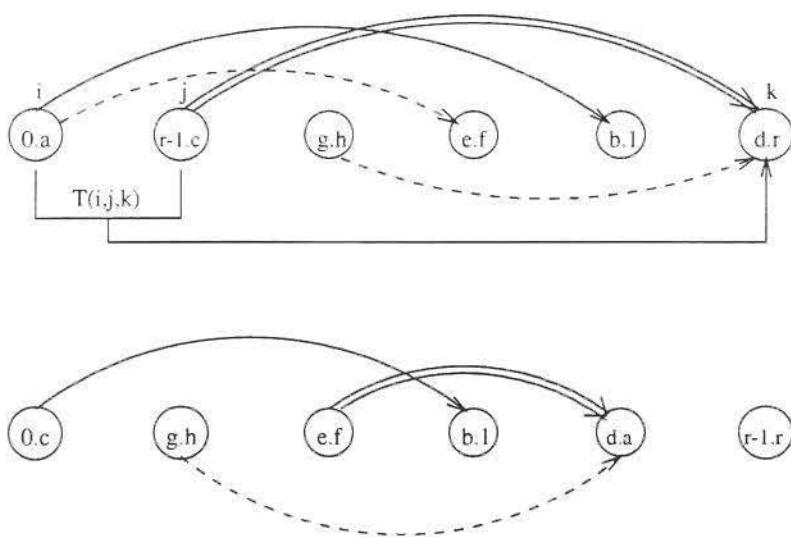


Figura 4.14: A -1 -transposição $T(i, j, k)$ indicada na figura gera um novo diagrama contendo uma -2 -transposição.

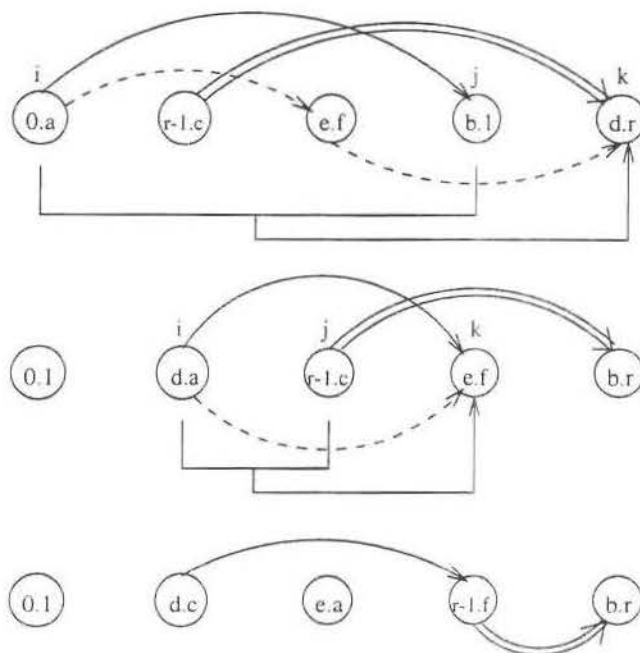


Figura 4.15: As duas -1 -transposições indicadas na figura geram um diagrama contendo uma -2 -transposição.

- $(g.h)$ entre $(e.f)$ e $(b.1)$: do Lema 4.1.7, temos o resultado desejado.
- $(g.h)$ entre $(b.1)$ e $(d.r)$: Temos necessariamente um nó $(l.m)$ entre $(g.h)$ e $(d.r)$. Então, tomando $i = (0.a)$, $j = (b.1)$ e $k = (l.m)$, a -1 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r - 1$ (o nó (0.1) é eliminado), onde temos uma -2 -transposição, $i = (g.h)$, $j = (r - 1.c)$ e $k = (d.r)$ (Figura 4.16).

□

Lema 4.1.9 Dada uma permutação π tal que $D(\pi, \iota)$:

- possui $|V| = r \geq 6$,
- não contém -2 -transposição nem -3 -transposição,
- possui nós $(0.a)$, $(b.1)$, $(r - 1.c)$ e $(d.r)$ e arestas $(0.a) \rightarrow (b.1)$, $(r - 1.c) \Rightarrow (d.r)$ tais que o nó $(r - 1.c)$ está no intervalo entre $(0.a)$ e $(b.1)$,

e existe a aresta $(0.a) \rightarrow (b.1)$ então é possível remover pelo menos quatro nós em três passos.

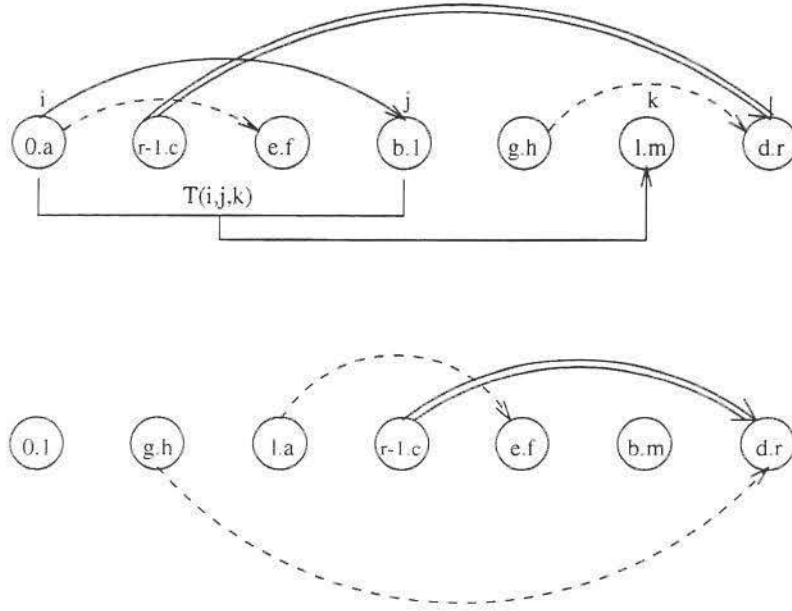


Figura 4.16: A -1 -transposição $T(i, j, k)$ indicada na figura gera um novo diagrama contendo uma -2 -transposição.

Prova:

Nas condições do lema, temos três possibilidades para $(g.h) \rightarrow (d.r)$: $(g.h) = (r-1.c)$, $(g.h)$ entre $(r-1.c)$ e $(b.l)$, e finalmente $(g.h)$ entre $(b.l)$ e $(d.r)$ (Figura 4.17):

- $(g.h) = (r-1.c)$: tomando $i = (0.a)$, $j = (b.l)$ e $k = (d.r)$, a -1 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r - 1$ (o nó $(0.l)$ é eliminado), contendo uma -3 -transposição, $i = (d.a)$, $j = (r-1.c)$ e $k = (b.r)$ (Figura 4.18).

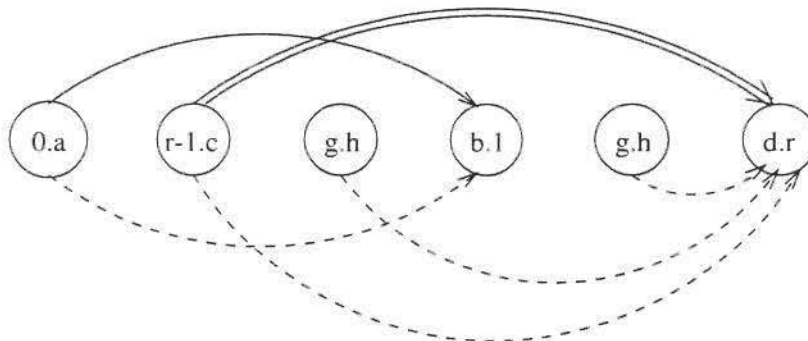


Figura 4.17: As formas gerais dos diagramas de pontos-de-quebra com $|V| \geq 6$, para as três possibilidades da aresta $(g.h) \rightarrow (d.r)$, quando $(e.f) = (b.l)$.

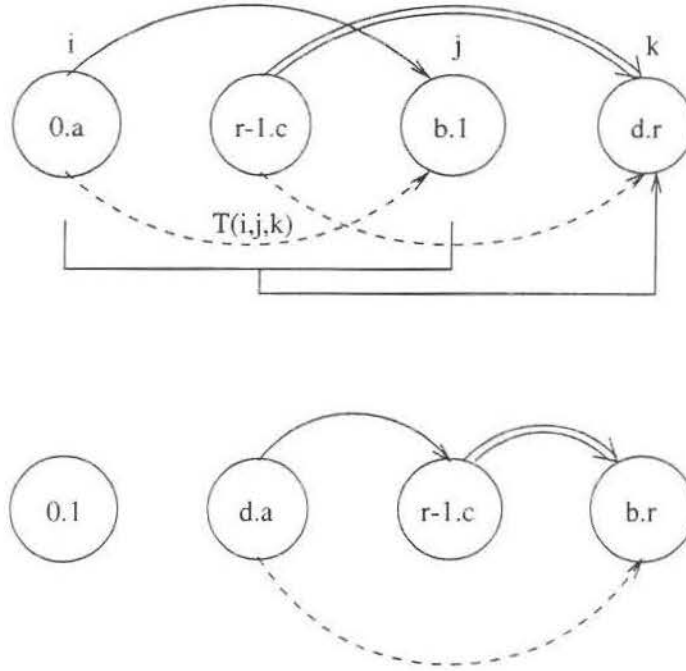


Figura 4.18: A -1 -transposição $T(i, j, k)$ indicada na figura gera um novo diagrama contendo uma -3 -transposição.

- $(g.h)$ entre $(r-1.c)$ e $(b.1)$: tomando $i = (0.a)$, $j = (r-1.c)$ e $k = (b.1)$, a -1 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r-1$ (o nó $(b.a)$ é eliminado), contendo uma -2 -transposição, $i = (0.c)$, $j = (r-1.1)$ e $k = (d.r)$ (Figura 4.19).
- $(g.h)$ entre $(b.1)$ e $(d.r)$: existe necessariamente um nó $k = (l.m)$ entre os nós $(g.h)$ e $(d.r)$. Então, tomando $i = (0.a)$ e $j = (b.1)$, a -1 -transposição $T(i, j, k)$ gera um diagrama com $|V| = r-1$ (o nó (0.1) é eliminado), contendo uma -2 -transposição, $i = (g.h)$, $j = (r-1.c)$ e $k = (d.r)$ (Figura 4.20).

□

Provamos agora que, se no diagrama de pontos-de-quebra existir unicamente a possibilidade de aplicar -1 -transposição, então podemos remover pelo menos quatro nós em três passos.

Teorema 4.1.10 *Dada uma permutação π , e $D(\pi, \iota)$, com $|V| = r \geq 6$, tais que não podemos aplicar -2 -transposições nem -3 -transposições, então é possível remover pelo menos quatro nós em três passos.*

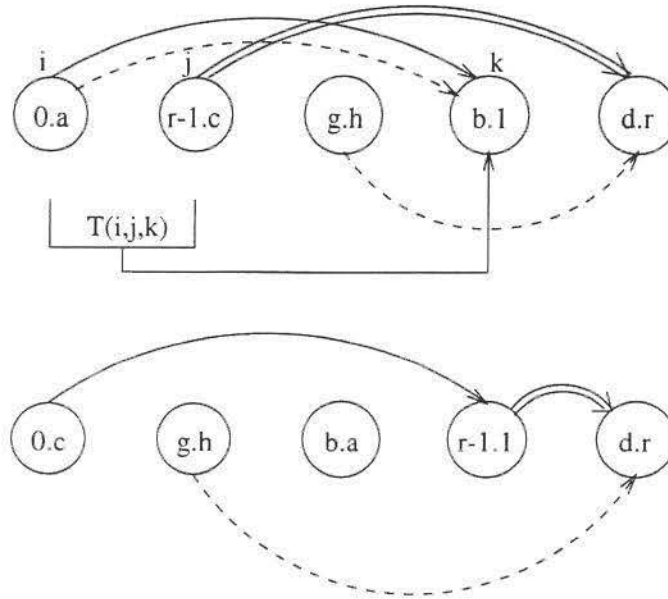


Figura 4.19: A -1 -transposição $T(i, j, k)$ indicada na figura gera um novo diagrama contendo uma -2 -transposição.

Prova:

Em todo diagrama de pontos-de-quebra com $|V| \geq 4$ sempre existe uma -1 -transposição (Lema 4.1.4).

Temos duas formas gerais de diagramas (Figura 4.21), uma forma onde o nó $(r-1.c)$ está no intervalo entre $(0.a)$ e $(b.1)$, e outra onde $(r-1.c)$ está no intervalo entre $(b.1)$ e $(d.r)$.

No primeiro caso, temos três possibilidades para a aresta $(0.a) \rightarrow (e.f)$ (Figura 4.22),

- quando $(e.f)$ está entre $(0.a)$ e $(r-1.c)$: temos o resultado do Lema 4.1.6.
- quando $(e.f)$ está entre $(r-1.c)$ e $(b.1)$: temos o resultado do Lema 4.1.8.
- quando $(e.f) = (b.1)$: temos o resultado do Lema 4.1.9.

O outro caso geral, quando $(r-1.c)$ está no intervalo entre os nós $(b.1)$ e $(d.r)$, reduz-se ao caso anterior da seguinte forma. Tomemos no intervalo entre os nós $(r-1.c)$ e $(d.r)$ o mínimo rótulo m no nó $(y.m)$. Então $m-1$ deve estar necessariamente num nó $(m-1.x)$ à esquerda de $(r-1.c)$, e deve existir uma aresta $(m-1.x) \rightarrow (y.m)$. Notemos aqui que $(y.m) \neq (r-1.c)$ senão existiria -2 -transposição $T(i, j, k)$, com $i = (m-1.x)$, $j = (r-1.c)$ e $k = (d.r)$. Além disso, se existisse um outro nó $(x-1.z)$

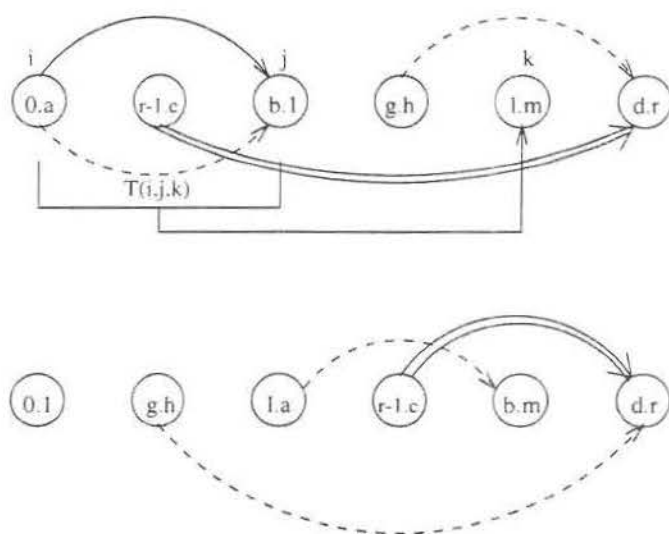


Figura 4.20: A -1 -transposição $T(i, j, k)$ indicada na figura gera um novo diagrama contendo uma -2 -transposição.

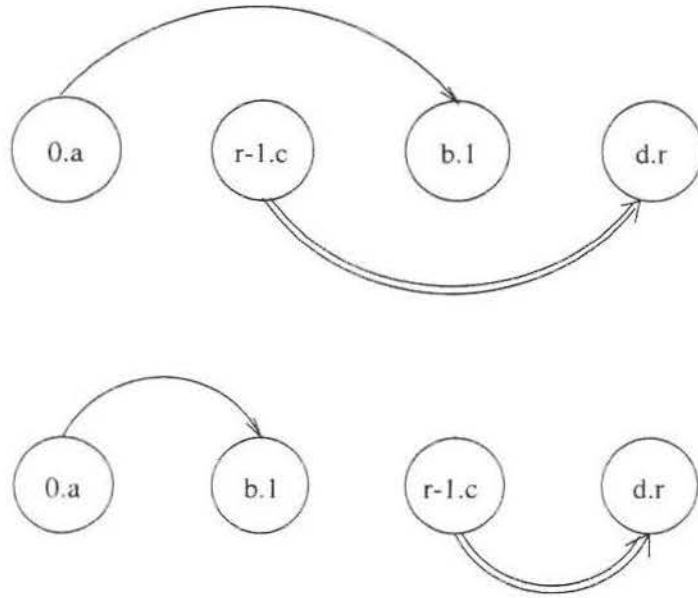


Figura 4.21: As duas únicas formas gerais de diagramas de pontos-de-quebra quando $|V| \geq 6$.

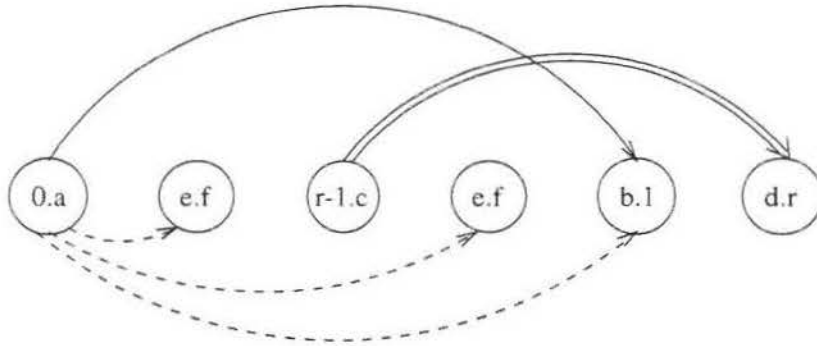


Figura 4.22: As formas gerais dos diagramas de pontos-de-quebra com $|V| = r \geq 6$, para os três casos possíveis da aresta $(0.a) \rightarrow (e.f)$, quando $(r-1.c)$ está no intervalo entre $(0.a)$ e $(b.l)$.

à esquerda de $(m-1.x)$, então o diagrama conteria a aresta $(x-1.z) \rightarrow (m-1.x)$, e portanto existiria uma -2 -transposição $T(i, j, k)$, com $i = (x-1.z)$, $j = (m-1.x)$ e $k = (y.m)$. Portanto, o nó $(x-1.z)$ está necessariamente à direita de $(m-1.x)$, de tal forma que existe a aresta $(m-1.x) \rightarrow (x-1.z)$. Podemos aplicar exatamente os casos descritos no caso geral anterior.

□

O Lema 4.1.4 implica imediatamente no próximo resultado.

Corolário 4.1.11 *Dada uma permutação π , se em $D(\pi, \iota)$ existe -3 -transposição ou -2 -transposição, então pode-se eliminar no mínimo quatro pontos-de-quebra em três passos.*

O Teorema 4.1.10 e o Corolário 4.1.11 fornecem um novo limite superior para o problema da ordenação por transposições.

Teorema 4.1.12 *Dada a permutação π , então*

$$d_t(\pi, \iota) \leq \frac{3}{4}b(\pi, \iota)$$

Utilizando o Teorema 4.1.10 e o Corolário 4.1.11, temos o algoritmo Aprox-2.25, mostrado na Figura 4.23.

Este algoritmo tem razão 2.25, conforme os Teoremas 4.1.1 e 4.1.12.

4.1.3 Resultados e análise dos experimentos

Para observar na prática a razão de aproximação, o Algoritmo Aprox-2.25 e um algoritmo exato (Figura 4.24) foram codificados em linguagem C. Observamos que o algoritmo exato também foi baseado no diagrama de pontos-de-quebra.

Foram realizados três tipos de experimentos.

O primeiro foi executar os dois algoritmos com todas as permutações de tamanho n , para n variando de 2 até 6. Em todos os casos, as distâncias de transposição encontradas por ambos os algoritmos foram iguais, exceto em seis casos, onde a distância era 3, enquanto o valor encontrado por Aprox foi 4, o que forneceu uma razão de aproximação 1.34.

O segundo tipo de testes envolveu inversas de permutações, isto é, permutações da forma

$$r_n = (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$$

Algoritmo *Aprox-2.25*

entrada: π

saída: u , uma sequência $t_1, t_2, \dots, t_u$
tal que $t_u \cdot \dots \cdot t_2 \cdot t_1 \cdot \pi = \iota$

$u \leftarrow 0$

gerar $D(\pi, \iota)$

enquanto $|V| \neq 0$ **faça**

$u \leftarrow u + 1$

se existe -3 -transposicao **então**

$t_u \leftarrow -3$ -transposicao

senão

se existe -2 -transposicao **então**

$t_u \leftarrow -2$ -transposicao

senão

$t_u \leftarrow -1$ -transposicao (conforme Teorema 4.1.10)

$\pi \leftarrow t_u \cdot \pi$

gerar $D(\pi, \iota)$

retorne $u, t_1, t_2, \dots, t_u$

Figura 4.23: O algoritmo de aproximação de razão 2.25.

Algoritmo exato**entrada:** π **saida:** u , uma seqüência $t_1, t_2, \dots, t_u$
tal que $t_u \cdot \dots \cdot t_2 \cdot t_1 \cdot \pi = \iota$, onde $u = d_\iota(\pi, \iota)$ busca(π, t_{corrente})se $\pi = \iota$ entãose $|t_{\text{corrente}}| < u$ então $u = |t_{\text{corrente}}|$ $t_{\text{melhor}} \leftarrow t_{\text{corrente}}$

senão

para todas as transposições t que podem ser aplicadas em π façase $(|t_{\text{corrente}}| + b(t \cdot \pi, \iota)/3) < u$ entãobusca($t_{\text{corrente}} \cdot \pi$, concatena(t_{corrente}, t)) $t_{\text{melhor}} \leftarrow$ lista-vazia $t_{\text{corrente}} \leftarrow$ lista-vazia $u \leftarrow b(\pi, \iota)$ busca(π, t_{corrente})**retorne** u, t_{melhor}

Figura 4.24: O algoritmo exato. Note que todos os parâmetros são passados por valor.

n	7	8	9	10	11	12	13	14	15	16	17
$d_t(r_n, \iota_n)$	4	5	5	6	6	7	7	8	8	9	9
Aprox	5	6	6	6	6	9	9	10	11	12	12
razão	1.3	1.2	1.2	1	1	1.3	1.3	1.3	1.4	1.4	1.4

Figura 4.25: Os resultados do algoritmo Aprox para permutações inversas.

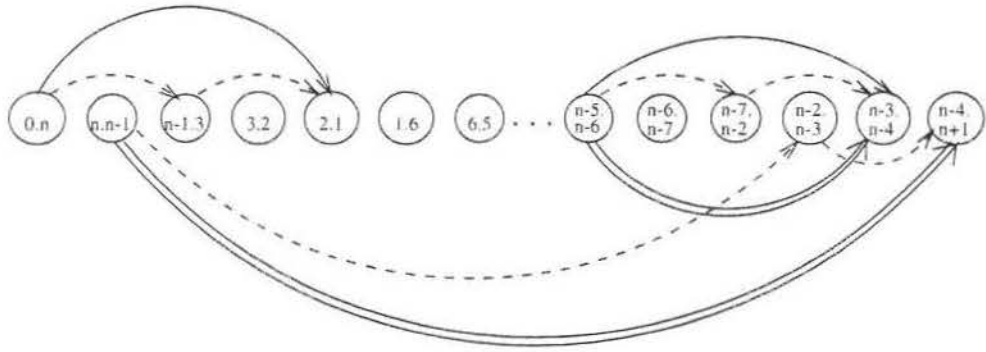


Figura 4.26: O diagrama de pontos-de-quebra geral $D(\pi_k, \iota)$, onde $\pi_k = (n \ (n-1) \ 3 \ 2 \ 1 \ 6 \ 5 \ 4 \ \dots \ (n-5) \ (n-6) \ (n-7) \ (n-2) \ (n-3) \ (n-4))$, sendo $n \geq 5+3i, i \geq 0$, e $k = c(\pi_k, \iota)$. Notamos que todos os ciclos possuem a mesma estrutura, têm tamanho três e não permitem -3 -transposições nem -2 -transposições.

Sabe-se que $d_t(r_n, \iota_n) = \left\lfloor \frac{n}{2} \right\rfloor + 1$ [30, 10]. O algoritmo Aprox funcionou conforme mostrado na tabela da Figura 4.25.

O terceiro tipo de experimentos foi aplicado numa particular família de permutações: $\pi_2 = (5 \ 4 \ 3 \ 2 \ 1)$, $\pi_3 = (8 \ 7 \ 3 \ 2 \ 1 \ 6 \ 5 \ 4)$, $\pi_4 = (11 \ 10 \ 3 \ 2 \ 1 \ 6 \ 5 \ 4 \ 9 \ 8 \ 7)$, $\pi_5 = (14 \ 13 \ 3 \ 2 \ 1 \ 6 \ 5 \ 4 \ 9 \ 8 \ 7 \ 12 \ 11 \ 10)$, ou de forma genérica $\pi_k = (n \ (n-1) \ 3 \ 2 \ 1 \ 6 \ 5 \ 4 \ \dots \ (n-5) \ (n-6) \ (n-7) \ (n-2) \ (n-3) \ (n-4))$, sendo $n = 5+3i, i \geq 0$, e $k = c(\pi_k, \iota)$ em $D(\pi_k, \iota)$ (Figura 4.26). Esta família é interessante porque suspeitamos que estas permutações produzam distâncias arbitrariamente maiores do que o limite inferior de Bafna e Pevzner [4], que envolve os ciclos do grafo de ciclos, conforme enunciado no Teorema 2.2.4 da Seção 2.2.

Executamos ambos os algoritmos com as primeiras três permutações acima como entrada, obtendo valores iguais para $k = 2, 3, 4$. Para π_5 , executamos apenas o algoritmo de aproximação, obtendo distância de transposição igual a 7. O algoritmo exato não pôde ser executado porque o tempo necessário era muito grande. Um limite inferior para $d_t(\pi_5, \iota_{14})$ é $\frac{b(\pi_5, \iota_{14})}{3} = 4$, mas ele não pode ser atingido porque se construirmos o diagrama de pontos-de-quebra correspondente, podemos observar

que não temos -3 -transposições. Portanto, $d_t(\pi_5, \iota_{14})$ é no mínimo 5, e o fator de aproximação é no máximo 1.4.

