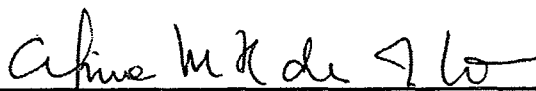


REARRANJOS DE GENOMAS: TEORIA E APLICAÇÕES

Rodrigo de Alencar Hausen

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

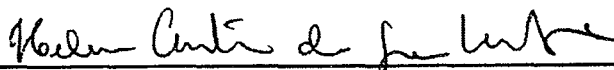
Aprovada por:



Prof. Celina Miraglia Herrera de Figueiredo, D.Sc.



Prof. Claudio Thomas Bornstein, Dr.Rer.Nat.



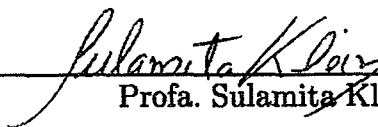
Prof. Helena Cristina da Gama Leitão, D.Sc.



Prof. Luiz Carlos Guimarães, Ph.D.



Prof. Luerbio Faria, D.Sc.



Prof. Sulamita Klein, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2007

HAUSEN, RODRIGO DE ALENCAR

Rearranjos de Genomas: Teoria e Aplicações[Rio de Janeiro] 2007

«««< .mine XII, 141 p. 29,7 cm
(COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação, 2007) ===== »»»>
.r96

XII, 141 p. 29,7 cm (COPPE/UFRJ,
D.Sc., Engenharia de Sistemas e Computação,
2007)

Tese – Universidade Federal do Rio de Janeiro, COPPE

- 1 - Rearranjos de genomas
- 2 - Ordenação por transposições
- 3 - Grafo tórico

I. COPPE/UFRJ II. Título (série)

À minha grande família.

Agradecimentos

Aos meus pais, Juvenal e Iolanda, que me deram a oportunidade de estar neste mundo e me apóiam desde os primeiros momentos, incansáveis. À minha irmã Moema e ao meu sobrinho Júlio que me alegam a vida.

Aos amigos que palmilham ao meu lado e que, espero, me desculpem por não poder citá-los todos aqui. A Janaína e Ricardo, cuja amizade já dura tantos anos que eu não me atrevo mais a contar; àqueles que me acompanharam desde o início do doutorado, e que me ouvem as piadas sem-graça desde então¹: Adriana, André, Bernardo, Cristiane, Danilo, Elias, Fabiano, Priscila, Rafael, Raphael, Raquel, Rosiane, Thatiana, Vinícius, . . . Àqueles que, mesmo quando estive longe, se mantiveram por perto através do pensamento, transmitindo força a mim e à minha família: Andréa Canella, Celina Maria, Deise Puga, Fábio Canella, João da Silva, Mônica, Sônia Regina, Vera Coutinho, Walteno, Wanderley, . . . E aos que me ajudaram no retorno: Vânia Butrucci, Fernando Ramos e o pessoal do Comunidade da UFRJ.

Àqueles que colaboraram ativamente no desenvolvimento deste trabalho e que se dispuseram a ajudar desde o início: ao Rafael Barbastefano pelo incentivo; a Luiz Carlos, com o qual tenho trabalhado desde a graduação e que tem apoiado a minha formação acadêmica; ao Luerbio e ao Luiz Antonio Kowada, que doaram mais do que tempo e atenção, mas também sugestões

¹Em ordem *estritamente* alfabética.

fundamentais para esta tese, como no capítulo 3, em que o Luerbio sugeriu a abordagem por teoria de grafos, e Kowada teve o brilhantismo de observar a estrutura do grafo tórico quando $n + 1$ é primo. Aos professores com os quais tive contato e troquei informações no desenvolvimento deste trabalho, Zanoni Dias e João Meidanis, da UNICAMP, e Maria Emília, da UNB.

Aos membros da banca que ainda não foram citados, Claudio Bornstein e Helena Cristina, que estiveram em meu exame de qualificação, além de Sulamita Klein que prontamente aceitou o convite de participar da defesa.

Para aqueles que estiveram comigo durante o período em que estive no Canadá, especialmente ao professor David Sankoff, por me receber em seu grupo de pesquisa, por me orientar no período de 09/2005 a 09/2006 e por sugerir o estudo das transposições encontradas entre os genomas humano e do chimpanzé. Aos colegas do grupo de pesquisa da Universidade de Ottawa, sempre gentis. Ao Steve Rothman, pelas instalações, e à Linda Ham e Korr Onguglo pela companhia. À Ariane, Daniel Panario e Lucia Moura, pelas inúmeras dicas de como sobreviver alhures, e pelas festas e conversas, quando o filho deste solo podia matar saudades da Pátria-mãe gentil.

Por último, mas não menos importante, à Celina, minha orientadora, que me aceitou como aluno no mestrado e agora, novamente, no doutorado. Durante o período em que estive sob a sua tutela, tive a oportunidade de receber inúmeras diretrizes, não só no que tange ao conteúdo desta tese, mas também sobre como tornar-me um pesquisador. Por ter acreditado em mim (muitas vezes mais do que eu mesmo), compreendido e ajudado-me a contornar muitas de minhas imperfeições e pelo exemplo de profissional e pessoa dedicada que é, merece aqui um lugar de destaque.

A todos esses, meus mais sinceros agradecimentos. *Thank you*. Muito obrigado.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

REARRANJOS DE GENOMAS: TEORIA E APLICAÇÕES

Rodrigo de Alencar Hausen

Dezembro/2007

Orientador: Celina Miraglia Herrera de Figueiredo

Programa: Engenharia de Sistemas e Computação

Este trabalho apresenta aspectos teóricos e práticos do problema de rearranjo de genomas, em particular para rearranjos por transposições. Na parte teórica, são demonstrados alguns resultados conhecidos baseados no diagrama de realidade e desejo (também denominado “grafo de pontos de quebra”), porém com provas mais rigorosas do que as encontradas em trabalhos anteriores de outros autores; também é abordado o problema de ordenação por transposições como um problema de teoria dos grafos, abordagem muito pouco utilizada na literatura, e para a qual são demonstrados alguns resultados novos sobre a estrutura do grafo teórico para o conjunto das permutações de n elementos, onde $n+1$ é primo. Na parte prática, é descrito um método para comparação entre os genomas humano e do chimpanzé, onde pela primeira vez são consideradas transposições. Com esse método, foram encontradas 542 reversões e 255 transposições.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

GENOME REARRANGEMENTS: THEORY AND APPLICATIONS

Rodrigo de Alencar Hausen

December/2007

Advisor: Celina Miraglia Herrera de Figueiredo

Department: Systems Engineering and Computer Science

This work presents theoretical and practical aspects of the genome rearrangement problem, in particular for rearrangements by transpositions. Regarding the theory, some known results based on the reality and desire diagram (also denominated as “breakpoint graph”) are shown, but with more scrupulous proofs than those found in previous papers by other authors; an oft-overlooked graph-theoretical approach to the problem of sorting by transpositions is also used, for which some new results are found on the structure of the toric graph for the set of permutations of n elements, where $n + 1$ is prime. On the practical side, a method to compare the human and chimpanzee genomes is described, where for the first time transpositions have been taken into consideration. By using this method, 542 reversals and 255 transpositions have been found.

Sumário

1	Introdução	1
1.1	Definições	5
1.2	Problemas de rearranjo	11
2	O diagrama de realidade e desejo	16
2.1	Um limite inferior por pontos de quebra	18
2.2	Determinação da distância de transposição vista como problema de decisão	21
2.3	O diagrama de realidade e desejo: um limite inferior mais justo	28
3	O grafo de rearranjos por transposições	49
3.1	Propriedades de $TRG(n)$	53
3.1.1	Número de caminhos mínimos entre dois vértices	67
3.2	Agrupando vértices no grafo de rearranjos por transposições	72
4	Uma aplicação: encontrando reversões e transposições entre os genomas humano e do chimpanzé	87
4.1	Formalização do problema	88
4.2	O método	94
4.3	Resultados	101

5 Conclusão	106
A Alguns parâmetros conhecidos	111
B Lista de reversões e transposições encontradas entre o ser humano e o chimpanzé	113
Referências bibliográficas	137

Lista de figuras

1.1	Cálculo da pontuação de um alinhamento entre duas cadeias de DNA.	2
1.2	Evolução entre o cloroplasto do tabaco e da <i>L. fervens</i> (adaptado de [2]).	3
2.1	Remoção de pontos de quebra por uma transposição.	21
2.2	Seqüência de transposições que ordena $\rho_{[n]}$, para n par.	30
2.3	Seqüência de transposições que ordena $\rho_{[n]}$, para n ímpar.	31
2.4	Diagramas de realidade e desejo: a) $RD(\rho_{[5]}, \iota_{[5]})$; b) $RD(\iota_{[5]}, \iota_{[5]})$ c) $RD((54321), (52143))$	33
2.5	Análise do subcaso 1.1. As linhas tracejadas representam um caminho (podendo ser uma aresta).	40
2.6	Análise do subcaso 1.2. As linhas tracejadas representam um caminho (podendo ser uma aresta).	40
2.7	Caso 2. As linhas tracejadas representam um caminho (podendo ser uma aresta).	41
2.8	Caso 3. As linhas tracejadas representam um caminho (podendo ser uma aresta).	41
2.9	Caso 4 (a) e caso 5 (b). As linhas tracejadas representam um caminho (podendo ser uma aresta).	42

2.10	diagrama de realidade e desejo $RD(\rho_{[7]}, \iota_{[7]})$	44
2.11	Análise do caso 1 do teorema 2.26.	46
2.12	Análise do caso 2 do teorema 2.26.	47
3.1	O grafo $TRG(n)$, para n entre 1 e 4.	51
3.2	O grafo $TRG(4)$ como 4 cópias de $TRG(3)$	59
3.3	Ordenações mínimas diferentes para $\rho_{[4]}$	67
3.4	Caminhos mínimos entre $\rho_{[4]}$ e $\iota_{[4]}$	68
3.5	Alterando a seqüência de transposições que transforma σ em $\rho_{[n]}$ para transformar $\iota_{[n-1]}$ em $\rho_{[n-1]}$ para o lema 3.18.	70
3.6	Efeito de uma transposição em uma permutação circular	75
3.7	Caso 1 do teorema 3.29.	79
3.8	Caso 2 do teorema 3.29.	80
3.9	Os grafos tóricos $T(3)$, $T(4)$ e $T(5)$	83
4.1	Exemplo de gráfico de pontos para um alinhamento com es- paços (<i>gaps</i>).	92
4.2	Exemplo de gráfico de pontos onde houve uma reversão na ordem dos elementos em uma das cadeias.	92
4.3	Exemplo de gráfico de pontos onde houve uma transposição na ordem dos elementos em uma das cadeias.	93
4.4	Diagrama esquemático do método adotado para identificar re- versões e transposições entre os genomas humano e do chim- panzé.	95

4.5	Agrupando segmentos alinhados em blocos (diagonais): (a) inicie com o segmento mais à esquerda; (b) adicionando segmentos ao bloco; (c) próximo alinhamento está fora da região; (d) um novo bloco é criado; (e) blocos encontrados após a execução do algoritmo; (f) identificação de reversões e transposições.	102
4.6	Distribuição dos tamanhos dos segmentos que sofreram reversão.	105
4.7	Distribuição dos tamanhos dos segmentos que sofreram transposição.	105
5.1	Uma seqüência de transposições que troca o primeiro elemento uma única vez.	109

Capítulo 1

Introdução

Avanços tecnológicos recentes permitiram a captação de uma quantidade crescente de informação sobre a biologia molecular dos organismos. Muitos genomas, desde bactérias até o ser humano, foram mapeados extensivamente. Tal quantidade de dados fomentou o interesse no problema de comparação de genomas, ou seja, determinar o quão relacionadas estão duas espécies distintas por meio da observação dos seus conjuntos de genes. A distância entre duas espécies é comumente usada como entrada em algoritmos de construção de árvores filogenéticas, úteis para o estudo da evolução das espécies.

Há duas abordagens principais para determinação de distância entre dois diferentes organismos pela comparação dos seus dados genéticos: alinhamento de cadeias e rearranjo de genomas.

A abordagem de alinhamento de cadeias consiste em determinar as diferenças e similaridades entre duas ou mais cadeias, atribuindo uma pontuação (*score*) para cada alinhamento ou não-alinhamento que ocorre (figura 1.1). A distância entre as duas cadeias é determinada ao se encontrar o alinhamento que possui pontuação total máxima, calculado sobre as cadeias de bases de DNA – no alfabeto {a, t, c, g} – ou cadeias de aminoácidos – no alfabeto

Seqüências alinhadas:

```
agactacttac
cga---gacgt
```

Matriz de similaridade S e penalidade de espaço d :

$$\begin{array}{ccccc}
 & \mathbf{a} & \mathbf{g} & \mathbf{c} & \mathbf{t} \\
 \mathbf{a} & 10 & -1 & -3 & -4 \\
 \mathbf{g} & -1 & 7 & -5 & -3 \\
 \mathbf{c} & -3 & -5 & 9 & 0 \\
 \mathbf{t} & -4 & -3 & 0 & 8
 \end{array}
 \quad d = -5$$

Pontuação:

$$\begin{aligned}
 S(a,c) + S(g,g) + S(a,a) + 3 \cdot d + S(g,g) + S(t,a) + S(t,c) + S(a,g) + S(c,t) = \\
 -3 + 7 + 10 + 3 \cdot (-5) + 7 + (-4) + 0 + (-1) + 0 = \\
 = 1
 \end{aligned}$$

Figura 1.1: Cálculo da pontuação de um alinhamento entre duas cadeias de DNA.

{A, V, L, I, P, M, F, W, G, S, T, C, N, Q, Y, D, E, K, R, H}. Essa abordagem é extremamente sensível a mutações pontuais entre as cadeias, e pode falhar na identificação de mutações grandes. Para maiores detalhes sobre esta abordagem, sugerimos a consulta ao capítulo 3 do livro de introdução à biologia molecular computacional de Setubal e Meidanis [21].

Na abordagem por rearranjo de genomas estamos interessados em determinar a distância entre dois genomas diferentes através da aplicação de mutações ao primeiro genoma de modo que, ao final, o segundo genoma seja encontrado. Esta abordagem difere da anterior no sentido em que não estamos mais comparando “letras” individuais, mas um conjunto de genes. A distância entre dois genomas é, então, o número mínimo de mutações que transforma um genoma em outro [2, 21] (fig. 1.2).

Apesar de alguns bons resultados haverem surgido no curso dos últimos anos para rearranjos por reversões [12] e até mesmo para casos em que são admitidos mais de um tipo de mutação, como reversões e intercâmbio de

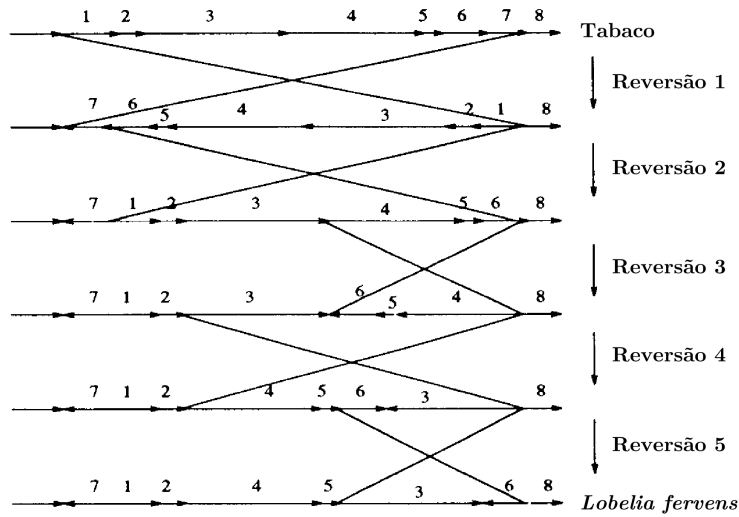


Figura 1.2: Evolução entre o cloroplasto do tabaco e da *L. fervens* (adaptado de [2]).

blocos [27], alguns eventos evolucionários ainda são difíceis de se lidar, no que tange à complexidade dos algoritmos atuais. Um desses eventos é a transposição.

Uma transposição é um deslocamento de um bloco contíguo em um genoma. Pode-se encontrar a distância de transposição, ou seja, a menor seqüência de transposições que transforma um genoma A em um genoma B, de maneira eficiente (em tempo polinomial)? Esta é uma questão ainda em aberto, a qual abordaremos neste texto.

Este documento está organizado da seguinte forma: definiremos a seguir os eventos de reversão, transposição e intercâmbio de blocos e os problemas de rearranjo, como cálculo da distância por reversões e por transposições; o capítulo 2 introduz algumas estruturas auxiliares para a resolução dos problemas, como o diagrama de realidade e desejo, e revisa alguns resultados conhecidos sobre limites para as distâncias; no capítulo 3 propomos o grafo

de rearranjos por transposições para reinterpretar alguns resultados da teoria usando essa estrutura; no capítulo 4, a parte mais aplicada de nosso trabalho, apresentamos os resultados obtidos durante o período em que o aluno esteve trabalhando sob a orientação do prof. David Sankoff, da Universidade de Ottawa, sobre a análise de rearranjos entre os genomas humano e do chimpanzé. Por fim, no capítulo 5, encerraremos com um sumário dos desenvolvimentos desta tese e apresentaremos prováveis desdobramentos futuros deste trabalho.

Esta tese contribui para o estudo de rearranjos de genomas tanto no aspecto teórico quanto no aspecto prático.

No aspecto teórico, o capítulo 2 revisa a teoria clássica de rearranjos de genomas por transposições, fornecendo provas mais detalhadas para os limites do diâmetro de transposição sem recorrer a demonstrações simplesmente através de diagramas. A crítica ao pouco formalismo das provas apresentadas na teoria clássica [1, 2, 3, 4] foi feita em pelo menos dois trabalhos [15, 17] onde uma teoria algébrica foi proposta como alternativa para obter um maior rigor. Já o capítulo 3 propõe o estudo de rearranjos por transposições como um problema de teoria de grafos. Encontramos apenas uma referência [10] que também usa teoria de grafos, através de um grafo auxiliar, denominado grafo tórico, que reduz o número de permutações. Descrevemos neste capítulo várias propriedades do problema de ordenação por transposições, já conhecidas ou novas, como propriedades de teoria de grafos. Em relação ao grafo tórico, também apresentamos novas propriedades.

No aspecto prático, o capítulo 4 propõe um método computacional usado para comparação dos genomas humano e do chimpanzé. Pela primeira vez, são consideradas transposições na comparação de genomas em larga escala. Como resultado, conseguimos identificar 542 reversões e 255 transposições.

1.1 Definições

Uma *permutação linear*, ou simplesmente uma permutação, de n elementos é uma função bijetiva π tal que

$$\begin{aligned}\pi : \{1, 2, \dots, n\} &\rightarrow \{1, 2, \dots, n\} \\ i &\mapsto \pi(i) = \pi_i,\end{aligned}$$

denotada usualmente por

$$\pi = \begin{pmatrix} 1 & 2 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix},$$

ou de forma mais sucinta,

$$\pi = (\pi_1 \pi_2 \dots \pi_i \dots \pi_j \dots \pi_n).$$

Cada elemento π_i é um inteiro no intervalo $1 \leq \pi_i \leq n$, e todos os elementos são distintos: $\pi_i \neq \pi_j$ se $i \neq j$.

Dois permutações, em particular, são importantes no estudo de rearranjos, uma por possuir todos os seus elementos em ordem crescente e outra por possuir todos em ordem decrescente.

Definição 1.1 (permutação identidade). *A permutação identidade é uma permutação que leva todos os elementos neles próprios, sendo denotada por*

$$\iota_{[n]} = (1 \ 2 \ \dots \ n).$$

Definição 1.2 (permutação reversa). *A permutação reversa é uma permutação que leva o elemento i no elemento $n - i + 1$, sendo denotada por*

$$\rho_{[n]} = (n \ n-1 \ \dots \ 2 \ 1).$$

Por brevidade, onde o valor de n estiver claramente definido, utilizaremos ι e ρ para denotar, respectivamente, $\iota_{[n]}$ e $\rho_{[n]}$.

Definição 1.3 (inversa de uma permutação). *A inversa da permutação π , denotada por π^{-1} , é uma permutação que mapeia cada elemento π_i de π em sua posição $\pi^{-1}(\pi_i) = i$, ou seja:*

$$\pi^{-1} = (\pi_1^{-1} \ \pi_2^{-1} \ \dots \ \pi_n^{-1}),$$

onde π_i^{-1} é a posição do elemento i em π .

Exemplo 1.4. *Seja $\pi = (35214)$. A sua inversa π^{-1} deve ser tal que $\pi^{-1}(3) = 1$, $\pi^{-1}(5) = 2$, $\pi^{-1}(2) = 3$, $\pi^{-1}(1) = 4$ e $\pi^{-1}(4) = 5$. Portanto, $\pi^{-1} = (43152)$.*

O produto (ou composição) entre duas permutações π e σ é uma permutação $\gamma = \sigma\pi$ onde $\gamma(i) = \sigma(\pi(i))$. Note que, em geral, $\sigma\pi$ difere de $\pi\sigma$. O produto de uma permutação pela sua inversa é a identidade: $\pi^{-1}\pi = \pi\pi^{-1} = \iota$ (neste caso específico, a operação produto é comutativa). Claramente, o produto de permutações é associativo: $\gamma(\sigma\pi) = (\gamma\sigma)\pi$.

As definições e propriedades mencionadas acima permitem-nos concluir que as permutações lineares de n elementos formam um grupo S_n , chamado de grupo simétrico, com ordem $n!$.

Um evento de mutação, ou operação de rearranjo, altera a ordem dos elementos de uma permutação de uma maneira bem definida. Entre as operações de rearranjo mais estudadas, estão as *reversões* e as *transposições*.

Definição 1.5 (reversão). [1, 13, 26] Uma reversão, denotada por $r(i, j)$, $1 \leq i \leq j \leq n$, é uma operação que inverte a ordem dos elementos de uma permutação entre as posições i e j , da forma ilustrada a seguir. Seja $\pi = (\pi_1 \pi_2 \dots \pi_i \dots \pi_j \dots \pi_n)$; a aplicação da reversão $r(i, j)$ à permutação π produz a seguinte permutação:

$$\pi \cdot r(i, j) = (\pi_1 \pi_2 \dots \pi_{i-1} \boxed{\pi_j \pi_{j-1} \dots \pi_{i+1} \pi_i} \pi_{j+1} \dots \pi_n).$$

O caso particular $r(1, n) \iota_{[n]} = \rho_{[n]}$ justifica o nome da permutação reversa.

Notamos que uma operação de reversão também pode ser vista como a seguinte permutação em S_n :

$$r(i, j) = (1 \ 2 \ \dots \ i-1 \ \boxed{j \ j-1 \ \dots \ i+1 \ i} \ j+1 \ \dots \ n).$$

Observe que aplicando-se $r(i, j)$ à permutação $\pi \cdot r(i, j)$, obtemos novamente π , ou seja, a inversa de uma dada reversão é ela própria: $\pi \cdot r(i, j) \cdot r(i, j) = \pi$ para qualquer π , portanto $r(i, j) \cdot r(i, j) = \iota$, logo $r(i, j) = (r(i, j))^{-1}$. Também é fácil ver que, dada uma permutação $\pi \in S_n$, sempre existe uma seqüência de reversões r_1, \dots, r_m tal que $\pi \cdot r_1 \cdot r_2 \cdot \dots \cdot r_m = \iota$. Diremos que uma tal seqüência *ordena* π .

Exemplo 1.6. Seja $\pi = (3 \ 2 \ 5 \ 4 \ 1)$ e considere a reversão $r(3, 5)$, aplicada à permutação π :

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

Neste caso, $r(3, 5)$ pode ser vista como a permutação $(1\ 2\ 5\ 4\ 3)$, e a aplicação de $r(3, 5)$ a π equivale ao produto:

$$\pi \cdot (1\ 2\ 5\ 4\ 3) = (3\ 2\ 1\ 4\ 5).$$

Vemos que a reversão $r(1, 3)$, quando aplicada a $(3\ 2\ 1\ 4\ 5)$, produz $(\boxed{1\ 2\ 3}\ 4\ 5)$, a permutação identidade.