4.2 Um limite inferior para o diâmetro de transposição

Nesta seção, mostramos um limite inferior para o diâmetro de transposição, computando a distância de transposição entre uma permutação $r_n = (n \ (n-1) \ \dots \ 2 \ 1)$ e sua inversa $\iota_n = (1 \ 2 \ \dots \ (n-1) \ n)$. Na seção 4.2.1 mostramos que esta distância é $d_t(r_n, \iota_n) = \lfloor \frac{n}{2} \rfloor + 1$ para todo $n > 2$. Na seção 4.2.2 apresentamos um algoritmo que computa uma sequência ótima de transposições para estas duas permutações.

4.2.1 Distância de transposição entre r_n e ι_n

Dadas as permutações $r_n = (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$ e $\iota_n = (1 \ 2 \ \dots \ (n-1) \ n)$ queremos computar a distância de transposição $d_t(r_n, \iota_n)$, para qualquer n . No Teorema 4.2.11 abaixo provamos que $d_t(r_n, \iota_n) = \lfloor \frac{n}{2} \rfloor + 1$ para todo $n > 2$, e que $d_t(r_n, \iota_n) = 1$ se $n = 2$.

Na prova deste teorema precisamos de alguns resultados auxiliares, que passamos a enunciar.

O seguinte lema pode ser provado facilmente. Bafna e Pevzner [4] mencionam parte deste resultado no seu trabalho.

Lema 4.2.1 *Seja c um ciclo e (x, y, z) uma tripla de c na representação canônica. Então temos*

(x, y, z) é orientada se e somente se $t(y, z, x)$ é um 2-movimento

e

(x, y, z) é não-orientada se e somente se $t(y, z, x)$ é um 0-movimento

Podemos verificar que, dada a permutação $\pi_2 = (2 \ 1)$, então $d_t(\pi_2, \iota_2) = 1$, bastando tomar a transposição $t(1, 2, 3)$.

Agora enunciamos e provamos um lema que estabelece um limite inferior para $d_t(r_n, \iota_n)$ quando $n \geq 3$, e n ímpar.

Lema 4.2.2 Dadas as permutações $r_n = (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$ e $\iota_n = (1 \ 2 \ \dots \ (n-1) \ n)$, temos para $n \geq 3$, n ímpar,

$$d_t(r_n, \iota_n) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1$$

Prova: Inicialmente, do trabalho de Bafna e Pevzner [4], dada $r_n = (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$, $n \geq 3$, n ímpar, temos $c_{\text{ímpar}}(r_n, \iota_n) = 0$ ou 2 . Aplicando o limite inferior dado pelo Teorema 2.2.4 para $d_t(r_n, \iota_n)$ temos

$$\frac{(n+1) - c_{\text{ímpar}}(r_n, \iota_n)}{2} = \frac{n+1}{2} \text{ ou } \frac{n-1}{2}$$

Observe que para obter o limite inferior, cada transposição aplicada deve aumentar o número de ciclos ímpares.

Temos portanto dois casos (Figura 4.27(a)):

- No caso de 0 ciclos ímpares:

$$\frac{n+1}{2} = \left\lfloor \frac{n}{2} \right\rfloor + 1$$

Neste caso, o limite inferior é exatamente igual ao valor desejado, e portanto $d_t(r_n, \iota_n) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1$ (Figura 4.27(b)).

- No caso de 2 ciclos ímpares:

$$\frac{n-1}{2} < \left\lfloor \frac{n}{2} \right\rfloor + 1$$

e a diferença é exatamente 1. Mas aqui temos dois ciclos ímpares não-orientados

$$c_1 = [n+1, n-1, \dots, 4, 2]$$

e

$$c_2 = [n, n-2, \dots, 3, 1]$$

Ambos os ciclos são formados por seqüências estritamente decrescentes. Um caso particular é mostrado na Figura 4.27(c). Portanto qualquer tripla (x, y, z) é não-orientada, implicando em que qualquer transposição $t(x, y, z)$ é um 0-movimento (Lema 4.2.1). Assim o próximo movimento não pode aumentar $c_{\text{ímpar}}(\pi, \iota)$, e portanto não podemos obter o limite inferior.

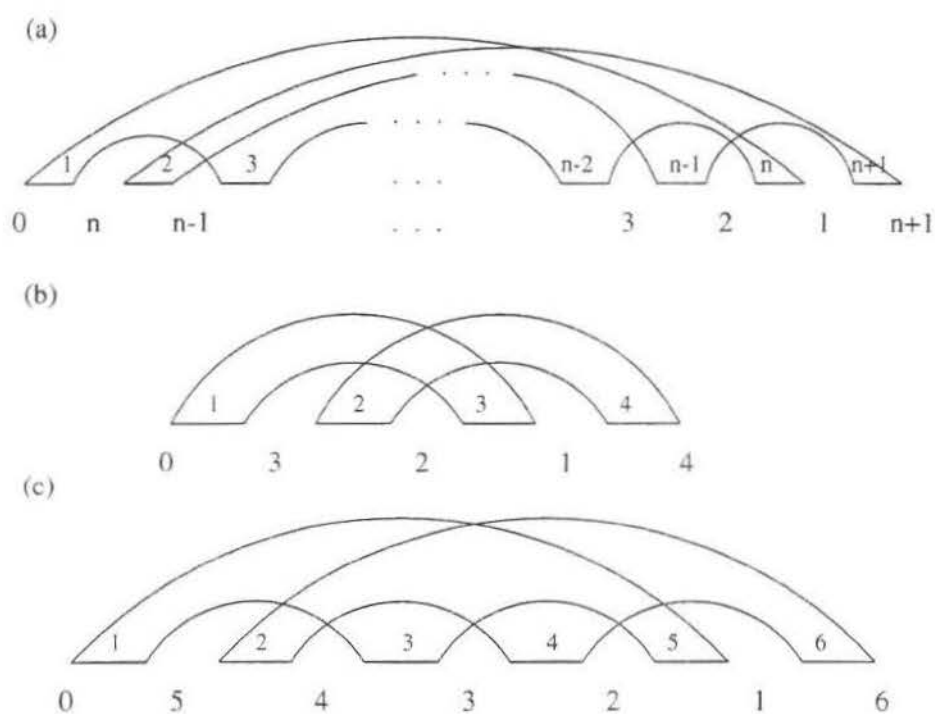


Figura 4.27: O grafo de ciclos gerado por r_n e t_n , com $n \geq 3$, n ímpar. (a) O grafo de ciclos para o caso geral. (b) O grafo para $n = 3$, onde $c_{\text{ímpar}} = 0$. (c) O grafo para $n = 5$, onde $c_{\text{ímpar}} = 2$.

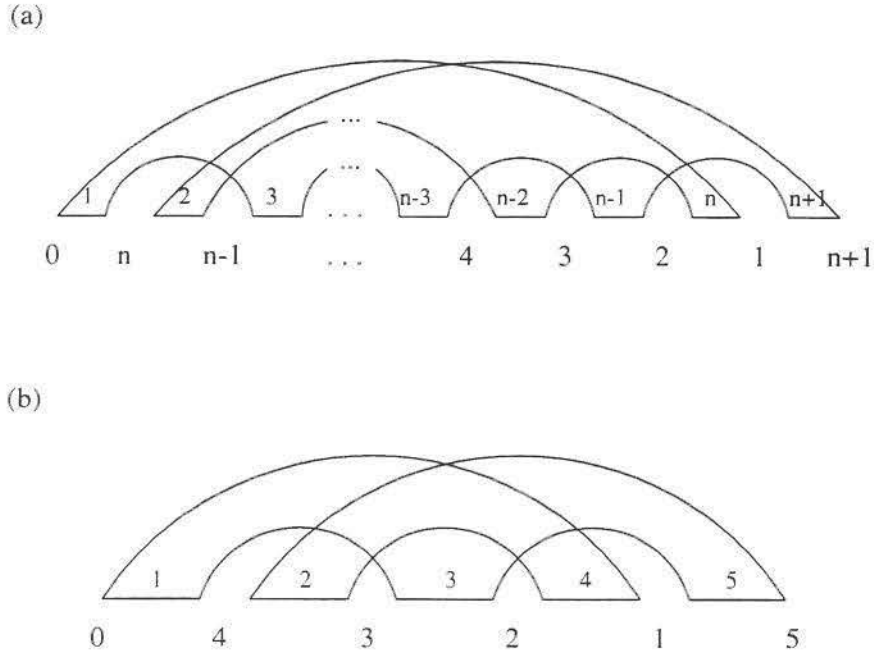


Figura 4.28: O grafo de ciclos gerado por r_n e ι_n com $n \geq 4$, n par. (a) O grafo de ciclos para o caso geral. Note que $c_{\text{ímpar}} = 1$. (b) O grafo para $n = 4$.

Então, quando n é ímpar temos $d_t(r_n, \iota_n) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1$.

□

Em seguida, enunciamos e provamos um teorema que estabelece um limite inferior para $d_t(r_n, \iota_n)$ quando $n \geq 4$, e n par. Antes de prová-lo, enunciamos alguns lemas auxiliares.

Lema 4.2.3 *Dadas as permutações $r_n = (n \ (n-1) \ \dots \ 1)$ e $\iota_n = (1 \ 2 \ \dots \ n)$, com $n \geq 4$ e n par, então no grafo de ciclos correspondente de r_n com relação a ι_n , os únicos 2-movimentos possíveis são aqueles obtidos de uma transposição $t(i, j, k)$ com i e k com a mesma paridade e j com paridade oposta a i e k .*

Prova: Inicialmente, observamos que o único ciclo do grafo de ciclos de r_n com relação a ι_n (Figura 4.28) é

$$[n+1, n-1, n-3, \dots, 1, n, n-2, \dots, 2]$$

A ordem das arestas realidade neste ciclo é tal que todos os rótulos ímpares aparecem juntos, em ordem decrescente, e todos os rótulos pares aparecem juntos, também em

ordem decrescente. Do Lema 4.2.1 sabemos que todo 2-movimento corresponde a uma tripla orientada neste único ciclo.

Podemos verificar que todas as triplas orientadas começando em uma aresta realidade *ímpar* são

$$(n+1-k, n-i, n-j), \text{ } k \text{ par } 0 \leq k \leq n, \\ i \text{ ímpar } k < i \leq n-1 \text{ e } j \text{ par } 0 \leq j < i$$

Podemos escrever os intervalos de k , i e j de forma mais apropriada, como k par $0 \leq k \leq n-2$, i ímpar $k+1 \leq i \leq n-1$ e j par $k \leq j \leq i-1$.

Podemos verificar também que todas as triplas orientadas começando com uma aresta realidade *par* são

$$(n+1-k, n-i, n-j)$$

com os intervalos k ímpar $1 \leq k \leq n-3$, i par $k+1 \leq i \leq n-2$ e j ímpar $k \leq j \leq i-1$

Podemos reunir ambos os casos em um único, de tal forma que todas as triplas orientadas do grafo de ciclos são

$$(n+1-k, n-i, n-j)$$

onde k e j têm a mesma paridade, i tem paridade oposta a k e j e $0 \leq k \leq n-2$, $k+1 \leq i \leq n-1$ e $k \leq j \leq i-1$.

Portanto, todos os 2-movimentos são

$$t(n-i, n-j, n+1-k) \cdot (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$$

Para computar esta transposição de maneira mais fácil, faremos algumas substituições nos seus parâmetros. Tomemos $i' = n-i$, $j' = n-j$ e $k' = n+1-k$. Observe que k' e j' têm paridades diferentes, i' e j' também têm paridades diferentes, e i' e k' têm a mesma paridade.

Então, se modificarmos $k = n+1-k'$, $i = n-i'$ e $j = n-j'$ nos intervalos acima teremos $0 \leq n+1-k' \leq n-2$, $n+1-k'+1 \leq n-i' \leq n-1$ e $n+1-k' \leq n-j' \leq n-i'-1$.

Agora devemos definir os intervalos de k' , i' e j' . Primeiro, tomando o produto do intervalo de k' com -1 , e depois adicionando $n+1$, obtemos o intervalo de k' , $n+1 \geq k' \geq 3$. Segundo, tomando o produto do intervalo de i' com -1 , e em seguida, adicionando n obtemos o intervalo de i' , $k'-2 \geq i' \geq 1$. Finalmente tomando o produto do intervalo de j' com -1 , e depois adicionando n , temos o intervalo de j' , $k'-1 \geq j' \geq i'+1$.

Segue que todos os possíveis 2-movimentos são da forma $t(i, j, k)$, onde $k \not\equiv j \pmod{2}$ e $k \equiv i \pmod{2}$, com os intervalos $3 \leq k \leq n+1$, $1 \leq i \leq k-2$ e $i+1 \leq j \leq k-1$.

Se unirmos estes intervalos então obteremos $1 \leq i < j < k \leq n+1$, com i e j com paridades opostas e k com a mesma paridade de i .

Portanto, os únicos 2-movimentos possíveis são aqueles obtidos de transposições $t(i, j, k)$ tais que i e k têm a mesma paridade e j tem a paridade oposta a i e k . $\square$

Os lemas seguintes mostram que, nos grafos de ciclos gerados por $r_n = (n \ (n-1) \ \dots \ 1)$ com relação a $\iota_n = (1 \ 2 \ \dots \ n)$, a aplicação de transposições 2-movimentos gera novos grafos de ciclos que contêm apenas ciclos não-orientados.

Lema 4.2.4 *Dadas as permutações $r_n = (n \ (n-1) \ \dots \ 1)$ e $\iota_n = (1 \ 2 \ \dots \ n)$, $n \geq 4$, n par, então a transposição $t(i, j, k)$ com $1 < i < j < k < n+1$, i e k com a mesma paridade e j com a paridade oposta a i e k , gera um grafo de ciclos contendo três ciclos não-orientados.*

Prova:

Primeiro, dadas r_n , ι_n , n , i , j e k nas condições da hipótese, a transposição $t(i, j, k)$ é um 2-movimento (Lema 4.2.3):

$$\begin{aligned} t(i, j, k) \cdot (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1) = \\ (n \ (n-1) \ \dots \ (n-i+2) \ (n-j+1) \ \dots \ (n-k+2) \\ (n-i+1) \ \dots \ (n-j+2) \ (n-k+1) \ \dots \ 1). \end{aligned}$$

As Figuras 4.29 e 4.30 mostram os grafos de ciclos gerados por transposições com estas características.

Vamos considerar inicialmente o caso de i par (Figura 4.29).

Podemos verificar que este grafo de ciclos possui três ciclos:

$$\begin{aligned} c_1 &= [n+1, n-1, \dots, k+1, i+(k-j)-1, i+(k-j)-3, \dots, i, i-2, \dots, 2] \\ c_2 &= [n, n-2, \dots, k, k-2, \dots, i+(k-j)+1, i-1, i-3, \dots, 1] \\ c_3 &= [k-1, k-3, \dots, i+(k-j), \dots, i+1] \end{aligned}$$

O ciclo c_1 é formado por uma sequência estritamente decrescente, pois $k+1 > i+(k-j)+1 > i$. Portanto, podemos verificar que c_1 contém apenas triplas não-orientadas, implicando em que c_1 seja um ciclo não-orientado. Usando o mesmo raciocínio para

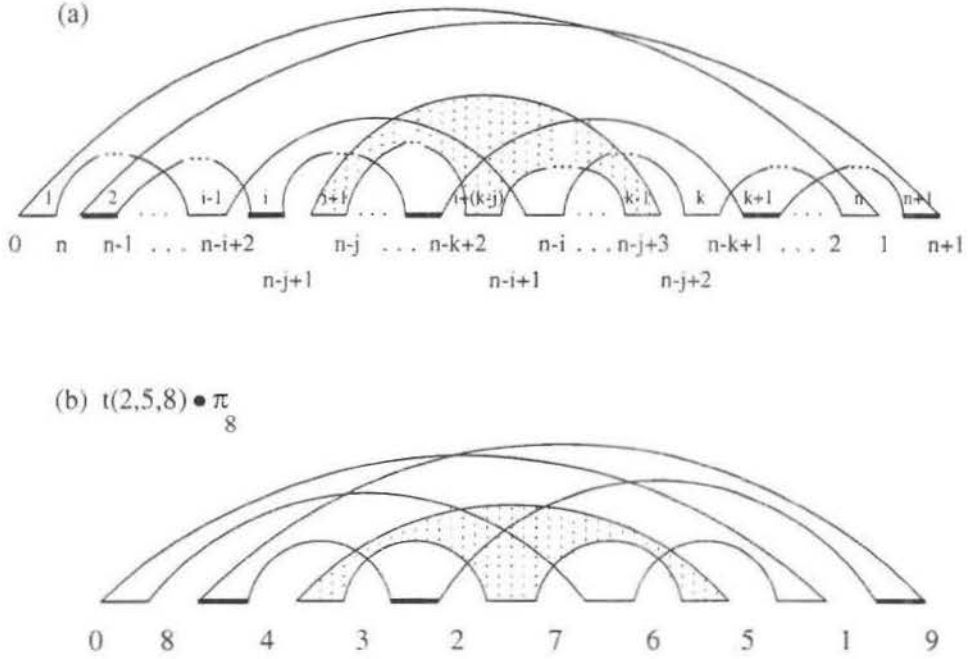


Figura 4.29: O grafo de ciclos gerado pela aplicação de uma transposição $t(i, j, k)$ a r_n , com n par ≥ 4 , i e k pares e j ímpar. Note que os três ciclos são não-orientados. (a) O caso genérico. (b) Um caso particular.

os ciclos c_2 e c_3 , concluímos que este grafo de ciclos é formado por três ciclos não-orientados.

Temos ainda a possibilidade de i ser ímpar (figura 4.30), mas a análise neste caso é idêntica à do caso par. Os ciclos neste caso são

$$c_1 = [n+1, n-1, \dots, k, k-2, \dots, i+(k-j)+1, i-1, i-3, \dots, 2]$$

$$c_2 = [n, n-2, \dots, k+1, i+(k-j)-1, \dots, i, i-2, \dots, 1]$$

$$c_3 = [k-1, k-3, \dots, i+(k-j), i+(k-j)-2, \dots, i+1]$$

Porém, em ambos os casos geramos um grafo de ciclos contendo três ciclos não-orientados.

□

Lema 4.2.5 Dadas as permutações $r_n = (n(n-1) \dots 1)$ e $\iota_n = (1\ 2 \dots n)$, $n \geq 4$, n par, então a transposição $t(1, j, k)$ com $1 < j < k < n+1$, k ímpar e j par gera um grafo de ciclos contendo três ciclos não-orientados.

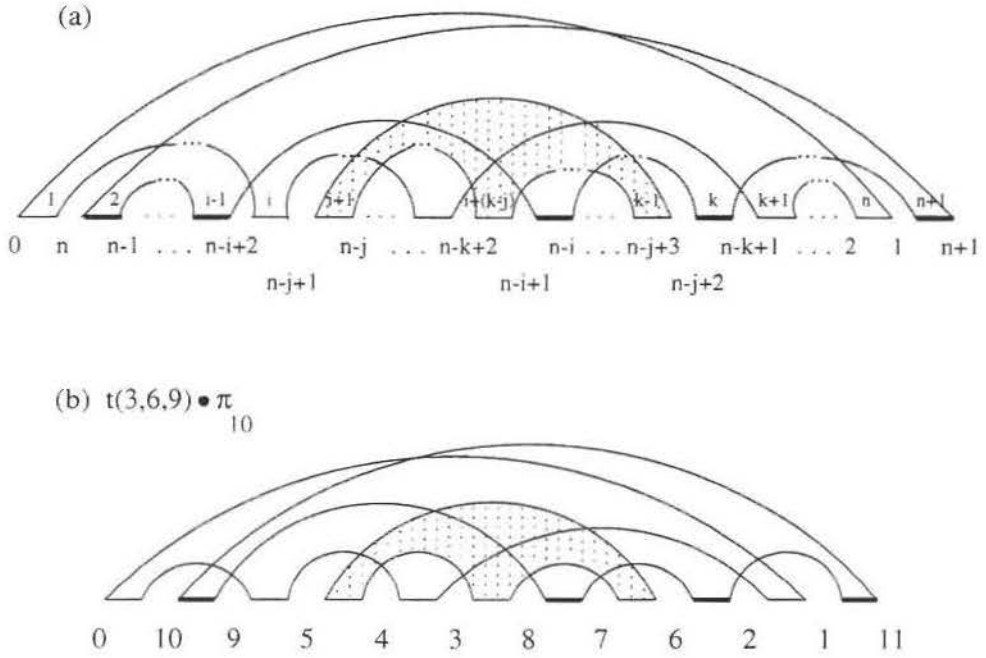


Figura 4.30: O grafo de ciclos gerado pela aplicação de uma transposição $t(i, j, k)$ a r_n , com n par ≥ 4 , i e k ímpares e j par. Note que os três ciclos são não-orientados. (a) O caso genérico. (b) Um caso particular.

Prova:

De forma análoga à prova do lema anterior, dadas r_n , t_n , n , j e k nas condições da hipótese, a transposição $t(1, j, k)$ é um 2-movimento (Lema 4.2.3):

$$\begin{aligned} t(1, j, k) \cdot (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1) = \\ ((n-j+1) \ \dots \ (n-k+2) \ n \ (n-1) \ \dots \ (n-j+2) \\ (n-k+1) \ \dots \ 2 \ 1). \end{aligned}$$

Os ciclos que compõem o grafo de ciclos são:

$$\begin{aligned} c_1 &= [n+1, n-1, \dots, k, k-2, \dots, k-j+2] \\ c_2 &= [n, n-2, \dots, k+1, k-j, k-j-2, \dots, 1] \\ c_3 &= [k-1, k-3, \dots, k-j+1, k-j-1, \dots, 2] \end{aligned}$$

Podemos verificar que os três ciclos são não-orientados.

A Figura 4.31 mostra o grafo de ciclos gerado por este 2-movimento.

□

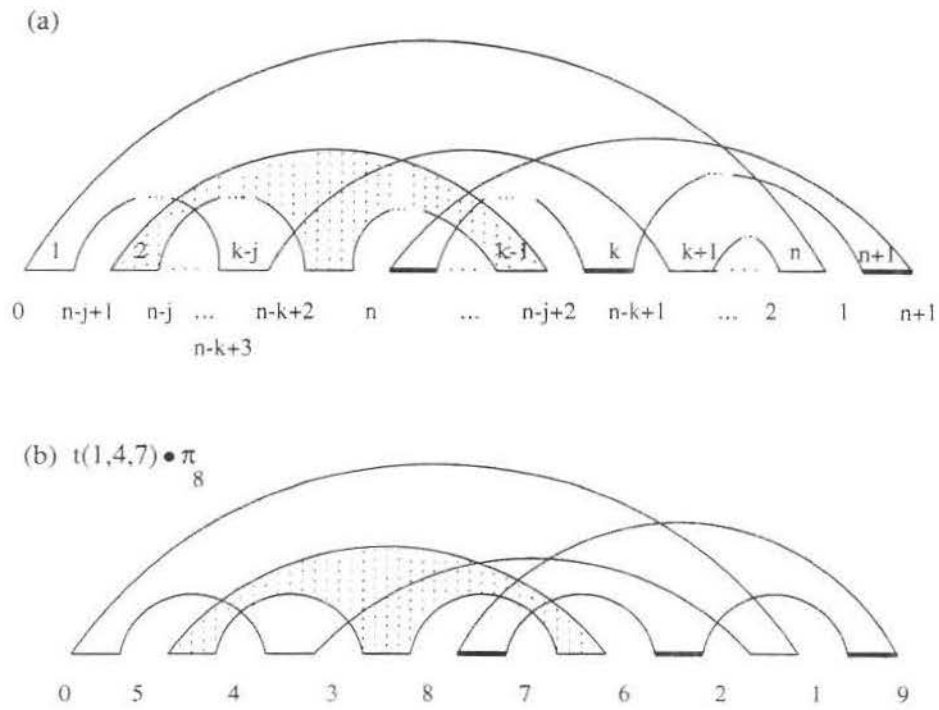


Figura 4.31: O grafo de ciclos gerado pela aplicação de transposições $t(1, j, k)$, com k ímpar e j par. Note que os três ciclos são não-orientados. (a) O grafo de ciclos geral gerado por $t(1, j, k)$. (b) Um caso particular.

Lema 4.2.6 *Dadas as permutações $r_n = (n (n-1) \dots 1)$ e $\iota_n = (1 2 \dots n)$, $n \geq 4$, n par, então a transposição $t(i, j, n+1)$ com $1 < i < j < n+1$, i ímpar e j par gera um grafo de ciclos contendo três ciclos não-orientados.*

Prova:

Novamente, dadas r_n, ι_n, n, i e j nas condições da hipótese, a transposição $t(i, j, n+1)$ é um 2-movimento (Lema 4.2.3):

$$\begin{aligned} t(i, j, n+1) \cdot (n (n-1) (n-2) \dots 2 1) = \\ (n (n-1) \dots (n-i+2) (n-j+1) \dots 2 1 \\ (n-i+1) \dots (n-j+2)). \end{aligned}$$

Os três ciclos que compõem o grafo de ciclos são:

$$\begin{aligned} c_1 &= [n+1, n-1, \dots, i+(n-j)+2, i-1, i-3, \dots, 2] \\ c_2 &= [n, n-2, \dots, i+(n-j)+1, i+(n-j)-1, \dots, i+1] \\ c_3 &= [i+(n-j), i+(n-j)-2, \dots, i, i-2, \dots, 1] \end{aligned}$$

Podemos verificar que os três ciclos são não-orientados.

A Figura 4.32 mostra o grafo de ciclos gerado por este 2-movimento.

□

Lema 4.2.7 *Dadas as permutações $r_n = (n (n-1) \dots 1)$ e $\iota_n = (1 2 \dots n)$, $n \geq 4$, n par, então a transposição $t(1, j, n+1)$ com $1 < j < n+1$ e j par gera um grafo de ciclos contendo três ciclos não-orientados.*

Prova:

De novo, dadas r_n, ι_n, n e j nas condições da hipótese, a transposição $t(1, j, n+1)$ é um 2-movimento (Lema 4.2.3):

$$\begin{aligned} t(1, j, n+1) \cdot (n (n-1) (n-2) \dots 2 1) = \\ ((n-j+1) \dots 2 1 n (n-1) \dots (n-j+2)). \end{aligned}$$

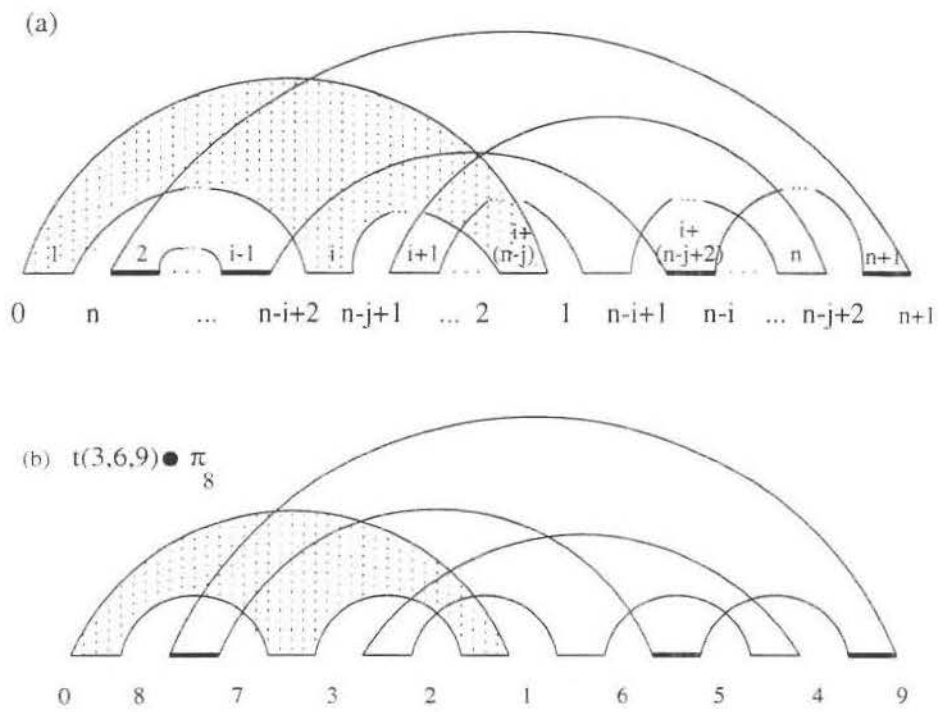


Figura 4.32: O grafo de ciclos gerado por transposições $t(i, j, n+1)$, i ímpar e j par. Note que os três ciclos são não-orientados, conforme descrito no texto. (a) O grafo de ciclos geral gerado por $t(i, j, n+1)$. (b) Um caso particular.

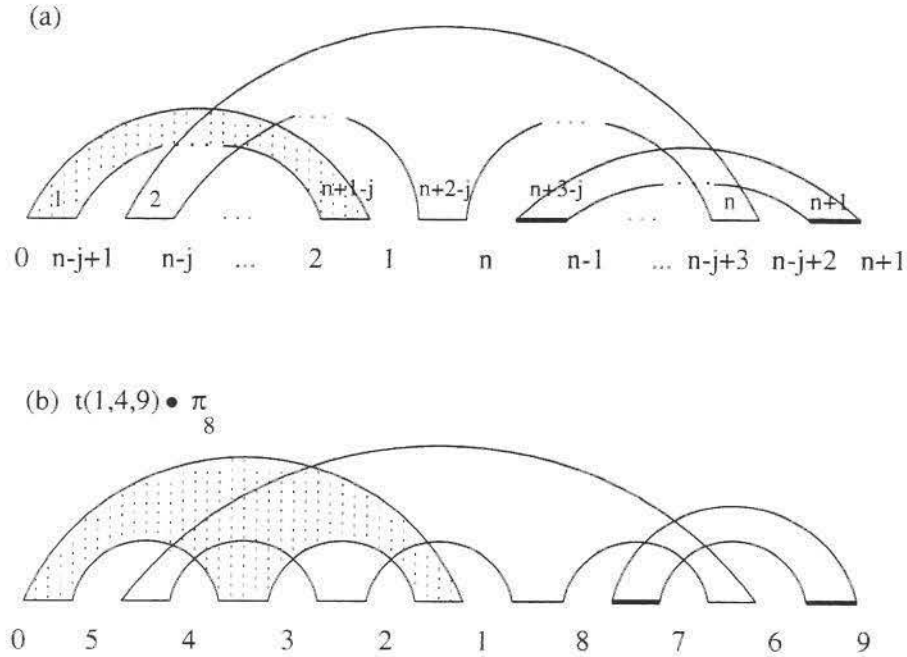


Figura 4.33: O grafo de ciclos gerado por aplicação de transposições $t(1, j, n+1)$, com j par. Os três ciclos são não-orientados, conforme descrito no texto. (a) O grafo de ciclos geral gerado por $t(1, j, n+1)$. (b) Um caso particular.

Os três ciclos que compõem o grafo de ciclos são:

$$\begin{aligned} c_1 &= [n+1, n-1, \dots, n+3-j] \\ c_2 &= [n, n-2, \dots, n+2-j, n-j, \dots, 2] \\ c_3 &= [n+1-j, n-j-1, \dots, 1] \end{aligned}$$

De novo, podemos verificar que os três ciclos são não-orientados.

A Figura 4.33 mostra o grafo de ciclos gerado por este 2-movimento.

□

Lema 4.2.8 Dadas as permutações $r_n = (n \ (n-1) \ \dots \ 1)$ e $\iota_n = (1 \ 2 \ \dots \ n)$, $n \geq 4$, n par, então as transposições $t(i, j, k)$, com i e k com a mesma paridade e j com paridade oposta a i e k são as únicas transposições 2-movimentos possíveis, e quando aplicadas a r_n geram grafos de ciclos contendo três ciclos não-orientados.

Prova: Dadas as permutações r_n e ι_n , nas condições do lema, as únicas transposições 2-movimentos possíveis são aquelas geradas por triplas orientadas. O Lema 4.2.3

mostra que os únicos 2-movimentos possíveis neste caso são aqueles onde i e k têm a mesma paridade e j tem paridade oposta a i e k .

Existem exatamente quatro casos que podem ocorrer com i , j e k nestas condições que são:

1. $1 < i < j < k < n + 1$: o Lema 4.2.4 mostra que o grafo de ciclos gerado pela transposição $t(i, j, k)$, com i, j e k nas condições do lema, produz um grafo de ciclos contendo três ciclos, todos não-orientados.
2. $1 = i < j < k < n + 1$: o Lema 4.2.5 mostra que o grafo de ciclos gerado pela transposição $t(i, j, k)$, com j e k nas condições do lema, produz um grafo de ciclos contendo três ciclos, todos não-orientados.
3. $1 < i < j < k = n + 1$: o Lema 4.2.6 mostra que o grafo de ciclos gerado pela transposição $t(i, j, k)$, com i e j nas condições do lema, produz um grafo de ciclos contendo três ciclos, todos não-orientados.
4. $1 = i < j < k = n + 1$: o Lema 4.2.7 mostra que o grafo de ciclos gerado pela transposição $t(i, j, k)$, com j nas condições do lema, produz um grafo de ciclos contendo três ciclos, todos não-orientados.

Temos então o resultado. □

Provamos agora o teorema que estabelece um limite inferior para $d_t(r_n, \iota_n)$, quando $n \geq 4$ par.

Teorema 4.2.9 *Dadas as permutações $r_n = (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$ e $\iota_n = (1 \ 2 \ \dots \ (n-1) \ n)$, temos para $n \geq 4$, n par,*

$$d_t(r_n, \iota_n) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1$$

Prova: Novamente, do trabalho de Bafna e Pevzner [4], dada $r_n = (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$, $n \geq 4$, n par, temos $c_{\text{impar}}(\pi, \iota) = 1$. Aplicando o limite inferior dado pelo Teorema 2.2.4 para $d_t(r_n, \iota_n)$ temos

$$\frac{(n+1) - c_{\text{impar}}(r_n, \iota_n)}{2} = \frac{(n+1) - 1}{2} = \frac{n}{2}.$$

Neste caso (Figura 4.28(a)):

$$\frac{n}{2} = \left\lfloor \frac{n}{2} \right\rfloor < \left\lfloor \frac{n}{2} \right\rfloor + 1$$

e a diferença é de exatamente 1. Neste caso, temos que provar que existe necessariamente uma transposição que não incrementará $c_{\text{impar}}(r_n, \iota_n)$ durante qualquer sequência ótima de transposições que transforma π em ι .

A primeira transposição é um 0-movimento ou um 2-movimento. Não podemos aplicar um -2 -movimento porque o primeiro grafo de ciclos é formado por apenas um ciclo.

Se aplicamos um 0-movimento, o único ciclo ímpar do grafo de ciclos é transformado em um outro ciclo ímpar, não incrementando c_{impar} .

Então, temos que investigar o que acontece se aplicarmos um 2-movimento. O Lema 4.2.8 mostra que as únicas transposições que são 2-movimentos geram grafos de ciclos que contêm apenas ciclos não-orientados.

Isto implica no resultado da seguinte forma. Temos duas possibilidades. Se o grafo de ciclos resultante tem um ciclo ímpar e dois ciclos pares, a primeira transposição não aumentou c_{impar} . Por outro lado, se obtivermos três ciclos ímpares, a segunda transposição da série não pode aumentar c_{impar} , pois todos os ciclos são não-orientados, implicando em termos no mínimo mais um 0-movimento.

□

O Lema 4.2.2 e o Teorema 4.2.9 implicam imediatamente no resultado abaixo.

Teorema 4.2.10 *Dadas as permutações $r_n = (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$ e $\iota_n = (1 \ 2 \ \dots \ (n-1) \ n)$, temos para todo $n \geq 3$,*

$$d_t(r_n, \iota_n) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1$$

Finalmente, o teorema abaixo mostra a distância de transposição entre uma permutação e sua inversa.

Teorema 4.2.11 *Dadas as permutações $r_n = (n \ (n-1) \ (n-2) \ \dots \ 2 \ 1)$ e $\iota_n = (1 \ 2 \ \dots \ (n-1) \ n)$, temos para $n \geq 2$*

$$d_t(r_n, \iota_n) = \begin{cases} 1 & \text{se } n = 2 \\ \left\lfloor \frac{n}{2} \right\rfloor + 1 & \text{se } n \geq 3 \end{cases}$$

Prova: Conforme afirmação anterior, sabemos que $d_t(r_n, \iota_n) = 1$ quando $n = 2$.

Do Teorema 4.2.10, temos que $d_t(r_n, \iota_n) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1$.

Algoritmo *DistTransp***entrada:** $r_n = (n \ (n-1) \ \dots \ 2 \ 1), n > 2$ **saída:** $\delta, t_1, t_2, \dots, t_\delta$, onde $\delta = d_t(r_n, \iota)$ $\pi_1 \leftarrow t_1(1, \lfloor \frac{n}{2} \rfloor, n) \cdot \pi$ $\delta \leftarrow 1$ $\pi_0 \leftarrow \pi$ **se** n **for** par **então** $\delta \leftarrow \delta + 1$ $\pi_2 \leftarrow t_\delta(\frac{n}{2}, \frac{n}{2} + 1, n + 1) \cdot \pi_1$ $k \leftarrow 1$ $p \leftarrow 1$ **se** n **for** ímpar **então** $k \leftarrow 0$ $p \leftarrow 0$ **enquanto** $k < \lfloor \frac{n}{2} \rfloor$ **faça** $\delta \leftarrow \delta + 1$ $\pi_\delta \leftarrow t_\delta(\lfloor \frac{n}{2} \rfloor - k, \lfloor \frac{n}{2} \rfloor - k + 2, n + 1 - k + p) \cdot \pi_{\delta-1}$ $k \leftarrow k + 1$ **retorne** $\delta, t_1, t_2, \dots, t_\delta$

Figura 4.34: Um algoritmo para computar a distância de transposição entre uma permutação e sua inversa.

Um limite superior é dado pelo algoritmo apresentado na próxima seção, que computa $d_t(r_n, \iota_n)$ em $\lfloor \frac{n}{2} \rfloor + 1$ passos. Portanto, $d_t(r_n, \iota_n) \leq \lfloor \frac{n}{2} \rfloor + 1$, para todo $n > 2$. Notamos que Bafna e Pevzner [4] mostraram este mesmo limite superior, para todo $n \geq 1$.

□

4.2.2 Um algoritmo para computar $d_t(r_n, \iota_n)$

Apresentamos nesta seção um algoritmo para computar a distância de transposição entre uma permutação e sua inversa. Observamos que este algoritmo executa sem utilizar o grafo de ciclos. Em seu lugar, ele usa uma determinada série de transposições que funcionam no caso particular destas duas permutações. A Figura 4.34 mostra o algoritmo, e a Figura 4.35 mostra exemplos de execuções para $n = 6$ e $n = 7$.

Os quatro passos iniciais criam, a partir da permutação inicial, uma nova permutação com duas subsequências decrescentes na sua extremidade esquerda, e uma sequência

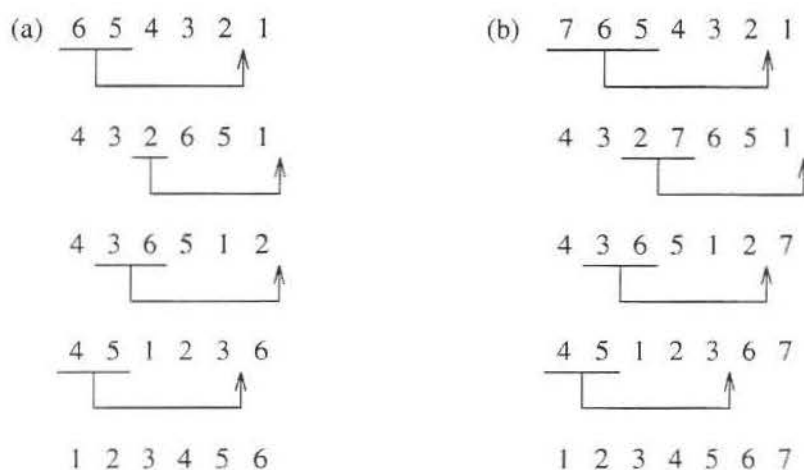


Figura 4.35: Duas execuções do algoritmo. (a) Exemplo com $n = 6$. (b) Exemplo com $n = 7$.

crescente, na sua extremidade direita. Se n for par temos

$$\pi_2 = \left(\left(\frac{n}{2} + 1 \right) \frac{n}{2} \dots 3 \right) \left(n(n-1) \dots \left(\frac{n}{2} + 2 \right) \right) (1 2)$$

Observemos que as duas subsequências decrescentes têm $\lceil \frac{n}{2} \rceil - 1$ elementos cada uma. Marcamos as subsequências com parênteses.

Se n for ímpar teremos

$$\pi_1 = \left(\left(\frac{n+1}{2} \right) \dots 3 2 \right) \left(n(n-1) \dots \left(\frac{n+1}{2} + 1 \right) \right) (1)$$

Analogamente, neste caso as duas primeiras subsequências também terão $\lceil \frac{n}{2} \rceil - 1$ elementos cada uma.

O comando de repetição no quinto passo move o último elemento da primeira subsequência e o primeiro elemento da segunda subsequência para a extremidade direita da permutação, onde duas outras subsequências são incrementadas à medida que o algoritmo executa. Genericamente, se n for par então teremos, após $k - 1$ iterações do comando,

$$\pi_{k+1} = \left(\left(\frac{n}{2} + 1 \right) \frac{n}{2} \dots (k+2) \right) \left((n-k+1) \dots \left(\frac{n}{2} + 2 \right) \right) (1 2 \dots (k+1)) \left((n-k+2) \dots n \right).$$

Se n for ímpar teremos, após k iterações,

$$\pi_{k+1} = \left(\left(\frac{n+1}{2} \right) \dots (k+2) \right) \left((n-k) \dots \left(\frac{n+1}{2} + 1 \right) \right) (1 2 \dots (k+1)) \left((n-k+1) \dots n \right).$$

Portanto, o algoritmo corretamente transforma a permutação na sua inversa, usando transposições. Além disso, o algoritmo executa em $\lceil \frac{n}{2} \rceil + 1$ passos, para $n > 2$.

4.3 Sumário e questões

Neste capítulo, inicialmente apresentamos um algoritmo de aproximação de raz ao 2.25 e um algoritmo exato, ambos baseados numa estrutura denominada de diagrama de pontos-de-quebra. Embora a razão seja elevada, quando comparado à melhor conhecida, experimentos demonstraram que esta razão é bastante menor quando executamos o algoritmo, sugerindo que ele possa ser útil na prática.

Algumas questões surgiram deste estudo. Primeiro, realizar mais testes com o algoritmo de aproximação, usando permutações significativamente maiores, permitiria avaliar a sua utilidade efetiva. Segundo, pesquisar se a razão do algoritmo pode ser diminuída, e verificar a possibilidade de decidir quando $d_t(\pi, \iota) = \frac{b(\pi)}{3}$ apenas estudando o diagrama, possivelmente forneceria parâmetros que poderiam ser empregados numa prova de complexidade para o problema da distância de transposição. Um outro ponto interessante seria investigar se existe uma permutação alcançando a razão 2.25. Por fim, o algoritmo de aproximação de Bafna e Pevzner [4] poderia ser implementado, permitindo comparar estes dois algoritmos.

Em seguida, demonstramos que a distância de transposição entre uma permutação e sua inversa (sem complementação) é $\lfloor \frac{n}{2} \rfloor + 1$ para todo $n > 2$, onde n é o tamanho da permutação. Além disso, apresentamos um algoritmo que encontra uma série ótima de transposições que ordenam a permutação, para o caso estudado. Portanto, temos um limite inferior para o diâmetro de transposição, $\lfloor \frac{n}{2} \rfloor + 1 \leq D_t(n)$. Bafna e Pevzner [4] demonstraram que $D_t(n) \leq \frac{3}{4}n$. Conjecturamos que o limite inferior seja de fato o valor do diâmetro de transposição.

Outras questões interessantes ainda não resolvidas são a complexidade do problema e um valor exato para o diâmetro de transposição.

Capítulo 5

Distância de reversão e transposição de cromossomos lineares

Atualmente, a análise de genomas evoluindo por diversos eventos mutacionais representam um grande desafio. Hannenhalli e co-autores [17] analisaram genomas evoluindo por diferentes eventos, particularmente reversões e transposições. Hannenhalli e Pevzner [20] apresentaram um algoritmo de tempo polinomial para comparar dois genomas evoluindo por reversões, translocações, fusões e fissões. Gu, Peng e Sudborough [14] mostraram algoritmos de aproximação para computar a distância entre dois genomas, permitindo três eventos, reversões, transposições e transversões.

Neste capítulo contribuímos para a análise de reversões e transposições agindo em um cromossomo.

Inicialmente, na Seção 5.1 apresentamos definições que serão utilizadas nas seções posteriores. Na Seção 5.2 estendemos a análise de transposições agindo em permutações sem sinais para transposições atuando em permutações com sinais, obtendo assim algoritmos de aproximação para computar a distância de reversão e transposição para ambas as permutações, com e sem sinais. Em seguida, na Seção 5.3 apresentamos um limite inferior para o diâmetro de reversão e transposição de permutações com sinais. Por fim, a Seção 5.4 traz um sumário e questões relativas a estes estudos.

$$0 \bullet 5 \bullet 1 \quad 2 \bullet 4 \bullet 7 \quad 6 \bullet 3 \bullet 9 \quad 8 \bullet 10$$

Figura 5.1: Os *pontos-de-quebra* e as *faixas* de uma permutação $\pi = (0 \ 5 \ 1 \ 2 \ 4 \ 7 \ 6 \ 3 \ 9 \ 8 \ 10)$ com relação a $\iota = (0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10)$. As faixas são as seqüências entre dois pontos-de-quebra consecutivos.

5.1 Definições

Nesta seção, primeiro enunciamos o problema de ordenar permutações lineares por reversões e transposições, e em seguida recordamos certos conceitos, que serão utilizados nas próximas seções.

No problema da distância de reversão e transposição de permutações lineares, sem perda de generalidade, podemos fixar σ como sendo a permutação identidade ι , isto é, $\iota = (1 \dots n)$ no caso sem sinais, e $\iota = (+1 \dots +n)$ no caso com sinais. Portanto, o problema estudado neste capítulo pode ser enunciado da seguinte forma. Dada a permutação π , queremos computar uma série de reversões e transposições que transformam π em ι , isto é, queremos encontrar $e_1, e_2, \dots, e_u$, onde cada e_i é uma reversão ou uma transposição, tais que $e_u \cdot e_{u-1} \cdot \dots \cdot e_2 \cdot e_1 \cdot \pi = \iota$ e u seja mínimo. Chamamos u de **distância de reversão e transposição** e o denotamos por $d_{rt}(\pi, \iota)$.

No que se segue, uma **operação** ϵ pode ser uma reversão ou uma transposição.

Estendemos a permutação π adicionando $\pi_0 = 0$ e $\pi_{n+1} = n+1$ no caso sem sinais, ou $\pi_0 = +0$ e $\pi_{n+1} = +(n+1)$ no caso com sinais. A permutação estendida será denotada por π .

Um **ponto-de-quebra** de uma permutação π é um par $x = (\pi_i, \pi_{i+1})$ tal que nem x nem $\bar{x} = (\bar{\pi}_{i+1}, \bar{\pi}_i)$ são da forma $(j, j+1)$ para algum j tal que $0 \leq j \leq n$. Então, para encontrar ι a partir de π , devemos ter pelo menos uma operação “separando” π_i de π_{i+1} . Como anteriormente, pontos-de-quebra serão indicados por um ponto ($\bullet$) entre π_i e π_{i+1} (Figura 5.1). Denotamos por $b(\pi, \iota)$ o número de pontos-de-quebra de π com relação a ι .

Pontos-de-quebra dividem uma permutação em **faixas**. Quando a permutação alvo é a identidade, ι , faixas são sempre seqüências de inteiros consecutivos (Figura 5.1). No caso sem sinais, uma faixa pode ser crescente ou decrescente como uma seqüência de inteiros, e será chamada respectivamente de *faixa crescente* ou de *faixa decrescente*. No caso com sinais, todas as faixas são crescentes, mas serão separadas em faixas *positivas* ou *negativas*, de acordo com o sinal dos seus elementos (todos os elementos numa faixa têm o mesmo sinal).

No Capítulo 2 mostramos o *grafo de pontos-de-quebra* para permutações com sinais, denotado por $G(\pi, \iota)$. Este diagrama é composto por um certo número de ciclos onde, em cada ciclo, as arestas pretas e cinzas são alternadas. O *comprimento* de um ciclo é o número de arestas pretas que compõem este ciclo, que é o mesmo que o número de arestas cinzas. Chamamos de *k-ciclo* um ciclo de comprimento k . Observamos que, no caso de permutações com sinais, a decomposição de $G(\pi, \iota)$ em ciclos é única, e denotamos por $c(\pi, \iota)$ o número de ciclos em $G(\pi, \iota)$.

Como descrito no Capítulo 2, ciclos em $G(\pi, \iota)$ são denotados entre colchetes, pelos rótulos associados às arestas pretas, onde o primeiro rótulo é o de maior valor. Além disso, associaremos a cada rótulo i pertencente a um ciclo c uma orientação $\overleftarrow{i}$ ou $\overrightarrow{i}$, definida em relação à orientação do maior rótulo r de c , que será convencionada como sendo $\overleftarrow{r}$.

5.2 Algoritmos de aproximação

Apresentamos agora algoritmos de aproximação para computar a distância de reversão e transposição entre duas permutações. Mostramos um 3-algoritmo de aproximação para o caso sem sinais e um 2-algoritmo de aproximação para o caso com sinais.

Iniciaremos com o caso **sem sinais**. Notamos que a única permutação que não contém pontos-de-quebra com relação a ι é exatamente ι , e então a sequência de reversões e transposições transformando π em ι deve reduzir o número de pontos-de-quebra de $b(\pi, \iota)$ para 0. Observamos ainda que as reversões podem reduzir no máximo dois pontos-de-quebra, e que as transposições podem reduzir no máximo três pontos-de-quebra. Esta observação implica imediatamente num limite inferior, conforme mostrado no próximo teorema.

Teorema 5.2.1 *Dadas duas permutações sem sinais π e ι temos*

$$\frac{b(\pi, \iota)}{3} \leq d_{\tau\iota}(\pi, \iota).$$

Teorema 5.2.2 *Dadas duas permutações sem sinais distintas π e ι , existe uma operação reduzindo pelo menos um ponto-de-quebra.*

Prova: A idéia intuitiva é aumentar a primeira faixa em cada operação.

A primeira faixa à esquerda é sempre crescente. Tomando o elemento máximo nesta primeira faixa, basta encontrar o seu sucessor, que está necessariamente à sua direita.

Teorema 5.2.3 *Dada uma permutação π e uma operação e ,*

$$\Delta c(\pi, e \cdot \pi) \in \{-2, -1, 0, 1, 2\}$$

Prova: Notemos inicialmente que e pode ser uma reversão ou uma transposição.

Cada reversão age em duas arestas pretas pertencendo a no máximo dois ciclos, criando ou destruindo no máximo um ciclo. Hannenhalli e Pevzner [19] mostraram que, para uma reversão, $\Delta c(\pi, e \cdot \pi) \in \{-1, 0, 1\}$.

Cada transposição age em três arestas pretas pertencendo a no máximo três ciclos. Bafna e Pevzner [4] mostraram que para o caso sem sinais $\Delta c(\pi, e \cdot \pi) \in \{-2, 0, 2\}$. Quando consideramos os sinais, este torna-se um caso particular, e corresponde a um grafo de pontos-de-quebra gerado por uma permutação composta apenas por faixas positivas. Então, no caso com sinais, temos também $\Delta c(\pi, e \cdot \pi) = -1$ ou $+1$ (Figura 5.3). $\square$

O seguinte teorema segue imediatamente do Teorema 5.2.3.

Teorema 5.2.4 *Dadas duas permutações com sinais π e ι então temos*

$$\frac{(n+1) - c(\pi, \iota)}{2} \leq d_{rt}(\pi, \iota)$$

Para $x \in \{-2, -1, 0, 1, 2\}$, definimos um x -**movimento** de π com relação a $e \cdot \pi$ como uma operação e tal que $\Delta c(\pi, e \cdot \pi) = x$.