Definição 1.7 (transposição). [4] Uma transposição¹, denotada por $t(i, j, k)$, onde $1 \leq i < j < k \leq n + 1$, “corta” os elementos entre as posições j e $k - 1$ (ambas inclusas) e “cola-os” imediatamente antes da posição i . Seja

$$\pi = (\pi_1 \pi_2 \dots \pi_{i-1} \pi_i \dots \pi_{j-1} \boxed{\pi_j \dots \pi_{k-1}} \pi_k \dots \pi_n),$$

então:

$$\pi \cdot t(i, j, k) = (\pi_1 \pi_2 \dots \pi_{i-1} \boxed{\pi_j \dots \pi_{k-1}} \pi_i \dots \pi_{j-1} \pi_k \dots \pi_n).$$

A transposição $t(i, j, k)$ pode ser vista como a permutação

$$t(i, j, k) = (1\ 2 \dots i-1 \boxed{j\ j+1 \dots k-1} i\ i+1 \dots j-1\ k\ k+1 \dots n).$$

Exemplo 1.8. $(\overline{8\ 7} \boxed{6\ 5} 1\ 4\ 3\ 2) \cdot t(1, 3, 5) = (\boxed{6\ 5} 8\ 7\ 1\ 4\ 3\ 2)$

A inversa de uma transposição é também uma transposição, como veremos na explicação que se segue. Seja $\pi' = \pi \cdot t(i, j, k)$, e queremos encontrar a permutação $(t(i, j, k))^{-1}$ tal que $\pi' \cdot (t(i, j, k))^{-1} = \pi$. Descrevamos cada

¹Chamamos a atenção do leitor familiarizado com Teoria de Grupos para o fato que, apesar da igualdade de nomes, o conceito de transposição usado neste texto difere do conceito clássico de transposição da Álgebra.

elemento π'_ℓ de π' , em função de i, j, k e ℓ :

$$\pi'_\ell = \begin{cases} \pi_\ell & \text{se } \ell < i \text{ ou } \ell \geq k \\ \pi_{\ell-i+j} & \text{se } i \leq \ell < i+k-j \\ \pi_{\ell-k+j} & \text{se } i+k-j \leq \ell < k, \end{cases} \quad (1.1)$$

e considere a aplicação da transposição $t(i, i+k-j, k)$ a π' , dando origem a $\pi'' = \pi' \cdot t(i, i+k-j, k)$, ou seja,

$$\pi'' = (\pi'_1 \pi'_2 \dots \pi'_{i-1} \boxed{\pi'_{i+k-j} \dots \pi'_{k-1}} \pi'_i \dots \pi'_{i+k-j-1} \pi'_k \dots \pi'_n),$$

onde cada elemento $\pi''_{\ell'}$ é tal que

$$\pi''_{\ell'} = \begin{cases} \pi'_{\ell'} & \text{se } \ell' < i \text{ ou } \ell' \geq k \\ \pi'_{\ell'+k-j} & \text{se } i \leq \ell' < j \\ \pi'_{\ell'+i-j} & \text{se } j \leq \ell' < k. \end{cases} \quad (1.2)$$

Para $\ell' < i$ ou $\ell' \geq k$, temos $\pi''_{\ell'} = \pi'_{\ell'} = \pi_{\ell'}$; no caso $i \leq \ell' < j$, obtemos que $\pi''_{\ell'} = \pi'_{\ell'+k-j}$ e, como neste caso $\ell' + k - j$ varia entre $i + k - j$ (incluso) e k (excluso), então $\pi''_{\ell'} = \pi'_{\ell'+k-j} = \pi_{(\ell'+k-j)-k+j} = \pi_{\ell'}$. Em última instância, $j \leq \ell' < k$, logo $\pi''_{\ell'} = \pi'_{\ell'+i-j} = \pi_{(\ell'+i-j)-i+j} = \pi_{\ell'}$, pois neste caso $i \leq \ell' + i - j < i + k - j$. Em todos os casos, temos que $\pi''_{\ell'} = \pi_{\ell'}$, logo $t(i, i+k-j, k)$ é a inversa de $t(i, j, k)$.

Exemplo 1.9. Considere a transposição $t(2, 5, 7)$ aplicada a $\iota_{[8]}$:

$$(\overline{1234} \boxed{56} 78) \cdot t(2, 5, 7) = (15623478).$$

A transposição $t(2, 2 + 7 - 5, 7) = t(2, 4, 7)$ é a inversa de $t(2, 5, 7)$, pois

$$(1 \overleftarrow{2 \ 3 \ 4} \boxed{5 \ 6} 7 \ 8) \cdot t(2, 5, 7) \cdot t(2, 4, 7) = (1 \overleftarrow{5 \ 6} \boxed{2 \ 3 \ 4} 7 \ 8) \cdot t(2, 4, 7) = (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8).$$

As definições das operações de reversão e transposição são motivadas por eventos de mutação observados na comparação entre os genomas de organismos distintos [11, 23]. Christie [6] define uma operação de rearranjo que não é motivada por razões biológicas, mas que é uma generalização de transposições, e cujo tratamento teórico é mais simples:

Definição 1.10 (intercâmbio de blocos). [6] *Um intercâmbio de blocos (ou block-interchange, também denominado block-swapping) com parâmetros (i, j, k, ℓ) , onde $1 \leq i < j \leq k < \ell \leq n + 1$, troca a posição dos blocos de elementos $\pi_i \dots \pi_{j-1}$ e $\pi_k \dots \pi_{\ell-1}$ em uma permutação:*

$$\pi \cdot b(i, j, k, \ell) = (\pi_1 \pi_2 \dots \pi_{i-1} \boxed{\pi_k \dots \pi_{\ell-1}} \pi_j \dots \pi_{k-1} \boxed{\pi_i \dots \pi_{j-1}} \pi_{\ell} \dots \pi_n).$$

A operação $b(i, j, k, \ell)$ pode ser descrita como a seguinte permutação:

$$(1 \ 2 \ \dots \ i-1 \ \boxed{k \ k+1 \ \dots \ \ell-1} \ j \ j+1 \ \dots \ k-1 \ \boxed{i \ i+1 \ \dots \ j-1} \ \ell \ \ell+1 \ \dots \ n).$$

Exemplo 1.11. $(8 \ 7 \ \boxed{6 \ 5} \ 1 \ \boxed{4 \ 3 \ 2}) \cdot b(3, 5, 6, 9) = (8 \ 7 \ \boxed{4 \ 3 \ 2} \ 1 \ \boxed{6 \ 5}).$

Uma transposição $t(i, j, k)$ equivale a um intercâmbio de blocos $b(i, j, j, k)$, onde dois blocos adjacentes trocam de posição. Assim como no caso das reversões e transposições, essas operações estão no grupo simétrico S_n , a inversa de um intercâmbio de blocos é um intercâmbio de blocos, e sempre existe uma seqüência de intercâmbios de blocos que ordena uma dada permutação de S_n .

Exemplo 1.12. A transposição do exemplo 1.8 pode ser descrita como o seguinte intercâmbio de blocos: $(\boxed{87}\boxed{65}1432)b(1,3,3,5) = (\boxed{65}\boxed{87}1432)$.

1.2 Problemas de rearranjo

Dadas duas permutações $\pi, \sigma \in S_n$, vê-se que sempre é possível transformar a primeira na segunda usando-se, no máximo, $n - 1$ transposições, como descrito a seguir: para cada elemento $\sigma_x = \sigma_1, \dots, \sigma_n$ que está na posição $\sigma^{-1}(\sigma_x) = x$ na permutação σ , de tal forma que a sua posição na permutação π seja diferente de sua posição em σ — ou seja, $\pi^{-1}(\sigma_x) \neq x$ — é possível, se $x < \pi^{-1}(\sigma_x)$ aplicar a π a transposição $t(x, \pi^{-1}(\sigma_x), \pi^{-1}(\sigma_x) + 1)$ a π , de tal forma que

$$\begin{aligned} \sigma &= (\sigma_1 \dots \sigma_x \dots \dots \dots \sigma_n) \\ \pi &= (\pi_1 \dots \pi_x \dots \boxed{\sigma_x} \dots \pi_n) \\ \pi \cdot t(x, \pi^{-1}(\sigma_x), \pi^{-1}(\sigma_x) + 1) &= (\pi_1 \dots \boxed{\sigma_x} \pi_x \dots \dots \pi_n), \end{aligned}$$

ou se $x > \pi^{-1}(\sigma_x)$ aplicar a π a transposição $t(\pi^{-1}(\sigma_x), \pi^{-1}(\sigma_x) + 1, x + 1)$, obtendo-se

$$\begin{aligned} \sigma &= (\sigma_1 \dots \dots \dots \sigma_x \dots \sigma_n) \\ \pi &= (\pi_1 \dots \sigma_x \boxed{\dots \pi_x} \dots \pi_n) \\ \pi \cdot t(\pi^{-1}(\sigma_x), \pi^{-1}(\sigma_x) + 1, x + 1) &= (\pi_1 \dots \boxed{\dots \pi_x} \sigma_x \dots \pi_n). \end{aligned}$$

Em ambos os casos a aplicação da transposição a π resulta em uma permutação que tem o elemento σ_x na x -ésima posição, que é a sua posição em σ . Vê-se facilmente que caso essa operação seja efetuada uma vez para cada elemento de σ que está em uma posição diferente em π , seguindo a ordem $\sigma_1, \sigma_2, \dots, \sigma_n$, ao cabo de $n - 1$ transposições, no pior caso, teremos trans-

formado π em σ . Esta seqüência de transposições apresentada, em muitos casos, não é ótima.

Definição 1.13 (distância de transposição). [4] A distância de transposição, $d_t(\pi, \sigma)$ entre duas permutações $\pi, \sigma \in S_n$ é o número mínimo q de transposições t_1, \dots, t_q tal que $\pi \cdot t_1 \cdot t_2 \cdot \dots \cdot t_q = \sigma$.

Como visto na discussão que precede à definição, se π e σ são permutações dos mesmos n elementos, sempre é possível encontrar uma seqüência com, no máximo, $n - 1$ transposições que transforma π em σ , logo a distância de transposição é limitada superiormente e está bem-definida.

Exemplo 1.14. Seja $\pi = (54321)$ e $\sigma = (12345)$. Existe uma seqüência de 3 transposições que transforma π em σ :

$$\begin{array}{c}
 \pi = (5 \ 4 \ 3 \ \boxed{2 \ 1}) \\
 \downarrow t(2, 4, 6) \\
 (5 \ 2 \ \boxed{1 \ 4} \ 3) \\
 \downarrow t(1, 3, 5) \\
 (1 \ 4 \ 5 \ \boxed{2 \ 3}) \\
 \downarrow t(2, 4, 6) \\
 \sigma = (1 \ 2 \ 3 \ 4 \ 5)
 \end{array}$$

Isto implica que $d_t(\pi, \sigma) \leq 3$.

A distância de transposição $d_t(\pi, \sigma)$ é nula se, e somente se, $\pi = \sigma$; é óbvio que, se não precisamos de nenhuma transposição para transformar uma permutação em outra, elas são idênticas. Por outro lado, se temos que t_1, t_2, \dots, t_q é uma seqüência mínima que transforma π em σ , a seqüência $t_q^{-1}, t_{q-1}^{-1}, \dots, t_1^{-1}$, onde a inversa t_ℓ^{-1} de uma transposição t_ℓ é também

uma transposição, transforma σ em π , e não é possível encontrar uma outra seqüência com menor número de transposições com essa característica (caso contrário, poderíamos invertê-la e transformar π em σ com menos transposições do que o mínimo, uma contradição). O leitor atento deve ter antecipado que provaremos que a distância de transposição obedece à desigualdade triangular em seguida; claramente, se temos $\pi, \gamma, \sigma \in S_n$, e tomarmos uma seqüência mínima de transposições t_1, \dots, t_q que transforma π em γ e outra seqüência mínima t'_1, \dots, t'_s que leva γ em σ , então a seqüência $t_1, \dots, t_q, t'_1, \dots, t'_s$, com $q+s$ transposições transforma π em σ (não necessariamente com um número mínimo de operações). O raciocínio deste parágrafo nos induz a concluir que a distância de transposição é uma métrica no conjunto das permutações de n elementos.

Definição 1.15 (ordenação mínima por transposições). *Dada uma permutação $\pi \in S_n$, o problema de ordenação mínima por transposições consiste em encontrar uma seqüência com comprimento mínimo de transposições t_1, \dots, t_q tal que $\pi \cdot t_1 \cdot t_2 \cdot \dots \cdot t_q = \iota$. O comprimento da menor seqüência de transposições que ordena π é denotado por $d_t(\pi)$.*

Como geralmente somente estamos interessados em encontrar uma seqüência mínima de transposições que transforma uma dada permutação na permutação identidade, a expressão *ordenação por transposições* será utilizada para designar o problema de ordenação *mínima* por transposições, salvo quando indicarmos explicitamente que a seqüência que ordena uma permutação não é mínima.

Os problemas de transformar uma permutação em outra por transposições e de ordenar uma permutação por transposições são equivalentes, como visto no corolário 1.17.

Lema 1.16. *Sejam $\gamma, \pi, \sigma \in S_n$. Então $d_t(\pi, \sigma) = d_t(\gamma\pi, \gamma\sigma)$.*

Prova. Seja t_1, \dots, t_q uma seqüência de transposições que transforma π em σ , ou seja, $\pi \cdot t_1 \cdot \dots \cdot t_q = \sigma$. Considere o produto $\gamma \cdot \pi \cdot t_1 \cdot \dots \cdot t_q$; pela associatividade do produto, temos que ele é equivalente a $\gamma \cdot (\pi \cdot t_1 \cdot \dots \cdot t_q) = \gamma \cdot \sigma$. Logo, a seqüência t_1, \dots, t_q também transforma $\gamma\pi$ em $\gamma\sigma$, portanto $d_t(\gamma\pi, \gamma\sigma) \leq d_t(\pi, \sigma)$. O argumento também é válido no sentido inverso, ou seja, que $d_t(\pi, \sigma) \leq d_t(\gamma\pi, \gamma\sigma)$. As duas desigualdades obtidas implicam que $d_t(\pi, \sigma) = d_t(\gamma\pi, \gamma\sigma)$. ■

Corolário 1.17. *Sejam $\pi, \sigma \in S_n$. Então $d_t(\pi, \sigma) = d_t(\sigma^{-1}\pi, \sigma^{-1}\sigma) = d_t(\sigma^{-1}\pi, \iota) = d_t(\sigma^{-1}\pi)$.*

Outro problema relacionado à determinação da distância de transposição é encontrar a maior distância entre dois elementos quaisquer do conjunto S_n .

Definição 1.18 (diâmetro de transposição). [4] *O diâmetro de transposição $D_t(n)$ do grupo simétrico S_n é a maior distância de transposição entre quaisquer dois elementos do grupo, ou seja:*

$$D_t(n) := \max\{d_t(\pi, \sigma); \pi, \sigma \in S_n\}.$$

Como $d_t(\pi, \sigma) = d_t(\sigma^{-1}\pi)$, podemos reescrever a definição acima como:

$$D_t(n) := \max\{d_t(\sigma); \sigma \in S_n\}.$$

As definições de distância, ordenação, e diâmetro podem ser aplicadas também a outros eventos de rearranjo, como para as reversões e para os intercâmbios de blocos, alterando-se as definições de maneira conveniente. Como exemplo, veremos como definir a distância entre duas permutações por meio de reversões.

Definição 1.19 (distância de reversão). [1, 13, 26] A distância de reversão, $d_r(\pi, \sigma)$ entre duas permutações $\pi, \sigma \in S_n$ é o número mínimo q de reversões r_1, \dots, r_q tal que $\pi \cdot r_1 \cdot \dots \cdot r_q = \sigma$.

Exemplo 1.20. Sejam $\pi = (7\ 1\ 2\ 3\ 5\ 4\ 6\ 8)$ e $\sigma = (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8)$. Existe uma seqüência de 3 reversões que transforma π em σ :

$$\begin{array}{c}
 \pi = (7\ 1\ 2\ 3\ \boxed{5\ 4}\ 6\ 8) \\
 \downarrow r(5, 6) \\
 (7\ \boxed{1\ 2\ 3\ 4\ 5\ 6}\ 8) \\
 \downarrow r(2, 7) \\
 (\boxed{7\ 6\ 5\ 4\ 3\ 2\ 1}\ 8) \\
 \downarrow r(1, 7) \\
 \sigma = (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8)
 \end{array}$$

Isto implica que $d_r(\pi, \sigma) \leq 3$.

Como esperado, $d_r(\pi, \sigma)$ é sempre bem-definida para quaisquer permutações π, σ com o mesmo número n de elementos e d_r é uma métrica no espaço S_n . Os argumentos são semelhantes aos utilizados para as transposições: sempre é possível transformar π em σ fazendo-se uma reversão que, a cada passo, leva um elemento para a sua correta posição em σ ; para provarmos que d_r é uma métrica podemos praticamente repetir o argumento visto anteriormente *ipsis litteris*, substituindo “transposição” por “reversão”.

A partir deste ponto, concentraremos o nosso foco sobre os problemas de ordenação por transposições e de determinação do diâmetro de transposição.

Capítulo 2

O diagrama de realidade e desejo

Apresentaremos, neste capítulo, estruturas auxiliares e resultados conhecidos, como limites para a distância e o diâmetro de transposição, iniciando com uma discussão do que já se conhece para os problemas de rearranjo.

Apesar de problemas de rearranjo específicos já serem estudados desde o início do século XX [23], somente em tempos recentes principiou-se a praticar uma abordagem mais formal para solucioná-los [18, 26].

As estruturas utilizadas atualmente foram primeiramente introduzidas por Nadeau e Taylor [18] e Bafna e Pevzner [1, 3], sendo que o primeiro algoritmo aproximativo polinomial a usá-las para ordenar uma permutação por reversões foi proposto por Kececioğlu e Sankoff em 1992 [13]. Hannenhalli, Bafna e Pevzner propuseram, em seguida, um algoritmo exato com complexidade temporal polinomial para um problema relacionado, de ordenação por reversões de permutações em que cada elemento possui um sinal que indica a sua orientação, e cada reversão inverte o sinal de todos os elementos afetados [12, 3]. O problema original de ordenação por reversões permaneceu em aberto até que Caprara [5] provou que determinar a distância de reversão é um problema NP-difícil.

Sobre o problema de ordenação por transposições, sabe-se muito menos que para ordenação por reversões. Desde os primeiros trabalhos formais [3] até os resultados mais recentes [9], busca-se sem sucesso uma forma de estimar de maneira exata a distância de transposição. Mesmo a estratégia de se adicionar informação sobre orientação aos elementos de uma permutação [25], como feito no problema de ordenação por reversões, não deu origem ainda a nenhum algoritmo exato. Além disso, ainda há uma grande diferença entre os limites superior e inferior demonstrados para a distância de transposição [14, 7].

Para este capítulo, seguiremos o seguinte roteiro: na seção 2.1, apresentaremos o conceito de pontos de quebra entre duas permutações e um limite inferior que pode ser obtido diretamente por meio dessa definição. Na seção 2.2, faremos uma mudança de ponto de vista, estudando o problema de decisão $d_t(\pi) \leq d$ para um inteiro fixo não-negativo d , o que nos dá um algoritmo para determinação exata para a distância de transposição, cuja complexidade é polinomial quando consideramos d como uma constante. Concluiremos o capítulo na seção 2.3 apresentando a demonstração de um limite inferior mais justo para a distância de transposição a partir do diagrama de realidade e desejo, definido nessa seção. Nossa demonstração é a mesma apresentada em [4], porém usamos um rigor maior no desenvolvimento da prova, sem usar argumentos puramente baseados em figuras. Desta forma, acreditamos ter contribuído para a compreensão das provas e sanado algumas das críticas [15, 17] ao pouco formalismo da teoria clássica.

2.1 Um limite inferior por pontos de quebra

Dadas duas permutações π e σ e uma operação de rearranjo que permitenos transformar π em σ com uma seqüência de operações, podemos estimar de maneira simples um limite inferior para o número mínimo de operações que devem ser aplicadas sucessivamente a π para transformá-la em σ .

Para o cálculo do número de operações necessárias, adicionaremos mais dois elementos às permutações que desejarmos calcular a distância, de forma a fixar os seus extremos.

Definição 2.1 (permutação estendida). *Dada $\pi = (\pi_1 \pi_2 \dots \pi_n)$ uma permutação de S_n , a sua permutação estendida é a permutação acrescida de dois elementos fixos, $\pi_0 = 0$ e $\pi_{n+1} = n + 1$, definida por $(0 \pi_1 \pi_2 \dots \pi_n n+1)$.*

Definição 2.2 (elementos subseqüentes). *Dada uma permutação π , os elementos π_i, π_j de sua permutação estendida $(0 \pi_1 \pi_2 \dots \pi_n n+1)$ são ditos subseqüentes se $j = i + 1$.*

Observa-se imediatamente que, se dois elementos π_i e π_{i+1} são subseqüentes em π , mas não em σ , então esses dois elementos necessariamente deverão tornar-se não-subseqüentes por alguma operação de uma seqüência que transforma π em σ .

Definição 2.3 (pontos de quebra, elementos adjacentes). *Dadas duas permutações π e σ , com suas correspondentes permutações estendidas, um par de elementos subseqüentes π_i, π_{i+1} na permutação estendida de π , mas não-subseqüentes na permutação estendida de σ é um ponto de quebra (do inglês, breakpoints) de π em relação a σ . Caso π_i, π_{i+1} sejam subseqüentes em π e em σ , diremos que esses elementos são adjacentes.*

Definição 2.4. Dadas duas permutações π e σ , o número $b(\pi, \sigma)$ de pontos de quebra de π em relação a σ é a quantidade de pares subseqüentes na permutação estendida de π que não são subseqüentes na permutação estendida de σ .

Exemplo 2.5. Sejam $\pi = (156234)$ e a permutação identidade $\iota = (123456)$. A permutação estendida de π e da identidade são, respectivamente

$$(0 \underline{1} \underline{5} \underline{6} \underline{2} \underline{3} \underline{4} \underline{7})$$

e

$$(0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7).$$

Os pares 1 5, 6 2 e 4 7 são os pontos de quebra de π em relação a ι , logo $b(\pi, \iota) = 3$.

Veremos que o número de pontos de quebra $b(\pi, \sigma)$ é nulo se, e somente se, $\pi = \sigma$. Com efeito, se $\pi = \sigma$, então todo par de elementos subseqüentes em π é adjacente com relação a σ , logo não há *pontos de quebra*. Por outro lado, se $b(\pi, \sigma) = 0$, os pares $0\pi_1, \pi_1\pi_2, \dots, \pi_{n-1}\pi_n, \pi_n n+1$ são todos adjacências com relação a σ ; mas, para isso, é necessário que $\pi_1 = \sigma_1, \pi_2 = \sigma_2, \dots, \pi_n = \sigma_n$, ou seja, $\pi = \sigma$.

Como temos $n + 1$ pares de elementos $0\pi_1, \pi_1\pi_2, \dots, \pi_{n-1}\pi_n, \pi_n n+1$, então $0 \leq b(\pi, \sigma) \leq n + 1$. Observe que o limite superior é atingido pela permutação reversa, com relação à permutação identidade: $b(\rho_{[n]}, \iota_{[n]}) = n + 1$. Portanto, para toda permutação $\pi \in S_n$, $b(\pi, \iota_{[n]}) \leq b(\rho_{[n]}, \iota_{[n]})$.

Podemos ver também que o número de pontos de quebra $b(\pi, \sigma)$ é igual a $b(\sigma^{-1}\pi, \iota)$, conforme o resultado a seguir.

Lema 2.6. Sejam $\gamma, \pi, \sigma \in S_n$. Então $b(\pi, \sigma) = b(\gamma\pi, \gamma\sigma)$.

Prova. Primeiramente, demonstraremos que, todo ponto de quebra de π , com relação a σ , corresponde a um ponto de quebra de $\gamma\pi$, com relação a $\gamma\sigma$, o que irá implicar a desigualdade $b(\pi, \sigma) \leq b(\gamma\pi, \gamma\sigma)$.

Seja $\pi_{\ell-1}\pi_\ell$ um ponto de quebra de π , com relação a σ , onde ℓ satisfaz $1 \leq \ell < n+1$. Como π e σ são permutações sobre o mesmo conjunto de elementos, existe um elemento σ_{m-1} tal que $\pi_{\ell-1} = \sigma_{m-1}$; Porém, $\sigma_m \neq \pi_\ell$, pois $\pi_{\ell-1}\pi_\ell$ é um ponto de quebra. Consideremos agora o par de elementos $\gamma(\pi_{\ell-1})\gamma(\pi_\ell)$, subseqüentes em $\gamma\pi$. Como $\pi_{\ell-1} = \sigma_{m-1}$, então o elemento $\gamma(\pi_{\ell-1})$ de $\gamma\pi$ é igual ao elemento $\gamma(\sigma_{m-1})$ de $\gamma\sigma$, mas $\gamma(\pi_\ell) \neq \gamma(\sigma_m)$, pois γ é bijetora e $\pi_\ell \neq \sigma_m$. O par de elementos $\gamma(\sigma_{m-1})\gamma(\sigma_m)$ não é subseqüente em $\gamma\sigma$, logo $\gamma(\pi_{\ell-1})\gamma(\pi_\ell)$ é um ponto de quebra em $\gamma\pi$ com relação a $\gamma\sigma$. Novamente usamos a característica de que γ é bijetora para concluir que cada ponto de quebra $\pi_{\ell-1}\pi_\ell$ é levado em um ponto de quebra distinto $\gamma(\pi_{\ell-1})\gamma(\pi_\ell)$, logo $b(\pi, \sigma) \leq b(\gamma\pi, \gamma\sigma)$.

A desigualdade $b(\pi, \sigma) \geq b(\gamma\pi, \gamma\sigma)$ é obtida fazendo-se $\pi' = \gamma\pi$, $\sigma' = \gamma\sigma$ e comparando-se $b(\pi', \sigma')$ e $b(\gamma^{-1}\pi', \gamma^{-1}\sigma')$, como fizemos no parágrafo anterior.

■

Corolário 2.7. *Sejam $\pi, \sigma \in S_n$. Então $b(\pi, \sigma) = b(\sigma^{-1}\pi, \sigma^{-1}\sigma) = b(\sigma^{-1}\pi, \iota)$.*

Com as definições apresentadas anteriormente, estamos prontos para estimar um primeiro limite inferior para a distância de transposição.

Teorema 2.8. [4] *Sejam π e σ duas permutações de n elementos. Então, $d_t(\pi, \sigma) \geq \lceil b(\pi, \sigma)/3 \rceil$.*

Prova. Sabemos que $b(\pi, \sigma) = 0$ se, e somente se, $\pi = \sigma$. Além disso, vemos que, se $\pi' = \pi \cdot t(i, j, k)$, para alguma transposição $t(i, j, k)$, então $b(\pi, \sigma) - b(\pi', \sigma) \leq 3$, ou seja, uma transposição diminui o número de pontos de quebra

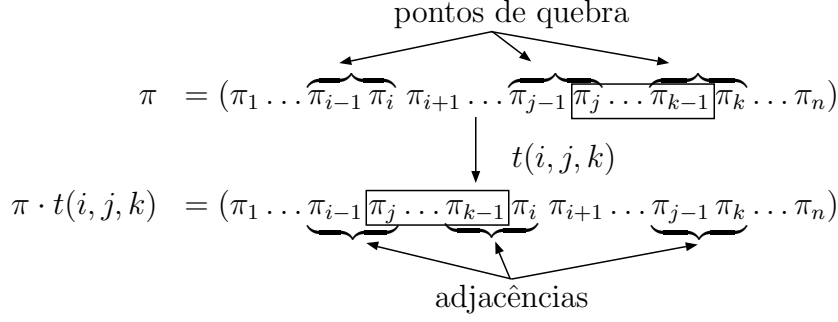


Figura 2.1: Remoção de pontos de quebra por uma transposição.

em, no máximo, 3 unidades (figura 2.1). Uma seqüência de transposições que transforma π em σ poderá, então, subtrair 3 unidades do número de pontos de quebra, na melhor das hipóteses, a cada transposição aplicada, até chegarmos a 0 pontos de quebra, donde obtemos a cota inferior $d_t(\pi, \sigma) \geq \lceil b(\pi, \sigma)/3 \rceil$.

■

Para a permutação reversa $\rho_{[n]}$, obtemos $d_t(\rho_{[n]}, \iota_{[n]}) \geq \lceil \frac{n+1}{3} \rceil$, e conseqüentemente um limite inferior para o diâmetro de transposição $D_t(n) = \max\{d_t(\sigma), \sigma \in S_n\} \geq \lceil \frac{n+1}{3} \rceil$.

A prova do teorema 2.8 usa que se $d_t(\pi, \sigma) \leq 1$, então $b(\pi, \sigma) \in \{0, 3\}$. A recíproca também é verdadeira, como veremos na seção 2.2, a seguir.

2.2 Determinação da distância de transposição vista como problema de decisão

O problema de encontrar a menor seqüência de transposições que transforma uma transposição π em σ pode ser transformado no problema de decisão equivalente, de determinar se $d_t(\pi, \sigma) \leq d$, para algum inteiro d fixo, onde $0 \leq d \leq n - 1$. Para $d = 0$, o problema reduz-se a determinar se $\pi = \sigma$,

o que pode ser feito trivialmente em tempo linear no número de elementos. Se $d = n - 1$, a resposta ao problema é sempre positiva, pois podemos tomar cada elemento de π individualmente e colocá-lo com uma transposição em sua posição correspondente em σ .

Também é fácil determinar se $d_t(\pi, \sigma) \leq 1$. Neste caso, precisamos determinar se $\pi = \sigma$ — ou, equivalentemente, que $b(\pi, \sigma) = 0$ — ou se $\sigma = \pi \cdot t(i, j, k)$, o que pode ser feito também em tempo linear, bastando-se percorrer as duas permutações e verificar se existem apenas 3 pontos de quebra em π com relação a σ , de acordo com o que nos diz o lema 2.9. O algoritmo 1 descreve o procedimento em maiores detalhes. Para deixar mais claro para o leitor qual a transposição que transforma σ em π , os parâmetros i, j e k da transposição são calculados, apesar de não serem necessários, pois apenas estamos interessados em determinar se a distância entre duas permutações é igual ou inferior a 1.

Lema 2.9. *Sejam π e σ duas permutações. Existe uma transposição $t(i, j, k)$ tal que $\pi \cdot t(i, j, k) = \sigma$ se, e somente se, $b(\pi, \sigma) = 3$.*

Prova. Se $\sigma = \pi \cdot t(i, j, k)$, então é trivial verificar que $b(\pi, \sigma) = 3$ (veja o diagrama da figura 2.1). A afirmação recíproca é verificada mais facilmente comparando-se $\sigma^{-1}\pi$ e ι . Pelo corolário 2.7, $b(\pi, \sigma) = b(\sigma^{-1}\pi, \iota) = 3$.

Seja π' definida por $\pi' := \sigma^{-1}\pi$. Existem três pares de elementos, os quais denominaremos $\pi'_{i-1}\pi'_i$, $\pi'_{j-1}\pi'_j$ e $\pi'_{k-1}\pi'_k$, tais que $\pi_{\ell-1}\pi_\ell$ são consecutivos em π , para $\ell \in \{i, j, k\}$, mas $\pi'_i \neq \pi'_{i-1} + 1$. Suponhamos, sem perda de generalidade, que $1 \leq i < j < k \leq n + 1$.

Como há apenas 3 pontos de quebra entre π' e ι , então devemos ter que:

- $\pi'_\ell = \iota_\ell = \ell$, $0 \leq \ell \leq i - 1$ ou $k \leq \ell \leq n + 1$;

Algoritmo 1 Determina se $d_t(\pi, \sigma) \leq 1$

Sejam: $\pi, \sigma \in S_n$.**Retorna:** VERDADEIRO, caso $d_t(\pi, \sigma) \leq 1$, FALSO caso contrário.

```
1:  $b \leftarrow 0$  { número de pontos de quebra }
2:  $i' \leftarrow 1$ 
3: enquanto  $i' \leq n$  faça
4:     se  $\pi_{i'} = \sigma_{i'}$  então
5:          $i' \leftarrow i' + 1$ 
6:     senão
7:          $b \leftarrow b + 1$ 
8:          $i \leftarrow i'$ 
9:          $j' \leftarrow i' + 1$ 
10:        enquanto  $\sigma_{j'} \neq \pi_{i'}$  faça
11:             $j' \leftarrow j' + 1$ 
12:             $j \leftarrow j'$ 
13:            enquanto  $\sigma_{j'} = \pi_{i'}$  faça
14:                 $j' \leftarrow j' + 1$ 
15:                 $i' \leftarrow i' + 1$ 
16:             $b \leftarrow b + 1$ 
17:             $j' \leftarrow i + 1$ 
18:            enquanto  $\sigma_{j'} = \pi_{i'}$  faça
19:                 $j' \leftarrow j' + 1$ 
20:                 $i' \leftarrow i' + 1$ 
21:             $k \leftarrow i'$  { neste ponto, se  $d_t(\pi, \sigma) = 1$  então  $\pi = \sigma \cdot t(i, j, k)$  }
22:             $b \leftarrow b + 1$ 
23: se  $b = 0$  ou  $b = 3$  então
24:     retorna VERDADEIRO
25: senão
26:     retorna FALSO
```

- $\pi'_i \neq i$, mas existe $i' \neq i$ tal que $\pi'_i = i'$ e, portanto, $\pi'_{i+1} = i' + 1$, $\pi'_{i+2} = i' + 2, \dots, \pi'_{i+(j-1-i)} = \pi'_{j-1} = i' + (j - 1 - i)$;
- $\pi'_j \neq j$, mas existe j' tal que $\pi'_j = j'$ e, portanto, $\pi'_{j+1} = j' + 1$, $\pi'_{j+2} = j' + 2, \dots, \pi'_{j+(k-1-j)} = \pi'_{k-1} = j' + (k - 1 - j)$;
- $\pi'_{k-1} \neq k - 1$.

Logo, a permutação estendida de π' possui a forma

$$(0 \ 1 \ \dots \ i-1 \ i' \ i'+1 \ \dots \ i'+(j-1-i) \ j' \ j'+1 \ \dots \ j'+(k-1-j) \ k \ k+1 \ \dots \ n \ n+1).$$

Como nas duas seqüências $i', i'+1, \dots, i'+(j-1-i)$ e $j', j'+1, \dots, j'+(k-1-j)$ temos $k - i$ elementos, cada elemento é 1 unidade maior do que o anterior, e todos os números $i, i + 1, \dots, k - 1$ devem estar presentes nessas duas seqüências, então $i' > j'$ pois, caso $i' < j'$, para que todos os números entre i e $k - 1$ estivessem em ambas as seqüências precisaríamos que i' fosse igual a i , uma contradição. Além disso, $j' = i$ e $i' = j' + (k - 1 - j) + 1$, novamente pelo fato de as duas seqüências serem crescentes e terem que conter todos os números entre i e $k - 1$. Concluimos então que a transposição $t(i, j, k)$ transforma

$$\sigma^{-1}\pi = (1 \ \dots \ i-1 \ i' \ i'+1 \ \dots \ i'+(j-1-i) \ \boxed{j' \ j'+1 \ \dots \ j'+(k-1-j)} \ k \ k+1 \ \dots \ n).$$

em

$$(1 \ \dots \ i-1 \ \boxed{j' \ j'+1 \ \dots \ j'+(k-1-j)} \ i' \ i'+1 \ \dots \ i'+(j-1-i) \ k \ k+1 \ \dots \ n).$$

que é igual a

$$\iota = (1 \dots i-1 \boxed{i \ i+1 \dots \ j-1} j \ j+1 \dots \ k-1 \ k \ k+1 \dots \ n).$$

Ou seja, $d_t(\sigma^{-1}\pi, \iota) = 1$. Mas, pelo corolário 1.17, $d_t(\pi, \sigma) = d_t(\sigma^{-1}\pi, \iota) = 1$.

■

Para decidir se duas permutações π e σ estão a uma distância fixa d , em geral, tomamos valores i, j, k tais que $1 \leq i < j < k \leq n + 1$, aplicamos a transposição $t(i, j, k)$ a π e verificamos se a permutação $\pi' = \pi \cdot t(i, j, k)$ possui $d_t(\pi', \sigma) \leq d - 1$; se a resposta for positiva, então $d_t(\pi, \sigma) \leq d$, caso contrário, escolhemos outros valores para i, j e k até encontrarmos π' tal que $d_t(\pi', \sigma) \leq d - 1$ ou até esgotarmos todas as possibilidades. O procedimento é descrito de maneira estruturada no algoritmo 2.

Determinar $d_t(\pi, \sigma) = 0$ ou $d_t(\pi, \sigma) \leq 1$ pode ser executado com uma quantidade de operações linear em n . A verificação de $d_t(\pi, \sigma) \leq 2$ requer que apliquemos todas as possíveis transposições a π , o que é feito no laço compreendido entre as linhas 12 e 17. Calculemos quantas escolhas admissíveis existem para i, j e k , de tal forma que $t(i, j, k)$ seja uma transposição válida e sem repetirmos transposições:

$$\begin{aligned} \sum_{i=1}^{n-1} \sum_{k=i+2}^{n+1} \sum_{j=i+1}^{k-1} 1 &= \sum_{i=1}^{n-1} \sum_{k=i+2}^{n+1} k - (i + 1) \\ &= \sum_{i=1}^{n-1} \left(-(i + 1)(n - i) + (i + 2 + n + 1) \frac{n - 1}{2} \right) \\ &= \sum_{i=1}^{n-1} \frac{n - i}{2} (-2i - 2 + i + 2 + n + 1) \\ &= \sum_{i=1}^{n-1} \frac{n - i}{2} (n - i + 1) \end{aligned}$$

(continua na página 27)

Algoritmo 2 Determina se $d_t(\pi, \sigma) \leq d$

Sejam: $\pi, \sigma \in S_n$, d inteiro não-negativo.

Retorna: VERDADEIRO, caso $d_t(\pi, \sigma) \leq d$, FALSO caso contrário.

Função distância(π, σ, d)

```
1: se  $d = 0$  então
2:     se  $\pi = \sigma$  então
3:         retorna VERDADEIRO
4:     senão
5:         retorna FALSO
6: se  $d = 1$  então
7:     se  $\sigma = \pi \cdot t(i, j, k)$  ou  $\pi = \sigma$  então {verificável em tempo linear}
8:         retorna VERDADEIRO
9:     senão
10:        retorna FALSO
11: se  $d > 1$  então
12:     para  $i = 1 \dots n - 1$  faça
13:         para  $j = i + 1 \dots n$  faça
14:             para  $k = j + 1 \dots n + 1$  faça
15:                  $\pi' \leftarrow \pi \cdot t(i, j, k)$ 
16:                 se distância( $\pi', \sigma, d - 1$ ) = VERDADEIRO então
17:                     retorna VERDADEIRO
18:     retorna FALSO
```

$$\begin{aligned}
&= \frac{1}{2} \sum_{i=1}^{n-1} n^2 + n - 2ni + i^2 - i \\
&= \frac{1}{2} \left((n-1)(n^2 + n) - 2n \sum_{i=1}^{n-1} i + \sum_{i=1}^{n-1} i^2 - \sum_{i=1}^{n-1} i \right) \\
&= \frac{1}{2} \left(n^3 - n^2 + n^2 - n - \underbrace{2n \frac{n^2 - n}{2}}_{-n^3 + n^2} - \frac{n^2 - n}{2} + \right. \\
&\quad \left. \underbrace{\frac{n^2 - n}{2}}_{\frac{n^2 - n}{2}} + \frac{2(n-1)^3 + 3(n-1)^2 + n - 1}{6} \right) \\
&= \frac{n^2 - n}{4} + \frac{2n^3 - 6n^2 + 6n - 2 + 3n^2 - 6n + 1 + n - 1}{12} \\
&= \frac{3n^2 - 3n}{12} + \frac{2n^3 - 3n^2 + n}{12} \\
&= \frac{n^3 - n}{6}.
\end{aligned}$$

Tal valor é usado para a análise do algoritmo de força bruta que, exaustivamente, procura transformar π em σ por transposições. Logo daremos maior ênfase a esse resultado no parágrafo a seguir.

Proposição 2.10. *Dada uma permutação $\pi \in S_n$, existem $\frac{n^3-n}{6}$ permutações π' tais que $\pi' = \pi \cdot t(i, j, k)$.*

Voltando à análise da complexidade do algoritmo 2, concluímos que, para $d = 2$, são efetuadas $\frac{n^3-n}{6}$ iterações, onde a cada iteração executamos o procedimento distância para $d = 1$, o que é feito em tempo $O(n)$. Logo, a execução do algoritmo consome, neste caso, tempo $O(n^4)$. Para valores fixos

e arbitrários de d , consumiremos um tempo $T(d)$, onde

$$\begin{aligned}
T(d) &= O\left(\frac{n^3 - n}{6}T(d-1)\right) \\
&= O\left(\frac{n^3 - n}{6}\frac{n^3 - n}{6}T(d-2)\right) \\
&= \dots \\
&= O\left(\left(\frac{n^3 - n}{6}\right)^{d-1}T(1)\right) \\
&= O(n^{3(d-1)}n) \\
&= O(n^{3d-2}).
\end{aligned}$$

Vemos assim que a complexidade do problema de determinação da distância de transposição é polinomial em n quando consideramos d uma constante. Porém, se não impusermos limites a d , como $d_t(\pi, \sigma) \geq \lceil b(\pi, \sigma)/3 \rceil$ e o número de pontos de quebra pode alcançar valores de até $n+1$, poderemos, no pior caso, depender um tempo exponencial na execução de nosso algoritmo.

2.3 O diagrama de realidade e desejo: um limite inferior mais justo

Analisemos o comportamento do limite inferior para a distância de transposição $d_t(\pi, \sigma) \geq \lceil b(\pi, \sigma)/3 \rceil$ e verifiquemos o quão próximo ele se encontra da distância real, usando a permutação reversa para essa análise.

Já se sabia desde o artigo de Bafna e Pevzner [4] que a permutação reversa $\rho_{[n]}$ pode ser ordenada com, no máximo, $\lfloor n/2 \rfloor + 1$ transposições. Meidanis, Walter e Dias [16] formalizaram o resultado de Bafna e Pevzner, exibindo uma seqüência com $\lfloor n/2 \rfloor + 1$ transposições que ordena a permutação reversa

e provaram que essa seqüência era mínima, comprovando então que vale a igualdade $d_t(\rho_{[n]}) = \lfloor n/2 \rfloor + 1$, estabelecendo o limite inferior $D_t(n) \geq \lfloor \frac{n}{2} \rfloor + 1$.

Apresentaremos, nesta seção, os resultados que permitem-nos calcular a distância de transposição exata entre a permutação reversa e a identidade, e concluiremos que o limite inferior $\lceil b(\pi, \sigma)/3 \rceil$ para a distância pode não ser muito próximo da distância real.

Proposição 2.11. [4, 16] *A permutação reversa $\rho_{[n]}$, $n \geq 3$ pode ser ordenada com $\lfloor n/2 \rfloor + 1$ transposições.*

Prova. Se $n = 2\ell$ é par, a seqüência com $\ell + 1$ transposições da figura 2.2 ordena $\rho_{[n]}$.

Para $n = 2\ell + 1$ ímpar, apresentamos uma seqüência com $\ell + 1$ transposições na figura 2.3. ■

A proposição 2.11 nos dá um limite superior $d_t(\rho_{[n]}) \leq \lfloor n/2 \rfloor + 1$ para a distância entre $\rho_{[n]}$ e $\iota_{[n]}$. O limite inferior obtido no teorema 2.8 pela análise dos pontos de quebra dá-nos $d_t(\rho_{[n]}) \geq \lceil b(\rho_{[n]}, \iota_{[n]})/3 \rceil = \lceil (n+1)/3 \rceil$. Há uma grande diferença entre o limite inferior e o superior, logo este limite inferior não é de muita valia para estimar a distância neste caso. Podemos obter limites melhores se utilizarmos uma estrutura que captura melhor a complexidade do problema.

Definição 2.12 (diagrama de realidade e desejo). [4, 21] *Dadas duas permutações $\pi, \sigma \in S_n$, o diagrama de realidade e desejo $RD(\pi, \sigma) = (V, R \cup D)$ é um multigrafo tal que:*

$$\begin{aligned} V &= \{0, -1, +1, -2, +2, \dots, -n, +n, -(n+1)\} \\ R &= \{(+\pi_i, -\pi_{i+1}) ; \pi_i \pi_{i+1} \text{ são elementos subseqüentes em } \pi\} \\ D &= \{(+\sigma_i, -\sigma_{i+1}) ; \sigma_i \sigma_{i+1} \text{ são elementos subseqüentes em } \sigma\} \end{aligned}$$

$$\begin{array}{c}
\left(n \ n-1 \overbrace{n-2 \ n-3 \ \dots \ \ell} \boxed{\ell-1 \ \dots \ 2 \ 1} \right) \\
\downarrow t(3, \ell+2, n+1) \\
\left(\overbrace{n \ n-1 \ \ell-1 \ \dots \ 2} \boxed{1 \ n-2} \ n-3 \ \dots \ \ell \right) \\
\downarrow t(1, \ell+1, \ell+3) \\
\left(1 \ \overbrace{n-2 \ n \ n-1 \ \ell-1 \ \dots} \boxed{2 \ n-3} \ \dots \ \ell \right) \\
\downarrow t(2, \ell+2, \ell+4) \\
\vdots \\
\downarrow t(\ell-1, n-1, n+1) \\
\left(1 \ 2 \ \dots \ \ell-1 \ \ell \ \dots \ n-3 \ n-2 \ \overbrace{n} \boxed{n-1} \right) \\
\downarrow t(n-1, n, n+1) \\
(1 \ 2 \ \dots \ \ell-1 \ \ell \ \dots \ n-3 \ n-2 \ n-1 \ n)
\end{array}$$

Figura 2.2: Seqüência de transposições que ordena $\rho_{[n]}$, para n par.

$$\begin{array}{c}
\left(\overbrace{n \ n-1 \ n-2 \ \dots \ \ell+1} \quad \boxed{\ell \ \dots \ 2 \ 1} \right) \\
\downarrow t(2, \ell+1, n+1) \\
\left(\overbrace{n \ \ell \ \dots \ 2} \quad \boxed{1 \ n-1} \quad n-2 \ \dots \ell+1 \right) \\
\downarrow t(1, \ell, \ell+2) \\
\left(\overbrace{1 \ n-1 \ n \ \ell \ \dots} \quad \boxed{2 \ n-2} \quad \dots \ell+1 \right) \\
\downarrow t(2, \ell+1, \ell+3) \\
\vdots \\
\downarrow t(\ell-1, n-2, n) \\
\left(1 \ 2 \ \dots \ \ell-1 \ \overbrace{\ell+1 \ \dots \ n-2 \ n-1 \ n} \quad \boxed{\ell \ \ell+1} \right) \\
\downarrow t(\ell, n-1, n+1) \\
(1 \ 2 \ \dots \ \ell-1 \ \ell \ \ell+1 \ \ell+1 \ \dots \ n-2 \ n-1 \ n)
\end{array}$$

Figura 2.3: Seqüência de transposições que ordena $\rho_{[n]}$, para n ímpar.

A figura 2.4 mostra alguns exemplos de diagramas de realidade e desejo $RD(\pi, \sigma)$, para diversas permutações π e σ . Representamos as arestas do conjunto R com linhas mais fortes. Os vértices estão dispostos segundo a ordem dos elementos em π .

Na literatura, o diagrama de realidade e desejo costuma ser apresentado também sob o nome de grafo de pontos de quebra (*breakpoint graph*), sendo esta última a denominação adotada originalmente¹ por Bafna e Pevzner [4]. Optamos por nos referir a esse multigrafo exclusivamente pelo nome de diagrama de realidade e desejo, como o fazem² Setubal e Meidanis em seu livro [21] por crer que ele descreve melhor a estrutura do grafo e a razão pela qual o definimos.

Cada aresta de $RD(\pi, \sigma)$ pertence a um dos conjuntos R ou D , aos quais nos referiremos pelos nomes “realidade” e “desejo”, respectivamente. As arestas de realidade descrevem a estrutura da permutação π , a permutação que deverá ser transformada por sucessivas transposições até chegarmos à permutação desejada σ , cuja estrutura é refletida pelas arestas de desejo. Se os conjuntos R e D contêm as mesmas arestas, então $\pi = \sigma$.

Pela definição de $RD(\pi, \sigma)$, ele é 2-regular, pois há exatamente 1 aresta de realidade e 1 aresta de desejo incidentes a cada vértice. Logo, podemos particioná-lo de maneira única em ciclos alternantes, ou seja, onde as arestas de R e D se alternam, e disjuntos em vértices entre si. Chamamos a atenção ao leitor para o fato de que, apesar da semelhança no nome, não se deve confundir a definição de ciclo no diagrama de realidade e desejo com a definição de ciclo (ou órbita) de uma permutação, encontrada na Teoria de Grupos.

¹Ressaltamos que a definição adotada por Bafna e Pevzner em [4] faz com que cada par de vértices $+i$ e $-i$ seja contraído em um só, além de serem utilizadas arestas direcionadas. Porém, as propriedades do grafo são as mesmas, como veremos mais adiante.

²Setubal e Meidanis definem o grafo no contexto do estudo de rearranjos por reversões. A estrutura do grafo, porém, é a mesma.