Como afirmamos acima, no caso de duas permutações com sinais π e ι , onde π contém apenas faixas positivas com relação a ι , o grafo de pontos-de-quebra $G(\pi, \iota)$ gerado é exatamente o grafo de ciclos de Bafna e Pevzner [4], descrito no Capítulo 2. Portanto, utilizaremos os resultados de Bafna e Pevzner [4] baseados no grafo de ciclos, e também descritos no Capítulo 2, para mostrar uma forma de obter um aumento no número de ciclos de pelo menos 2 em dois movimentos consecutivos, aplicando uma reversão ou uma transposição numa permutação com sinais.

Teorema 5.2.5 *Dadas duas permutações com sinais π e ι , existe um 1-movimento, ou um 2-movimento ou um 0-movimento seguido por um 2-movimento.*

Prova: Temos dois casos.

- existem faixas negativas em π com relação a ι : notamos que a faixa mais à esquerda é uma faixa positiva, e que as faixas negativas de π com relação a ι geram quatro formas gerais de ciclos em $G(\pi, \iota)$ (Figura 5.4).

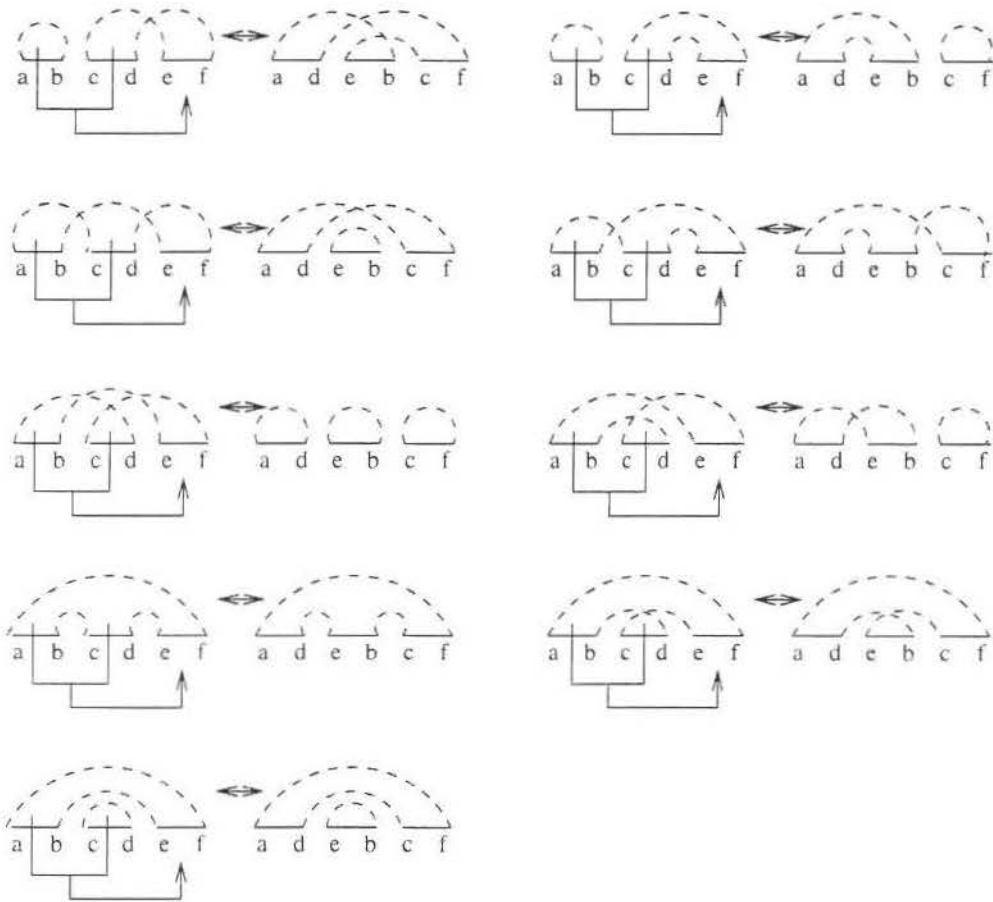


Figura 5.3: Todos os casos possíveis de uma transposição agindo numa permutação com sinais, onde apenas os ciclos afetados são mostrados. Em cada caso, uma transposição transforma as arestas pretas (b, a) , (d, c) e (f, e) nas arestas (d, a) , (b, e) e (f, c) . As linhas pontilhadas denotam um caminho que pode ser formado por uma ou mais arestas cinzas/pretas. Como a inversa de uma transposição é uma transposição, todas as transformações são inversíveis. Note que existe apenas um padrão correspondendo a um 2-movimento (-2 -movimento), e apenas três padrões correspondendo a um 1-movimento (-1 -movimento).

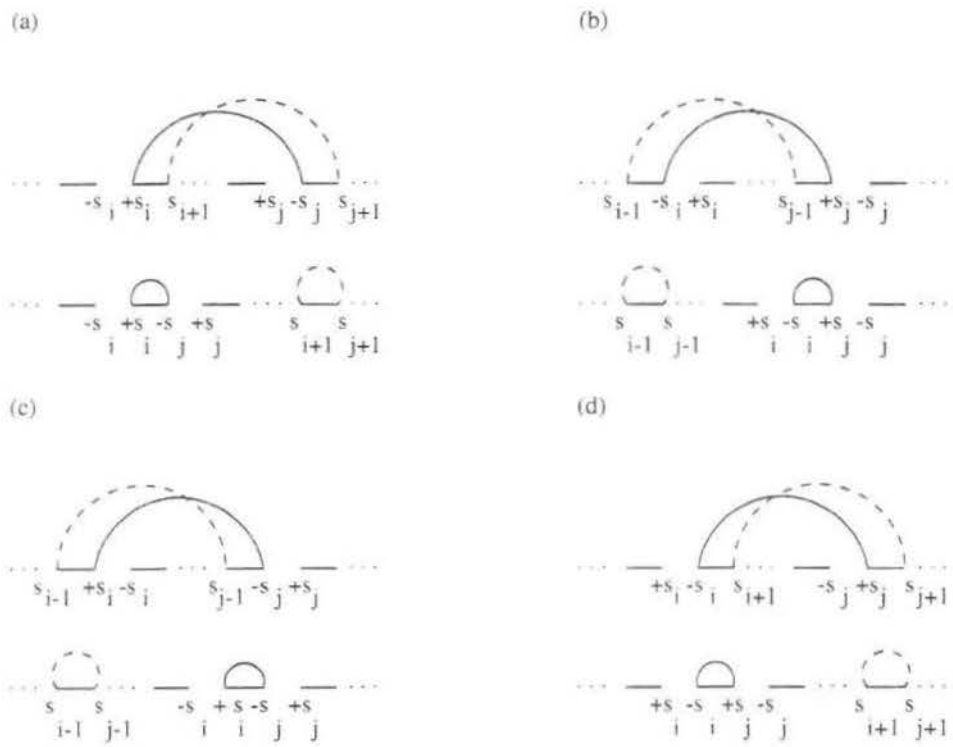


Figura 5.4: As quatro possíveis formas gerais de ciclos nos quais se pode aplicar reversões que aumentam o número de 1-ciclos.

No caso (a) desta figura aplicamos uma reversão da aresta preta $(s_{i+1}, +s_i)$ até a aresta preta $(s_{j+1}, -s_j)$, ambas pertencendo ao mesmo ciclo c , para obter um outro grafo de pontos-de-quebra com dois ciclos diferentes, c' , um 1-ciclo composto pela aresta preta $(-s_j, +s_i)$, e c'' , um k -ciclo $k \geq 1$ contendo a aresta preta (s_{j+1}, s_{i+1}) . Esta reversão não altera os comprimentos e rótulos dos outros ciclos e portanto temos um 1-movimento. Os outros casos são inteiramente análogos.

Portanto, aplicamos uma reversão que aumentou de 1 o número de 1-ciclos, isto é, obtivemos um 1-movimento.

- existem apenas faixas positivas em π com relação a ι : temos um grafo de pontos-de-quebra no qual não podemos aplicar reversões aumentando o número de 1-ciclos. De acordo com resultados de Hannenhalli e Pevzner [19] este é um grafo formado unicamente por obstáculos. Mas, como dissemos anteriormente, neste caso podemos visualizar esta permutação como sendo uma permutação sem sinais, e o grafo de pontos-de-quebra gerado é exatamente o grafo de ciclos descrito no trabalho de Bafna e Pevzner [4]. Aplicamos então o Teorema 2.2.7 de Bafna e Pevzner [4], que garante a existência de um 2-movimento, ou de um 0-movimento seguido por um 2-movimento.

□

Do Teorema 5.2.5 podemos deduzir um limite superior para distância de reversão e transposição.

Teorema 5.2.6 *Dadas duas permutações com sinais π e ι temos*

$$d_{rt}(\pi, \iota) \leq (n + 1) - c(\pi, \iota)$$

Dada uma permutação π a ser transformada em ι , a idéia intuitiva do algoritmo é, enquanto existirem faixas negativas com relação a ι aplicamos reversões, conforme descrito no Teorema 5.2.5. Se não pudermos aplicar reversões deste tipo, e esta seqüência de reversões não tiver transformado π em ι , então o grafo de pontos-de-quebra foi gerado por uma permutação contendo apenas faixas positivas com relação a ι . Neste caso, usamos o Teorema 2.2.7 de Bafna e Pevzner [4] para descobrir a seqüência de transposições a ser aplicada. Observamos que estas transposições sempre geram grafos de pontos-de-quebra construídos a partir de permutações contendo somente faixas positivas com relação a ι . A Figura 5.5 mostra um exemplo deste processo.

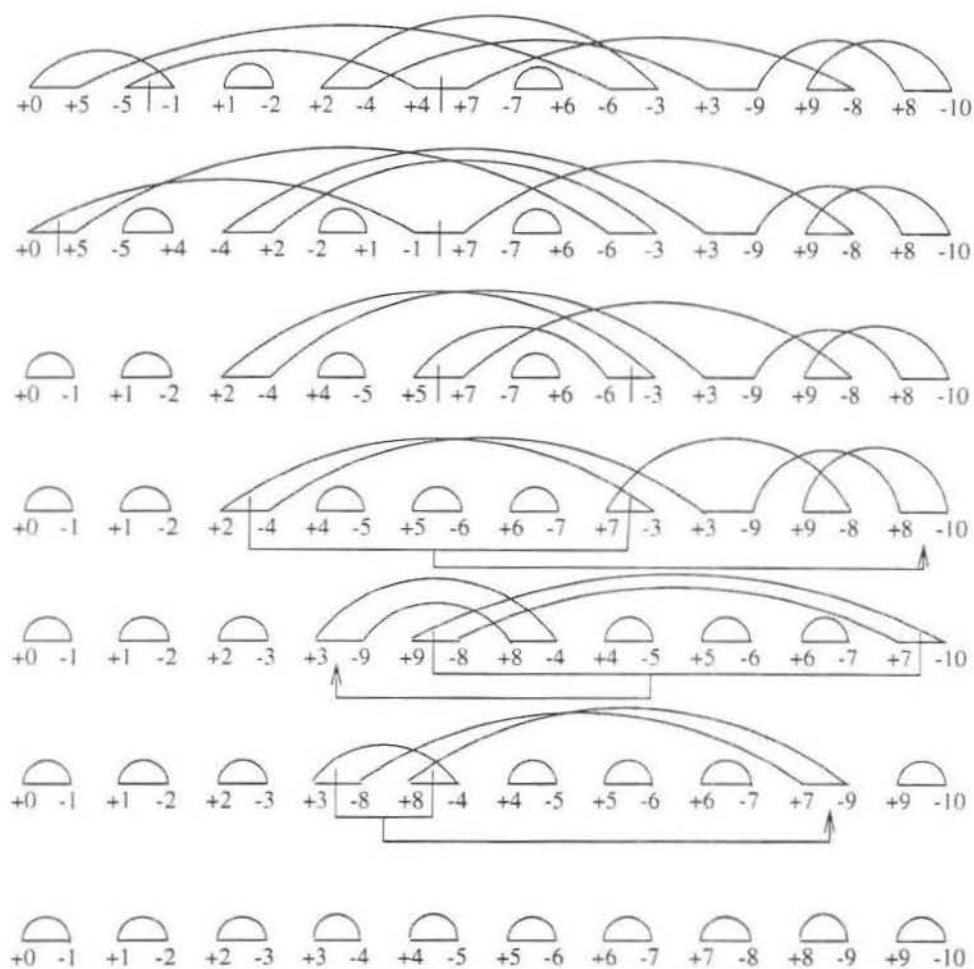


Figura 5.5: O processo de transformar π em ι usando reversões e transposições.

Na Figura 5.6 mostramos um 2-algoritmo de aproximação para este problema, denominado *DistRT2*. A correção deste algoritmo é garantida pelos Teoremas 2.2.7 e 5.2.5.

Descrevemos agora como encontrar uma reversão criando um 1-ciclo. Primeiro associamos a cada inteiro i da aresta preta (π_i, π_{i-1}) , com $1 \leq i \leq n+1$, um rótulo I para π_{i-1} e F para π_i . Depois, temos apenas que encontrar uma aresta cinza incidindo em duas arestas pretas ambas rotuladas por I ou ambas rotuladas por F . Isto leva no máximo $O(n)$. A construção de $G(\pi, \iota)$ e a computação de $c(\pi, \iota)$ cada uma leva $O(n)$ (conforme Berman e Hannenhalli [5]). Portanto a complexidade do algoritmo é $O(n^2)$, onde n é o tamanho das permutações.

Isto estabelece um 2-algoritmo de aproximação para computar a distância de reversão e transposição de permutações com sinais, com complexidade de tempo $O(n^2)$, onde n é o tamanho das permutações.

5.3 Um limite inferior para o diâmetro de reversão e transposição

De forma genérica, uma solução para o problema da distância é encontrar caminhos mínimos num grafo orientado G , no qual um vértice corresponde a uma permutação π , e existe uma aresta (π, σ) quando existe um evento (reversão ou transposição) e , tal que $\sigma = e \cdot \pi$.

O objetivo é encontrar caminhos orientados mínimos de π até ι , onde o comprimento de um caminho é exatamente o número de arestas neste caminho. Mais ainda, podemos associar pesos às arestas de um caminho de forma a nos auxiliar a investigar o problema. Assim, a seguir definiremos peso de uma aresta e peso de um caminho, neste grafo.

Definição 5.3.1 *Dadas as permutações π e σ , de comprimento n , e ι_n , e um evento e (reversão ou transposição) tais que (π, σ) é uma aresta de G , então em G definimos peso de uma aresta como sendo*

$$p(\pi, \sigma) = 2 + c(\pi, \iota_n) - c(\sigma, \iota_n).$$

Observamos que $p(\pi, \sigma) \geq 0$ (Teorema 5.2.3). O peso de uma aresta $p(\pi, \sigma)$ pode também ser escrito como $2 - \Delta c(\pi, \sigma)$, tomando e como o evento que transforma π em σ . Como 2 é o maior valor que $\Delta c(\pi, \sigma)$ pode assumir, e sabendo que os maiores

Algoritmo *DistRT2*

entrada: permutações distintas π e ι
saída: $u, e_1, e_2, \dots, e_u$, onde $e_i = r$ ou t
// cria o grafo de pontos-de-quebra de π e ι
 $u \leftarrow 0$
 $\pi_0 \leftarrow \pi$
 cria $G(\pi_0, \iota)$
 computa $c(\pi_0, \iota)$
// cria 1-ciclos enquanto houver possibilidade
enquanto $c(\pi_u, \iota) < n + 1$ e existe uma reversão r
aumentando o numero de 1-ciclos
(conforme Teorema 5.2.5) **faça**
 $u \leftarrow u + 1$
 $e_u \leftarrow r$
 $\pi_u \leftarrow e_u \cdot \pi_{u-1}$
cria $G(\pi_u, \iota)$
computa $c(\pi_u, \iota)$
// se π não foi transformado em ι , o grafo de pontos-de-quebra gerado
// tem a mesma forma geral descrita na teoria de Bafna e Pevzner
enquanto $c(\pi_u, \iota) < n + 1$ **faça**
chame uma rotina que implementa a teoria de Bafna e Pevzner
para encontrar a transposição t a ser aplicada
(conforme Teorema 2.2.7)
 $u \leftarrow u + 1$
 $e_u \leftarrow t$
 $\pi_u \leftarrow e_u \cdot \pi_{u-1}$
cria $G(\pi_u, \iota)$
computa $c(\pi_u, \iota)$
retorne $u, e_1, e_2, \dots, e_u$

Figura 5.6: Um 2-algoritmo de aproximação para computar $d_{rt}(\pi, \iota)$, para permutações com sinais.

valores de $\Delta c(\pi, \sigma)$ nos aproximam de i_n , então podemos pensar no peso como uma medida de “gasto” em cada evento aplicado.

Definição 5.3.2 *Dadas as permutações de comprimento n , $\pi_0, \pi_1, \dots, \pi_k$, tais que (π_{i-1}, π_i) , com $1 \leq i \leq k$, é uma aresta do grafo orientado G , então em G , tomando o caminho c de π a ι_n , $c = \pi_0 \pi_1 \pi_2 \dots \pi_{k-1} \pi_k$, definimos **peso de um caminho** como sendo*

$$p(c) = \sum_{i=1}^k p(\pi_{i-1}, \pi_i).$$

Observamos que $p(c) \geq 0$ para qualquer caminho. Podemos agora relacionar o comprimento de um caminho com o peso do mesmo caminho, com importantes consequências para a distância. Denotamos $|c|$ como o comprimento de um caminho c .

Lema 5.3.3 *Seja um caminho $c = \pi_0 \pi_1 \pi_2 \dots \pi_{k-1} \pi_k$. Então,*

$$p(c) = 2|c| + c(\pi_0, \iota_n) - c(\pi_k, \iota_n)$$

Prova: Seja $p(c) = \sum_{i=1}^k p(\pi_{i-1}, \pi_i)$. A prova será por indução em k .

Quando $k = 1$, pela Definição 5.3.1 e por $|c| = 1$, temos,

$$p(c) = p(\pi_0, \pi_1) = 2 + c(\pi_0, \iota_n) - c(\pi_1, \iota_n) = 2|c| + c(\pi_0, \iota_n) - c(\pi_1, \iota_n)$$

Assumindo que o resultado seja verdadeiro para todos os valores $< k$,

$$\begin{aligned} p(c) &= \sum_{i=1}^k p(\pi_{i-1}, \pi_i) = p(\pi_{k-1}, \pi_k) + \sum_{i=1}^{k-1} p(\pi_{i-1}, \pi_i) = \\ &= p(\pi_{k-1}, \pi_k) + \{2(k-1) + c(\pi_0, \iota_n) - c(\pi_{k-1}, \iota_n)\} = \\ &= 2 + c(\pi_{k-1}, \iota_n) - c(\pi_k, \iota_n) + 2(k-1) + c(\pi_0, \iota_n) - c(\pi_{k-1}, \iota_n) = \\ &= 2k - c(\pi_k, \iota_n) + c(\pi_0, \iota_n) = 2|c| - c(\pi_k, \iota_n) + c(\pi_0, \iota_n) \end{aligned}$$

□

Um importante corolário segue.

Corolário 5.3.4 *Para qualquer permutação π , de comprimento n , e ι_n , e qualquer caminho mínimo c de π a ι_n temos*

$$d_{rt}(\pi, \iota_n) = \frac{p(c) - c(\pi, \iota_n) + (n+1)}{2}$$

A prova é imediata do teorema anterior, tomando $\pi_0 = \pi$ e $\pi_k = \iota_n$.

Tomando S_n como o conjunto de todas as permutações com sinais de tamanho n , definimos

$$D_{rt}(n) = \max_{\pi \in S_n} d_{rt}(\pi, \iota_n)$$

como sendo o diâmetro de reversão e transposição deste conjunto.

Nesta seção mostramos um limite inferior para computar $D_{rt}(n)$, baseado nas distâncias de reversões e transposições de permutações específicas com relação a identidade, para cada inteiro n .

Estas permutações são

$$\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$$

e

$$\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$$

Calcularemos a distância de reversão e transposição entre elas, para cada n , o que fornecerá um limite inferior para o diâmetro $D_{rt}(n)$.

Começaremos mostrando um limite superior para $d_{rt}(\pi_n, \iota_n)$, para todo $n \geq 3$.

Teorema 5.3.5 *Tomando $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, para todo $n \geq 3$, então temos*

$$d_{rt}(\pi_n, \iota_n) \leq \left\lfloor \frac{n}{2} \right\rfloor + 2$$

Prova:

Inicialmente, aplicamos uma reversão $r(1, n)$ em π_n , obtendo $\pi = r(1, n) \cdot \pi_n = (+n \ +(n-1) \ \dots \ +2 \ +1)$, uma permutação com sinais positivos apenas.

Em seguida, usamos o Teorema 4.2.11, provado no Capítulo 4, obtido de forma independente por Christie [10], que determina a distância de transposição $d_t(\pi, \iota_n) = \left\lfloor \frac{n}{2} \right\rfloor + 1$, para $n > 2$.

O número total de operações é então $\left\lfloor \frac{n}{2} \right\rfloor + 2$, que é um limite superior para $d_{rt}(\pi_n, \iota_n)$, com $n \geq 3$.

□

Nossa estratégia para mostrar que este limite superior é também um limite inferior é provar que todo caminho c de π_n até ι_n satisfaz $p(c) \geq 3$. Então, pelo Corolário 5.3.4

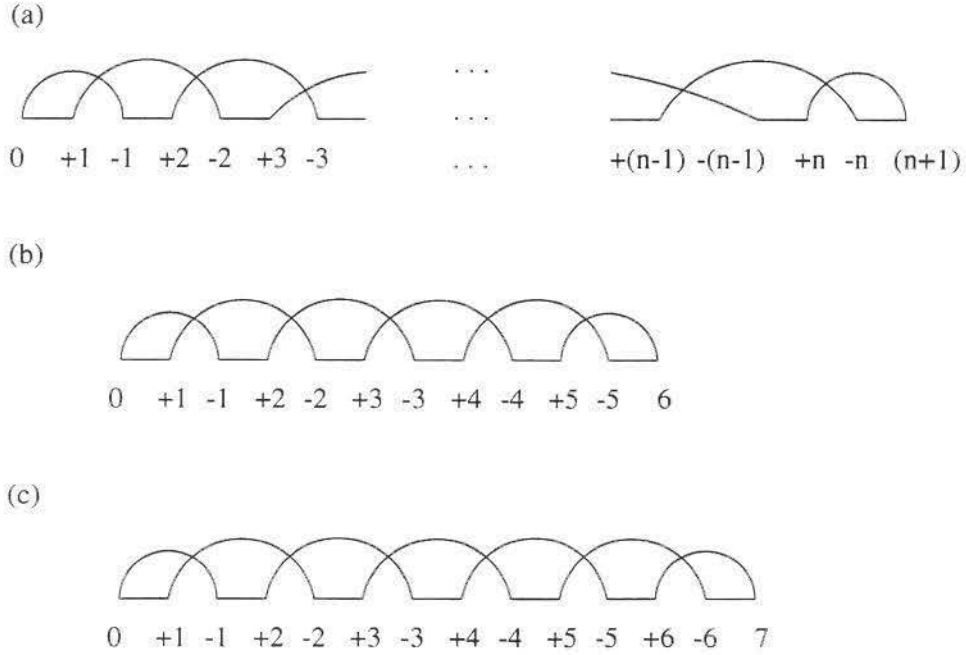


Figura 5.7: Os grafos de pontos-de-quebra gerados por $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ com relação a $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$ para todo n . (a) O grafo de pontos-de-quebra geral. (b) Um caso particular para $n = 5$. (c) Um caso particular para $n = 6$.

e pelo Teorema 5.3.5, teremos o resultado desejado, conforme enunciado no Teorema 5.3.25.

A forma genérica do grafo de pontos-de-quebra gerado por estas duas permutações é dada na Figura 5.7. O lema abaixo mostra que este grafo contém um único ciclo.

Lema 5.3.6

Sejam $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$. Então, em $G(\pi_n, \iota_n)$, $c(\pi_n, \iota_n) = 1$ para todo n .

Precisamos de resultados auxiliares para provar o teorema que estabelece o limite inferior. Um deles é uma condição suficiente para que um grafo de pontos-de-quebra não admita 2-movimentos. Podemos observar que o formato dos ciclos em $G(\pi_n, \iota_n)$ é (Lema 5.3.6):

$$c = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{2}, \overrightarrow{1}, \overrightarrow{3}, \dots, \overrightarrow{n}]$$

quando n é ímpar, e

$$c = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{3}, \overleftarrow{1}, \overrightarrow{2}, \dots, \overrightarrow{n}]$$

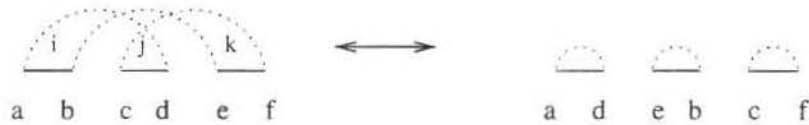


Figura 5.8: A forma genérica de um ciclo no qual podemos aplicar uma transposição criando dois ciclos.

quando n é par. Observe que a não ser pela paridade de n estas duas seqüências são formadas por duas subseqüências, $s_1 = \overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{2}$ e $s_2 = \overrightarrow{1}, \overrightarrow{3}, \dots, \overrightarrow{n}$, tais que s_1 é decrescente e s_2 é crescente. Chamaremos de **bimonótonos** os ciclos formados por duas subseqüências, a primeira decrescente e formada por rótulos com orientação $\leftarrow$ e a segunda crescente e formada por rótulos com orientação $\rightarrow$. Estes ciclos não admitem 2-movimentos como mostram os resultados seguintes.

No lema a seguir, mostramos a forma genérica de um grafo de pontos-de-quebra que admite uma transposição 2-movimento (Figura 5.8). Este lema decorre imediatamente do Teorema 5.2.3.