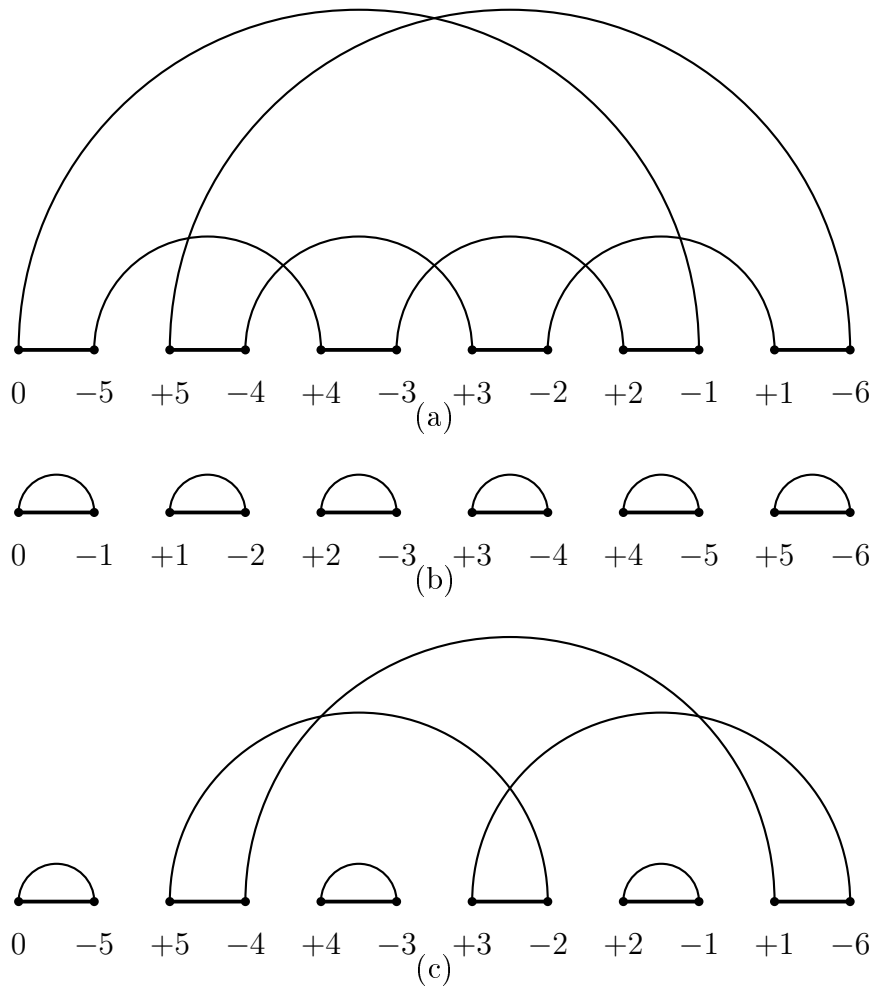


Figura 2.4: Diagramas de realidade e desejo: a) $RD(\rho_{[5]}, \iota_{[5]})$; b) $RD(\iota_{[5]}, \iota_{[5]})$
 c) $RD((54321), (52143))$

Definição 2.13 (número de ciclos). Denotaremos por $c(\pi, \sigma)$ o número de ciclos de $RD(\pi, \sigma)$.

Exemplo 2.14. Observando a figura 2.4a, vemos que $c(\rho_{[5]}, \iota_{[5]}) = 2$.

Exemplo 2.15. O número de ciclos da permutação identidade, relativamente a ela mesma, é $c(\iota_{[n]}, \iota_{[n]}) = n + 1$.

Exemplo 2.16. O número de ciclos da permutação reversa, relativamente à identidade é

$$c(\rho_n, \iota_{[n]}) = \begin{cases} 1 & \text{se } n \text{ é par,} \\ 2 & \text{se } n \text{ é ímpar.} \end{cases}$$

Proposição 2.17. Se $c(\pi, \sigma) = n + 1$, então $\pi = \sigma$.

A prova da proposição anterior é imediata. Se o número de ciclos $c(\pi, \sigma)$ for $n + 1$, os conjuntos de realidade e desejo possuem as mesmas arestas, o que somente ocorre quando $\pi = \sigma$.

Vemos que $c(\pi, \sigma)$ pode variar entre 1 – sempre há um ciclo, pois o grafo é finito e 2-regular – e $n + 1$. O processo de aplicação de uma seqüência t_1, t_2, \dots, t_q de transposições que transforma π em σ deverá, então, aumentar o número de ciclos de $c(\pi, \sigma)$ até $c(\pi t_1 t_2 \dots t_q, \sigma) = c(\sigma, \sigma) = n + 1$. Veremos a seguir que há um limite para a alteração no número de ciclos que uma transposição causa no diagrama de realidade e desejo, o que irá impor um limite inferior à distância de transposição.

Lema 2.18. [4] Sejam $\pi, \sigma \in S_n$ e $t(i, j, k)$ uma transposição. Então:

$$c(\pi \cdot t(i, j, k), \sigma) - c(\pi, \sigma) \in \{-2, 0, 2\}.$$

Em outras palavras, uma transposição altera em no máximo duas unidades o número de ciclos no diagrama de realidade e desejo.

Adiaremos a prova do lema 2.18 para apresentar um resultado importante no estudo da teoria de rearranjos por transposições, que decorre facilmente do lema 2.18, devido a Bafna e Pevzner.

Teorema 2.19. [4] *A distância de transposição entre duas permutações π e σ obedece ao seguinte limite inferior:*

$$d_t(\pi, \sigma) \geq \left\lceil \frac{n + 1 - c(\pi, \sigma)}{2} \right\rceil$$

Prova. Seja t_1, t_2, \dots, t_q uma seqüência de transposições que transforma π em σ . Pelo lema 2.18, a cada transposição aplicada a π o número de ciclos do diagrama de realidade e desejo aumenta de, no máximo, 2 unidades, ou seja, $c(\pi t_1, \sigma) \leq c(\pi, \sigma) + 2$, $c(\pi t_1 t_2, \sigma) \leq c(\pi t_1, \sigma) + 2$, \dots . Por indução, vemos que a progressão do número de ciclos se dá, então, da seguinte forma:

$$\begin{aligned} c(\pi, \sigma) & \\ c(\pi t_1, \sigma) & \leq c(\pi, \sigma) + 2 \\ c(\pi t_1 t_2, \sigma) & \leq c(\pi, \sigma) + 4 \\ & \vdots \\ c(\pi t_1 t_2 \dots t_q, \sigma) & \leq c(\pi, \sigma) + 2q \end{aligned}$$

Como devemos ter $c(\pi t_1 t_2 \dots t_q, \sigma) = n + 1$ para que $\pi = \sigma$, temos que $2q \geq n + 1 - c(\pi, \sigma)$. ■

Prova do lema 2.18. Sejam $(+\pi_{i-1}, -\pi_i)$, $(+\pi_{j-1}, -\pi_j)$, $(+\pi_{k-1}, -\pi_k)$ três arestas de realidade de $RD(\pi, \sigma) = (V, R \cup D)$. Verifiquemos a estrutura do diagrama de realidade e desejo $RD(\pi \cdot t(i, j, k), \sigma) = (V, R' \cup D')$.

Claramente, os conjuntos de arestas de desejo D e D' são idênticos, já que a ordem dos elementos de σ permanece inalterada. Relembremos como

uma transposição $t(i, j, k)$ altera uma permutação π para observarmos como R' é alterado em relação a R :

$$\begin{aligned} \pi &= (\pi_1 \dots \pi_{i-1} \pi_i \dots \pi_{j-1} \boxed{\pi_j \dots \pi_{k-1}} \pi_k \dots \pi_n) \\ &\quad \downarrow t(i, j, k) \\ \pi' = \pi \cdot t(i, j, k) &= (\pi_1 \dots \pi_{i-1} \boxed{\pi_j \dots \pi_{k-1}} \pi_i \dots \pi_{j-1} \pi_k \dots \pi_n) \end{aligned}$$

Podemos obter as seguintes relações entre os pares de elementos de π'

- $\pi_{i-1}\pi_i$, $\pi_{j-1}\pi_j$ e $\pi_{k-1}\pi_k$ deixam de ser subseqüentes;
- $\pi_{i-1}\pi_j$, $\pi_{k-1}\pi_i$ e $\pi_{j-1}\pi_k$ tornam-se subseqüentes;
- as demais relações entre os pares de elementos são mantidas em relação a π .

Deste modo, podemos descrever o conjunto de arestas de realidade R' em relação a R :

$$\begin{aligned} R' &= R \setminus \{ (+\pi_{i-1}, -\pi_i), (+\pi_{j-1}, -\pi_j), (+\pi_{k-1}, -\pi_k) \} \cup \\ &\quad \cup \{ (+\pi_{i-1}, -\pi_j), (+\pi_{k-1}, -\pi_i), (+\pi_{j-1}, -\pi_k) \}. \end{aligned}$$

Determinaremos agora de que forma os ciclos de $RD(\pi', \sigma)$ são alterados em relação a $RD(\pi, \sigma)$, dividindo a análise em casos segundo as relações entre as arestas $(+\pi_{i-1}, -\pi_i)$, $(+\pi_{j-1}, -\pi_j)$ e $(+\pi_{k-1}, -\pi_k)$:

- *Caso 1*: todas as arestas $(+\pi_{i-1}, -\pi_i)$, $(+\pi_{j-1}, -\pi_j)$ e $(+\pi_{k-1}, -\pi_k)$ estão em um único ciclo C .

- *Caso 2*: as arestas $(+\pi_{i-1}, -\pi_i)$, $(+\pi_{j-1}, -\pi_j)$ e $(+\pi_{k-1}, -\pi_k)$ estão em ciclos distintos C_1, C_2, C_3 .
- *Caso 3*: as arestas $(+\pi_{i-1}, -\pi_i)$ e $(+\pi_{j-1}, -\pi_j)$ estão em um ciclo C_1 e $(+\pi_{k-1}, -\pi_k)$ está em outro ciclo C_2 .
- *Caso 4*: a aresta $(+\pi_{i-1}, -\pi_i)$ está em um ciclo C_1 e $(+\pi_{j-1}, -\pi_j)$, $(+\pi_{k-1}, -\pi_k)$ estão em outro ciclo C_2 .
- *Caso 5*: as arestas $(+\pi_{i-1}, -\pi_i)$, $(+\pi_{k-1}, -\pi_k)$ estão em um ciclo C_1 e $(+\pi_{j-1}, -\pi_j)$ está em outro ciclo C_2 .

Notemos que, como as únicas arestas de $RD(\pi, \sigma)$ alteradas são $(+\pi_{i-1}, -\pi_i)$, $(+\pi_{j-1}, -\pi_j)$ e $(+\pi_{k-1}, -\pi_k)$, os demais ciclos que não as contêm permanecem inalterados em $RD(\pi', \sigma)$.

Análise do caso 1: Dividiremos a análise deste caso em 2 subcasos:

- *Subcaso 1.1*: O caminho alternante $-\pi_i, +\pi_{i-1}, \dots, -\pi_j, +\pi_{j-1}$ não contém a aresta $(+\pi_{k-1}, -\pi_k)$ (fig. 2.5). Neste caso, após aplicarmos a transposição $t(i, j, k)$, o subcaminho $+\pi_{i-1}, \dots, -\pi_j$ permanece inalterado, pois ele não contém nenhuma das arestas afetadas $(+\pi_{i-1}, -\pi_i)$, $(+\pi_{j-1}, -\pi_j)$ (pois as retiramos do caminho para compor o subcaminho) ou $(+\pi_{k-1}, -\pi_k)$ (por hipótese), e temos uma nova aresta $(+\pi_{i-1}, -\pi_j)$ que forma um ciclo passando por $+\pi_{i-1}$ e $-\pi_j$. Consideremos agora o caminho $+\pi_{i-1}, -\pi_i, \dots, +\pi_{k-1}, -\pi_k$; a hipótese nos permite dizer que este caminho não contém a aresta $(+\pi_{j-1}, -\pi_j)$ (caso contrário, violaríamos a hipótese deste subcaso, e teríamos assim uma situação como a da figura 2.6, relativa ao subcaso 1.2), logo o subcaminho $-\pi_i, \dots, +\pi_{k-1}$ não contém nenhuma aresta afetada pela transposição e a inclusão da aresta $(+\pi_{k-1}, -\pi_i)$ forma outro ciclo. De

maneira semelhante, temos que $+\pi_{j-1}, \dots, -\pi_k, +\pi_{j-1}$ é um ciclo de $RD(\pi', \sigma)$. Logo, $c(\pi \cdot t(i, j, k), \sigma) = c(\pi, \sigma) - 1 + 3$, o que implica $c(\pi \cdot t(i, j, k), \sigma) - c(\pi, \sigma) = +2$.

- *Subcaso 1.2:* O caminho alternante $-\pi_i, +\pi_{i-1}, \dots, -\pi_j, +\pi_{j-1}$ contém a aresta $(+\pi_{k-1}, -\pi_k)$ (fig. 2.6). Neste caso, o caminho $+\pi_{i-1}, -\pi_i, \dots, +\pi_{k-1}, -\pi_k$ deve conter a aresta $(+\pi_{j-1}, -\pi_j)$ e o caminho $-\pi_j, +\pi_{j-1}, \dots, -\pi_k, +\pi_{k-1}$ deve conter a aresta $(+\pi_{i-1}, -\pi_i)$. Observemos então os seguintes caminhos:

- (i) $+\pi_{i-1}, -\pi_i, \dots, +\pi_{j-1}, -\pi_j$, que não contém $(+\pi_{k-1}, -\pi_k)$;
- (ii) $-\pi_i, +\pi_{i-1}, \dots, -\pi_k, +\pi_{k-1}$, que não contém $(+\pi_{j-1}, -\pi_j)$;
- (iii) $+\pi_{j-1}, -\pi_j, \dots, +\pi_{k-1}, -\pi_k$, que não contém $(+\pi_{i-1}, -\pi_i)$.

Os respectivos subcaminhos $-\pi_i, \dots, +\pi_{j-1}$, $+\pi_{i-1}, \dots, -\pi_k, -\pi_j, \dots, +\pi_{k-1}$ não contêm nenhuma aresta afetada pela transposição, logo $+\pi_{i-1}, -\pi_j, \dots, +\pi_{k-1}, -\pi_i, \dots, +\pi_{j-1}, -\pi_k, +\pi_{i-1}$ é um ciclo em $RD(\pi', \sigma)$. Logo, $c(\pi', \sigma) = c(\pi, \sigma)$, portanto $c(\pi \cdot t(i, j, k), \sigma) - c(\pi, \sigma) = 0$.

Análise do caso 2 (fig. 2.7): Neste caso, cada aresta está em um ciclo distinto (este caso é o “contrário” do subcaso 1.1). A transposição $t(i, j, k)$ transforma 3 ciclos em um, logo $c(\pi \cdot t(i, j, k), \sigma) - c(\pi, \sigma) = -2$.

Análise do caso 3 (fig. 2.8): Por hipótese, temos que:

- (i) o ciclo $+\pi_{i-1}, -\pi_i, \dots, +\pi_{j-1}, -\pi_j, \dots, +\pi_{i-1}$ não contém a aresta $(+\pi_{k-1}, -\pi_k)$;
- (ii) o ciclo $+\pi_{k-1}, -\pi_k, \dots, +\pi_{k-1}$ não contém a aresta $(+\pi_{i-1}, -\pi_i)$ nem a aresta $(+\pi_{i-1}, -\pi_i)$.

Os subcaminhos $-\pi_i, \dots, +\pi_{j-1}, -\pi_j, \dots, +\pi_{i-1}$ (obtidos a partir do ciclo i) e $-\pi_k, \dots, +\pi_{k-1}$ (obtido a partir do ciclo ii) não contêm nenhuma das arestas afetadas pela transposição. A transposição $t(i, j, k)$ desfaz, então, os ciclos i e ii e cria dois novos ciclos $+\pi_{k-1}, -\pi_i, \dots, +\pi_{j-1}, -\pi_k, \dots, +\pi_{k-1}$ e $+\pi_{i-1}, -\pi_k, \dots, +\pi_{i-1}$, logo $c(\pi \cdot t(i, j, k), \sigma) = c(\pi, \sigma) - 2 + 2$, implicando que $c(\pi \cdot t(i, j, k), \sigma) - c(\pi, \sigma) = 0$.

Análise dos casos 4 e 5: estes casos são semelhantes ao caso 3 (ver figura 2.9). ■

O resultado do lema 2.18 pode ser melhorado para fornecer um limite inferior maior para a distância de transposição.

Definição 2.20 (comprimento de um ciclo). *Um ciclo de comprimento k , ou um k -ciclo, no diagrama de realidade e desejo $RD(\pi, \sigma)$ é um ciclo que possui k arestas de realidade (ou, equivalentemente, k arestas de desejo).*

Definição 2.21 (ciclo par, ciclo ímpar). *Um ciclo ímpar (par) é um k -ciclo tal que k é ímpar (par).*

Definição 2.22 (número de ciclos ímpares). *Denotaremos por $c_{odd}(\pi, \sigma)$ o número de ciclos ímpares de $RD(\pi, \sigma)$.*

Lema 2.23. *Sejam $\pi, \sigma \in S_n$ e $t(i, j, k)$ uma transposição. Então:*

$$c_{odd}(\pi \cdot t(i, j, k), \sigma) - c_{odd}(\pi, \sigma) \in \{-2, 0, 2\}.$$

Prova. Dados um diagrama de realidade e desejo $RD(\pi, \sigma)$ e uma transposição $t(i, j, k)$, diremos que um ciclo C é alterado por $t(i, j, k)$ se pelo menos uma das arestas $(+\pi_{i-1}, -\pi_i)$, $(+\pi_{j-1}, -\pi_j)$ ou $(+\pi_{k-1}, -\pi_k)$ pertence a C . Por definição, uma transposição pode alterar, no máximo, 3 ciclos.

Na análise dos casos do lema 2.18, podemos levar em conta a paridade dos ciclos para alcançarmos o resultado enunciado neste lema. Sejam C_1, \dots, C_m

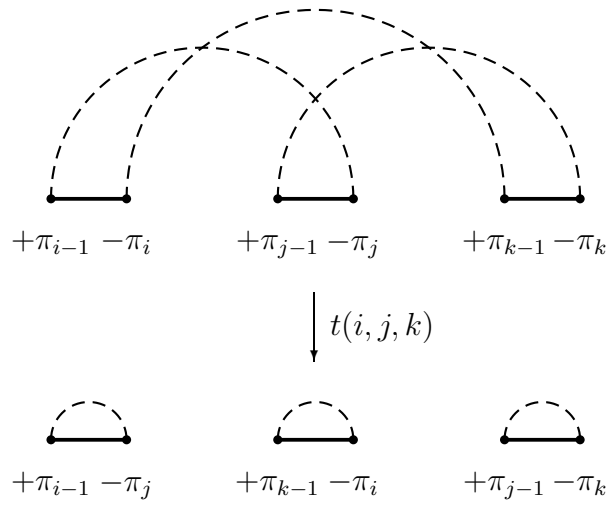


Figura 2.5: Análise do subcaso 1.1. As linhas tracejadas representam um caminho (podendo ser uma aresta).

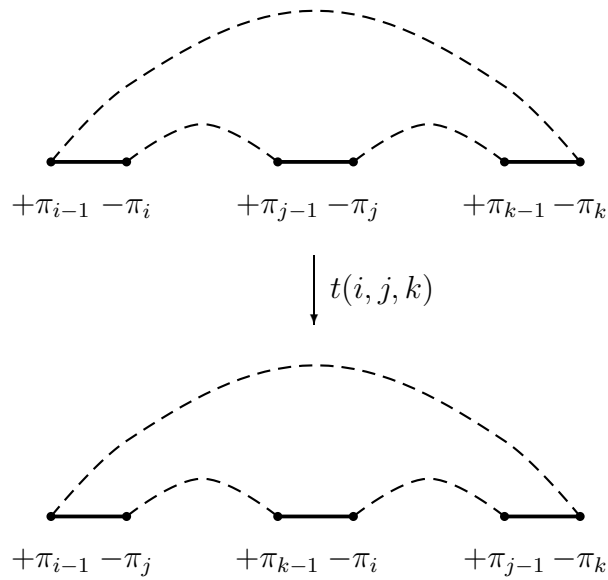


Figura 2.6: Análise do subcaso 1.2. As linhas tracejadas representam um caminho (podendo ser uma aresta).

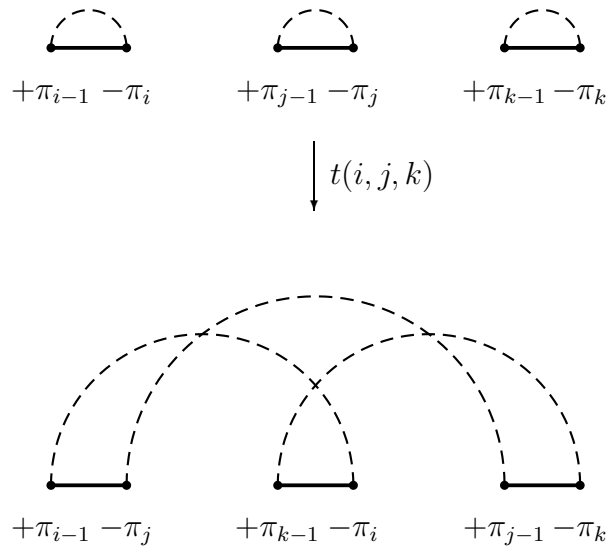


Figura 2.7: Caso 2. As linhas tracejadas representam um caminho (podendo ser uma aresta).

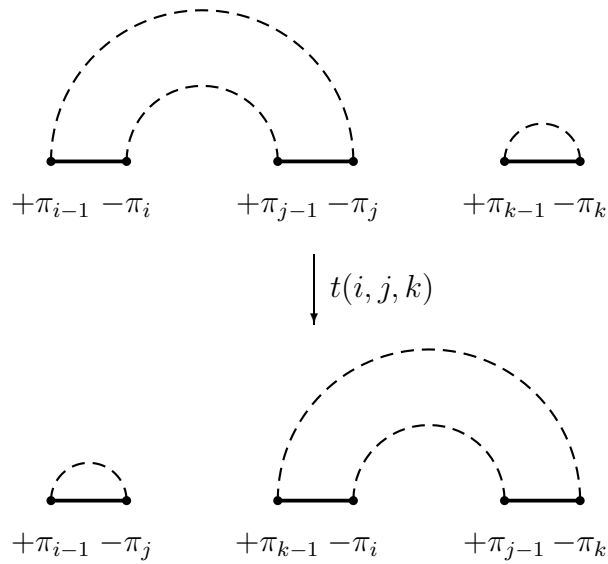


Figura 2.8: Caso 3. As linhas tracejadas representam um caminho (podendo ser uma aresta).

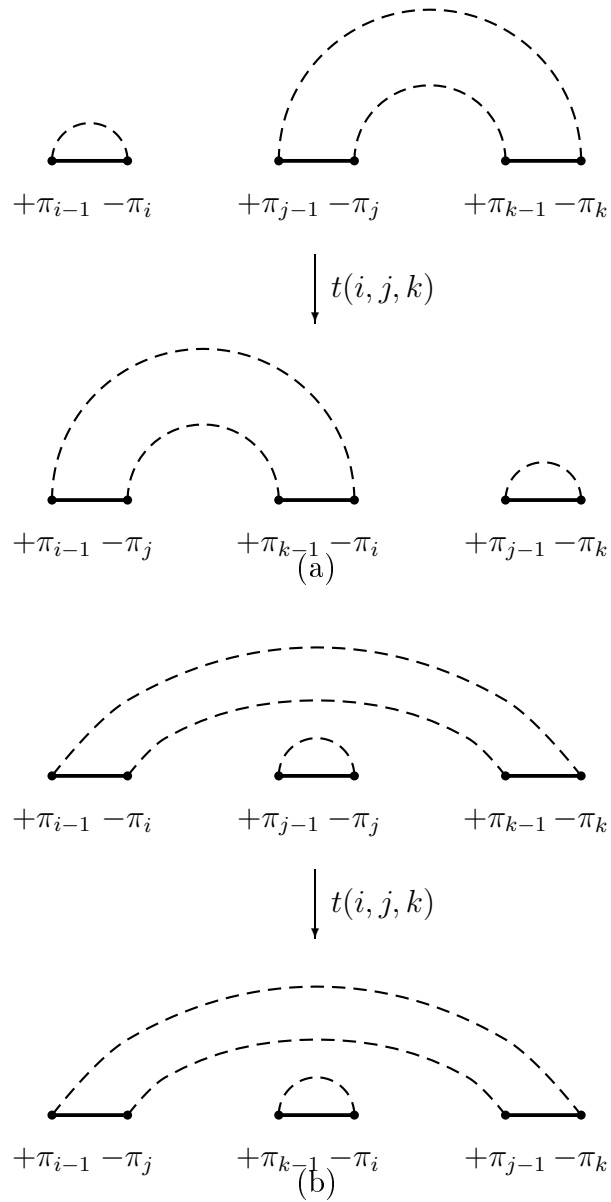


Figura 2.9: Caso 4 (a) e caso 5 (b). As linhas tracejadas representam um caminho (podendo ser uma aresta).

ciclos de $RD(\pi, \sigma)$ afetados por uma transposição $t(i, j, k)$, que são transformados em ciclos $C'_1, \dots, C'_{m'}$ de $RD(\pi \cdot t(i, j, k), \sigma)$. Como a transposição remove 3 arestas de realidade de $RD(\pi, \sigma)$ e acrescenta outras 3 arestas de realidade a $RD(\pi \cdot t(i, j, k), \sigma)$, o número total de arestas não se altera e, conseqüentemente, a soma dos comprimentos de $C'_1, \dots, C'_{m'}$ é igual à soma dos comprimentos de C_1, \dots, C_m .

Análise do subcaso 1.1: Neste caso, um k -ciclo C_1 é dividido em 3 ciclos C_1, C_2 e C_3 , com seus respectivos comprimentos k_1, k_2 e k_3 . Se k é ímpar, então, para que $k_1 + k_2 + k_3$ seja ímpar, temos que os três valores são ímpares, e portanto,

$$c_{odd}(\pi \cdot t(i, j, k), \sigma) = c_{odd}(\pi, \sigma) - 1 + 3 = c_{odd}(\pi, \sigma) + 2,$$

ou apenas um dos valores é ímpar, implicando

$$c_{odd}(\pi \cdot t(i, j, k), \sigma) = c_{odd}(\pi, \sigma) - 1 + 1 = c_{odd}(\pi, \sigma).$$

A análise dos demais casos é bem semelhante, sendo praticamente uma repetição da prova do lema 2.18 com o cuidado adicional da análise da paridade dos ciclos, motivo pelo qual pouparemos o leitor de uma prova mais extensa, e que, acreditamos, não trará maiores ganhos para a compreensão do resultado. ■

Corolário 2.24. [4] *A distância de transposição entre duas permutações π e σ obedece ao seguinte limite inferior:*

$$d_t(\pi, \sigma) \geq \left\lceil \frac{n + 1 - c_{odd}(\pi, \sigma)}{2} \right\rceil$$

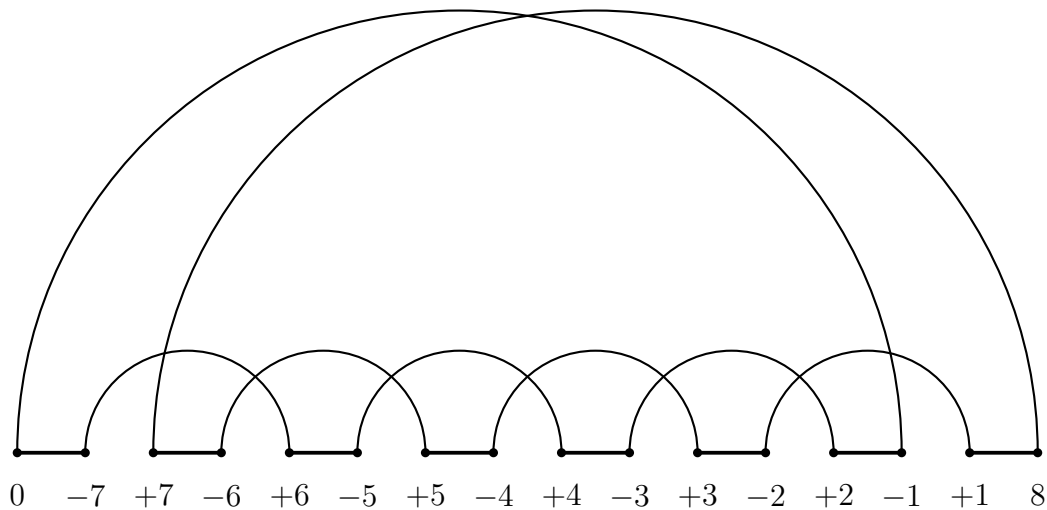


Figura 2.10: diagrama de realidade e desejo $RD(\rho_{[7]}, \iota_{[7]})$.

O limite inferior do corolário 2.24 é um pouco maior do que o obtido através do teorema 2.19, como podemos verificar no exemplo a seguir.

Exemplo 2.25. *O grafo $RD(\rho_{[7]}, \iota_{[7]})$ possui 2 ciclos pares e nenhum ciclo ímpar (fig. 2.10). Logo, $d_t(\rho_{[7]}, \iota_{[7]}) \geq \lceil \frac{7+1-0}{2} \rceil = 4$. Como a proposição 2.11 nos dá uma seqüência com 4 transposições que ordena $\rho_{[7]}$, então essa seqüência deve ser necessariamente mínima.*

Escolhemos cuidadosamente a permutação do exemplo 2.25 para demonstrar um caso em que o limite inferior obtido no corolário 2.24 dá-nos de maneira direta a distância, o que nem sempre ocorre. Por exemplo, para a permutação $\rho_{[5]}$ (fig. 2.4a), o limite inferior é $\lceil \frac{5+1-2}{2} \rceil = 2$, enquanto que o limite superior é 3. Porém, veremos adiante que nenhuma transposição que pode ser aplicada a $\rho_{[5]}$ consegue aumentar o número de ciclos ímpares no diagrama de realidade e desejo, donde concluimos que será necessária uma transposição a mais do que o limite inferior para ordenar $\rho_{[5]}$.

Teorema 2.26. [16] São necessárias ao menos $\lfloor n/2 \rfloor + 1$ transposições para ordenar a permutação $\rho_{[n]}$, $n \geq 3$.

Prova. (Esboço) Dividiremos a análise deste teorema nos casos a seguir:

- *Caso 1:* n é ímpar;
 - *Caso 1.1:* $n = 4k + 3$, para $k \in \mathbb{N}$;
 - *Caso 1.2:* $n = 4k + 1$, para $k \in \mathbb{N}$;
- *Caso 2:* n é par.

Como o número total de arestas de Desejo do diagrama de realidade e desejo é sempre $n + 1$, para o caso 1, $RD(\rho_{[n]}, \iota_{[n]})$ possuirá um número par de arestas de Desejo; já no caso 2, teremos um número ímpar de arestas.

No caso 1, um dos ciclos do diagrama de realidade e desejo é $0, -1, +2, \dots, -i, +(i + 1), \dots, +(n - 1), -n, 0$ e o outro ciclo é $+n, -(n + 1), +1, -2, \dots, +i, -(i + 1), \dots, -(n - 1), +n$; ambos possuem o mesmo tamanho (figura 2.11).

Análise do subcaso 1.1: Como o número de arestas de Desejo é $n + 1 = 4k + 4$, o comprimento de cada ciclo é $2k + 2$, ou seja, par. O limite inferior do corolário 2.24 nos dá $d_t(\rho_{[n]}, \iota_{[n]}) \geq \lceil \frac{n+1-0}{2} \rceil = \lceil \frac{4k+4}{2} \rceil = 2k + 2 = \frac{4k-2}{2} + 1 = \lfloor n/2 \rfloor + 1$.

Análise do subcaso 1.2: O número de arestas de Desejo é $n + 1 = 4k + 2$ e o comprimento de cada ciclo é $2k + 1$, ou seja, ímpar. Pelo corolário 2.24 temos $d_t(\rho_{[n]}, \iota_{[n]}) \geq \lceil \frac{n+1-2}{2} \rceil = \lceil \frac{4k}{2} \rceil = 2k = \frac{4k}{2} = \lfloor (n - 1)/2 \rfloor$. Porém, qualquer transposição que afeta um dos dois ciclos é uma transposição do mesmo tipo do subcaso 1.2 do lema 2.18, que não incrementa o número de ciclos; logo, precisaremos de 1 transposição para cada ciclo de modo a transformá-los em ciclos do mesmo tipo analisado no subcaso 1.1 do lema 2.18, para os

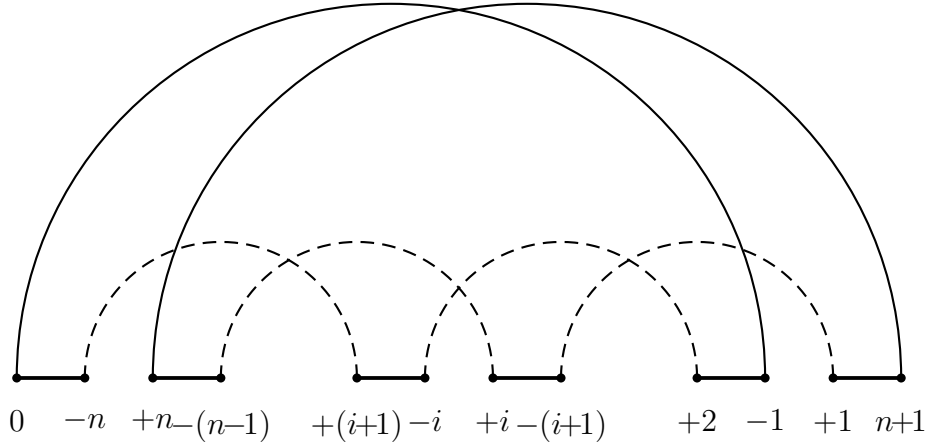


Figura 2.11: Análise do caso 1 do teorema 2.26.

quais existe transposição que aumenta o número de ciclos. Logo, teremos que $d_t(\rho_{[n]}, \iota_{[n]}) \geq \lfloor (n-1)/2 \rfloor + 2 = \lfloor n/2 \rfloor + 1$.

Análise do caso 2: Neste caso, temos apenas 1 ciclo de tamanho ímpar no diagrama de realidade e desejo: $0, -1, +2, \dots, -(n-1), +n, -(n+1), +1, -2, \dots, +(n-1), -n, 0$ (figura 2.12). Pelo corolário 2.24 temos

$$d_t(\rho_{[n]}, \iota_{[n]}) \geq \left\lceil \frac{n+1-1}{2} \right\rceil = \frac{n}{2}.$$

Podemos aplicar uma transposição que quebre esse ciclo em três ciclos ímpares, porém Meidanis, Walter e Dias [16] demonstraram que, após a primeira transposição, não há nenhuma transposição que possa ser aplicada em seguida que aumente o número de ciclos ímpares. Dessa forma, é necessário aplicar, pelo menos, uma transposição que não aumente o número de ciclos, o que nos dá um limite inferior de $\frac{n}{2} + 1$ para este caso. ■

A proposição 2.11 e o teorema 2.26 nos dão o resultado principal deste capítulo, devido a Meidanis, Walter e Dias: a distância de transposição exata entre a permutação reversa e a identidade, destacada no corolário 2.27.

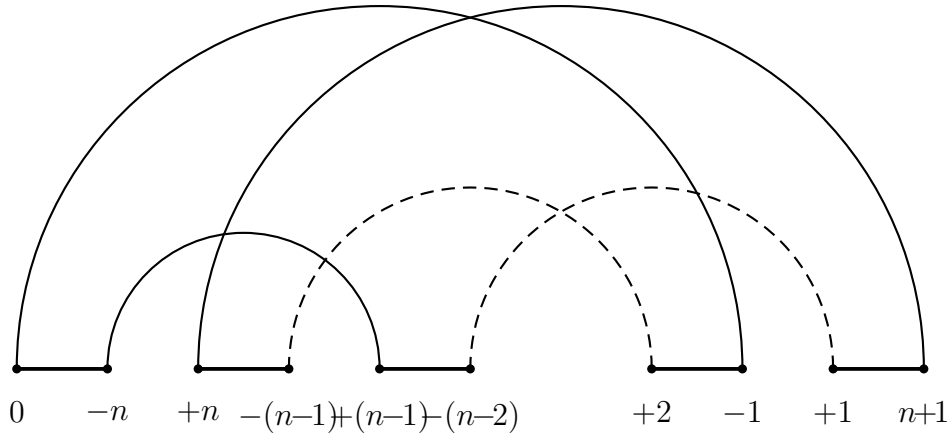


Figura 2.12: Análise do caso 2 do teorema 2.26.

Corolário 2.27. [16] *A distância de transposição entre a permutação reversa $\rho_{[n]}$ e a identidade $\iota_{[n]}$ é*

$$d_t(\rho_{[n]}, \iota_{[n]}) = \left\lfloor \frac{n}{2} \right\rfloor + 1.$$

Note que, pelo fato de sabermos qual é a distância entre a permutação reversa $\rho_{[n]}$ e a identidade, temos o seguinte limite para o diâmetro de transposição $D_t(n) = \max\{d_t(\pi), \pi \in S_n\} \geq d_t(\sigma_{[n]}) = \left\lfloor \frac{n}{2} \right\rfloor + 1$. Meidanis, Walter e Dias [16] conjecturaram que esse limite inferior era de fato igual ao diâmetro de transposição, uma vez que eles não conseguiram encontrar nenhuma permutação mais distante da identidade do que a permutação $\rho_{[n]}$ para todos os valores de n até 12, que eram os valores para os quais era possível encontrar a distância entre $\iota_{[n]}$ e todas as $n!$ permutações de S_n . Eriksson *et al.* [10] demonstraram que a conjectura não era válida, usando uma estrutura denominada grafo tórico – a qual descrevemos no capítulo 3 – para diminuir o número de permutações que deveriam ser testadas, encontrando

a permutação de 13 elementos

$$(43215131211109876),$$

que está a uma distância $8 > d_t(\rho_{[13]}, \iota_{[13]}) = 7$ da permutação identidade. Vê-se que é possível transformar (43215131211109876) na permutação identidade ordenando-se as partes 4321 e 131211109876 separadamente, usando respectivamente $\lfloor 4/2 \rfloor + 1$ e $\lfloor 8/2 \rfloor + 1$ transposições, num total de $3 + 5 = 8$ transposições. É computacionalmente viável usar o algoritmo 2 para verificar que a distância entre essa permutação de 13 elementos e a identidade é maior do que 7.

Capítulo 3

O grafo de rearranjos por transposições

No nosso estudo dos rearranjos de permutações por transposições, introduziremos um grafo que indica as relações entre as permutações de S_n segundo essa operação. Desta forma, propomos nesta tese o estudo de rearranjos por transposições como um problema de teoria dos grafos.

Definição 3.1 (grafo de rearranjos por transposições). *Denotaremos por $TRG(n)$ o grafo cujo conjunto de vértices é S_n , possuindo como arestas os pares $\pi\sigma$ para os quais $\sigma = \pi \cdot t(i, j, k)$ ou, equivalentemente, $\pi = \sigma \cdot t(i', j', k')$, onde $t(i, j, k)$ e $t(i', j', k')$ são transposições.*

Pela definição, $TRG(n)$ possui $n!$ vértices. A figura 3.1 exhibe os grafos $TRG(n)$ para valores de n até 4. Observe que cada um dos quatro grafos é regular. Representamos o grafo $TRG(3)$ de maneira planar; optamos por desenhar o grafo não-planar $TRG(4)$ em camadas, à exceção dos vértices (2413) e (3412), para facilitar a visualização da distância entre os vértices (1234) e (4321). Os vértices (2413) e (3412) foram destacados pois eles são

os únicos que não se encontram em nenhum caminho mínimo entre (1234) e (4321). Lembramos que o comprimento de um caminho é o seu número de arestas e que, dados dois vértices v e w em um grafo, a distância $d(v, w)$ é o comprimento de um caminho mínimo entre v e w . Sendo assim, a distância entre os vértices (1234) e (4321) no grafo $TRG(4)$ é $d((1234), (4321)) = 3$. Veja que em $TRG(4)$ temos $d(v, w) \leq 3$ para todo par de vértices v, w . Para permitir uma melhor compreensão, colorimos as arestas entre os vértices que são vizinhos da identidade em vermelho, as arestas entre os vizinhos da permutação reversa em verde, as arestas incidentes à permutação (2413) em magenta e em azul as arestas incidentes a (3412).

Definição 3.2 (excentricidade). *Dados $G = (V, E)$ um grafo e um vértice $v \in V$, a excentricidade de v , denotada por $ecc(v)$, é definida como*

$$ecc(v) := \max\{d(v, w); w \in V\},$$

onde $d(v, w)$ é o comprimento de um caminho mínimo entre v e w .

Definição 3.3 (diâmetro). *O diâmetro de um grafo G é o valor máximo que a excentricidade atinge para os seus vértices:*

$$D(G) := \max\{ecc(v); v \in V\}.$$

Em outras palavras, $D(G)$ é o menor inteiro que satisfaz $d(v, w) \leq D(G)$ para todo par de vértices v, w de G .

Definição 3.4 (par diametral). *Dizemos que dois vértices u, v de um grafo G formam um par diametral se o comprimento $d(u, v)$ do caminho mínimo é $D(G)$.*

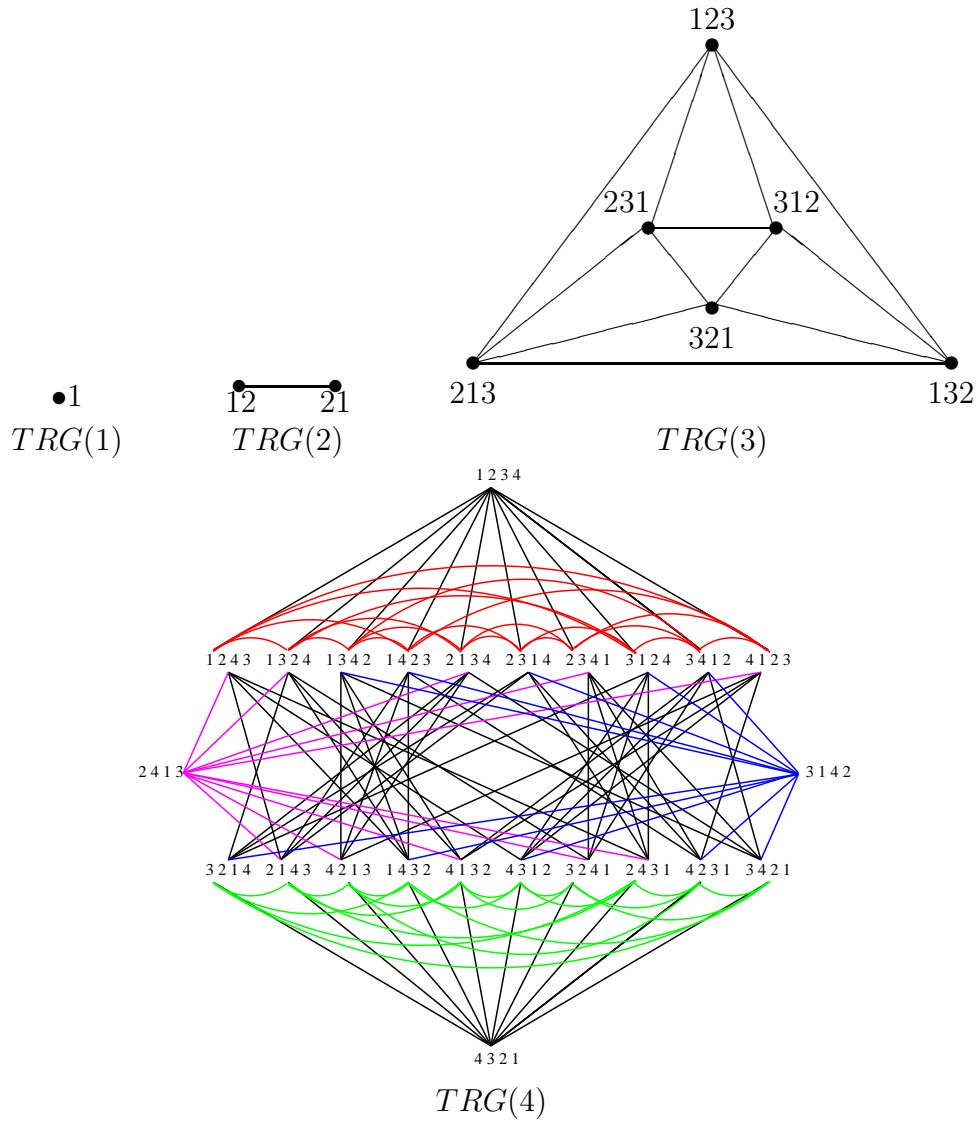


Figura 3.1: O grafo $TRG(n)$, para n entre 1 e 4.

Definição 3.5 (geodésica). *Um caminho mínimo entre dois vértices de um par diametral é denominado geodésica.*

Por definição, as geodésicas de um grafo G são os caminhos mínimos de comprimento $D(G)$.

Pela maneira como definimos $TRG(n)$, o caminho mínimo entre dois vértices π e σ desse grafo corresponde a uma seqüência mínima de transposições que transforma a permutação π em σ ; mais ainda, há uma equivalência entre o diâmetro $D(TRG(n))$ e o diâmetro de transposição $D_t(n)$ para S_n . Portanto, valem as seguintes identidades entre as definições estabelecidas nos capítulos 1 e 2, comparadas com as do capítulo 3. Sejam $\pi, \sigma \in S_n$, respectivamente correspondentes aos vértices π, σ de $TRG(n)$. Temos $d_t(\pi, \sigma) = d(\pi, \sigma)$ e $D_t(n) = D(TRG(n))$.

Seguiremos o seguinte roteiro neste capítulo: primeiramente, estabeleceremos uma série de propriedades de $TRG(n)$ que nos permitirão caracterizar alguns caminhos mínimos e, conseqüentemente, seqüências de transposições que ordenam uma determinada permutação, como por exemplo, aquela que denominamos “princípio de atração dos $TRGs$ ”. Em seguida, apresentaremos uma simplificação do $TRG(n)$, proposta por Eriksson *et al.* [10] e extrairemos propriedades não mencionadas no artigo original, como por exemplo o número exato de vértices do grafo simplificado em alguns casos. Acreditamos que essas propriedades podem ser úteis em futuros desenvolvimentos, como discutiremos na conclusão da tese.

3.1 Propriedades de $TRG(n)$

Iniciaremos esta seção com uma pequena discussão informal sobre propriedades que podemos observar nos grafos da figura 3.1, após o que daremos definições que nos permitirão formalizar esses resultados.

Nota-se uma certa simetria em todos os grafos da figura 3.1. No $TRG(3)$ sempre é possível encontrar um caminho mínimo de comprimento 2 — o diâmetro de $TRG(3)$ — a partir de qualquer vértice; o caminho mínimo (123), (132), (321), por exemplo, pode ser transformado no caminho mínimo (213), (321), (312) através de uma “rotação”. Podemos dizer algo similar sobre o $TRG(4)$, cujo diâmetro é 3. Por exemplo, a geodésica (1234), (1243), (1432), (4321) pode ser transformada na geodésica (2413), (2431), (2314), (3142). Através de uma rotulação conveniente para os vértices, podemos transformar uma geodésica a partir de (1234) em uma geodésica a partir de qualquer outro vértice, o que nos permite dizer que qualquer vértice de $TRG(4)$ pertence a um par diametral.

As simetrias descritas no parágrafo anterior não se dão por acaso, mas por propriedades do TRG que passaremos a formalizar a partir deste ponto.

Definição 3.6 (automorfismo). *Seja $G = (V, E)$ um grafo e Φ uma função bijetiva no conjunto de vértices de G :*

$$\begin{aligned}\Phi : V &\rightarrow V \\ v &\mapsto \Phi(v) = w.\end{aligned}$$

Diremos que a função Φ é um automorfismo se preserva as relações de adjacência entre os vértices, ou seja, existe uma aresta entre os vértices u e v se, e somente se, os vértices $\Phi(u)$ e $\Phi(v)$ são adjacentes.