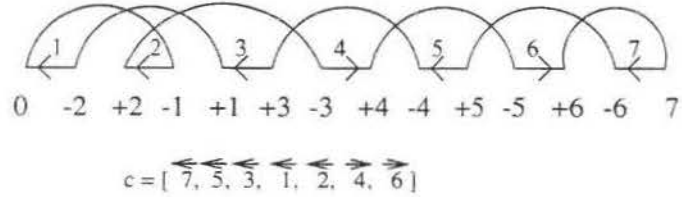
Lema 5.3.7 *Dadas duas permutações com sinais π e σ , e $G(\pi, \sigma)$, pode-se obter um 2-movimento se e somente se existirem arestas pretas $i = (b, a)$, $j = (d, c)$ e $k = (f, e)$, nesta ordem, pertencentes a um único ciclo tais que a aplicação da transposição $t(i, j, k)$ resulta em três novos ciclos, um contendo a aresta preta (a, d) , outro contendo (b, e) , e o terceiro contendo (c, f) .*

Podemos caracterizar mais precisamente os 2-movimentos gerados por uma transposição, por meio da ordem dos rótulos em um ciclo, juntamente com a orientação de cada rótulo.

Lema 5.3.8 *Dadas duas permutações com sinais, π de comprimento n , e ι_n , então $G(\pi, \iota_n)$ admite 2-movimento se e somente se existem três rótulos de arestas pretas i, j e k tais que $i < j < k$, os três pertencem a um mesmo ciclo e a ordem em que eles aparecem no ciclo é k, i, j (ou i, j, k ou j, k, i), com a orientação $\leftarrow$, ou k, j, i (ou j, i, k ou i, k, j), com a orientação $\rightarrow$.*

Prova: O Lema 5.3.7 mostra que existe uma única forma geral de ciclo no qual se pode obter um 2-movimento. Nesta forma (Figura 5.8), podemos verificar que, tomando os três rótulos de arestas pretas (pertencentes ao mesmo ciclo) i, j e k tais que $i < j < k$, e associando ao rótulo k a orientação $\overleftarrow{k}$, forçamos as orientações de i e j como sendo respectivamente $\overleftarrow{i}$ e $\overleftarrow{j}$, implicando em que os três rótulos apareçam

$$(a) \pi = (+2 +1 -3 -4 -5 -6)$$



$$(b) \pi = (-3 -4 -5 -1 -2 -6)$$

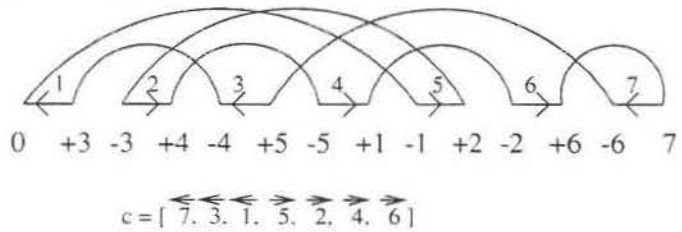


Figura 5.9: Exemplos de transposições aplicadas numa permutação π com relação a ι_n . Observe que para uma transposição ser 2-movimento, as orientações dos rótulos devem ser exatamente aquelas descritas no Lema 5.3.8. (a) As transposições 2-movimentos são $t(1, 2, 3) \cdot \pi$, $t(1, 2, 5) \cdot \pi$ e $t(1, 2, 7) \cdot \pi$. (b) A transposição $t(2, 4, 5) \cdot \pi$ não é 2-movimento.

no ciclo na ordem k, i, j (ou i, j, k ou j, k, i), e todos com a mesma orientação de k . Analogamente, se associarmos a k a orientação $\overrightarrow{k}$, as orientações de i e j ficam sendo $\overrightarrow{i}$ e $\overrightarrow{j}$, implicando na ordem k, j, i (ou j, i, k ou i, k, j), e i, j e k com a mesma orientação de k .

A prova no outro sentido é imediata. Aplicamos $t(i, j, k)$ em π , com i, j e k seguindo as condições do lema, e obtemos o resultado desejado.

A Figura 5.9 mostra exemplos.

□

Teorema 5.3.9 *Seja π uma permutação para a qual todos os ciclos no grafo de ciclos $G(\pi, \iota)$ são bimonótonos. Então, $p(\pi, e \cdot \pi) \geq 1$ para qualquer evento e .*

Prova: Claramente, $p(\pi, e \cdot \pi) = 0$ é equivalente a dizer que e é um 2-movimento. Um 2-movimento deve ser uma transposição, e agindo em três arestas pretas do mesmo ciclo.

Entretanto, pela bimonotonicidade dos ciclos de π , não podemos escolher três rótulos seguindo as condições do Lema 5.3.8, considerando apenas uma das subsequências.

Uma outra forma de tomar os rótulos seria escolhê-los de ambas as subsequências. Mas então os rótulos não teriam a mesma orientação, e portanto também neste caso não teríamos as condições do Lema 5.3.8. $\square$

Neste ponto, provamos certos lemas que serão utilizados na prova do Teorema 5.3.24.

Os Lemas 5.3.14 e 5.3.15 mostram que quando a primeira operação aplicada em π_n é uma reversão r 1-movimento então não teremos 2-movimentos no grafo de pontos-de-quebra gerado por $r \cdot \pi_n$ e ι_n .

Vamos mostrar agora quais são as reversões 1-movimentos e 0-movimentos que podem ser aplicadas em $G(\pi_n, \iota_n)$.

Lema 5.3.10 *Dadas as permutações $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, com n ímpar, $n \geq 3$, e $G(\pi_n, \iota_n)$, a reversão $r(i, j)$, com $1 \leq i \leq j \leq n$ e $(j-i)$ par, constitui-se num 1-movimento.*

Prova: Inicialmente, $G(\pi_n, \iota_n)$, com n ímpar, é composto por um único ciclo de tamanho $n+1$ (Figura 5.7).

Queremos provar que a aplicação de $r(i, j)$, com $1 \leq i \leq j \leq n$ e $(j-i)$ par constitui-se num 1-movimento. Assim,

$$\begin{aligned} \pi_n &= (-1 \ -2 \ \dots \ -(i-1) \ -i \ -(i+1) \ \dots \ -(j-1) \ -j \ -(j+1) \ \dots \ -(n-1) \ -n) \\ r(i, j) \cdot \pi_n &= \\ &(-1 \ -2 \ \dots \ -(i-1) \ +j \ +(j-1) \ \dots \ -(i+1) \ +i \ -(j+1) \ \dots \ -(n-1) \ -n) \\ &\text{com } 1 \leq i \leq j \leq n \text{ e } (j-i) \text{ par.} \end{aligned}$$

Nestas condições, devemos ter i e j com a mesma paridade.

No caso de i e j ímpares, temos o grafo de pontos-de-quebra mostrado na Figura 5.10, formado por dois ciclos, c_1 e c_2 . Considerando os rótulos associados às arestas pretas do grafo de pontos-de-quebra, esta operação corresponde a reverter as arestas pretas entre i e $j+1$, incluindo ambas. O c_1 -ciclo tem $i + \frac{j-i}{2}$ arestas, e o c_2 -ciclo tem $n-j+1 + \frac{j-i}{2}$ arestas, sendo c_1 e c_2 ambos pares ou ambos ímpares. Nestas condições, o ciclo c_1 é

$$c_1 = [\overleftarrow{j}, \overleftarrow{j-2}, \dots, \overleftarrow{i+2}, \overleftarrow{i}, \overleftarrow{i-2}, \dots, \overleftarrow{3}, \overleftarrow{1}, \overrightarrow{2}, \overrightarrow{4}, \dots, \overrightarrow{i-1}]$$

e o ciclo c_2 é

$$c_2 = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{j+1}, \overleftarrow{j-1}, \overleftarrow{j-3}, \dots, \overleftarrow{i+1}, \overrightarrow{j+2}, \overrightarrow{j+4}, \dots, \overrightarrow{n}]$$

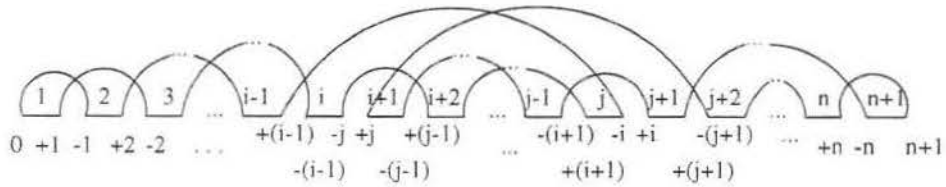


Figura 5.10: O grafo de pontos-de-quebra gerado pela aplicação de uma reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $j - i$ par, em π_n , $n \geq 3$, ímpar. Com estas condições, i e j devem ter a mesma paridade. O grafo genérico para o caso de i e j ímpares.

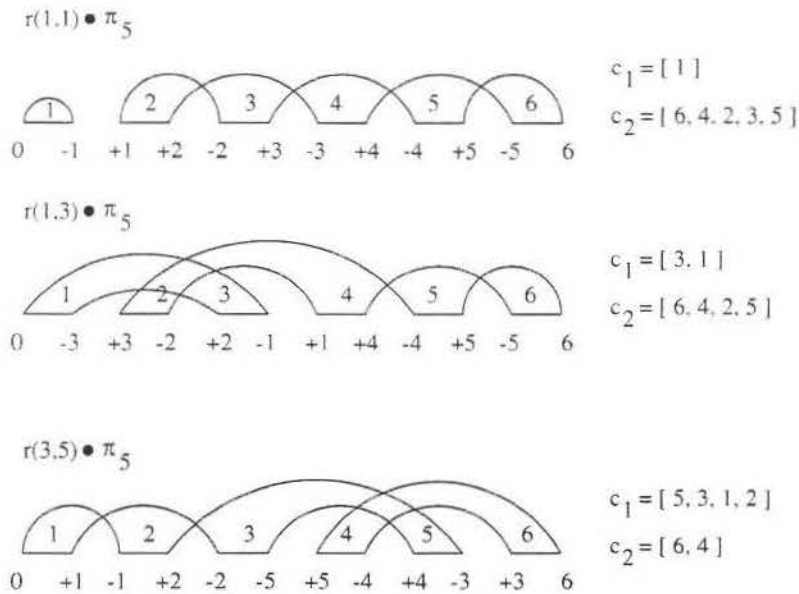


Figura 5.11: Exemplos de grafos de pontos-de-quebra gerados pela aplicação de uma reversão $r(i, j)$, com $1 \leq i \leq j \leq n$, $j - i$ par e i, j ímpares, em π_n , $n \geq 3$, ímpar. As reversões estão descritas na figura.

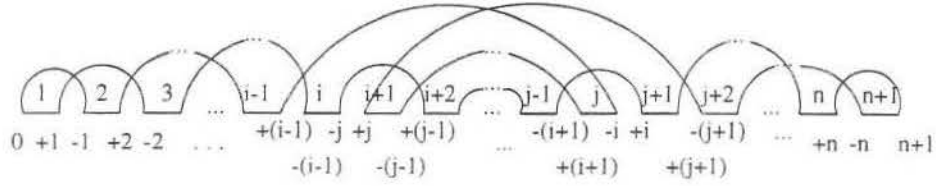


Figura 5.12: O grafo de pontos-de-quebra gerado pela aplicação de uma reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $j - i$ ímpar, em π_n , $n \geq 3$, ímpar. Com estas condições, i e j devem ter paridades opostas. O grafo genérico para o caso de i ímpar e j par.

Portanto, esta reversão é um 1-movimento. A Figura 5.11 mostra alguns exemplos. O caso de i e j pares é inteiramente análogo.

□

Lema 5.3.11 Dadas as permutações $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, com n ímpar, $n \geq 3$, e $G(\pi_n, \iota_n)$, a reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ ímpar constitui-se num 0-movimento.

Prova: Novamente, $G(\pi_n, \iota_n)$, com n ímpar, é composto por um único ciclo de tamanho $n + 1$ (Figura 5.7).

Queremos provar que a aplicação de $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ ímpar constitui-se num 0-movimento. De novo,

$$r(i, j) \cdot \pi_n =$$

$$(-1 \ -2 \ \dots \ -(i-1) \ +j \ +(j-1) \ \dots \ -(i+1) \ +i \ -(j+1) \ \dots \ -(n-1) \ -n)$$

com $1 \leq i \leq j \leq n$ e $(j - i)$ ímpar

Neste caso, temos i, j com paridades opostas.

No caso de i ímpar e j par, temos o grafo de pontos-de-quebra mostrado na Figura 5.12, formado por um único ciclo de tamanho $n + 1$. Temos o $(n + 1)$ -ciclo:

$$c = [\overrightarrow{n+1}, \overrightarrow{n-1}, \dots, \overrightarrow{j+2}, \overrightarrow{i+1}, \overrightarrow{i+3}, \dots, \overrightarrow{j}, \overrightarrow{i-1}, \overrightarrow{i-3}, \dots, \overrightarrow{2},$$

$$\overrightarrow{1}, \overrightarrow{3}, \dots, \overrightarrow{i}, \overrightarrow{i+2}, \dots, \overrightarrow{j-1}, \overrightarrow{j+1}, \overrightarrow{j+3}, \dots, \overrightarrow{n}].$$

Portanto, esta reversão é um 0-movimento. A Figura 5.13 mostra alguns exemplos.

O caso de i par e j ímpar é inteiramente análogo.

□

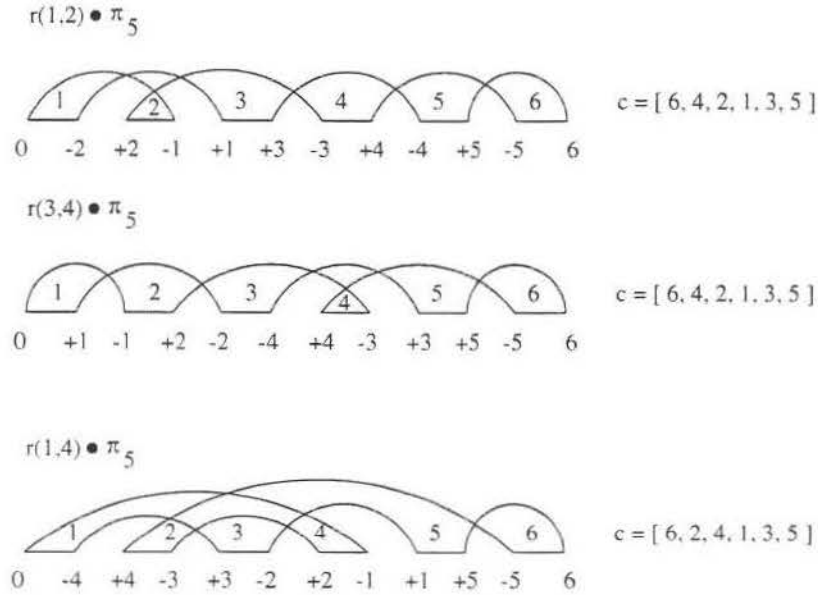


Figura 5.13: Exemplos de grafos de pontos-de-quebra gerados pela aplicação de uma reversão $r(i, j)$ com $1 \leq i \leq j \leq n$, $j - i$ ímpar, i ímpar e j par, em π_n , $n \geq 3$, ímpar. As reversões estão descritas na figura.

Lema 5.3.12 Dadas as permutações $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, com n par, $n \geq 4$, e $G(\pi_n, \iota_n)$, a reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ par constitui-se num 1-movimento.

Prova: O grafo de pontos-de-quebra $G(\pi_n, \iota_n)$, com n par, é composto por um único ciclo de tamanho $n + 1$ (Figura 5.7).

Queremos provar que a aplicação de $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ par constitui-se num 1-movimento. Assim,

$$r(i, j) \cdot \pi_n =$$

$$(-1 \ -2 \ \dots \ -(i-1) \ +j \ +(j-1) \ \dots \ -(i+1) \ +i \ -(j+1) \ \dots \ -(n-1) \ -n)$$

com $1 \leq i \leq j \leq n$ e $(j - i) = 0$ ou $(j - i)$ par.

Nestas condições, temos i e j com a mesma paridade.

No caso de i, j ímpares, temos o grafo de pontos-de-quebra mostrado na Figura 5.14, formado por dois ciclos, c_1 e c_2 . Considerando os rótulos associados às arestas pretas do grafo de pontos-de-quebra, esta operação corresponde a reverter as arestas pretas entre i e $j + 1$, incluindo ambas. O c_1 -ciclo tem $i + \frac{j-i}{2}$ arestas, e o c_2 -ciclo tem $n - j + 1 + \frac{j-i}{2}$ arestas, sendo c_1 e c_2 de paridades opostas. Neste caso, o ciclo c_1 é

$$c_1 = [\overleftarrow{j}, \overleftarrow{j-2}, \dots, \overleftarrow{i+2}, \overleftarrow{i}, \overleftarrow{i-2}, \dots, \overleftarrow{3}, \overleftarrow{1}, \overrightarrow{2}, \overrightarrow{4}, \dots, \overrightarrow{i-1}]$$

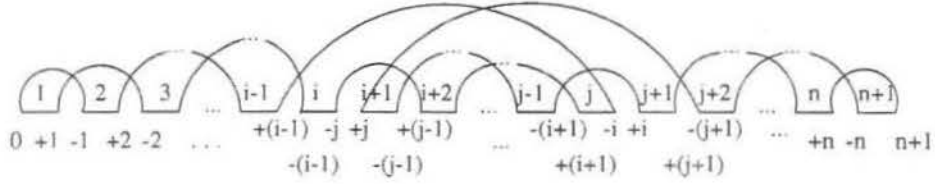


Figura 5.14: O grafo de pontos-de-quebra gerado pela aplicação de uma reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ par, em π_n , $n \geq 4$, par. Com estas condições, i e j devem ter a mesma paridade. O grafo genérico para o caso de i e j ímpares.

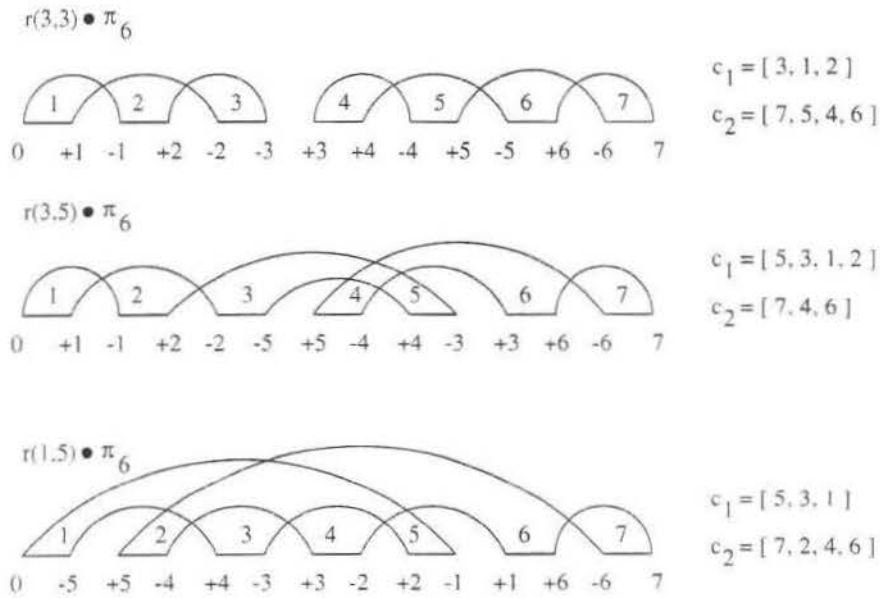


Figura 5.15: Exemplos de grafos de pontos-de-quebra gerados pela aplicação de uma reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ par, i e j ímpares, em π_n , $n \geq 4$, par. As reversões estão descritas na figura.

e o ciclo c_2 é

$$c_2 = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{j+2}, \overrightarrow{i+1}, \overrightarrow{i+3}, \dots, \overrightarrow{j-1}, \overrightarrow{j+1}, \overrightarrow{j+3}, \dots, \overrightarrow{n}]$$

Portanto, esta reversão é um 1-movimento. A Figura 5.15 mostra alguns exemplos. O caso de i, j pares é inteiramente análogo.

□

Lema 5.3.13 Dadas as permutações $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, com n par, $n \geq 4$, e $G(\pi_n, \iota_n)$, a reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ ímpar constitui-se num 0-movimento.

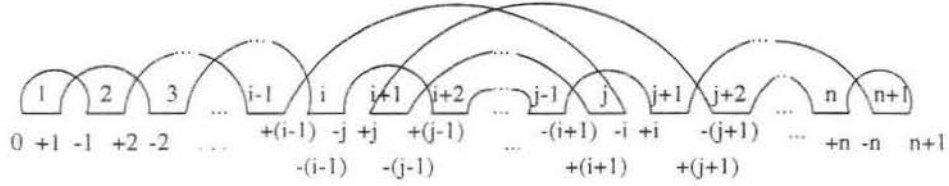


Figura 5.16: O grafo de pontos-de-quebra gerado pela aplicação de uma reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $j - i$ ímpar, em π_n , $n \geq 4$, par. Com estas condições, i e j devem ter paridades opostas. O grafo genérico para o caso de i ímpar e j par.

Prova: O grafo de pontos-de-quebra $G(\pi_n, \iota_n)$, com n par, é composto por um único ciclo de tamanho $n + 1$ (Figura 5.7).

Queremos provar que a aplicação de $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ ímpar constitui-se num 0-movimento. De novo,

$$r(i, j) \cdot \pi_n =$$

$$(-1 \ -2 \ \dots \ -(i-1) \ +j \ +(j-1) \ \dots \ -(i+1) \ +i \ -(j+1) \ \dots \ -(n-1) \ -n)$$

com $1 \leq i \leq j \leq n$ e $(j - i)$ ímpar

Neste caso, temos i, j com paridades opostas. No caso de i ímpar e j par, temos o grafo de pontos-de-quebra mostrado na Figura 5.16, formado por um único ciclo de tamanho $n + 1$. Neste caso, temos o $(n + 1)$ -ciclo:

$$c = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{j+1}, \overleftarrow{j-1}, \dots, \overleftarrow{i}, \overleftarrow{i-2}, \dots, \overleftarrow{3}, \overleftarrow{1}, \overrightarrow{2}, \overrightarrow{4}, \dots, \overrightarrow{i-1},$$

$$\overrightarrow{j}, \overrightarrow{j-2}, \dots, \overrightarrow{i+1}, \overrightarrow{j+2}, \overrightarrow{j+4}, \dots, \overrightarrow{n}]$$

Portanto, esta reversão é um 0-movimento. A Figura 5.17 mostra alguns exemplos. O caso de i par e j ímpar é inteiramente análogo.

□

Os lemas seguintes mostram que em $G(\pi_n, \iota_n)$ a aplicação de qualquer reversão 1-movimento gera grafos que não admitem 2-movimentos.

Lema 5.3.14 Dadas as permutações π_n e ι_n , com n ímpar, $n \geq 3$, e a reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ par, $r(i, j) \cdot \pi_n$ produz um grafo de pontos-de-quebra que não contém 2-movimentos.

Prova:

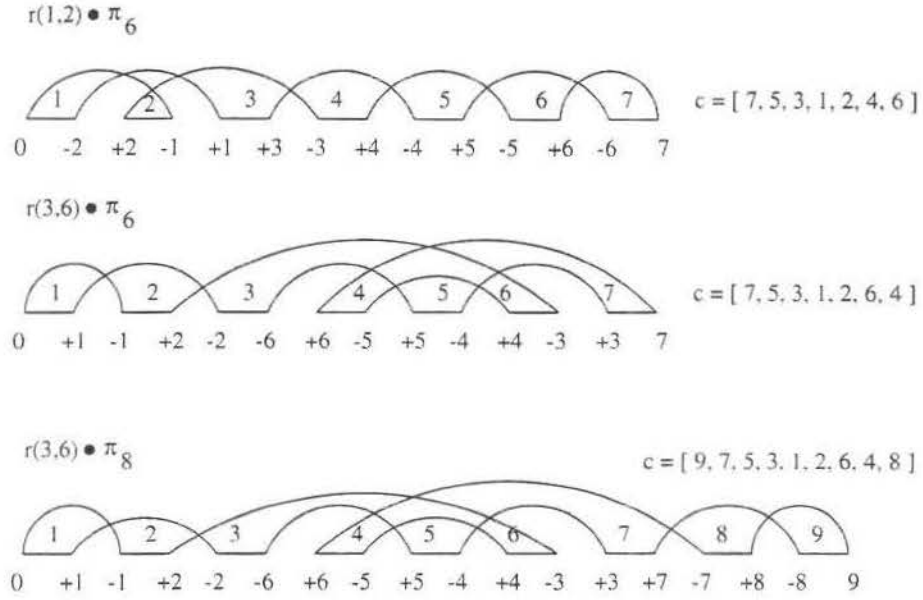


Figura 5.17: Exemplos de grafos de pontos-de-quebra gerados pela aplicação de uma reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $j - i$ ímpar, i ímpar e j par, em π_n , $n \geq 4$, par. As reversões estão descritas na figura.

Do Lema 5.3.10, quando aplicamos $r(i, j)$ nas condições da hipótese em π_n , n ímpar, $n \geq 3$, e para i e j ímpares, temos os seguintes ciclos em $G(r(i, j) \cdot \pi_n, \iota_n)$:

$$c_1 = [\overleftarrow{j}, \overleftarrow{j-2}, \dots, \overleftarrow{i+2}, \overleftarrow{i}, \overleftarrow{i-2}, \dots, \overleftarrow{3}, \overleftarrow{1}, \overrightarrow{2}, \overrightarrow{4}, \dots, \overrightarrow{i-1}]$$

e

$$c_2 = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{j+1}, \overleftarrow{j-1}, \overleftarrow{j-3}, \dots, \overleftarrow{i+1}, \overrightarrow{j+2}, \overrightarrow{j+4}, \dots, \overrightarrow{n}]$$

Podemos observar que c_1 e c_2 são bimonótonos e portanto, pelo Teorema 5.3.9, não existem 2-movimentos possíveis.

O caso de i e j pares é análogo.

Assim, não temos 2-movimentos possíveis no grafo de pontos-de-quebra gerado por uma reversão, nas condições deste lema, aplicada em π_n , n ímpar, $n \geq 3$. $\square$

A prova para o caso em que n é par é inteiramente análoga, bastando utilizar o Lema 5.3.12.