Definição 3.7 (grafo vértice-transitivo). *Um grafo G é vértice-transitivo se, para qualquer par de vértices v e w de G , existe um automorfismo Φ tal que $\Phi(v) = w$.*

Dados quaisquer dois vértices π, σ , podemos definir um automorfismo $\Phi_{\pi, \sigma}$ em $TRG(n)$ que transforma o vértice π no vértice σ :

$$\begin{aligned} \Phi_{\pi, \sigma} : V(TRG(n)) &\rightarrow V(TRG(n)) \\ \gamma &\mapsto \Phi_{\pi, \sigma}(\gamma) = \sigma\pi^{-1}\gamma. \end{aligned}$$

Verifica-se imediatamente que $\Phi_{\pi, \sigma}(\pi) = \sigma$. Se $\pi\pi'$ é uma aresta de $TRG(n)$, então $1 = d(\pi, \pi') = d_t(\pi, \pi')$. Pelo lema 1.16, $d_t(\pi, \pi') = d_t((\sigma\pi^{-1})\pi, (\sigma\pi^{-1})\pi')$. Como as permutações $(\sigma\pi^{-1})\pi$ e $(\sigma\pi^{-1})\pi'$ correspondem respectivamente aos vértices $\Phi_{\pi, \sigma}(\pi)$ e $\Phi_{\pi, \sigma}(\pi')$, a distância de transposição $d_t((\sigma\pi^{-1})\pi, (\sigma\pi^{-1})\pi')$ equivale à distância $d(\Phi_{\pi, \sigma}(\pi), \Phi_{\pi, \sigma}(\pi'))$, portanto $d(\Phi_{\pi, \sigma}(\pi), \Phi_{\pi, \sigma}(\pi')) = d(\pi, \pi') = 1$ ou seja, $\Phi_{\pi, \sigma}(\pi)$ e $\Phi_{\pi, \sigma}(\pi')$ são adjacentes. Logo, o grafo $TRG(n)$ é vértice-transitivo.

Como conseqüência da transitividade dos vértices de um TRG , temos que o grafo é “visto” por cada vértice da mesma maneira. Mais formalmente, seja $\pi, \pi', \pi'', \dots, \pi^{(\ell)}$ um caminho mínimo de comprimento ℓ em um TRG e seja σ um vértice qualquer. Existe um caminho de comprimento ℓ partindo de $\sigma = \Phi_{\pi, \sigma}(\pi)$ em direção a $\Phi_{\pi, \sigma}(\pi^{(\ell)})$, a saber, $\sigma, \Phi_{\pi, \sigma}(\pi'), \Phi_{\pi, \sigma}(\pi''), \dots, \Phi_{\pi, \sigma}(\pi^{(\ell)})$; tal caminho é mínimo, pois caso houvesse um caminho $\sigma, \sigma', \sigma'', \dots, \sigma^{(m)} = \Phi_{\pi, \sigma}(\pi^{(\ell)})$ de comprimento $m < \ell$, então poderíamos transformá-lo em um caminho $\Phi^{-1}(\sigma) = \pi, \dots, \Phi^{-1}(\sigma^{(m)}) = \pi^{(\ell)}$ com comprimento menor que o do caminho mínimo entre π e $\pi^{(\ell)}$. Em particular, uma geodésica a partir de um vértice π pode ser transformada em uma geodésica a partir de qualquer outro vértice de $TRG(n)$, logo o diâmetro de $TRG(n)$ pode ser calculado a partir

da excentricidade de qualquer vértice, ou seja, todos os vértices possuem a mesma excentricidade.

Continuando o paralelo entre as definições estabelecidas nos capítulos 1 e 2, comparadas com as deste capítulo, temos que $d_t(\sigma) = d_t(\sigma, \iota) = d(\sigma, \iota)$, para σ uma permutação de S_n correspondente a um vértice σ de $TRG(n)$. O problema central, o cálculo do diâmetro de transposição $D_t(n)$ é, portanto:

$$D_t(n) = \max_{\pi \in S_n} d_t(\sigma) = ecc(\iota) = D(TRG(n)).$$

Outra consequência do fato de o grafo de rearranjos por transposições ser vértice-transitivo é a sua regularidade, isto é, todos os vértices do grafo possuem o mesmo grau. Podemos calcular este grau utilizando a proposição 2.10, que nos indica de quantas formas distintas podemos aplicar uma transposição a uma permutação π , e o teorema a seguir decorre diretamente desse resultado.

Proposição 3.8. *O grafo $TRG(n)$ é regular, com grau $\frac{n^3-n}{6}$.*

A proposição 3.8, e o fato de que $TRG(n)$ possui um número exponencial de vértices, indicam-nos que estamos tratando com grafos cuja densidade decai exponencialmente conforme n aumenta.

Observemos o conjunto de vértices $TRG(n)$ que representam as permutações na forma $(x \pi_2 \dots \pi_n)$, para algum x fixo entre 1 e n ; tal conjunto de vértices corresponde a $(n-1)!$ permutações, e o subgrafo induzido por este conjunto de vértices será denotado por $TRG_x(n)$. Podemos particionar os vértices de $TRG(n)$ em n subconjuntos disjuntos, que são respectivamente os vértices de $TRG_1(n), TRG_2(n), \dots, TRG_n(n)$, sendo que cada um corresponde a um subgrafo isomorfo a $TRG(n-1)$, como visto no teorema 3.11.

Para a prova do teorema seguinte, necessitaremos definir a operação de inserção de um elemento à esquerda em uma permutação:

Definição 3.9 (inserção de elemento à esquerda). *Seja $\pi = (\pi_1 \dots \pi_{n-1})$ uma permutação de $n - 1$ elementos e x um inteiro entre 1 e n . Definiremos $x \oplus \pi$ como sendo a permutação $(x \pi'_2 \pi'_3 \dots \pi'_n)$ onde*

$$\pi'_i = \begin{cases} \pi_{i-1} & \text{se } \pi_{i-1} < x \\ \pi_{i-1} + 1 & \text{se } \pi_{i-1} \geq x \end{cases}.$$

Ou seja, a operação de inserção de um elemento x à esquerda em uma permutação π consiste em inserir esse elemento na primeira posição e renumerar os demais elementos.

Exemplo 3.10. *A inserção do elemento 3 à esquerda da permutação (3241) resulta em $3 \oplus (3241) = (34251)$.*

Teorema 3.11. *Existem n cópias disjuntas de $TRG(n - 1)$ em $TRG(n)$.*

Prova. Sejam $\pi = (\pi_1 \dots \pi_{n-1})$ e $\sigma = (\sigma_1 \dots \sigma_{n-1})$ dois vértices distintos de $TRG(n - 1)$. Considere a permutação $x \oplus \pi$, sendo x um inteiro entre 1 e n . Veremos a seguir que $x \oplus \pi$ é um vértice de $TRG_x(n)$ e que σ, π são adjacentes se e somente se $x \oplus \sigma, x \oplus \pi$ são adjacentes.

Seja $t(i, j, k)$ uma transposição tal que $\sigma = \pi \cdot t(i, j, k)$, ou seja,

$$\sigma = (\pi_1 \dots \pi_{i-1} \pi_j \dots \pi_{k-1} \pi_i \dots \pi_{j-1} \pi_k \dots \pi_n),$$

onde o ℓ -ésimo elemento de σ é

$$\sigma_\ell = \begin{cases} \pi_\ell & \text{se } \ell < i \text{ ou } \ell \geq k \\ \pi_{\ell+j-i} & \text{se } i \leq \ell < i + k - j \\ \pi_{\ell-k+j} & \text{se } i + k - j \leq \ell < k \end{cases}.$$

Cada elemento da permutação $\sigma' = x \oplus \sigma$ é descrito da seguinte forma:

$$\sigma'_\ell = \begin{cases} x & \text{se } \ell = 1 \\ \sigma_{\ell-1} & \text{se } \sigma_{\ell-1} < x \\ \sigma_{\ell-1} + 1 & \text{se } \sigma_{\ell-1} \geq x \end{cases} .$$

Aplicando-se $t(i, j, k)$ a $\pi' = x \oplus \pi$, obtemos $\pi'' = \pi' \cdot t(i+1, j+1, k+1)$, descrita por:

$$\pi''_\ell = \begin{cases} \pi'_\ell & \text{se } \ell < i+1 \text{ ou } \ell \geq k+1 \\ \pi'_{\ell+j-i} & \text{se } i+1 \leq \ell < (i+1) + k - j \\ \pi'_{\ell-k+j} & \text{se } (i+1) + k - j \leq \ell < k+1 \end{cases} .$$

Comparemos um elemento σ'_ℓ com um elemento π''_ℓ . Se $\ell = 1$, então $\sigma'_1 = \pi''_1 = x$. Sendo $\ell > 1$, dividamos a análise em três casos:

- $\ell < i+1$ ou $\ell \geq k+1$:

$$\begin{aligned} \pi''_\ell = \pi'_\ell &= \begin{cases} \pi_{\ell-1} & \text{se } \pi_{\ell-1} < x \\ \pi_{\ell-1} + 1 & \text{se } \pi_{\ell-1} \geq x \end{cases} \\ &= \begin{cases} \sigma_{\ell-1} & \text{se } \sigma_{\ell-1} < x \\ \sigma_{\ell-1} + 1 & \text{se } \sigma_{\ell-1} \geq x \end{cases} \\ &\quad (\text{pois } \sigma_{\ell-1} = \pi_{\ell-1} \text{ para } \ell-1 < i \text{ ou } \ell-1 \geq k) \\ &= \sigma'_\ell \end{aligned}$$

- $i + 1 \leq \ell < (i + 1) + k - j$:

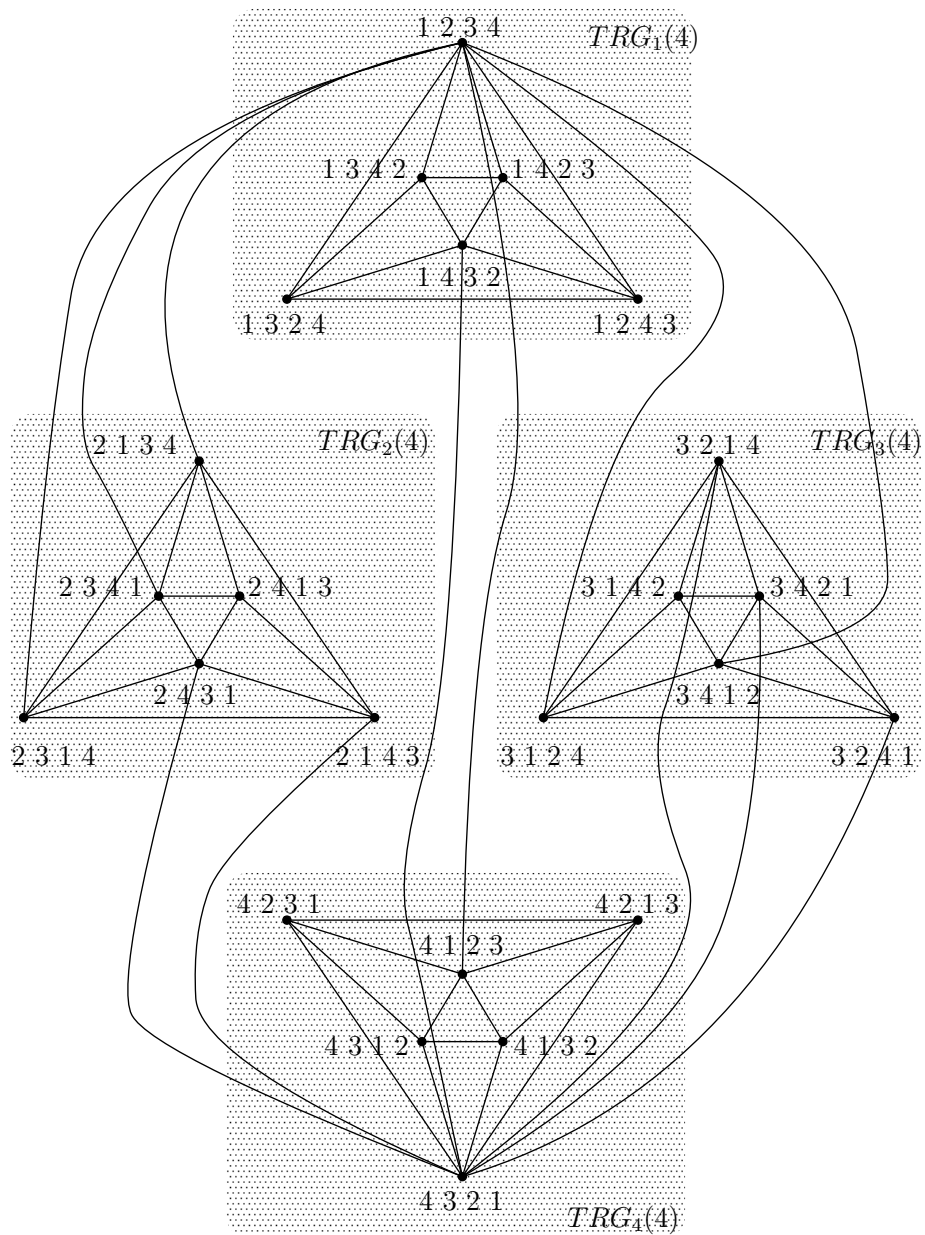
$$\begin{aligned}
\pi''_\ell = \pi'_{\ell+j-i} &= \begin{cases} \pi_{\ell+j-i-1} & \text{se } \pi_{\ell+j-1} < x \\ \pi_{\ell+j-i-1} + 1 & \text{se } \pi_{\ell+j-1} \geq x \end{cases} \\
&= \begin{cases} \sigma_{\ell-1} & \text{se } \sigma_{\ell-1} < x \\ \sigma_{\ell-1} + 1 & \text{se } \sigma_{\ell-1} \geq x \end{cases} \\
&\quad (\text{pois } \sigma_{\ell-1} = \pi_{\ell-1+j-i} \text{ para } i \leq \ell - 1 < i + k - j) \\
&= \sigma'_\ell
\end{aligned}$$

- $(i + 1) + k - j \leq \ell < k + 1$: analogamente ao caso anterior, chegamos também à conclusão que $\pi''_\ell = \sigma'_\ell$.

Logo, temos que todo e qualquer elemento de $\pi'' = (x \oplus \pi) \cdot t(i + 1, j + 1, k + 1)$ é igual a $\sigma' = x \oplus \sigma$, ou seja, existe uma transposição que transforma $x \oplus \pi$ em $x \oplus \sigma$, o que indica que estas duas permutações são adjacentes em $TRG(n)$. ■

A figura 3.2 mostra as 4 cópias de $TRG(3)$ contidas em um $TRG(4)$; para facilitar a visualização, dentre as arestas entre cópias distintas de $TRG(3)$, exibimos apenas aquelas que são incidentes à permutação identidade, um vértice de $TRG_1(4)$, ou à permutação reversa, um vértice de $TRG_4(4)$, omitindo as demais.

Concentraremos agora nossos esforços nas relações de adjacência entre cópias distintas de $TRG(n - 1)$ contidas em um $TRG(n)$. Observemos, na figura 3.2, os vértices adjacentes a (1234) que estão em $TRG_2(4) - (2134)$, (2314) e (2341) – induzem uma clique, ou seja, um subgrafo onde todos os vértices são adjacentes dois a dois. O mesmo pode ser dito para os vértices adjacentes a (1234) que estão em $TRG_3(4) - (3124)$ e (3412) induzem uma



(Obs.: as arestas representadas são apenas aquelas entre vértices de cada cópia de $TRG(3)$, entre a permutação identidade e seus vizinhos e entre a permutação reversa e seus vizinhos.)

Figura 3.2: O grafo $TRG(4)$ como 4 cópias de $TRG(3)$.

clique de 2 vértices – e em $TRG_4(4) - (4123)$, que induz trivialmente uma clique de apenas 1 vértice. Generalizamos esta observação no teorema 3.12.

Teorema 3.12. *Seja π um vértice de $TRG_x(n)$. O subconjunto de vértices adjacentes a π e contidos em $TRG_y(n)$, $y \neq x$ induz uma clique.*

Prova. O vértice π , pertencente a $TRG_x(n)$, pode ser escrito da seguinte forma:

$$(\mathbf{x} \pi_2 \dots \pi_{j-1} \mathbf{y} \pi_{j+1} \dots \pi_n),$$

onde x e y são inteiros entre 1 e n , distintos entre si e dos demais π_i . Sejam σ', σ'' vértices de $TRG_y(n)$, expressos da seguinte forma:

$$\begin{aligned} \sigma' &= (\mathbf{y} \sigma'_2 \dots \sigma'_n), \\ \sigma'' &= (\mathbf{y} \sigma''_2 \dots \sigma''_n). \end{aligned}$$

Sendo σ' e σ'' vizinhos de π , existem transposições t' e t'' tais que $\sigma' = \pi \cdot t'$ e $\sigma'' = \pi \cdot t''$.

Podemos afirmar que $t' = t(1, j, k')$ e $t'' = t(1, j, k'')$, onde $j = \pi^{-1}(y)$ (ou seja, j é a posição de y em π) e $j + 1 \leq k', k'' \leq n + 1$. Logo, σ' e σ'' são iguais a

$$\begin{aligned} \sigma' &= (\mathbf{y} \pi_{j+1} \dots \pi_{k'-1} \mathbf{x} \pi_2 \dots \pi_{j-1} \pi_{k'} \dots \dots \dots \pi_n) \\ \sigma'' &= (\mathbf{y} \pi_{j+1} \dots \dots \dots \pi_{k''-1} \mathbf{x} \pi_2 \dots \pi_{j-1} \pi_{k''} \dots \pi_n). \end{aligned}$$

Suponha, sem perda de generalidade, que $k'' > k'$. Então, a transposição $t(k' - j + 1, k', k'')$ transforma σ' em σ'' , o que permite-nos dizer que σ' e σ'' são adjacentes. ■

O teorema anterior nos dá alguma informação sobre a estrutura dos caminhos mínimos no grafo. Por exemplo, seja

$\pi, \pi', \dots, \pi^{(\ell-1)}, \pi^{(\ell)}, \pi^{(\ell+1)}, \dots, \pi^{(d)} = \sigma$ um caminho mínimo. Consideremos as seguintes hipóteses, mutuamente exclusivas, sobre três vértices consecutivos $\pi^{(\ell-1)}, \pi^{(\ell)}, \pi^{(\ell+1)}$ neste caminho:

- (i) todos os três vértices pertencem ao mesmo $TRG_x(n)$;
- (ii) os vértices pertencem a três TRGs distintos $TRG_x(n)$, $TRG_y(n)$ e $TRG_z(n)$;
- (iii) $\pi^{(\ell-1)}$ e $\pi^{(\ell)}$ pertencem ao mesmo $TRG_x(n)$ e $\pi^{(\ell+1)}$ pertence a um TRG distinto $TRG_y(n)$;
- (iv) $\pi^{(\ell)}$ e $\pi^{(\ell+1)}$ pertencem ao mesmo $TRG_x(n)$ e $\pi^{(\ell-1)}$ pertence a um TRG distinto $TRG_y(n)$;
- (v) $\pi^{(\ell-1)}$ e $\pi^{(\ell+1)}$ pertencem ao mesmo $TRG_x(n)$ e $\pi^{(\ell)}$ pertence a um TRG distinto $TRG_y(n)$;

À exceção da hipótese (v), todas as outras podem ocorrer. A última hipótese não é válida pois, como $\pi^{(\ell-1)}, \pi^{(\ell)}, \pi^{(\ell+1)}$ são consecutivos, temos as arestas $\pi^{(\ell-1)}\pi^{(\ell)}$ e $\pi^{(\ell)}\pi^{(\ell+1)}$; como, por hipótese, $\pi^{(\ell-1)}$ e $\pi^{(\ell+1)}$ pertencem ao mesmo $TRG_x(n)$ e $\pi^{(\ell)}$ encontra-se em outro TRG, então o teorema 3.12 nos diz que $\pi^{(\ell-1)}$ e $\pi^{(\ell+1)}$ são adjacentes, portanto poderíamos diminuir o comprimento do caminho, indo de $\pi^{(\ell-1)}$ a $\pi^{(\ell+1)}$ sem passar por $\pi^{(\ell)}$, o que contradiz o fato de que o caminho é mínimo.

O próximo lema nos dará a quantidade exata de vizinhos de um vértice π de $TRG_x(n)$ cujas permutações correspondentes têm o elemento π_i como primeiro elemento.

Lema 3.13. *Existem exatamente $n - j + 1$ vértices na vizinhança de um vértice π de $TRG_{\pi_1}(n)$ que estão em $TRG_{\pi_j}(n)$, $2 \leq j \leq n$.*

Prova. Dado j , escolha k tal que $j < k \leq n + 1$. A transposição $t(1, j, k)$ coloca o elemento π_j na primeira posição de $\sigma = \pi \cdot t(1, j, k)$. Logo σ é um vértice de $TRG_{\pi_j}(n)$. Vê-se facilmente que qualquer transposição que leva o elemento π_j para a primeira posição é da forma $t(1, j, k)$ e que existem $n - j + 1$ valores possíveis para k . ■

Em outras palavras, o lema anterior indica-nos que um dado vértice possui um número distinto, entre 1 e $n - 1$, de vizinhos em cópias distintas de um mesmo $TRG(n - 1)$ fixo em um $TRG(n)$.

Resta-nos calcular o número de vizinhos de um vértice π que estão no mesmo $TRG_{\pi_1}(\pi)$. Como este subgrafo é isomorfo a $TRG(n - 1)$ (teorema 3.11), há $\frac{(n-1)^3 - (n-1)}{6}$ vértices adjacentes a π em $TRG_{\pi_1}(n)$, segundo a proposição 3.8.

Tomemos o $TRG(3)$ como exemplo e vejamos as relações de vizinhança entre a permutação identidade (123), a permutação reversa (321) e os demais vértices do grafo. Os vértices adjacentes à permutação reversa e à identidade são:

- 1 vértice em $TRG_1(3)$: (132);
- 2 vértices em $TRG_2(3)$: (213) e (231); e
- 1 vértice em $TRG_3(3)$: (312).

Vemos que os vértices (123) e (321) possuem os mesmos vizinhos em $TRG(3)$. Diremos que esses dois vértices são gêmeos.

Definição 3.14 (vértices gêmeos). *Dados dois vértices u e v de um grafo G , seja $N(u)$ o conjunto dos vértices de G que são adjacentes a u e $N(v)$ o conjunto dos vértices que são adjacentes a v . Diremos que u e v são gêmeos se $N(u) = N(v)$.*

No grafo $TRG(3)$ podemos identificar os vértices vizinhos aos gêmeos (123) e (321) , unindo-os em um único vértice w . O grafo resultante — composto pelos vértices (123) , w e (321) , com arestas entre (123) e w e entre (321) e w — possui características similares ao $TRG(3)$ original; em especial o seu diâmetro também é 2.

Infelizmente, a estratégia de identificação dos vizinhos de vértices gêmeos não funciona para $TRG(n)$ em geral, com $n \geq 3$, pois o único caso em que há gêmeos o grafo é para $n = 3$, como veremos a seguir.

Proposição 3.15. *Não há vértices gêmeos em $TRG(n)$, $n \geq 4$.*

Prova. Provaremos a proposição por indução em n . No caso base, $n = 4$, consideremos o vértice (1234) e provaremos que ele não possui nenhum gêmeo. Consideremos um vértice $\sigma \neq (1234)$. Em um $TRG(4)$, somente existem três possibilidades para σ (ver $TRG(4)$ na figura 3.1), a saber: (i) σ é vizinho de (4321) ; (ii) σ é vizinho de (1234) ; ou (iii) $\sigma = (3142)$ ou $\sigma = (2413)$.

No caso (i), σ não pode ser gêmeo de (1234) pois o vértice (4321) não está na vizinhança de (1234) . Nos casos em que σ satisfaz (ii) e (iii), existem vizinhos de σ que estão na vizinhança de (4321) — novamente, isto pode ser visto claramente na figura 3.1 — portanto, esses vizinhos não são adjacentes a (1234) . Logo, qualquer vértice que não seja (1234) possui vizinhança diferente deste.

Como $TRG(4)$ é vértice-transitivo, podemos transformar (1234) em um outro vértice qualquer π desse grafo por meio de um automorfismo Φ . Dado que Φ preserva as relações de vizinhança, temos que π não possui gêmeos, pois (1234) não possui gêmeos.

Para $n > 4$, a hipótese de indução é que não existem vértices gêmeos em $TRG(n - 1)$. Tome um par de vértices π , σ . Se π e σ pertencem ambos ao mesmo $TRG_{\pi_1}(n)$, então eles não podem ser gêmeos, pois $TRG_x(n)$ é

isomorfo a $TRG(n-1)$. Se π e σ pertencem a TRGs diferentes, respectivamente $TRG_{\pi_1}(n)$ e $TRG_{\sigma_1}(n)$, com $\sigma_1 \neq \pi_1$, então π possui, pelo lema 3.13, no máximo $n-1$ vizinhos em $TRG_{\sigma_1}(n)$, ao passo que σ possui $\frac{(n-1)^3-(n-1)}{6}$ vizinhos nesse mesmo subgrafo. ■

O próximo resultado fornecer-nos-á uma primeira visão sobre a estrutura dos caminhos mínimos em um TRG .