Lema 5.3.15 *Dadas as permutações π_n e ι_n , com n par, $n \geq 4$, e a reversão $r(i, j)$ com $1 \leq i \leq j \leq n$ e $(j - i)$ par, $r(i, j) \cdot \pi_n$ produz um grafo de pontos-de-quebra que não contém 2-movimentos.*

Os Lemas 5.3.20 e 5.3.21 mostram que quando a primeira operação aplicada a π_n é uma transposição t 1-movimento então não temos 2-movimentos possíveis no grafo de pontos-de-quebra gerado por $t \cdot \pi_n$ e ι_n .

Vamos mostrar agora quais são as transposições 1-movimentos e 0-movimentos que podem ser aplicadas em $G(\pi_n, \iota_n)$.

Lema 5.3.16 *Dadas as permutações $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, com n ímpar, $n \geq 3$, e $G(\pi_n, \iota_n)$, a transposição $t(i, j, k)$, com $1 \leq i < j < k \leq n+1$, e $(j-i)$ e $(k-j)$ ambos ímpares, constitui-se num 1-movimento.*

Prova: De novo, $G(\pi_n, \iota_n)$, com n ímpar, é composto por um único ciclo de tamanho $n+1$ (Figura 5.7).

Queremos provar que a aplicação de $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $(j-i)$ e $(k-j)$ ambos ímpares constitui-se num 1-movimento. Assim, temos

$$\begin{aligned} \pi_n = \\ (-1 \ -2 \ \dots \ -(i-1) \ -i \ -(i+1) \ \dots \ -(j-1) \\ -j \ -(j+1) \ \dots \ -(k-1) \ -k \ -(k+1) \ \dots \ -n) \end{aligned}$$

e então

$$\begin{aligned} t(i, j, k) \cdot \pi_n = \\ (-1 \ -2 \ \dots \ -(i-1) \ -j \ -(j+1) \ \dots \ -(k-1) \\ -i \ -(i+1) \ \dots \ -(j-1) \ -k \ -(k+1) \ \dots \ -n) \end{aligned}$$

com $1 \leq i < j < k \leq n+1$ e $(j-i)$ e $(k-j)$ ambos ímpares.

Neste caso, i e k devem ter a mesma paridade e j deve ter a paridade oposta a i e k .

No caso de i e k ambos ímpares e j par, temos o grafo de pontos-de-quebra mostrado na Figura 5.18, formado por dois ciclos. Neste caso, temos os ciclos:

$$c_1 = [\overleftarrow{k-1}, \overleftarrow{k-3}, \dots, \overleftarrow{j}, \overleftarrow{j-2}, \dots, \overleftarrow{i+1}]$$

e

$$\begin{aligned} c_2 = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{k+1}, \overleftarrow{i+(k-1-j)}, \overleftarrow{i+(k-j-3)}, \dots, \overleftarrow{i+2}, \overleftarrow{i}, \overleftarrow{i-2}, \\ \dots, \overleftarrow{3}, \overleftarrow{1}, \overrightarrow{2}, \overrightarrow{4}, \dots, \overrightarrow{i-1}, \overrightarrow{i+(k-j+1)}, \overrightarrow{i+(k-j+3)}, \dots, \overrightarrow{k}, \overrightarrow{k+2}, \dots, \overrightarrow{n}] \end{aligned}$$

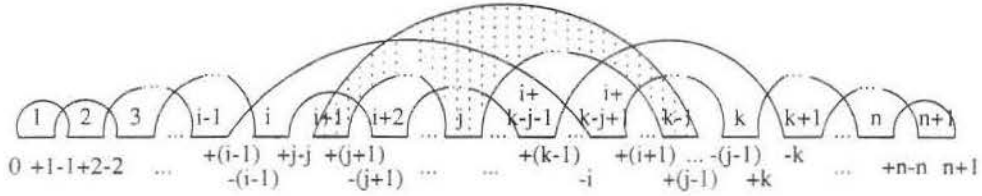


Figura 5.18: O grafo de pontos-de-quebra gerado pela aplicação de uma transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $j-i$ e $k-j$ ambos ímpares, em π_n , $n \geq 3$, ímpar. Com estas condições, i e k devem ter a mesma paridade e j deve ter paridade oposta a i e k . O grafo genérico para o caso de i e k ímpares e j par.

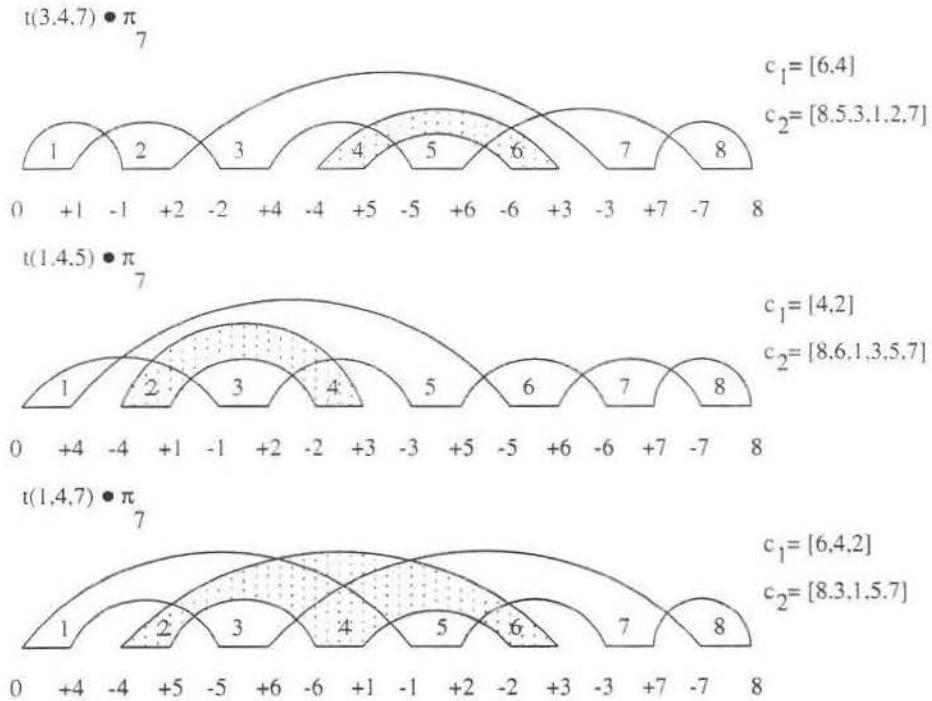


Figura 5.19: Exemplos de grafos de pontos-de-quebra gerados pela aplicação de uma transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $j-i$ e $k-j$ ambos ímpares, em π_n , $n \geq 3$, ímpar. As transposições estão descritas na figura.

Portanto, esta transposição é um 1-movimento. A Figura 5.19 mostra alguns exemplos.

O caso para i e k ambos pares e j ímpar é inteiramente análogo.

□

Lema 5.3.17 *Dadas as permutações $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, com n ímpar, $n \geq 3$, e $G(\pi_n, \iota_n)$, a transposição $t(i, j, k)$, com $1 \leq i < j < k \leq n+1$, e $(j-i)$ ou $(k-j)$ par, constitui-se num 0-movimento.*

Prova: O grafo de pontos-de-quebra $G(\pi_n, \iota_n)$, com n ímpar, é composto por um único ciclo de tamanho $n+1$ (Figura 5.7).

Queremos provar que a aplicação de $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $(j-i)$ ou $(k-j)$ par constitui-se num 0-movimento. Novamente,

$$t(i, j, k) \cdot \pi_n =$$

$$(-1 \ -2 \ \dots \ -(i-1) \ -j \ -(j+1) \ \dots \ -(k-1)$$

$$-i \ -(i+1) \ \dots \ -(j-1) \ -k \ -(k+1) \ \dots \ -n)$$

com $1 \leq i < j < k \leq n+1$ e $(j-i)$ ou $(k-j)$ par.

Neste caso, j deve ter paridade igual a pelo menos um dos outros dois.

No caso de j ímpar, i ímpar e k par, temos o grafo de pontos-de-quebra mostrado na Figura 5.20, formado por um único ciclo. Neste caso, temos o ciclo:

$$c = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{k+2}, \overleftarrow{k}, \overleftarrow{k-2}, \dots, \overleftarrow{i+1}, \overleftarrow{k-1}, \overleftarrow{k-3}, \dots, \overleftarrow{i+(k-j)+1},$$

$$\overleftarrow{i-1}, \dots, \overleftarrow{2}, \overrightarrow{1}, \overrightarrow{3}, \dots, \overrightarrow{i}, \overrightarrow{i+2}, \dots, \overrightarrow{i+(k-j-1)}, \overrightarrow{k+1}, \overrightarrow{k+3}, \dots, \overrightarrow{n}]$$

Portanto, esta transposição é um 0-movimento. A Figura 5.21 mostra alguns exemplos.

Os outros casos são inteiramente análogos.

□

Lema 5.3.18 *Dadas as permutações $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, com n par, $n \geq 4$, e $G(\pi_n, \iota_n)$, a transposição $t(i, j, k)$, com $1 \leq i < j < k \leq n+1$, e $(j-i)$ e $(k-j)$ ambos ímpares, constitui-se num 1-movimento.*

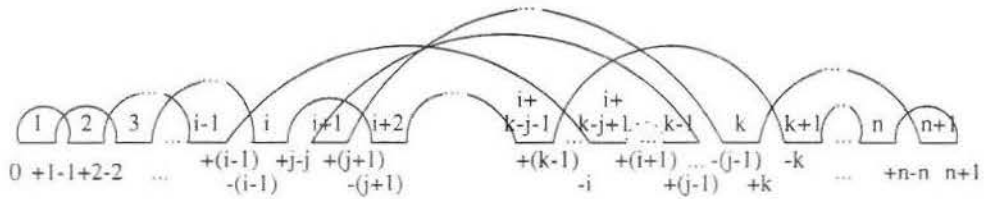


Figura 5.20: O grafo de pontos-de-quebra gerado pela aplicação de uma transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $j-i$ ou $k-j$ par, em π_n , $n \geq 3$, ímpar. Com estas condições, j deve ter paridade igual a um dos outros dois. O grafo genérico para o caso de j ímpar, i ímpar e k par.

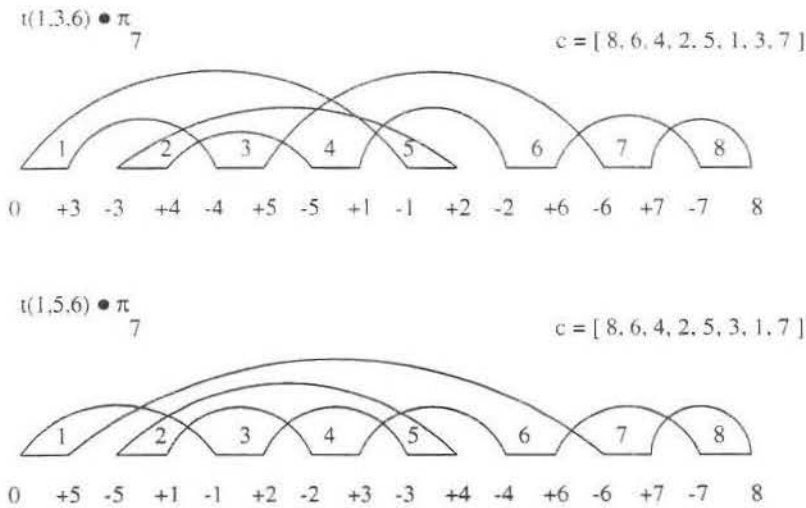


Figura 5.21: Exemplos de grafos de pontos-de-quebra gerados pela aplicação de uma transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $j-i$ ou $k-j$ par, em π_n , $n \geq 3$, ímpar. As transposições estão descritas na figura.

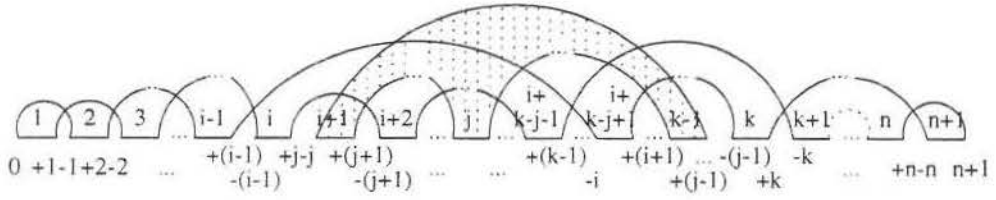


Figura 5.22: O grafo de pontos-de-quebra gerado pela aplicação de uma transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $j-i$ e $k-j$ ambos ímpares, em π_n , $n \geq 4$ par. Com estas condições, i e k devem ter a mesma paridade e k deve ter paridade oposta a i e k . O grafo de pontos-de-quebra para o caso de i e k ímpares e j par.

Prova: De novo, temos $G(\pi_n, \iota_n)$, com n ímpar, composto por um único ciclo de tamanho $n+1$ (Figura 5.7).

Queremos provar que a aplicação de $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $(j-i)$ e $(k-j)$ ambos ímpares constitui-se num 1-movimento. Assim,

$$\begin{aligned} t(i, j, k) \cdot \pi_n = \\ (-1 -2 \dots -(i-1) -j -(j+1) \dots -(k-1) \\ -i -(i+1) \dots -(j-1) -k -(k+1) \dots -n) \end{aligned}$$

com $1 \leq i < j < k \leq n+1$ e $(j-i)$ e $(k-j)$ ambos ímpares.

Neste caso, temos i e k com a mesma paridade e j com a paridade oposta a i e k .

No caso de i e k ambos ímpares e j par, temos o grafo de pontos-de-quebra mostrado na Figura 5.22, formado por dois ciclos. Neste caso, temos os ciclos:

$$c_1 = [\overleftarrow{k-1}, \overleftarrow{k-3}, \dots, \overrightarrow{j}, \overrightarrow{j-2}, \dots, \overrightarrow{i+1}]$$

e

$$\begin{aligned} c_2 = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{k+2}, \overleftarrow{k}, \overleftarrow{k-2}, \dots, \overleftarrow{i+(k-j+1)}, \\ \overleftarrow{i-1}, \dots, \overleftarrow{2}, \overrightarrow{1}, \overrightarrow{3}, \dots, \overrightarrow{i}, \overrightarrow{i+2}, \dots, \overrightarrow{i+(k-j-1)}, \\ \overrightarrow{k+1}, \overrightarrow{k+3}, \dots, \overrightarrow{n}] \end{aligned}$$

Portanto, esta transposição é um 1-movimento. A Figura 5.23 mostra alguns exemplos.

O caso para i, k ambos pares e j ímpar é inteiramente análogo.

□

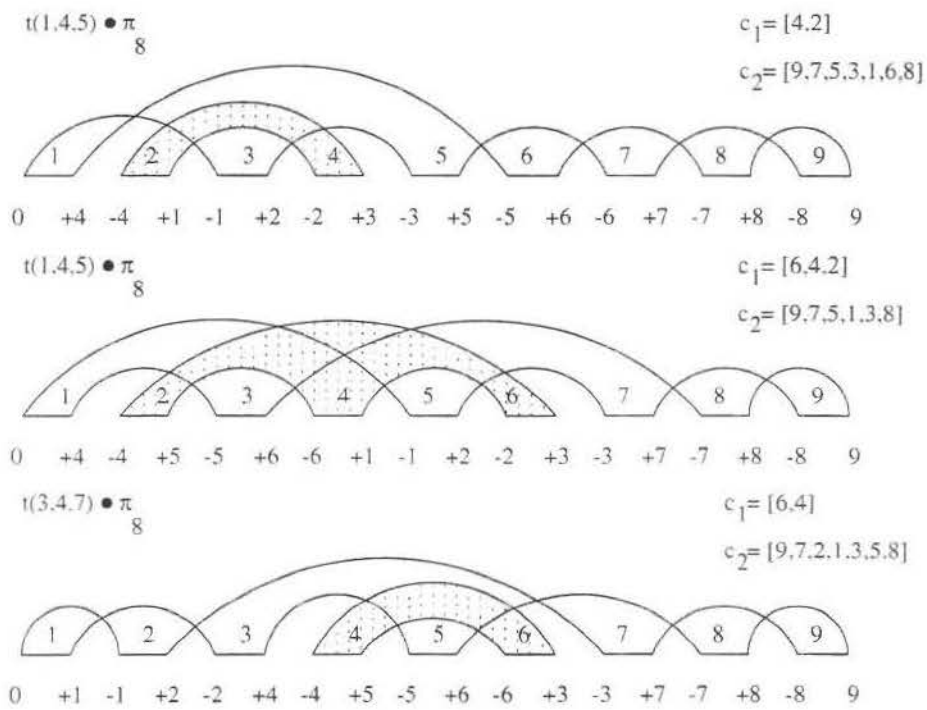


Figura 5.23: Exemplos de grafos de pontos-de-quebra gerados pela aplicação de uma transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n + 1$ e $j - i$ e $k - j$ ambos ímpares, i e k ímpares e j par, em π_n , $n \geq 4$, par. As transposições estão descritas na figura.

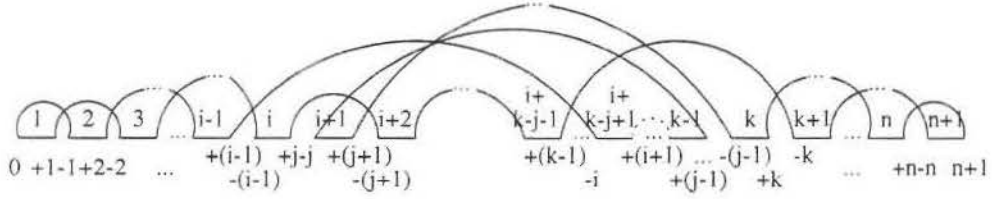


Figura 5.24: O grafo de pontos-de-quebra gerado pela aplicação de uma transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $j-i$ ou $k-j$ par, em π_n , $n \geq 4$, par. Com estas condições, j deve ter paridade igual a um dos outros dois. O grafo genérico para o caso de j ímpar, i ímpar e k par.

Lema 5.3.19 Dadas as permutações $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, com n par, $n \geq 4$, e $G(\pi_n, \iota_n)$, a transposição $t(i, j, k)$, com $1 \leq i < j < k \leq n+1$, e $(j-i)$ ou $(k-j)$ par, constitui-se num 0-movimento.

Prova: De novo, temos $G(\pi_n, \iota_n)$, com n par, composto por um único ciclo de tamanho $n+1$ (Figura 5.7).

Queremos provar que a aplicação de $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $(j-i)$ ou $(k-j)$ par constitui-se num 0-movimento. Assim,

$$t(i, j, k) \cdot \pi_n =$$

$$(-1 \ -2 \ \dots \ -(i-1) \ -j \ -(j+1) \ \dots \ -(k-1)$$

$$-i \ -(i+1) \ \dots \ -(j-1) \ -k \ -(k+1) \ \dots \ -n)$$

com $1 \leq i < j < k \leq n+1$ e $(j-i)$ ou $(k-j)$ par.

Neste caso, j deve ter paridade igual a um dos outros dois.

No caso de j ímpar, i ímpar e k par, temos o grafo de pontos-de-quebra mostrado na Figura 5.24, formado por um único ciclo. Neste caso, temos o ciclo:

$$c = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{k+1}, \overleftarrow{i+(k-j-1)}, \overleftarrow{i+(k-j-3)}, \dots, \overleftarrow{i}, \overleftarrow{i-2}, \dots, \overleftarrow{3}, \overleftarrow{1},$$

$$\overrightarrow{2}, \overrightarrow{4}, \dots, \overrightarrow{i-1}, \overrightarrow{i+(k-j+1)}, \overrightarrow{i+(k-j+3)}, \dots, \overrightarrow{k-1},$$

$$\overrightarrow{i+1}, \overrightarrow{i+3}, \dots, \overrightarrow{i+(k-j)}, \overrightarrow{i+(k-j+2)}, \dots, \overrightarrow{k}, \overrightarrow{k+2}, \dots, \overrightarrow{n}]$$

Portanto, esta transposição é um 0-movimento. A Figura 5.25 mostra alguns exemplos.

Os outros casos são inteiramente análogos.

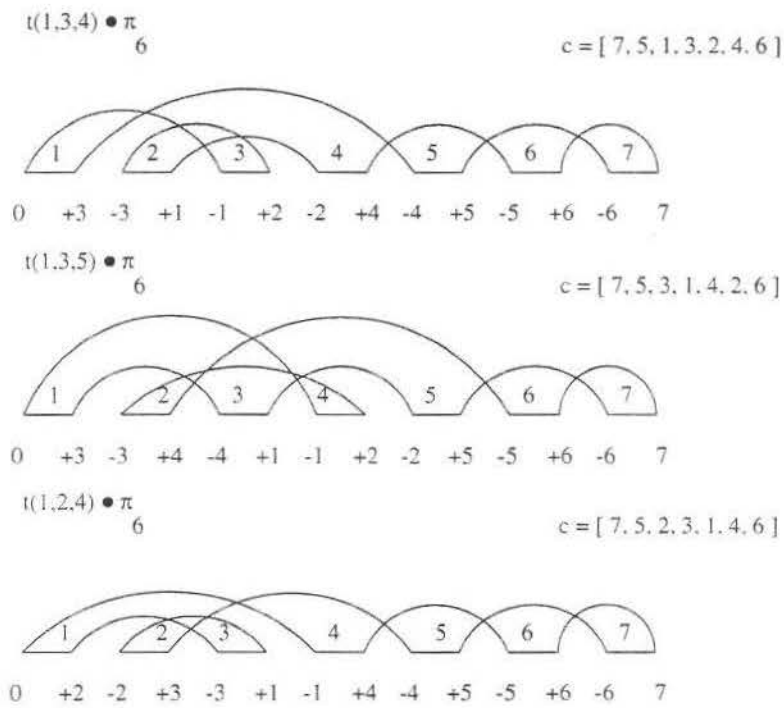


Figura 5.25: Exemplos de grafos de pontos-de-quebra gerados pela aplicação de uma transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $j-i$ ou $k-j$ par, em π_n , $n \geq 4$, par. As transposições estão descritas na figura.

□

Os lemas seguintes mostram que em $G(\pi_n, \iota_n)$ a aplicação de qualquer transposição 1-movimento gera grafos de pontos-de-quebra que não admitem 2-movimentos.

Lema 5.3.20 *Dadas as permutações π_n e ι_n , com n ímpar, $n \geq 3$, e a transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $(j-i)$ e $(k-j)$ ambos ímpares, $t(i, j, k) \cdot \pi_n$ produz um grafo de pontos-de-quebra que não contém 2-movimentos.*

Prova: Do Lema 5.3.16, sabemos que a aplicação de $t(i, j, k)$ nas condições da hipótese em π_n , n ímpar, $n \geq 3$, constitui-se num 1-movimento, e os ciclos, para i, k ímpares e j par, são:

$$c_1 = [\overleftarrow{k-1}, \overleftarrow{k-3}, \dots, \overleftarrow{j}, \overleftarrow{j-2}, \dots, \overleftarrow{i+1}]$$

e

$$c_2 = [\overleftarrow{n+1}, \overleftarrow{n-1}, \dots, \overleftarrow{k+1}, \overleftarrow{i+(k-1-j)}, \\ \overleftarrow{i+(k-j-3)}, \dots, \overleftarrow{i+2}, \overleftarrow{i}, \overleftarrow{i-2}, \dots, \overleftarrow{3}, \overleftarrow{1}, \\ \overrightarrow{2}, \overrightarrow{4}, \dots, \overrightarrow{i-1}, \overrightarrow{i+(k-j+1)}, \overrightarrow{i+(k-j+3)}, \dots, \overrightarrow{k}, \overrightarrow{k+2}, \dots, \overrightarrow{n}]$$

Podemos observar que c_1 é uma seqüência estritamente decrescente e c_2 é bimonótono, e portanto, pelo Teorema 5.3.9, não existem 2-movimentos possíveis.

O caso para i, k pares e j ímpar é análogo.

Assim, não temos 2-movimentos possíveis no grafo de pontos-de-quebra gerado por uma transposição, nas condições deste lema, aplicada em π_n , n ímpar, $n \geq 3$. □

A prova para o caso em que n é par é inteiramente análoga, bastando utilizar o Lema 5.3.18.

Lema 5.3.21 *Dadas as permutações π_n e ι_n , com n par, $n \geq 4$, e a transposição $t(i, j, k)$ com $1 \leq i < j < k \leq n+1$ e $(j-i)$ e $(k-j)$ ambos ímpares, $t(i, j, k) \cdot \pi_n$ produz um grafo de pontos-de-quebra que não contém 2-movimentos.*

Os lemas seguintes também serão utilizados na prova do Teorema 5.3.24.

O próximo lema mostra quais são as únicas operações que geram uma permutação contendo apenas faixas positivas, nos dois primeiros passos. A Figura 5.26 mostra exemplos.

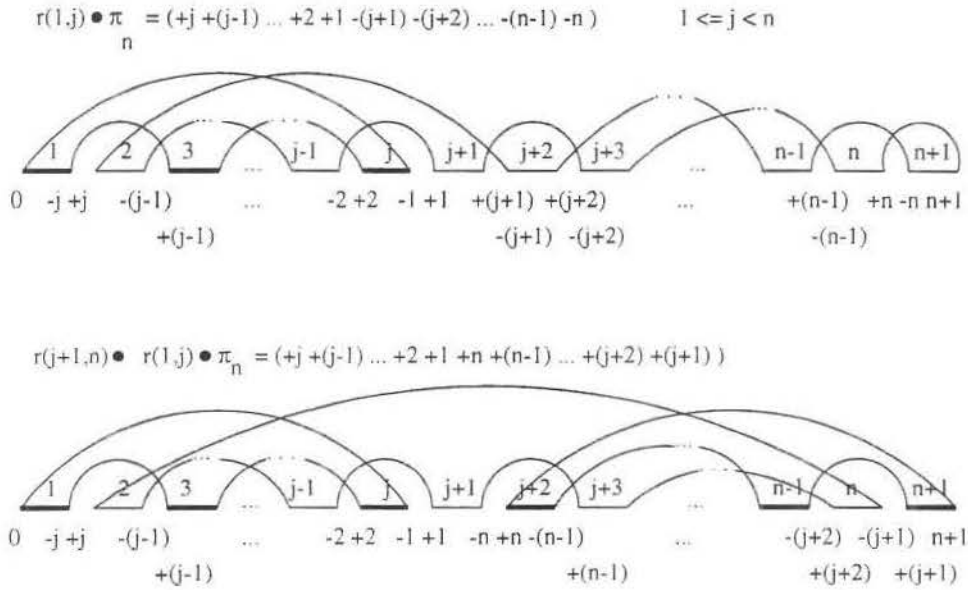


Figura 5.26: Os grafos de pontos-de-quebra gerados por reversões, nas condições do Lema 5.3.22.

Lema 5.3.22 *Dadas as permutações π_n e ι_n , $n \geq 3$, uma permutação π_2 contendo apenas faixas positivas é gerada nos dois primeiros passos se e somente se uma das quatro condições ocorre:*

1. $\pi_2 = r(i+1, n) \cdot r(1, i) \cdot \pi_n, 1 \leq i < n$
2. $\pi_2 = r(1, i) \cdot r(i+1, n) \cdot \pi_n, 1 \leq i < n$
3. $\pi_2 = t(i, j, k) \cdot r(1, n) \cdot \pi_n, 1 \leq i < j < k \leq n+1$
4. $\pi_2 = r(1, n) \cdot t(i, j, k) \cdot \pi_n, 1 \leq i < j < k \leq n+1$.

O lema abaixo baseia-se nos fatos de que, quando uma permutação tem faixas negativas com relação a uma outra permutação, devemos necessariamente aplicar reversões para tornar estas faixas positivas, e de que uma reversão pode criar no máximo 1 ciclo.

Lema 5.3.23 *Dadas as permutações π , de comprimento n , e ι_n , tais que π contém alguma faixa negativa com relação a ι_n , então qualquer caminho c de π a ι_n satisfaz $p(c) \geq 1$.*

Vamos provar agora o principal teorema desta seção.

Teorema 5.3.24 *Seja $c = \sigma_0\sigma_1\sigma_2\ldots\sigma_{k-1}\sigma_k$ qualquer caminho de $\pi_n = \sigma_0$ até $\iota_n = \sigma_k$. Então temos para $n \geq 3$:*

1. $p(\sigma_0\sigma_1) \geq 1$
2. se $p(\sigma_0\sigma_1) = 1$ então $p(\sigma_1\sigma_2) \geq 1$
3. se $p(\sigma_0\sigma_1\sigma_2) = 2$ então $p(\sigma_2\sigma_3\sigma_4\ldots\sigma_{k-1}\sigma_k) \geq 1$.

Prova:

A primeira afirmação é verdadeira porque o peso $p(\sigma_0, \sigma_1)$ é sempre maior ou igual a zero, e é zero apenas se o evento é um 2-movimento. Entretanto, $\sigma_0 = \pi_n$ tem apenas um ciclo, e este ciclo é bimonótono. A afirmação segue do Teorema 5.3.9.

Para a segunda afirmação, observamos que $p(\sigma_0\sigma_1) = 1$ exatamente quando a operação e que age em $\sigma_0 = \pi_n$ é 1-movimento. Ambos os eventos, reversões e transposições, podem ser 1-movimentos. A afirmação segue dos Lemas 5.3.14, 5.3.15, 5.3.20 e 5.3.21.

Vamos agora analisar a terceira afirmação. Dividiremos a prova em dois casos. Existem elementos negativos em σ_2 ou todos os elementos de σ_2 são positivos.

Se existem elementos negativos em σ_2 , então do Lema 5.3.23 temos que qualquer caminho $c = \sigma_2\sigma_3\sigma_4\ldots\sigma_{k-1}\sigma_k$ satisfaz $p(c) \geq 1$.

Vamos verificar o caso em que σ_2 tem todos os elementos positivos. Do Lema 5.3.22, temos que existem apenas quatro formas de obter uma permutação com todos os elementos positivos, em dois passos. Neste lema, os primeiros dois casos são os mesmos pois $r(1, i)$ e $r(i + 1, n)$ comutam. De fato, podemos mostrar que todos os casos podem ser reduzidos ao Caso 3. Para isto basta observar que toda transposição pode ser escrita como o produto de três reversões:

$$t(i, j, k) = r(i, k - 1) \cdot r(i, j - 1) \cdot r(j, k - 1)$$

Isto pode ser facilmente verificado a partir das definições. Se substituirmos $i = 1$ e $k = n + 1$ nesta equação, obtemos:

$$t(1, j, n + 1) = r(1, n) \cdot r(1, j - 1) \cdot r(j, n)$$

mostrando que os casos 1 e 2 são ainda casos particulares do Caso 4 (basta lembrar que $r(1, n)^2 = I$). Por outro lado,

$$r(1, n) \cdot t(i, j, k) \cdot r(1, n) = t(k', j', i')$$

onde $i' = n + 2 - i$, $j' = n + 2 - j$ e $k' = n + 2 - k$, o que mostra que o Caso 4 pode ser reduzido ao Caso 3.

Vamos portanto nos concentrar no Caso 3. Observe que neste caso σ_1 é a permutação $(+n \ + (n-1) \ \dots \ +2 \ +1)$. Uma consequência do trabalho de Meidanis, Walter e Dias [30] e de Christie [10], que computa a distância de transposição desta permutação é tal que $p(\sigma_1\sigma_2 \dots \iota_n) \geq 2$ para qualquer caminho constituído somente por transposições. Então, se $p(\sigma_2\sigma_3 \dots \iota_n) = 0$, isto iria significar um caminho usando apenas transposições, e então concluiríamos que $p(\sigma_1\sigma_2) = 2$ e que $p(\sigma_0\sigma_1) = 0$, uma contradição pois no primeiro passo $r(1, n)$ foi uma reversão. Segue portanto que $p(\sigma_2\sigma_3 \dots \iota_n) \geq 1$.

□

O Teorema 5.3.24 garante que $p(c) \geq 3$ para qualquer caminho c de π_n até ι_n . Utilizando este fato na fórmula do Corolário 5.3.4 concluímos que

$$d_{rt}(\pi_n, \iota_n) \geq \frac{n+3}{2},$$

o que implica que $d_{rt}(\pi_n, \iota_n) \geq \lfloor \frac{n}{2} \rfloor + 2$ pois $d_{rt}(\pi_n, \iota_n)$ é um número inteiro.

O teorema seguinte decorre imediatamente deste resultado e do Teorema 5.3.5.

Teorema 5.3.25 *Tomando $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ e $\iota_n = (+1 \ +2 \ \dots \ +(n-1) \ +n)$, para todo n , então temos*

$$d_{rt}(\pi_n, \iota_n) = \begin{cases} \lfloor \frac{n}{2} \rfloor + 1 & \text{se } n = 1, 2 \\ \lfloor \frac{n}{2} \rfloor + 2 & \text{se } n \geq 3 \end{cases}$$

Prova: Para $n = 1$ é óbvio que $d_{rt}(\pi_n, \iota_n) = 1$ pois $\pi_n \neq \iota_n$ e uma reversão transforma π_n em ι_n . Para $n = 2$ uma série mínima de operações transformando π_n em ι_n consiste de duas operações. Para $n \geq 3$ o resultado segue dos Teoremas 5.3.5 e do resultado $d_{rt}(\pi_n, \iota_n) \geq \lfloor \frac{n}{2} \rfloor + 2$. □

Neste ponto, observamos que um algoritmo para obter uma seqüência mínima de operações para transformar $\pi_n = (-1 \ -2 \ \dots \ -(n-1) \ -n)$ na identidade ι_n é: inicialmente aplique a reversão $r(1, n)$ em π_n para obter $r(1, n) \cdot \pi_n = (+n \ +(n-1) \ \dots \ +2 \ +1)$, e depois use o Algoritmo *DistTransp*, descrito na Seção 4.2.2, para obter as transposições para computar $d_{rt}(r(1, n) \cdot \pi_n, \iota_n)$.

Então temos um limite inferior para o diâmetro

$$\left\lfloor \frac{n}{2} \right\rfloor + 2 \leq D_{rt}(n).$$

Bafna e Pevzner [4] estabeleceram um limite superior para $D_{rt}(n)$,

$$D_{rt}(n) \leq \frac{3}{4}n.$$

Finalmente, podemos verificar que $D_{rt}(n) = \lfloor \frac{n}{2} \rfloor + 2$ para $n = 3, 4, 5, 6$.

5.4 Sumário e questões

Neste capítulo inicialmente apresentamos algoritmos de aproximação para computar a distância de reversão e transposição. Para o caso de permutações sem sinais mostramos um algoritmo baseado na noção de pontos-de-quebra, e para o caso com sinais, o algoritmo proposto foi baseado nos ciclos do grafo de pontos-de-quebra. Para o caso com sinais nosso algoritmo usa um tipo específico de reversão enquanto for possível, e em seguida usa parte dos resultados de Bafna e Pevzner [4] para encontrar as transposições a serem aplicadas.

Os limites inferiores usados para estimar a razão de aproximação destes algoritmos foram simples, mas levaram a um resultado mais profundo, que é o cálculo da distância exata para a permutação $(-1 \ -2 \ \dots \ -(n-1) \ -n)$ com relação à permutação identidade. Esta prova é baseada no número de ciclos que podem ser criados nos grafos gerados por qualquer seqüência de operações transformando π_n em ι_n . Claramente este resultado estabelece um limite inferior para o diâmetro, que conjecturamos ser também o limite superior.

Um tópico a ser estudado futuramente é como a inclusão de um terceiro evento, a transversão, afetaria os algoritmos e o cálculo do diâmetro. Apenas para lembrar, este evento transfere um bloco de genes de um local para outro dentro do cromossomo, mas com os genes invertidos. Este evento é natural se considerado sob o ponto de vista biológico.

Uma outra linha de pesquisa é considerar diferentes pesos para transposições e reversões. Com pesos iguais, como feito neste trabalho, o caminho mínimo é formado predominantemente por transposições. Seria interessante usar pesos sugeridos por observações feitas na prática pelos biólogos.

Capítulo 6

Conclusões

Dentre os diversos eventos de rearranjo de genomas já descobertos em Biologia Molecular, particularmente, nesta tese, estudamos os eventos de *reversão* e de *transposição* agindo num único cromossomo.

No primeiro capítulo, apresentamos formalismos para diversos eventos de rearranjo de genomas, e para os problemas relacionados a eles. Em seguida, por meio de uma tabela, descrevemos o estado destes problemas com relação à complexidade e aos melhores algoritmos conhecidos.

Dos problemas apresentados no Capítulo 1, mais especificamente, estudamos três: o problema da distância de reversão de cromossomos circulares com sinais, o problema da distância de transposição de cromossomos lineares sem sinais e finalmente o problema da distância de reversão e transposição de cromossomos lineares com sinais.

Mostramos inicialmente que os problemas da distância de reversão de cromossomos com sinais, circulares e lineares, eram equivalentes. Como consequência propusemos um algoritmo polinomial para resolvê-lo, baseado num algoritmo polinomial que resolve o mesmo problema para cromossomos lineares. Em seguida, calculamos o diâmetro de reversão de cromossomos com sinais, lineares e circulares. Depois, mostramos relações entre as distâncias de reversão de cromossomos circulares e lineares, do mesmo tamanho.

Para o problema da distância de transposição de cromossomos lineares sem sinais, apresentamos uma estrutura denominada de diagrama de pontos-de-quebra, e um algoritmo de aproximação baseado neste diagrama. Apesar da razão 2.25, superior à do melhor algoritmo conhecido de razão 1.5, proposto por Bafna e Pevzner [4], apresentou um desempenho bastante melhor quando utilizado na prática, conforme mostraram os testes realizados. Mostramos ainda um limite inferior para o diâmetro

de transposição.

Por último, para o problema da distância de reversão e transposição agindo em cromossomos lineares com sinais, apresentamos um algoritmo de aproximação de razão 2, que embora sendo um resultado simples, levou a um resultado mais complexo, um limite inferior para o diâmetro de reversão e transposição.

Uma razão prática para estudar o diâmetro, para cada um dos eventos estudados, é que este valor fornece um referencial para valores de distância encontrados pelos algoritmos, no seguinte sentido. Quanto mais próxima a distância estiver do diâmetro, menor a chance dos dois organismos estarem relacionados em termos de evolução genética.

Extensões e aprimoramentos dos nossos resultados foram propostos diretamente nos sumários dos capítulos.

De uma forma geral, como pudemos verificar nos nossos estudos, a área de rearranjo de genomas não possui uma teoria estabelecida e uniforme. Nesta área, as estruturas e algoritmos propostos são completamente independentes umas das outras, inclusive na notação, e as estratégias de soluções vêm de diferentes áreas como Teoria da Computação, Estatística e Matemática. Neste contexto, julgamos que seria interessante investigar a possibilidade de formalização dos eventos e dos problemas relacionados a um único evento ou múltiplos eventos de uma forma mais padronizada. Assim, passamos agora a sugerir novas direções de pesquisa que possivelmente propiciem novos e significativos avanços no sentido de uma padronização.

Uma direção seria utilizar Álgebra. Mais particularmente, sugerimos “Grupos de Weyl”, que possui formalismos para gerar um grupo de permutações, sem e com sinais, utilizando como operador um tipo específico de reversão. Sugerimos ainda Álgebra Linear, considerando que uma sequência de permutações que fornece uma distância mínima poderia ser encontrada se fosse possível obter uma base para o espaço vetorial de todas as permutações, considerando como operação um determinado evento de rearranjo.

Outra direção de pesquisa seria estudar grafos de permutações, dentro da área de Teoria dos Grafos. Existe um grande ferramental teórico já disponível na literatura.

Por último, estudar problemas de distância entre dois cromossomos que evoluem por múltiplos eventos de rearranjo de genomas representa um desafio considerável atualmente. Um outro problema importante hoje é pesquisar as complexidades dos diversos problemas de distância de eventos de rearranjo de genomas, formados por um ou mais cromossomos.

Apêndice A

Glossário de definições e notações

Este apêndice apresenta definições e notações utilizadas nesta tese, para facilitar a leitura do texto.

Observamos que as definições estão separados por assunto, e após a definição há uma referência para a página do texto na qual o termo é referenciado pela primeira vez.

Eventos de rearranjo

reversão inverte a ordem e a orientação de um bloco de genes num cromossomo (pág. 6).

transposição move um bloco de genes de um local para o outro dentro do cromossomo (pág. 6).

transversão move um bloco de genes de um local para o outro dentro do cromossomo, mas inverte a ordem e a orientação dos genes movidos (pág. 6).

translocação troca blocos de genes de dois cromossomos diferentes (pág. 6).

Grafo de pontos-de-quebra

ciclo bom ciclo c no qual podemos aplicar pelo menos uma reversão que quebra c , isto é, conseguimos aumentar o número de ciclos do grafo de pontos-de-quebra de 1 (pág. 27).

ciclo ruim ciclo c no qual a aplicação de qualquer reversão não quebra c , isto é, não conseguimos aumentar o número de ciclos do grafo de pontos-de-quebra de 1 (pág. 27).

ciclos entrelaçados ciclos que possuem arestas cinzas que se cruzam (pág. 27).

componente conjunto de ciclos entrelaçados (pág. 28).

obstáculo componente composta apenas por ciclos ruins, que não separam quaisquer duas outras componentes (pág. 28).

super-obstáculo obstáculo, cuja remoção transforma uma outra componente, que não é um obstáculo, num obstáculo (pág. 29).

fortaleza grafo de pontos-de-quebra que contém um número ímpar de obstáculos, e todos eles são super-obstáculos. Numa fortaleza, existe pelo menos um ciclo que não pertence a um obstáculo (pág. 29).

Cromossomos circulares

rotação operação que move os elementos de uma permutação uma posição à esquerda (pág. 40).

reflexão operação que inverte a ordem e os sinais de todos os elementos de uma permutação (pág. 40).

reversão circular inverte o mesmo bloco de elementos, qualquer que seja a permutação escolhida da classe de equivalência que representa o cromossomo circular (pág. 44).

representante canônica uma particular permutação de uma classe de equivalência, tendo como primeiro elemento o bloco 1, com orientação + (pág. 45).

Ponto-de-quebra

em permutação com sinais/reversão par de elementos adjacentes $x = (\pi_i, \pi_{i+1})$ tal que nem x nem $\bar{x} = (\bar{\pi}_{i+1}, \bar{\pi}_i)$ são da forma $(j, j+1)$ para algum j , $0 \leq j \leq n$ (pág. 107).

em permutação sem sinais/transposição par de elementos adjacentes π_i, π_{i+1} , $0 \leq i \leq n$ tal que $\pi_{i+1} - \pi_i \neq 1$ (pág. 66).

faixa todos os elementos de uma permutação que estão entre dois pontos-de-quebra consecutivos (pág. 66).

Em seguida relacionamos as notações utilizadas nesta tese.

$r(i, j) \cdot \pi$ reversão aplicada à permutação π .

$t(i, j, k) \cdot \pi$ transposição aplicada à permutação π .

$\bar{l}(i, j, k) \cdot \pi$ transversão aplicada à permutação π .

$Tr(\pi, \sigma, i, j) \cdot \Pi$ translocação aplicada às permutações π e σ do genoma Π .

$\varrho = d_r(\pi, \sigma)$ número mínimo de reversões que transformam π em σ .

$\delta = d_t(\pi, \sigma)$ número mínimo de transposições que transformam π em σ .

$\bar{\delta} = d_{\bar{l}}(\pi, \sigma)$ número mínimo de transversões que transformam π em σ .

$\tau = d_{Tr}(\Pi, \Gamma)$ número mínimo de translocações que transformam Π em Γ .

$\varrho\delta = d_{rt}(\pi, \sigma)$ número mínimo de reversões e transposições que transformam π em σ .

$\varrho\delta\bar{\delta} = d_{rt\bar{l}}(\pi, \sigma)$ número mínimo de reversões, transposições e transversões que transformam π em σ .

$\varrho\tau = d_{rTr}(\Pi, \Gamma)$ número mínimo de reversões e translocações que transformam Π em Γ .

$\delta\tau = d_{tTr}(\Pi, \Gamma)$ número mínimo de transposições e translocações que transformam Π em Γ .

$G(\pi, \sigma)$ grafo de pontos-de-quebra de π com relação a σ .

$c(\pi, \sigma)$ número de ciclos de $G(\pi, \sigma)$.

$h(\pi, \sigma)$ número de obstáculos de $G(\pi, \sigma)$.

$f(\pi, \sigma)$ indica se o grafo $G(\pi, \sigma)$ é ou não uma fortaleza.

rot (π) rotação aplicada à permutação π .

refl (π) reflexão aplicada à permutação π .

$can(A)$ representante canônica de uma classe de equivalência A .

S_n conjunto de todas as permutações lineares com sinais de n elementos.

S_n^c conjunto de todas as classes de equivalência das permutações lineares com sinais de n elementos.

R_n conjunto de todas as reversões lineares agindo nas permutações de S_n .

R_n^c conjunto de todas as reversões circulares agindo nas classes de equivalência de S_n^c .

$\pi_i \bullet \pi_j$ ponto-de-quebra entre π_i e π_j .

$\pi \prec \pi_j$ indica que $\pi_j - \pi_i = 1$.

$D(\pi, \iota)$ diagrama de pontos-de-quebra de π com relação a ι .

$T(i, j, k)$ transposição aplicada aos vértices i, j e k de $D(\pi, \iota)$.

Apêndice B

Referências bibliográficas comentadas

Este apêndice traz um resumo dos artigos referenciados na bibliografia, sendo ressaltados certos pontos importantes para a área de rearranjo de genomas.

[1]

Neste artigo, Aigner e West estudaram o problema de ordenar uma permutação, considerando como operação a reinserção do primeiro elemento na permutação. O diâmetro neste caso é $n - 1$, onde n é o número de elementos da permutação.

[2]

Neste trabalho é estudado o problema de ordenação por reversões de permutações lineares. Inicialmente, é introduzida uma estrutura chamada de *grafo de pontos-de-quebra* de uma permutação inicial em relação a uma permutação alvo, que permitiu estabelecer um limite inferior mais preciso para a distância de reversão considerando um novo parâmetro, baseado numa decomposição de ciclos alternantes de máxima cardinalidade deste grafo. Em seguida, com base nesta construção, foi apresentada uma prova da conjectura de Gollan, a saber, existem apenas duas permutações cuja distância de reversão é $n - 1$, onde n é o número de elementos da permutação. Depois, mostraram que a distância de reversão esperada entre duas permutações randômicas aproxima-se muito do diâmetro de reversão, indicando que a distância de reversão fornece uma boa separação entre seqüências de genes relacionados e não-relacionados. Para o problema de ordenação por reversões de permutações lineares com sinais, apresentaram um algoritmo com razão de aproximação 1.5, baseado no fato de que a decomposição em ciclos do grafo de pontos-de-quebra é única para permutações com sinais. Finalmente, utilizaram o algoritmo anterior e estruturas especiais obtidas

a partir do grafo de pontos-de-quebra para desenvolver um algoritmo com razão de aproximação 1.75 para o problema de ordenação por reversões de permutações sem sinais.

[3]

Neste trabalho, Bafna e Pevzner mostram os resultados obtidos de um programa que constrói um cenário para rearranjo de genomas (para DNA linear e circular). As aplicações foram feitas em organelas de plantas e no cromossomo X humano e de camundongos. O programa implementa um algoritmo de aproximação para ordenação por reversões, apresentado em trabalho anterior destes mesmos autores [2].

[4]

Neste trabalho, Bafna e Pevzner apresentaram dois algoritmos de aproximação, com razões 1.75 e 1.5, respectivamente, para o problema de ordenação por transposição de uma permutação linear sem sinais. Os limites inferiores foram encontrados a partir de definições de ciclos alternantes de uma estrutura denominada de *grafo de ciclos*, gerada a partir da permutação inicial com relação à permutação identidade. Os limites superiores foram encontrados com base em definições de ciclos que se entrelaçam neste grafo de ciclos. Por fim, considerando estes algoritmos, encontraram um limite superior de $\frac{3}{4}n$ para o diâmetro de transposição.

[5]

Neste trabalho foi apresentada uma implementação do algoritmo polinomial para ordenação por reversões de permutações lineares com sinais de Hannenhalli e Pevzner [19], reduzindo a complexidade de tempo de execução deste algoritmo de $O(n^4)$ para $O(n^2\alpha(n))$, onde $\alpha(n)$ é a inversa da função de Ackerman. Para reduzir a complexidade, foram modificados dois procedimentos: as componentes conexas do grafo de entrelaçamento são computadas em tempo $O(n\alpha(n))$ por meio da estrutura de dados floresta de conjuntos disjuntos; as reversões que não criam componentes não-orientadas são encontradas em tempo $O(n^2\alpha(n))$ considerando certas propriedades específicas do grafo de entrelaçamento.

[6]

Neste artigo, é apresentada a prova de que o problema de ordenação por reversões de uma permutação linear sem sinais é NP-difícil. A prova é feita em dois passos. Primeiro, Caprara **mostra** uma transformação polinomial do problema de *particionar o conjunto de arestas de um grafo euleriano num número máximo de ciclos* no problema de *encontrar um número máximo de ciclos alternantes aresta-disjuntos num grafo bi-colorido nas arestas*. Caprara mostra neste mesmo artigo que o primeiro problema é NP-completo utilizando resultados publicados em trabalhos anteriores, e

define o grafo bi-colorido de forma tal que o grafo possui todas as propriedades de um grafo de pontos-de-quebra. No segundo passo, é feita uma transformação polinomial deste último problema no de *ordenar uma permutação sem sinais por reversões*.

[7]

Neste artigo, Caprara, Lancia e Ng propõem, para o problema da ordenação por reversões, um algoritmo exato branch-and-bound. Neste algoritmo, a estimativa do limite inferior é baseada em resultados anteriores de Bafna e Pevzner [2], e ainda em Programação Linear, na qual utilizaram uma técnica específica chamada “geração de colunas”. Para obter o limite superior os autores usaram certas heurísticas. O algoritmo mostrou-se muito eficiente na prática, permitindo resolver problemas de tamanho significativo, quando comparado ao algoritmo de Kececioğlu e Sankoff [25], o outro algoritmo exato conhecido para solucionar o mesmo problema. Por último, mostram resultados dos experimentos realizados.

[8]

Christie estuda o problema da ordenação de permutações por uma operação chamada “troca de blocos”, que troca de lugar dois blocos de elementos não necessariamente adjacentes na permutação, onde os blocos podem ter qualquer tamanho. Ele descreve um algoritmo polinomial, bastante simples, para o problema, e computa o diâmetro (maior distância entre duas permutações quaisquer).

[9]

Neste artigo, Christie apresenta, para o problema da ordenação por reversões, um algoritmo de aproximação de razão 1.5, menor do que a razão mais baixa conhecida anteriormente, de 1.75, de Bafna e Pevzner [2].

[10]

Nesta tese, Christie estuda os eventos de reversão, transposição e troca de blocos. Para o problema da ordenação por reversões, entre outros resultados, ele apresenta um algoritmo de aproximação de razão 1.5. Depois, propõe um algoritmo polinomial para o problema da ordenação por troca de blocos. Para o problema da ordenação por transposições, ele descreve um novo algoritmo de aproximação de razão 1.5, e apresenta resultados de experimentos realizados executando este algoritmo. No último capítulo, Christie define problemas de distância de reversão e transposição em strings, mostra alguns resultados para estes problemas, e prova que o problema da distância de reversão em strings é NP-difícil, mesmo com strings formadas a partir de um alfabeto binário.

[11]

Copeland e co-autores apresentam mapas de ligação genética de alta-resolução para o camundongo, utilizando avanços tecnológicos recentes. Eles apontam diversas aplicações para mapas deste tipo. Citaremos apenas o mapeamento comparativo (que é mostrado no artigo) entre os cromossomos humanos e do camundongo. Fizeram ainda uma estimativa de que aproximadamente 150 eventos de rearranjo ocorreram desde a divergência entre estas duas linhagens há 80 milhões de anos atrás, confirmando outras pesquisas feitas anteriormente em Biologia Molecular.