Teorema 3.16 (Princípio de atração dos TRGs). *Entre dois vértices de $TRG_x(n)$, sempre existe um caminho mínimo que passa apenas por vértices de $TRG_x(n)$.*

Prova. Sejam π e σ dois vértices de $TRG_x(n)$, e $\pi, \pi', \pi'', \dots, \pi^{(t)}$, com $\pi^{(t)} = \sigma$, um caminho mínimo entre π e σ que possui vértices fora de $TRG_x(n)$ (estamos supondo que tal caminho existe, pois caso qualquer caminho mínimo somente possuísse vértices em $TRG_x(n)$, a prova do teorema seria imediata). Considere os seguintes vértices pertencentes ao caminho:

- $\pi^{(a)}$ vértice de $TRG_x(n)$ tal que o vértice seguinte no caminho, $\pi^{(a+1)}$, não está em $TRG_x(n)$;
- $\pi^{(b)}$ vértice de $TRG_x(n)$ tal que o vértice anterior no caminho, $\pi^{(b-1)}$, não está em $TRG_x(n)$;
- todos os vértices $\pi^{(a+1)}, \dots, \pi^{(b-1)}$ não estão em $TRG_x(n)$.

Definiremos a operação de remoção de um elemento x de uma permutação π de n elementos, que dá origem à permutação denotada por $\pi \ominus x = (\gamma_1 \gamma_2 \dots \gamma_{n-1})$, onde cada elemento γ_i da permutação resultante

é

$$\gamma_i = \begin{cases} \pi_i & \text{se } i < \pi^{-1}(x) \text{ e } \pi_i < x \\ \pi_i - 1 & \text{se } i < \pi^{-1}(x) \text{ e } \pi_i > x \\ \pi_{i-1} & \text{se } i > \pi^{-1}(x) \text{ e } \pi_i < x \\ \pi_{i-1} - 1 & \text{se } i > \pi^{-1}(x) \text{ e } \pi_i > x \end{cases}.$$

Ou seja, $\pi \ominus x$ é a permutação obtida a partir de π , donde se apaga o elemento x e renumeram-se os demais elementos maiores do que x para obtermos uma permutação de $n - 1$ elementos.

Seja $\gamma^{(i)} = \pi^{(i)} \ominus x$, para $i = a, \dots, b$. A seqüência $\gamma^{(a)}, \gamma^{(a+1)}, \dots, \gamma^{(b)}$ é um caminho entre $\gamma^{(a)}$ e $\gamma^{(b)}$ em $TRG(n-1)$ – observe que, pela minimalidade do caminho entre $\pi^{(a)}$ e $\pi^{(b)}$, não há vértices repetidos na seqüência. Observamos que $x \oplus \gamma^{(a)}, x \oplus \gamma^{(a+1)}, \dots, x \oplus \gamma^{(b)}$ é um caminho com o mesmo número de vértices que o caminho mínimo entre $x \oplus \gamma^{(a)} = \pi^{(a)}$ e $x \oplus \gamma^{(b)} = \pi^{(b)}$, sendo que este passa apenas por vértices de $TRG_x(n)$, logo, podemos substituir o trecho $\pi^{(a+1)}, \dots, \pi^{(b-1)}$ por $x \oplus \gamma^{(a+1)}, \dots, x \oplus \gamma^{(b-1)}$.

Repetindo-se a substituição descrita acima para todos os trechos do caminho que passam por vértices exteriores a $TRG_x(n)$, transformaremos-lo em um caminho contido em $TRG_x(n)$. ■

Uma interpretação do princípio de atração dos TRGs nos diz que, no processo de transformar uma permutação σ em π , sendo $\sigma, \pi \in S_n$, com um número mínimo de transposições, podemos considerar que, no momento em que a seqüência de transposições der origem a uma permutação π' cujo primeiro elemento é π_1 , então as permutações que se seguirem possuirão todas o mesmo elemento π_1 na primeira posição. Podemos assim descartar o primeiro elemento de π_1 e encontrar a distância entre duas permutações $\pi' \ominus \pi_1$ e $\pi \ominus \pi_1$, ambas com $n - 1$ elementos.

A estratégia de remoção de um elemento utilizada na prova do princípio de atração dos TRGs dá-nos também a oportunidade de limitar o diâmetro $D(TRG(n))$ em função do diâmetro $D(TRG(n-1))$.

Teorema 3.17. *A seguinte relação entre os diâmetros de $TRG(n)$ e $TRG(n-1)$ é válida: $D(TRG(n-1)) \leq D(TRG(n)) \leq D(TRG(n-1)) + 1$.*

Prova. Demonstraremos primeiro que $D(TRG(n)) \leq D(TRG(n-1)) + 1$. Seja π, σ um par diametral em $TRG(n)$; suponha, sem perda de generalidade, que $\pi \in TRG_x(n)$ e $\sigma \in TRG_y(n)$, com $x \neq y$ (caso π e σ estivessem no mesmo $TRG_x(n)$, então a distância entre esses dois vértices seria, no máximo, $D(TRG(n-1))$, pois $TRG_x(n-1)$ é isomorfo a um $TRG(n-1)$). Existe pelo menos um vértice π' de $TRG_x(n)$ que é adjacente a σ , a saber, o vértice $\pi' = \sigma \cdot t(1, \sigma^{-1}(x), \sigma^{-1}(x) + 1)$, ou seja, aquele que é obtido a partir de σ pela transposição do elemento x para a primeira posição. Logo, $d_t(\pi, \sigma) \leq d_t(\pi, \pi') + d_t(\pi', \sigma) \leq D(TRG(n-1)) + 1$.

A desigualdade $D(TRG(n-1)) \leq D(TRG(n))$ é válida pois, caso $D(TRG(n)) < D(TRG(n-1))$, poderíamos tomar qualquer par diametral π, σ no mesmo $TRG_x(n)$, de tal forma que um caminho mínimo $\pi, \pi', \pi'', \dots, \pi^{(d)} = \sigma$ teria comprimento $d < D(TRG(n-1))$. Pelo princípio de atração dos TRGs, podemos considerar que tal caminho mínimo está contido em $TRG_x(n)$. Como $TRG_x(n)$ é isomorfo a $TRG(n-1)$, o caminho mínimo entre o par diametral π e σ seria correspondente a um caminho mínimo em $TRG(n-1)$, mas isso seria uma contradição com o fato de que um caminho mínimo entre um par diametral de $TRG(n-1)$ possui comprimento $D(TRG(n-1))$. ■

A desigualdade $D(TRG(n)) \leq D(TRG(n-1)) + 1$ permite-nos dar um limite superior para a distância através da análise do TRG: como $D(TRG(2)) = 1$, teremos que $D(TRG(n)) \leq n - 1$, $n \geq 2$. Este limite

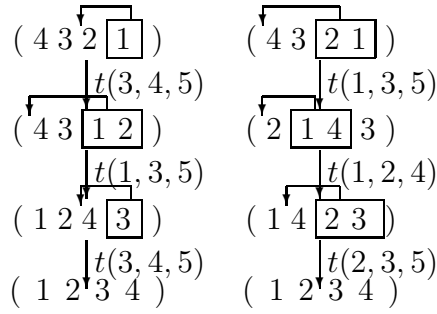


Figura 3.3: Ordenações mínimas diferentes para $\rho_{[4]}$.

é igual ao limite obtido no capítulo 1, ao colocarmos cada elemento de uma permutação no seu lugar, usando no máximo $n - 1$ transposições.

3.1.1 Número de caminhos mínimos entre dois vértices

Passemos a analisar de quantas formas distintas possíveis podemos transformar uma permutação π em σ com um número mínimo de transposições. Usando a estrutura do grafo de rearranjos por transposições, transformamos o problema em encontrar o número de caminhos mínimos entre os vértices π e σ em $TRG(n)$.

Começaremos com um exemplo pequeno, o de transformar a permutação reversa de 4 elementos $\rho_{[4]} = (4321)$ na permutação identidade $\iota_{[4]} = (1234)$. Sabemos, pelo teorema 2.26, que pelo menos 3 transposições são necessárias para ordenar $\rho_{[4]}$; tanto a seqüência $t(3, 4, 5), t(1, 3, 5), t(3, 4, 5)$ quanto $t(1, 3, 5), t(1, 2, 4), t(2, 3, 5)$ ordenam $\rho_{[4]}$ (fig. 3.3) satisfazendo à condição de minimalidade.

Observando-se o subgrafo de $TRG(4)$ cujos vértices e arestas estão em uma geodésica a partir de $\iota_{[4]}$ (figura 3.4), vemos que, dado um vizinho v de $\iota_{[4]}$, este possui exatamente 4 vértices adjacentes w_1, w_2, w_3, w_4 que estão na

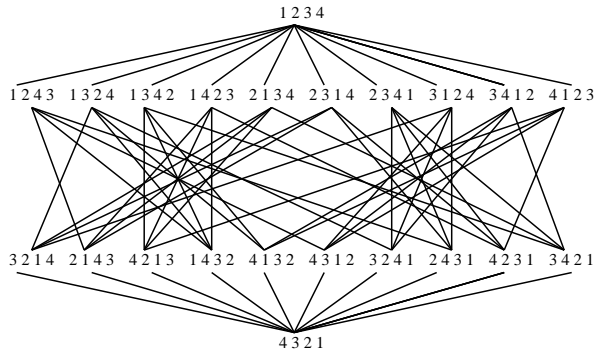


Figura 3.4: Caminhos mínimos entre $\rho_{[4]}$ e $\iota_{[4]}$.

vizinhança de $\rho_{[4]}$. Como existem 10 vizinhos de $\iota_{[4]}$, teremos um total de $4 \cdot 10 = 40$ combinações – isto é, geodésicas – entre $\rho_{[4]}$ e $\iota_{[4]}$. Além disso, vemos que o único par diametral que contém a identidade é $\rho_{[4]}, \iota_{[4]}$.

Passaremos à análise em $TRG(5)$ e veremos que a quantidade de geodésicas entre $\iota_{[5]}$ e $\rho_{[5]}$ é bem menor. Por outro lado, há mais de 1 par diametral contendo $\iota_{[5]}$.

Vemos que, entre os 20 vizinhos de $\iota_{[5]}$, aqueles 16 que correspondem a permutações obtidas da identidade pela transposição de um único elemento não são adjacentes a nenhum dos vizinhos de $\rho_{[5]}$. Por simetria, podemos também afirmar que, entre os vizinhos de $\rho_{[5]}$, todos aqueles obtidos por uma transposição de um único elemento não são adjacentes a vizinhos de $\iota_{[5]}$. As afirmativas anteriores são justificadas pelo lema que demonstraremos a seguir, que nos diz que qualquer permutação obtida a partir de $\iota_{[5]}$ por uma transposição $t(i, j, j + 1)$ – que move 1 elemento para a esquerda – ou $t(i, i + 1, k)$ – que move um elemento para a direita – está a uma distância de $\rho_{[5]}$ de, no mínimo, $\lfloor \frac{5-1}{2} \rfloor + 1 = 3$.

Lema 3.18. *Seja $\sigma^0 = \iota_{[n]} \cdot t(i, i + 1, k)$ ou $\sigma^0 = \iota_{[n]} \cdot t(i, j, j + 1)$, onde $1 \leq i < j < k \leq n + 1$. Então a distância de σ^0 a $\rho_{[n]}$ é, pelo menos, $\lfloor \frac{n-1}{2} \rfloor + 1$.*

Prova. Apresentaremos a prova apenas para o caso $\sigma^0 = \iota_{[n]} \cdot t(i, i + 1, k)$, uma vez que o raciocínio é bem similar para o outro caso. A permutação σ^0 pode ser escrita como $(1 \dots i - 1 \ i + 1 \dots k - 1 \ i \ k \dots n)$. Suponha que exista uma seqüência $t(i_1, j_1, k_1), t(i_2, j_2, k_2), \dots, t(i_l, j_l, k_l)$ com $l < \lfloor \frac{n-1}{2} \rfloor + 1$ transposições que transforma σ em $\rho_{[n]}$.

Considere agora as permutações $\sigma^0 \ominus i$ e $\rho_{[n]} \ominus i$ obtidas pela remoção do elemento i e renumerando-se os demais elementos. Vemos que $\sigma^0 \ominus i = \iota_{[n-1]}$ e $\rho_{[n]} \ominus i = \rho_{[n-1]}$. A seqüência de transposições $t(i_1, j_1, k_1), \dots, t(i_l, j_l, k_l)$ pode ser alterada para uma seqüência com $l' \leq l$ transposições, $t(i'_1, j'_1, k'_1), \dots, t(i'_l, j'_l, k'_l)$ que transforma $\iota_{[n-1]}$ em $\rho_{[n-1]}$ (ver figura 3.5). Mas isso contradiz o fato de que precisamos de $\lfloor \frac{n-1}{2} \rfloor + 1$ transposições para ordenar $\rho_{[n-1]}$. ■

Resta-nos examinar apenas as relações de adjacência entre os vizinhos de $\iota_{[5]}$ e de $\rho_{[5]}$ que correspondem a permutações obtidas por transposições de mais de 1 elemento. Para os vizinhos da identidade, somente existem quatro permutações que satisfazem essas condições, a saber: $(34125), (14523), (34512)$ e (45123) . Para os vizinhos de $\rho_{[5]}$, temos: $(32541), (52143), (21543)$ e (32154) . Vemos que apenas existem as adjacências $(14523), (52143)$ e $(34125), (32541)$. Destarte, somente há 2 caminhos entre $\iota_{[5]}$ e $\rho_{[5]}$,

$$(12345), (14523), (52143), (54321) \quad \text{e} \quad (12345), (34125), (32541), (54321).$$

$$\begin{array}{ll}
\sigma^0 = (1 \dots i-1 \ i+1 \dots k-1 \ i \ k+1 \dots n) & \sigma^0 \ominus i = (1 \dots i-1 \ i \dots k-1 \ k \dots n) \\
\sigma^1 = \sigma^0 \cdot t(i_1, j_1, k_1) & \sigma^1 \ominus i = (\sigma^0 \ominus i) \cdot t(i'_1, j'_1, k'_1) \\
\sigma^2 = \sigma^1 \cdot t(i_2, j_2, k_2) & \sigma^2 \ominus i = (\sigma^1 \ominus i) \cdot t(i'_1, j'_1, k'_1) \\
\vdots & \vdots \\
r_n = \sigma^l = \sigma^{l-1} \cdot t(i_l, j_l, k_l) & r_{n-1} = \sigma^l \ominus i = (\sigma^{l-1} \ominus i) \cdot t(i'_l, j'_l, k'_l)
\end{array}$$

Caso 1:

$$\begin{array}{l}
\sigma^{m-1} = (\sigma_1^{m-1} \sigma_2^{m-1} \dots \sigma_{i-1}^{m-1} \sigma_i^{m-1} \dots \sigma_{j-1}^{m-1} \boxed{\sigma_j^{m-1} \dots \sigma_{k-1}^{m-1}} \sigma_k^{m-1} \dots i \dots \sigma_n^{m-1}) \\
\sigma^{m-1} \ominus i = (\sigma_1^{m-1} \sigma_2^{m-1} \dots \sigma_{i-1}^{m-1} \sigma_i^{m-1} \dots \sigma_{j-1}^{m-1} \boxed{\sigma_j^{m-1} \dots \sigma_{k-1}^{m-1}} \sigma_k^{m-1} \dots \cancel{i} \dots \sigma_n^{m-1})
\end{array}$$

Caso 2:

$$\begin{array}{l}
\sigma^{m-1} = (\sigma_1^{m-1} \sigma_2^{m-1} \dots \sigma_{i-1}^{m-1} \sigma_i^{m-1} \dots \sigma_{j-1}^{m-1} \boxed{\sigma_j^{m-1} \dots i \dots \sigma_{k-1}^{m-1}} \sigma_k^{m-1} \dots \sigma_n^{m-1}) \\
\sigma^{m-1} \ominus i = (\sigma_1^{m-1} \sigma_2^{m-1} \dots \sigma_{i-1}^{m-1} \sigma_i^{m-1} \dots \sigma_{j-1}^{m-1} \boxed{\sigma_j^{m-1} \dots \cancel{i} \dots \sigma_{k-1}^{m-1}} \sigma_k^{m-1} \dots \sigma_n^{m-1})
\end{array}$$

Caso 3:

$$\begin{array}{l}
\sigma^{m-1} = (\sigma_1^{m-1} \sigma_2^{m-1} \dots \sigma_{i-1}^{m-1} \sigma_i^{m-1} \dots i \dots \sigma_{j-1}^{m-1} \boxed{\sigma_j^{m-1} \dots \sigma_{k-1}^{m-1}} \sigma_k^{m-1} \dots \sigma_n^{m-1}) \\
\sigma^{m-1} \ominus i = (\sigma_1^{m-1} \sigma_2^{m-1} \dots \sigma_{i-1}^{m-1} \sigma_i^{m-1} \dots \cancel{i} \dots \sigma_{j-1}^{m-1} \boxed{\sigma_j^{m-1} \dots \sigma_{k-1}^{m-1}} \sigma_k^{m-1} \dots \sigma_n^{m-1})
\end{array}$$

Caso 4:

$$\begin{array}{l}
\sigma^{m-1} = (\sigma_1^{m-1} \sigma_2^{m-1} \dots i \dots \sigma_{i-1}^{m-1} \sigma_i^{m-1} \dots \sigma_{j-1}^{m-1} \boxed{\sigma_j^{m-1} \dots \sigma_{k-1}^{m-1}} \sigma_k^{m-1} \dots \sigma_n^{m-1}) \\
\sigma^{m-1} \ominus i = (\sigma_1^{m-1} \sigma_2^{m-1} \dots \cancel{i} \dots \sigma_{i-1}^{m-1} \sigma_i^{m-1} \dots \sigma_{j-1}^{m-1} \boxed{\sigma_j^{m-1} \dots \sigma_{k-1}^{m-1}} \sigma_k^{m-1} \dots \sigma_n^{m-1})
\end{array}$$

$$(i'_m, j'_m, k'_m) = \begin{cases} (i_m, j_m, k_m) & \text{se } \sigma^{m-1}-1(i) \geq k_m & \text{(caso 1)} \\ (i_m, j_m, k_m - 1) & \text{se } j_m \leq \sigma^{i-1}-1(i) \leq j_m & \text{(caso 2)} \\ (i_m, j_m - 1, k_m - 1) & \text{se } i_m \leq \sigma^{i-1}-1(i) \leq j_m & \text{(caso 3)} \\ (i_m - 1, j_m - 1, k_m - 1) & \text{se } \sigma^{i-1}-1(i) < i_m & \text{(caso 4)} \end{cases}$$

Figura 3.5: Alterando a seqüência de transposições que transforma σ em $\rho_{[n]}$ para transformar $\iota_{[n-1]}$ em $\rho_{[n-1]}$ para o lema 3.18.

Aproveitaremos a abordagem adotada no lema anterior para estimar a distância entre uma série de permutações de n elementos e a permutação identidade $\iota_{[n]}$ a partir da distância entre uma permutação de $n-1$ elementos e $\iota_{[n-1]}$.

Lema 3.19. *Para cada permutação π de S_{n-1} , existem n permutações $\sigma^1, \dots, \sigma^n$ pertencentes a S_n tais que $d(\iota_{[n-1]}, \pi) \leq d(\iota_{[n]}, \sigma^i) \leq d(\iota_{[n-1]}, \pi) + 1$.*

Prova. Considere as $n + 1$ permutações $\sigma^1 = (n\pi_1\pi_2 \dots \pi_{n-1})$, $\sigma^2 = (\pi_1 n \pi_2 \dots \pi_{n-1})$, \dots , $\sigma^i = (\pi_1 \dots \pi_{i-1} n \pi_i \dots \pi_{n-1})$, \dots , $\sigma^n = (\pi_1 \pi_2 \dots \pi_{n-1} n)$.

A desigualdade $d(\iota_{[n]}, \sigma^i) \leq d(\iota_{[n-1]}, \pi) + 1$ é obtida através da construção do caminho

$$\iota_{[n]} = \iota_{[n-1]} \oplus n, \quad \gamma^1 \oplus n, \quad \dots, \quad \gamma^d \oplus n, \quad t((\sigma^i)^{-1}(n), n, n+1)(\gamma^d \oplus n) = \sigma^i$$

a partir do caminho mínimo

$$\iota_{[n-1]}, \quad \gamma^1, \quad \dots, \quad \gamma^d = \pi$$

onde $d = d(\iota_{[n-1]}, \pi)$.

O fato de que $d(\iota_{[n-1]}, \pi) \leq d(\iota_{[n]}, \sigma^i)$ advém da seguinte propriedade: se pudéssemos encontrar um caminho

$$\iota_{[n]}, \gamma^1, \dots, \gamma^\ell = \sigma^i$$

de comprimento $\ell < d(\iota_{[n-1]}, \pi)$, então poderíamos construir um caminho

$$\iota_{[n]} \ominus n = \iota_{[n-1]}, \quad \gamma^1 \ominus n, \quad \dots, \quad \gamma^\ell \ominus n = \pi$$

entre $\iota_{[n-1]}$ e π com comprimento menor do que o caminho mínimo entre esses dois vértices, o que seria uma contradição. ■

Através do lema 3.19, podemos estimar a distância de transposição entre um conjunto de permutações de n elementos e a permutação identidade $\iota_{[n]}$, desde que saibamos resolver o problema de determinar a distância entre uma permutação de $n-1$ elementos e a identidade $\iota_{[n-1]}$. Idealmente, gostaríamos de poder agrupar as permutações segundo coleções que possuam a mesma distância em relação à permutação identidade, reduzindo assim o número de vértices do TRG e tornando computacionalmente possível a tarefa de se encontrarem caminhos mínimos, mas infelizmente o lema 3.19 nos fornece uma estimativa da distância com uma margem de erro de 1 transposição, e não a distância exata. Os resultados que veremos a seguir, devidos a Eriksson *et al.* [10], são um desenvolvimento nessa direção: é possível separar as permutações em classes de equivalência de tal forma que, se π e σ pertencem à mesma classe, então $d_t(\pi) = d_t(\sigma)$.

3.2 Agrupando vértices no grafo de rearranjos por transposições

Nosso objetivo, nesta seção, será descrever o resultado de Eriksson *et al.* [10], que encontraram classes de permutações cuja distância em relação à permutação identidade é a mesma. Com isso, eles conseguiram descrever um grafo com menos vértices que o grafo de rearranjos por transposições, mas com o mesmo diâmetro. Ao final da seção, a partir do teorema 3.32, apresentamos contribuições originais derivadas da abordagem de Eriksson *et al.*

Definição 3.20 (permutação circular). [10] Uma permutação circular π^c de n elementos, representada por

$$\pi^c = (\pi_0 \pi_1 \pi_2 \dots \pi_n)^c,$$

onde $\pi_i \in \{0, \dots, n\}$ e $\pi_i \neq \pi_j$ se $i \neq j$, é uma função bijetiva tal que $\pi^c(\pi_i) = \pi_{i+1}$ para $0 \leq i < n$ e $\pi^c(\pi_n) = \pi_0$.

Definição 3.21 (permutação circular identidade). [10] A permutação circular identidade $\iota_{[n]}^c$ de n elementos é

$$\iota_{[n]}^c = (0 \ 1 \ 2 \ \dots \ n)^c.$$

Definição 3.22. Denotaremos por S_n^c o conjunto das permutações circulares de n elementos.

Podemos observar que a permutação circular $(\pi_0 \ \pi_1 \ \pi_2 \ \dots \ \pi_n)^c$ possui as seguintes representações equivalentes

$$\begin{aligned} & (\pi_1 \ \pi_2 \ \dots \ \pi_n \ \pi_0)^c \\ & (\pi_2 \ \dots \ \pi_n \ \pi_0 \ \pi_1)^c \\ & \quad \vdots \\ & (\pi_n \ \pi_0 \ \pi_1 \ \dots \ \pi_{n-1})^c, \end{aligned}$$

ou seja, as permutações obtidas por rotações cíclicas de uma permutação são todas equivalentes entre si. Diremos que duas permutações circulares π^c, σ^c são ciclicamente equivalentes, $\pi^c \equiv^c \sigma^c$, se σ^c pode ser representada por uma rotação cíclica de π^c . Como existe mais de uma forma para representar

a mesma permutação circular, adotaremos, por padrão, a representação a seguir.

Definição 3.23 (representação canônica). [10] *A representação canônica de uma permutação circular é aquela cujo primeiro elemento é 0.*

Representaremos cada classe de equivalência em S_n^c por um de seus elementos, expresso em sua representação canônica.

Definição 3.24 (circularização). [10] *Dada uma permutação linear $\pi = (\pi_1 \pi_2 \dots \pi_n) \in S_n$, a circularização de π é uma permutação circular $\pi^c \in S_n^c$ tal que*

$$\pi^c = (0 \pi_1 \pi_2 \dots \pi_n)^c.$$

Vemos imediatamente que, dada uma permutação circular π^c , existe apenas 1 permutação linear π tal que a sua circularização seja π^c . Deste modo, podemos definir sem ambigüidade a operação de transposição para permutações circulares da maneira a seguir.