[12]

Even e Goldreich provaram que dado um conjunto de geradores de um grupo de permutações e uma permutação alvo, determinar uma sequência de geradores com tamanho mínimo que gera a permutação alvo é NP-completo.

[13]

Neste trabalho, Gates e Papadimitriou estudaram o problema de ordenar por *reversão de prefixos* uma permutação linear sem sinais: reversões da forma $r[1, i]$. Foi provado que o diâmetro de reversão por prefixos é $\leq \frac{5}{3}n + \frac{5}{3}$ e $\geq \frac{17}{16}n$ para n suficientemente grande. Considerando a hipótese de que cada inteiro da permutação é revertido um número par de vezes, eles mostraram que $\frac{3}{2}n - 1 \leq$ diâmetro de reversão por prefixos $\leq 2n + 3$.

[14]

Gu, Peng e Sudborough estudaram o problema da distância de reversão, transposição e transversão, apresentando dois algoritmos de aproximação: o primeiro de razão 2, que pode não ser polinomial em casos específicos, e o segundo, que possui tempo polinomial, tem razão $2(1 + \frac{1}{k})$, onde $k > 3$ é qualquer inteiro fixado.

[15]

Neste trabalho, Guyer, Heath e Vergara estudaram o problema da ordenação por transposições. Apresentaram certas heurísticas e formularam algoritmos correspondentes à estas heurísticas, baseadas nos conceitos de faixas crescentes e subsequências. Implementaram ainda algoritmos branch-and-bound para determinar exatamente as distâncias de transposição de permutações de tamanho pequeno. Por fim, descreveram e analisaram resultados experimentais baseados nos algoritmos implementados.

[16]

Este artigo traz um algoritmo polinomial para o problema da distância de translocação de permutações lineares com sinais, com complexidade de tempo $O(n^3)$. Este algoritmo baseou-se num teorema da dualidade com seis parâmetros, que incluem o número de genes e o de cromossomos, além de determinadas características associ-

adas a uma estrutura denominada de *grafo de ciclos*, gerada a partir do genoma inicial com respeito ao genoma alvo. É interessante notar que o grafo de ciclos baseia-se no grafo de pontos-de-quebra para permutações com sinais, de Hannenhalli e Pevzner [19], e que este algoritmo trata genomas com o mesmo número de genes e de cromossomos, e admite apenas translocações recíprocas.

[17]

Hannenhalli e co-autores apresentaram neste trabalho um método para reconstrução de árvores filogenéticas utilizando distância de reversão e distância de reversão e transposição. O método foi aplicado na análise da evolução do vírus da herpes, pois os genomas destes vírus já foram quase que totalmente seqüenciados.

[18]

Este artigo traz descrições de problemas combinatoriais motivados por rearranjos de genomas, apresenta um levantamento de algoritmos desenvolvidos para comparação de genomas evoluindo por eventos de rearranjo (alguns deles bastante detalhadamente), e cita aplicações já desenvolvidas para estes algoritmos para analisar rearranjos em vírus da herpes, organelas de plantas e cromossomos de mamíferos.

[19]

Este artigo traz um algoritmo polinomial para o problema da ordenação por reversões de permutações lineares com sinais, com complexidade de tempo $O(n^4)$. Este algoritmo baseou-se num teorema da dualidade com quatro parâmetros, que incluem os pontos-de-quebra da permutação inicial com relação à permutação identidade, e certas características associadas a uma estrutura denominada de *grafo de pontos-de-quebra* gerada a partir da permutação inicial com relação à permutação identidade. Notamos que este foi o primeiro algoritmo polinomial para este problema.

[20]

Este artigo traz um algoritmo polinomial para o problema da distância de reversões e translocações de permutações lineares com sinais, com complexidade de tempo $O(n^4)$. Este algoritmo trata genomas com diferentes números de genes e de cromossomos, e admite translocações internas, fusões e fissões. Este algoritmo foi desenvolvido com base num teorema da dualidade com sete parâmetros, que refletem certas características de uma estrutura, que é basicamente um grafo de pontos-de-quebra com determinadas propriedades, gerada a partir do genoma inicial com relação ao genoma alvo. Genericamente, o algoritmo pode ser visualizado em dois passos. No primeiro passo, o algoritmo encontra uma *concatenação ótima* dos cromossomos de um genoma em relação ao outro, gerando um grafo de pontos-de-quebra. No segundo passo, é utilizado o algoritmo polinomial de Hannenhalli e Pevzner [19], para ordenar por

reversões permutações lineares com sinais, de tal forma que as reversões efetuadas por este algoritmo simulam reversões, translocações, fusões e fissões do problema da distância de reversões e translocações.

[21]

Neste trabalho, Hannenhalli e Pevzner estudam o problema de ordenação por reversões de permutações lineares sem sinais. Inicialmente, é apresentado um algoritmo polinomial para o problema, com complexidade de tempo de $O(n^4)$, para permutações com $O(\log n)$ faixas unitárias, onde n é o tamanho da permutação. Provam ainda duas conjecturas de Kececioglu e Sankoff [25], para toda permutação existe uma ordenação ótima por reversões que não corta faixas longas, e para toda permutação existe uma ordenação ótima que não aumenta o número de pontos-de-quebra. Finalmente, aplicam este algoritmo para analisar rearranjos em milho e algas verdes.

[22]

Este artigo mostra que o problema *dados um grupo de permutações descrito como um conjunto de geradores, e uma permutação alvo que é membro do grupo, encontrar a menor seqüência de geradores cuja composição é a permutação alvo* é Pespaço-completo. Apresenta ainda casos particulares deste problema que podem ser computados em tempo polinomial, dos quais citamos todas as transposições adjacentes (as permutações $(1, 2), (2, 3), \dots, (n-1, n)$ do conjunto $\{1, 2, \dots, n\}$) e todas as transposições adjacentes ciclicamente (o mesmo conjunto de geradores do primeiro caso e a transposição $(1, n)$).

[23]

Neste trabalho, Kaplan e co-autores apresentaram um algoritmo polinomial com complexidade de tempo $O(n^2)$ para o problema de ordenação por reversões de permutações lineares com sinais, baseado no algoritmo polinomial de Hannenhalli e Pevzner [19]. Basicamente, eles utilizam uma estrutura chamada de grafo de sobreposições para computar as componentes conexas do grafo de pontos-de-quebra em vez do grafo de entrelaçamento, sendo que a computação das componentes conexas é feita uma única vez utilizando o algoritmo de Berman e Hannenhalli [5]; tratam as componentes orientadas e não-orientadas do grafo de sobreposições separadamente; e simplificam tanto a pesquisa para encontrar uma reversão que não cria componentes não-orientadas quanto a atualização do grafo de sobreposições.

[24]

Este artigo apresenta inicialmente um algoritmo com razão de aproximação 2 para o problema da distância de translocação de permutações lineares, com complexidade de tempo de $O(n^2)$, onde n é o tamanho da permutação. Além disso, traz dois

algoritmos de aproximação para o problema da distância de reversões e translocações para permutações lineares: um com razão de aproximação 2 para permutações sem sinais e o outro com razão de aproximação 1.5 para permutações com sinais, ambos com complexidade de tempo $O(n^2)$.

[25]

Kececioğlu e Sankoff apresentam dois algoritmos para o problema de ordenação por reversões de permutações lineares sem sinais. O primeiro é um algoritmo com razão de aproximação 2 e complexidade de tempo $O(n^2)$, que tenta em cada passo tirar o máximo número de pontos-de-quebra (utiliza uma estratégia gulosa). O segundo é um algoritmo exato branch-and-bound que busca uma solução ótima em tempo $O(mL(n, n))$, onde m é o tamanho da árvore de busca e $L(n, n)$ é o tempo para resolver um problema de programação linear com n variáveis e n restrições.

[26]

Kececioğlu e Sankoff apresentam neste artigo um algoritmo branch-and-bound para o problema da distância de reversão para permutações circulares com sinais. O algoritmo, que usa procedimentos muito simples para encontrar os limites inferiores ($O(n)$) e superiores ($O(n^2)$), mostrou limites extremamente precisos para a distância de reversão, em uma série de experimentos. Os autores relatam que não encontraram motivos para que estes limites fossem tão próximos.

[27]

Este livro traz um estudo bastante aprofundado de Genética, apresentando inúmeros exemplos e ilustrações, que auxiliam muito no entendimento do texto. Descreve detalhadamente o DNA (como o DNA é traduzido em proteínas, formas de replicação e manutenção); descreve a organização do genoma dos eucariotos, não só em termos dos seus genes, mas também discute a estrutura dos cromossomos; mostra ainda estudos sobre a expressão dos genes nos eucariotos e funções do RNA. Particularmente, apresenta um capítulo sobre pedaços do genoma que têm a capacidade de “mover-se”, os *transposons*, que podem causar mutações, e portanto implicar na evolução deste genoma. Mutações afetando porções grandes do genoma, os eventos de rearranjo, podem ocorrer por inserções, duplicações, remoções, inversões ou transposições, modificando apenas um cromossomo, ou por translocações, modificando cromossomos diferentes.

[28]

Neste livro é feita uma ampla cobertura dos problemas estudados atualmente em Biologia Molecular Computacional. Para cada problema, Meidanis e Setubal apresentam conceitos e formalismos empregados para solucioná-los, exemplos, além de

indicar vasta bibliografia relacionada ao assunto. Em particular, dedicam um capítulo para o problema de rearranjo de genomas.

[29]

Meidanis, Walter e Dias estudam o problema da distância de reversão de cromossomos circulares com sinais, isto é, cromossomos circulares em que são conhecidas a ordem e a orientação dos genes. Inicialmente, apresentam uma formalização para o problema, onde o cromossomo circular é representado por uma classe de equivalência. Depois, definem reversão circular agindo num cromossomo circular por meio de reversão linear agindo numa particular permutação da classe de equivalência que representa o cromossomo circular. Como consequência, eles obtêm um algoritmo polinomial para o problema circular, utilizando o algoritmo polinomial de Kaplan, Shamir e Tarjan [23] que resolve o problema linear. Justificam, também, uma prática comum, em artigos onde se calculam distâncias de reversão, de escolher uma particular permutação para representar o cromossomo. Finalmente, calculam o diâmetro de reversão (maior distância entre dois cromossomos quaisquer), linear e circular.

[30]

Meidanis, Walter e Dias computam a distância de transposição entre uma permutação e sua inversa (sem complementação). Claramente, esta distância é um limite inferior para o problema da ordenação por transposições. Finalmente, apresentam um algoritmo polinomial que funciona para este caso particular.

[31]

Nadeau e Taylor estudaram mapas de ligação genética de homens e camundongos, fazendo uma estimativa de que aproximadamente 178 ± 39 eventos de rearranjo ocorreram desde a separação destas duas linhagens há 80 milhões de anos atrás. Esta foi uma tentativa pioneira no estudo de rearranjos de genomas em mamíferos.

[32]

Este artigo mostra estudos de Palmer e Herbon em DNA mitocondrial de plantas (gêneros *Brassica* e *Raphanus*). Eles descobriram que os genes eram quase idênticos em seqüência (99% a 99.9%) mas muito diferentes em ordem, descobrindo então um novo padrão de mudança evolucionária nos genomas mitocondriais de espécies do gênero *Brassica*. Este foi um trabalho pioneiro em rearranjo de genomas ocorrendo em DNA mitocondrial de plantas.

[33]

Palmer, Osorio e Thompson apresentam resultados de experimentos realizados em DNAs de cloroplastos de legumes, como tabaco, espinafre e ervilha. Eles constata-

que os genomas dos organismos das diferentes espécies estudadas evoluíram um a partir do outro por reversões, confirmando trabalhos anteriores, em Biologia Molecular, nos quais foi descoberto que mutações ocorrendo em porções grandes dos genomas (em particular, reversões) constituem um modo comum de evolução molecular.

[34]

Neste trabalho são discutidos resultados encontrados na literatura, e descritos problemas combinatoriais ainda não-solucionados até aquele momento (vários deles ainda não foram hoje), na área de Biologia Molecular Computacional. Particularmente, uma seção é dedicada aos problemas de rearranjo de genomas.

[35]

Sankoff introduz uma teoria de invariantes (com respeito a medidas de tempo) associados aos ramos de árvores filogenéticas. Esta teoria permite estudar o relacionamento evolutivo entre espécies, considerando eventos de rearranjo ocorridos nos seus genomas.

[36]

Neste artigo, Sankoff define medidas de certos eventos de rearranjo de genomas com um único cromossomo (reversão, remoção e transposição). Estas medidas são utilizadas na definição de *distância de edição* de rearranjos entre dois genomas, isto é, o número mínimo de eventos de rearranjo necessários para converter um no outro. A computação da distância de edição é feita por um programa chamado DERANGE, e o artigo comenta de forma genérica alguns procedimentos realizados por este programa. Por fim, relata resultados obtidos pela execução do programa em dados de genomas mitocondriais de organismos de vários gêneros.

[37]

Sankoff e co-autores apresentam modelos probabilísticos para embaralhamento de genes em genomas (particularmente por transposições), aplicando os modelos para descobrir relacionamentos entre genomas de bactérias. O DNA pode ser linear ou circular, mas os modelos os tratam uniformemente como sendo circulares.

[38]

Sankoff e co-autores apresentam um método para reconstruir árvores filogenéticas para genomas mitocondriais. Inicialmente, eles descrevem a construção de uma base de dados para armazenar a seqüência dos genes. Em seguida, propõem uma medida de rearranjo baseada num conjunto minimal de eventos de rearranjo (reversões, transposições, inserções, e remoções) necessários para transformar um genoma em outro. Descrevem então de forma genérica as técnicas empregadas para o desenvolvimento

do programa, que utiliza os dados da base e calcula a distância de rearranjo. Finalmente, apresentam os resultados obtidos a partir deste programa: as distâncias entre 16 genomas mitocondriais de organismos de diferentes gêneros e a reconstrução de uma árvore filogenética baseada nestas distâncias.

[39]

Este livro traz uma introdução à citogenética humana. Apresenta com detalhes a estrutura e comportamento de cromossomos humanos. Descreve também o processo da reprodução, além de diversos tipos de mutações que podem ocorrer nos cromossomos, em diferentes etapas da vida celular (em particular, citamos *translocações*). Como certos mecanismos são encontrados de forma similar em outros organismos eucariotos, os autores mostram inúmeros exemplos de fenômenos ocorrendo em outras espécies, diferentes da espécie humana.

[40]

Walter, Dias e Meidanis apresentam inicialmente algoritmos de aproximação para o problema da distância de reversão e transposição de cromossomos lineares. Para o caso com sinais, o algoritmo de aproximação de razão 3 utiliza o conceito de ponto-de-quebra. Para o caso com sinais, o algoritmo de razão 2 utiliza propriedades do grafo de pontos-de-quebra de Bafna e Pevzner [2] para aplicar reversões, e o algoritmo de aproximação de Bafna e Pevzner [4] para aplicar transposições. Em seguida mostram um limite inferior para o diâmetro de reversão e transposição, por meio do cálculo da distância de reversão e transposição entre as permutações $\pi = (-1 - 2 \dots -(n-1) - n)$ e a identidade $\iota = (+1 + 2 \dots + (n-1) + n)$.

[41]

Este é o primeiro livro na área de Biologia Computacional. Neste livro, Waterman descreve problemas estudados nas diversas áreas de Biologia Computacional, apresentando formalismos empregados para solucioná-los e exemplos. É enfatizada a análise estatística para diversos problemas nesta área.

[42]

Este artigo apresenta um algoritmo bastante simples para ordenar por reversões permutações circulares sem sinais, estabelecendo um limite inferior (número de pontos-de-quebra/2), e um limite superior $(n - 2)$ para a distância de reversão. Apresenta também um algoritmo estocástico para o mesmo problema.

[43]

Neste artigo, Whiting e co-autores relatam os resultados obtidos por experimentos realizados em *Drosophila melanogaster* e *Drosophila virilis*, que basicamente indicam

que os genes destas duas espécies são quase idênticos em seqüências, mas bastante diferentes em ordem. Esta pesquisa confirma proposições, feitas em trabalhos anteriores, de que espécies do gênero *Drosophila* evoluíram provavelmente por meio de reversões (com maior freqüência) e fusões (com menor freqüência).

Apêndice C

Glossário de termos de Biologia Molecular

Este apêndice apresenta alguns termos empregados em Biologia Molecular, utilizados nesta tese.

bactéria organismo constituído por uma única célula com um único cromossomo, limitada por uma membrana que a protege do ambiente externo.

cloroplasto organela celular de plantas, que tem como função realizar a fotossíntese.

eucariotos nestes organismos as células contêm um *núcleo* que guarda o material genético, envolvido por um *citoplasma*, que por sua vez é envolvido por uma *membrana* que marca os limites da célula. O citoplasma contém outras divisões, também envolvidas por membranas, as *organelas celulares*.

genes homólogos genes que têm origem evolucionária comum.

mitocôndria organela celular de organismos, com a capacidade de sintetizar todos os seus ácidos nucleicos e algumas das suas proteínas, e tem a função de produzir energia.

procariotos organismos muito simples constituídos por células que não possuem o núcleo celular, ou seja, as células são formadas por uma única divisão.

proteínas moléculas compostas por uma cadeia de moléculas mais simples, chamadas aminoácidos, responsáveis por funções vitais na vida de um organismo, como as enzimas, que catalisam reações químicas.

RNA molécula constituída de forma similar ao DNA, com certas diferenças: o açúcar é a ribose (e não desoxirribose como no DNA); a base T (timina) não é encontrada, sendo substituída pela base U (uracil) que é pareada com a base A (adenina); a estrutura espacial varia (não é dupla hélice como no DNA); existem diferentes tipos de RNA, que realizam diferentes funções (o DNA realiza basicamente uma função, a de codificar a informação genética).

vírus constituído por DNA ou RNA encapsulado por uma estrutura simétrica composta por proteínas. Vírus são diferentes de bactérias pois não apresentam reações químicas (metabolismo), mas podem interferir de forma mortal no metabolismo do organismo infetado por ele.

Bibliografia

- [1] M. Aigner and D. B. West. Sorting by insertion of leading element. *Journal of Combinatorial Theory*, 45:306–309, 1987.
- [2] V. Bafna and P. Pevzner. Genome rearrangements and sorting by reversals. In *34th Annual IEEE Symposium on Foundations of Computer Science*, pages 148–157, 1993. Versao estendida foi publicada com o mesmo titulo em *SIAM Journal of Computing*, 25(2):272-289,1996.
- [3] V. Bafna and P. Pevzner. Sorting by reversals: Genome rearrangements in plant organelles and evolutionary history of X chromosome. *Mol. Biol. and Evol.*, 12:239–246, 1995.
- [4] V. Bafna and P. Pevzner. Sorting by transpositions. In *Proc. 6th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 614–623, 1995.
- [5] P. Berman and S. Hannenhalli. Fast sorting by reversals. In *Proceedings of Combinatorial Pattern Matching - CPM'96*, 1996.
- [6] A. Caprara. Sorting by reversals is difficult. Technical report, DEIS,University of Bologna, abril 1996.
- [7] A. Caprara, G. Lancia, and S.-K. Ng. A column-generation based branch-and-bound algorithm for sorting by reversals. In *4th DIMACS International Algorithm Implementation Challenge*, 1995.
- [8] D. Christie. Sorting permutations by block-interchange. *Information Processing Letters*, 60(4):165–169, November 1996.
- [9] D. A. Christie. A $3/2$ -approximation algorithm for sorting by reversals. *Proc. 9th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 244–252, 1998.
- [10] D. A. Christie. *Genome rearrangement problems*. PhD thesis, Glasgow University, 1998.

- [11] N. G. Copeland, N. Jenkins, D. Gilbert, J. Eppig, L. Maltals, J. Miller, W. Dietrich, A. Weaver, S. Lincoln, R. Steen, L. Steen, J. Nadeau, and E. Lander. A genetic linkage map of the mouse: current applications and future prospects. *Science*, 262:57–65, 1993.
- [12] S. Even and O. Goldreich. The minimum-length generator sequence problem is NP-hard. *Journal of Algorithms*, 2:311–313, 1981.
- [13] W. H. Gates and C. H. Papadimitriou. Bounds for sorting by prefix reversals. *Discrete Mathematics*, 27:47–57, 1979.
- [14] Q.-P. Gu, S. Peng, and H. Sudbournough. Approximating algorithms for genome rearrangements. *Proc. 7th Workshop on Genome Informatics - GIW'96*, 1996.
- [15] S.A. Guyer, L.S. Heath, and J.P.C. Vergara. Subsequence and run heuristics for sorting by transpositions. In *4th DIMACS International Algorithm Implementation Challenge*, 1995.
- [16] S. Hannenhalli. Polynomial-time algorithm for computing translocation distance between genomes. In *Proc. 6th Annual Symposium of Computational Pattern Matching-CPM'95*, 1995.
- [17] S. Hannenhalli, C. Chappey, E. V. Koonin, and P. Pevzner. Genome sequence comparison and scenarios for gene rearrangements: a test case. *Genomics*, 30:299–311, 1995.
- [18] S. Hannenhalli and P. Pevzner. Towards a computational theory of genome rearrangements. In Jan van Leeuwen, editor, *Lecture Notes in Computer Science*, volume 1000, pages 184–202. Springer-Verlag, 1995.
- [19] S. Hannenhalli and P. Pevzner. Transforming cabbage into turnip (polynomial algorithm for sorting signed permutations by reversals). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, pages 178–189, 1995.
- [20] S. Hannenhalli and P. Pevzner. Transforming men into mice (polynomial algorithmic for genomic distance problem). *36th Annual IEEE Symposium on Foundations of Computer Science*, pages 581–592, 1995.
- [21] S. Hannenhalli and P. Pevzner. To cut... or not to cut (applications of comparative physical maps in molecular evolution). In *6th ACM-SIAM Symposium on Discrete Algorithms*, 1996.

- [22] M. Jerrum. The complexity of finding minimum-length generator sequences. *Theoretical Computer Science*, 36:265–289, 1985.
- [23] H. Kaplan, R. Shamir, and R. E. Tarjan. Faster and simpler algorithm for sorting signed permutations by reversals. *SODA '97*, 1997.
- [24] J. Kececioğlu and R. Ravi. Of mice and men: Evolutionary distances between genomes under translocation. In *Proc. 6th Annual ACM-SIAM Symposium on Discrete Algorithms - SODA '95*, pages 604–613, 1995.
- [25] J. Kececioğlu and D. Sankoff. Exact and approximation algorithms for the inversion distance between two permutations. In *Combinatorial Pattern Matching, Proc. 4th Annual Symposium (CPM'93) de Lecture Notes in Computer Science*, volume 684, pages 87–105, Berlin, 1993. Springer-Verlag. Versao estendida com titulo: “Exact and approximation algorithms for the inversion distance between two permutations” foi publicada em *Algorithmica*, 13: 180-210, 1995.
- [26] J. Kececioğlu and D. Sankoff. Efficient bounds for oriented chromosome inversion distance. *Lecture Notes in Computer Science*, 807:307–325, 1994.
- [27] B. Lewin. *Genes V*. Oxford University Press, 1994.
- [28] J. Meidanis and J. C. Setubal. *Introduction to Computational Molecular Biology*. PWS Publishing Co, 1997.
- [29] J. Meidanis, M.E.M.T. Walter, and Z. Dias. Distancia de reversao de cromossomos lineares com sinais. *Proc. XXIV Seminario Integrado de Software e Hardware - SEMISH'97*, pages 70–79, 1997.
- [30] J. Meidanis, M.E.M.T. Walter, and Z. Dias. Transposition distance between a permutation and its reverse. *Proc. 4th South American Workshop on String Processing - WSP'97*, pages 70–79, 1997.
- [31] J. H. Nadeau and B. A. Taylor. Lengths of chromosomal segments conserved since divergence of man and mouse. In *Proc. Natl. Acad.Sci. USA*, volume 81, pages 814–818, 1984.
- [32] J. D. Palmer and L. A. Herbon. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, 27:87–97, 1988.
- [33] J.D. Palmer, B. Osorio, and W.F. Thompson. Evolutionary significance of inversions in legume chloroplast DNAs. *Current Genetics*, 14:65–74, 1988.

- [34] P. Pevzner and M. S. Waterman. Open combinatorial problems in computational molecular biology. In *3rd Israel Symposium on Theory of Computing and Systems*, pages 158–163. IEEE Computer Society Press, 1995.
- [35] D. Sankoff. Analytical approaches to genomic evolution. *Biochimie*, 75:409–413, 1990.
- [36] D. Sankoff. Edit distance for genome comparison based on non-local operations. In *Combinatorial Pattern Matching, Proc. 3rd Annual Symposium(CPM'92)-Lecture Notes in Computer Science*, volume 644, pages 121–135, Berlin, 1992. Springer-Verlag.
- [37] D. Sankoff, R. Cedergren, and Y. Abel. Genomic divergence through gene rearrangement. In Russell F. Doolittle, editor, *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*, pages 428–438. Academic Press, 1990.
- [38] D. Sankoff, G. Leduc, N. Antoine, B. Paquin, B. F. Lang, and R. Cedergren. Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome. *Proc. Natl. Acad. Sci.*, 89:6575–6579, 1992.
- [39] E. Therman and M. Susman. *Human Chromosomes: Structure, Behavior, and Effects*. Springer-Verlag, 3rd edition, 1993.
- [40] M.E.M.T. Walter, Z. Dias, and J. Meidanis. Reversal and transposition distance of linear chromosomes. *SPIRE'98*, 1998.
- [41] M. S. Waterman. *Introduction to computational biology: maps, sequences and genomes*. Chapman and Hall, 1995.
- [42] G. A. Watterson, W. J. Ewens, and T. E. Hall. The chromosome inversion problem. *J. Theor. Biol.*, 99:1–7, 1982.
- [43] J. Whiting, M. Pliley, J. Farmer, and D. Jeffrey. In situ hybridization analysis of chromosomal homologies in drosophila melanogaster and drosophila virilis. *Genetics*, 122:99–109, 1989.