Definição 3.25 (transposição aplicada a uma permutação circular). *Dados $\pi^c \in S_n^c$ e i, j, k inteiros tais que $1 \leq i < j < k \leq n+1$, definimos a operação de transposição $t(i, j, k)$ aplicada a π^c como $\pi^c \cdot t(i, j, k) = \sigma^c$, onde σ^c é a circularização de $\sigma = \pi \cdot t(i, j, k)$.*

A figura 3.6 mostra como uma transposição aplicada a uma permutação circular “corta” essa permutação em três partes e cola-as de maneira diferente.

O problema original de ordenação de uma permutação linear por transposições pode ser transformado no problema equivalente de transformar uma permutação circular na permutação identidade ι^c usando a menor seqüência possível de transposições.

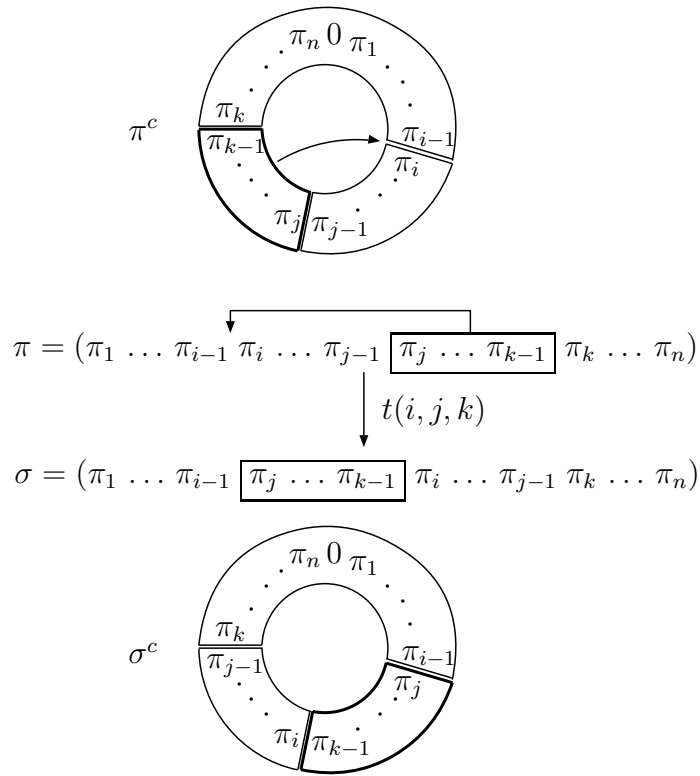


Figura 3.6: Efeito de uma transposição em uma permutação circular

O leitor poderá indagar-se sobre a utilidade de se introduzirem permutações circulares no estudo do problema da distância de transposição, visto que, até aqui, elas não trazem aparentemente nenhum resultado novo. O que veremos a seguir é que, a partir das permutações circulares, produziremos uma classe de equivalência para as permutações lineares de tal forma que as permutações que estiverem na mesma classe terão a mesma distância com relação à permutação identidade.

Definição 3.26 (deslocamento cíclico). [10] *Dados uma permutação circular $\pi^c = (0 \ \pi_1 \ \dots \ \pi_n)^c$ e um número $m \in \mathbb{N}$, usaremos a notação $m + \pi^c$ para denotar a permutação*

$$m + \pi^c := (\overline{m} \ \overline{\pi_1 + m} \ \dots \ \overline{\pi_n + m})^c,$$

onde \overline{x} é o resto da divisão (módulo) de x por $n + 1$.¹

Qualquer deslocamento cíclico da permutação circular identidade $\iota_{[n]}^c$ produz a própria permutação circular identidade. O mesmo vale para a permutação circular reversa $\rho_{[n]}^c$, ou seja, $m + \rho_{[n]}^c = \rho_{[n]}^c$.

Exemplo 3.27. *Seja $\pi^c = (01243)^c$. As permutações que podem ser obtidas a partir de um deslocamento cíclico de π^c são*

$$1 + \pi^c = (12304)^c = (04123)^c$$

$$2 + \pi^c = (23410)^c = (02341)^c$$

$$3 + \pi^c = (34021)^c = (02134)^c$$

$$4 + \pi^c = (40132)^c = (01324)^c.$$

¹N.B.: Não confunda a operação de deslocamento cíclico com a operação de inserção de um elemento da definição 3.9.

Definição 3.28 (equivalência tórica). [10] *Dadas duas permutações circulares π^c e σ^c , diremos que σ^c é toricamente equivalente a π^c , $\sigma^c \equiv^t \pi^c$, se existe um inteiro m tal que $\sigma^c = m + \pi^c$.*

Estenderemos a definição 3.28 para permutações lineares, dizendo que duas permutações lineares π, σ são toricamente equivalentes, $\pi \equiv^t \sigma$, se vale $\pi^c \equiv^t \sigma^c$ para as suas respectivas circularizações π^c, σ^c .

A relação de equivalência tórica é simétrica pois a operação de soma modular é inversível. Ou seja, existe um número natural, denotado por $-m$ e denominado inverso aditivo de m , para o qual $-m + m \equiv 0 \pmod{n+1}$. Logo, $\sigma^c \equiv^t \pi^c \Rightarrow$ existe m tal que $\sigma^c = m + \pi^c \Rightarrow$ existe $-m$ tal que $-m + m \equiv 0 \pmod{n+1} \Rightarrow -m + \sigma^c = -m + m + \pi^c = \pi^c \Rightarrow \pi^c \equiv^t \sigma^c$. Claramente, a relação de equivalência tórica é também reflexiva e transitiva.

Já vimos que $m + \iota_{[n]}^c = \iota_{[n]}^c$ para qualquer valor fixo de m , assim como $m + \rho_{[n]}^c = \rho_{[n]}^c$. Ou seja, a permutação $\iota_{[n]}$ é o único elemento da classe de equivalência que a contém, segundo a relação \equiv^t ; a mesma observação é válida para $\rho_{[n]}$.

Voltemos ao exemplo 3.27 e verifiquemos que a permutação $\pi = (1\ 2\ 4\ 3)$ está a distância 1 da permutação identidade, pois $(1\ 2\ 3\ \boxed{4}) \cdot t(3, 4, 5) = (1\ 2\ 4\ 3)$. O mesmo ocorre com as permutações toricamente equivalentes a π :

$$\begin{aligned} 1 + \pi^c &= (12304)^c, \text{ que equivale a } (4123) = (1\ 2\ 3\ \boxed{4}) \cdot t(1, 4, 5) \\ 2 + \pi^c &= (23410)^c, \text{ que equivale a } (2341) = (1\ \boxed{2}\ 3\ 4) \cdot t(1, 2, 5) \\ 3 + \pi^c &= (34021)^c, \text{ que equivale a } (2134) = (1\ \boxed{2}\ 3\ 4) \cdot t(1, 4, 5) \\ 4 + \pi^c &= (40132)^c, \text{ que equivale a } (1324) = (1\ 2\ \boxed{3}\ 4) \cdot t(1, 4, 5). \end{aligned}$$

O seguinte teorema, um dos resultados principais desta seção, generaliza o comportamento observado no parágrafo anterior e nos permitirá determinar a distância de transposição para uma coleção de permutações que sejam toricamente equivalentes, dado que saibamos calcular a distância de uma permutação que esteja nessa mesma classe de equivalência.

Teorema 3.29. [10] *Sejam $\pi, \sigma \in S_n$ duas permutações tais que $\pi \equiv^t \sigma$. Então $d_t(\pi) = d_t(\sigma)$.*

Prova. A prova do teorema consiste em transformar qualquer seqüência que ordena π em uma seqüência que ordena σ , resultando em $d_t(\sigma) \leq d_t(\pi)$. De maneira similar, podemos transformar qualquer seqüência que ordena σ em uma seqüência que ordena π , logo $d_t(\pi) \leq d_t(\sigma)$. As duas desigualdades implicam que $d_t(\sigma) = d_t(\pi)$. Dedicaremos o restante da prova a descrever como transformar uma seqüência de transposições que ordena π em uma seqüência que ordena σ . Utilizaremos indução na distância $d_t(\pi)$ para isso.

Primeiramente, suporemos que $d_t(\pi) = 1$ e, queremos demonstrar que, sendo σ toricamente equivalente a π , então $d_t(\sigma) = 1$. Em outras palavras, se existe uma transposição $t(i, j, k)$ tal que $\pi = \iota \cdot t(i, j, k)$, então existe uma transposição $t(i', j', k')$ tal que $\sigma = m + \pi = (m + \iota) \cdot t(i', j', k') = \iota \cdot t(i', j', k')$.

Descreveremos os parâmetros i', j', k' em função de i, j, k e de m , sendo que m é o inteiro tal que $\sigma^c = m + \pi^c$. Sendo $-m$ o inverso de m módulo $n + 1$, com $1 \leq -m \leq n$, consideraremos os seguintes casos:

- caso 1: $-m < i$ (fig. 3.7);
- caso 2: $i \leq -m < j$ (fig. 3.8);
- caso 3: $j \leq -m < k$; e
- caso 4: $k \leq -m \leq n$.

Caso 1: $-m < i$. Descrevamos a circularização π^c de $\pi = \iota \cdot t(i, j, k)$ e $\sigma^c = m + \pi^c$:

$$\begin{aligned} \pi^c &= (0 \quad 1 \quad \dots \quad -m \quad \dots \quad i-1 \quad \boxed{j \quad \dots \quad k-1} \quad \boxed{i \quad \dots \quad j-1} \quad k \quad \dots \quad n)^c \\ \sigma^c &= (\overline{m} \quad \overline{m+1} \quad \dots \quad \overline{m+(-m)} \quad \dots \quad \overline{m+i-1} \quad \overline{m+j} \quad \dots \quad \overline{m+k-1} \quad \overline{m+i} \quad \dots \quad \overline{m+j-1} \quad \overline{m+k} \quad \dots \quad \overline{m+n})^c \\ &= (\overline{m} \quad \overline{m+1} \quad \dots \quad 0 \quad \dots \quad \overline{m+i-1} \quad \overline{m+j} \quad \dots \quad \overline{m+k-1} \quad \overline{m+i} \quad \dots \quad \overline{m+j-1} \quad \overline{m+k} \quad \dots \quad \overline{m+n})^c \end{aligned}$$

Colocando σ^c na forma canônica, temos:

$$\sigma^c = \left(0 \quad \overline{m+(-m)+1} \quad \dots \quad \overline{m+i-1} \quad \overline{m+j} \quad \dots \quad \overline{m+k-1} \quad \overline{m+i} \quad \dots \quad \overline{m+j-1} \quad \overline{m+k} \quad \dots \quad \overline{m+n} \quad \overline{m+1} \quad \dots \quad \overline{m+(-m)-1} \right)^c,$$

donde obtemos que a permutação linear σ é

$$\sigma = \left(1 \ 2 \ \dots \ i - (-m) - 1 \quad \boxed{j - (-m) \ \dots \ k - (-m) - 1} \quad \boxed{i - (-m) \ \dots \ j - (-m) - 1} \quad k - (-m) \ \dots \ n \right)^c.$$

Podemos ver imediatamente que $\sigma = \iota \cdot t(i - (-m), j - (-m), k - (-m))$.

Caso 2: $i \leq -m < j$. Descrevamos a circularização π^c de $\pi = \iota \cdot t(i, j, k)$ e $\sigma^c = m + \pi^c$:

$$\begin{aligned} \pi^c &= (0 \quad 1 \quad \dots \quad i-1 \quad \boxed{j \quad \dots \quad k-1} \quad \boxed{i \quad \dots \quad -m \quad \dots \quad j-1} \quad k \quad \dots \quad n)^c \\ \sigma^c &= (\overline{m} \quad \overline{m+1} \quad \dots \quad \overline{m+i-1} \quad \overline{m+j} \quad \dots \quad \overline{m+k-1} \quad \overline{m+i} \quad \dots \quad \overline{m+(-m)} \quad \dots \quad \overline{m+j-1} \quad \overline{m+k} \quad \dots \quad \overline{m+n})^c \\ &= (\overline{m} \quad \overline{m+1} \quad \dots \quad \overline{m+i-1} \quad \overline{m+j} \quad \dots \quad \overline{m+k-1} \quad \overline{m+i} \quad \dots \quad 0 \quad \dots \quad \overline{m+j-1} \quad \overline{m+k} \quad \dots \quad \overline{m+n})^c \end{aligned}$$

Colocando σ^c na forma canônica, temos:

$$\sigma^c = \left(0 \quad 1 \quad \dots \quad \overline{m+j-1} \quad \overline{m+k} \quad \dots \quad \overline{m+n} \quad \overline{m} \quad \overline{m+1} \quad \dots \quad \overline{m+i-1} \quad \overline{m+j} \quad \dots \quad \overline{m+k-1} \quad \overline{m+i} \quad \dots \quad \overline{m+(-m)-1} \right)^c,$$

donde obtemos que a permutação linear σ é

$$\sigma = \left(1 \quad 2 \quad \dots \quad \overline{j-(-m)-1} \quad \boxed{\overline{k-(-m)} \quad \dots \quad \overline{i-(-m)-1}} \quad \boxed{\overline{j-(-m)} \quad \dots \quad \overline{k-(-m)-1}} \quad \overline{i-(-m)} \quad \dots \quad n \right)^c.$$

Podemos ver imediatamente que $\sigma = \iota \cdot t(\overline{j-(-m)}, \overline{k-(-m)}, \overline{i-(-m)}) = \iota \cdot t(j-(-m), k-(-m), i+m)$.

Nas figuras 3.7 e 3.8 mostramos a análise para os casos 1 e 2, respectivamente. Omitiremos a análise para os casos 3 e 4, notando que eles são similares aos casos anteriores. Em todos os casos, podemos descrever σ como $\iota \cdot t(i', j', k')$.

Tendo provado a base da indução, $d_t(\sigma) \leq d_t(\pi)$, assumiremos agora a hipótese $d_t(\sigma') \leq d_t(\pi')$ para quaisquer permutações toricamente equivalentes π' e σ' , onde $d_t(\pi') = d$ para algum d fixo, e provaremos que a hipótese permanece válida para σ e π toricamente equivalentes tais que $d_t(\pi) = d + 1$. Para isso, encontramos uma permutação π' , com $d_t(\pi') = d_t(\pi) - 1 = d$, tal que $\pi' = \pi \cdot t(i, j, k)$ para alguma transposição $t(i, j, k)$. Uma análise em casos idêntica à efetuada para a base de indução, nos permite encontrar uma transposição $t(i', j', k')$ tal que $\sigma' = \pi' \cdot t(i', j', k') = \sigma \cdot t(i', j', k')$, logo $d_t(\sigma') = d_t(\pi') = d$, portanto $d_t(\sigma) \leq d + 1$. ■

O teorema 3.29 permite-nos simplificar o grafo de rearranjos por transposições, agrupando vértices de maneira que as distâncias sejam preservadas. Chamaremos esse grafo simplificado de *grafo tórico*.

Definição 3.30 (grafo tórico). [10] *Seja o grafo $Tor(n) = (V, E)$, denominado grafo tórico de ordem n , onde cada vértice de V corresponde a uma classe de equivalência tórica das permutações de n elementos, e cujo conjunto de arestas é:*

$$E = \{\Pi\Sigma; \Pi, \Sigma \in V \text{ e existe aresta } \pi\sigma \text{ em } TRG(n), \text{ onde } \pi \in \Pi \text{ e } \sigma \in \Sigma\}$$

Corolário 3.31. [10] *Seja Π um vértice de $Tor(n)$, $\pi \in \Pi$ uma permutação que está nessa classe de equivalência, e I a classe de equivalência da permutação $\iota_{[n]}$. Então a distância $d(\Pi, I)$, em $Tor(n)$, equivale à distância $d(\pi, \iota_{[n]})$ em $TRG(n)$.*

Prova. Imediata, a partir do teorema 3.29 e da definição 3.30. ■

Usaremos $[\pi_1 \pi_2 \dots \pi_n]_{\circ}^{\circ}$ para denotar a classe equivalência tórica de $\pi = (\pi_1 \pi_2 \dots \pi_n)$. Notemos que, em geral, uma classe de equivalência tórica pode ser denotada de mais de uma maneira; usaremos, então, a que se apresentar mais conveniente em cada caso. A figura 3.9 exhibe os grafos tóricos para n entre 3 e 5.

Analisemos o tamanho do grafo tórico $Tor(n)$, relativamente ao grafo de rearranjos por transposições $TRG(n)$. Pela definição 3.26, um deslocamento cíclico $m + \pi$ pode gerar, no máximo, $n + 1$ permutações diferentes (contando-se π como uma dessas permutações). Portanto, cada classe de equivalência tórica possui, no máximo, $n + 1$ elementos. Como a permutação identidade e a permutação reversa estão em duas classes que possuem apenas 1 elemento, um limitante inferior para o número de vértices de $Tor(n)$ é $\frac{n! - 2}{n + 1} + 2$. Infelizmente, o crescimento do número de vértices é exponencial em n , porém, mesmo assim, a redução do número de vértices torna possível analisar computacionalmente o grafo $Tor(n)$ em tempo menor do que $TRG(n)$, o que nos permite alcançar valores maiores para n .

Como já sabemos, nem todas as classes de equivalência tórica possuem $n + 1$ elementos, como, por exemplo, a classe da permutação identidade e a classe da permutação reversa. Porém, o número de elementos de uma classe de equivalência tórica deve ser, necessariamente, um divisor de $n + 1$, como demonstrado a seguir.

Teorema 3.32. *Sejam $\pi \in S_n$ e a sua respectiva classe de equivalência tórica π_{\circ}° . O número de elementos que pertencem a essa classe de equivalência tórica é um divisor de $n + 1$.*

Prova. Pela definição de deslocamento cíclico, é óbvio que $(n + 1) + \pi = \pi$. Considere o menor número natural m' tal que $m' + \pi = \pi$. Observe que o

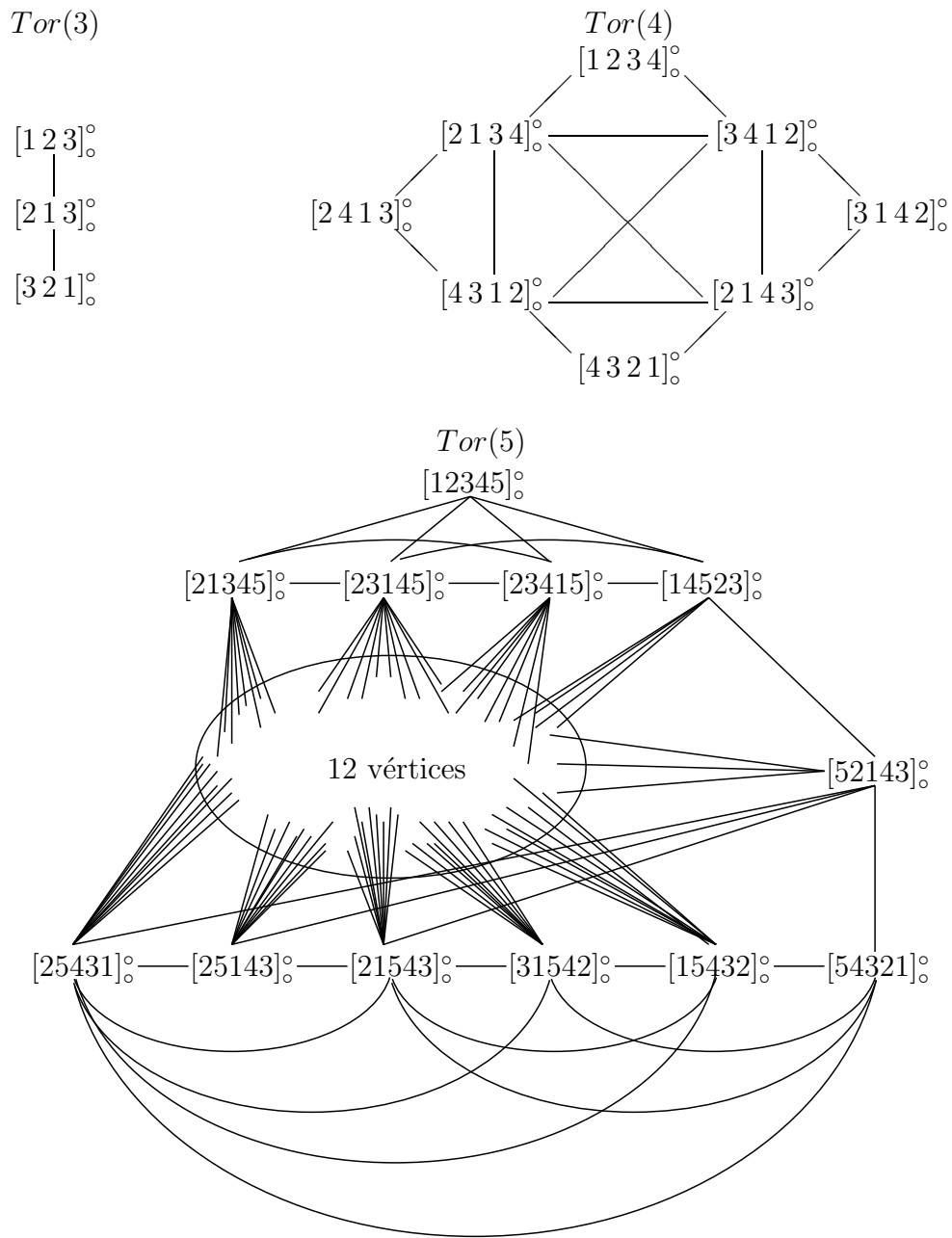


Figura 3.9: Os grafos tóricos $T(3)$, $T(4)$ e $T(5)$.

deslocamento cíclico de π por um múltiplo km' de m' é igual a π , pois

$$\begin{aligned} km' + \pi &= \underbrace{m' + \dots + m'}_k + \pi = \underbrace{m' + \dots + m'}_{k-1} + (m' + \pi) \\ &= \underbrace{m' + \dots + m'}_{k-1} + \pi = \dots \\ &= \pi. \end{aligned}$$

Suponha que m' não é divisor de $n + 1$; ou seja, $n + 1 = km' + r$ para inteiros k, r , onde $0 < r < m'$. Então,

$$(n + 1) + \pi = (r + km') + \pi = r + (km' + \pi) = r + \pi.$$

Como $(n + 1) + \pi = \pi$, chegamos à conclusão que $r + \pi = \pi$, mas isto contradiz o fato que m' é o menor valor tal que $m' + \pi = \pi$. Logo, m' deve necessariamente ser um divisor de $n + 1$. ■

Corolário 3.33. *Se $n + 1$ é primo, as classes de equivalência tórica de S_n possuem 1 ou $n + 1$ elementos.*

No caso em que $n + 1$ é primo, também é possível caracterizar as classes que possuem apenas 1 elemento. Diremos que uma classe de equivalência tórica é *unitária* se ela possui apenas 1 elemento. Por exemplo, para $n = 4$, existem exatamente 4 classes unitárias, a saber, aquelas que contêm uma das seguintes permutações: (1234), (2413), (3124) e (4321).

Teorema 3.34. *Se $n + 1$ é primo, uma classe de equivalência tórica é unitária se, e somente se, ela contém uma permutação da forma*

$$(0 \ \bar{k} \ \overline{2k} \ \overline{3k} \ \dots \ \overline{nk})^c,$$

onde k é um inteiro, $1 \leq k \leq n$.

Prova. Se $n+1$ é primo, então $\overline{\ell k} \not\equiv 0 \pmod{n+1}$. É fácil ver que uma classe de equivalência tórica que contém a permutação $(\overline{0} \ \overline{k} \ \overline{2k} \ \overline{3k} \ \dots \ \overline{nk})^c$ é unitária. Consideremos a permutação $1 + (\overline{0} \ \overline{k} \ \overline{2k} \ \overline{3k} \ \dots \ \overline{nk})^c$, e considere que $\overline{\ell k + 1} = 0$, para algum inteiro ℓ entre 1 e n :

$$\begin{aligned}
1 + (\overline{0} \ \overline{k} \ \overline{2k} \ \dots \ \overline{nk})^c &= \\
&= (\overline{0+1} \ \overline{k+1} \ \overline{2k+1} \ \dots \ \overline{\ell k+1} \ \overline{(\ell+1)k+1} \ \dots \ \overline{nk+1})^c \\
&= (\overline{\ell k+1} \ \overline{(\ell+1)k+1} \ \dots \ \overline{nk+1} \ \overline{0+1} \ \overline{k+1} \ \overline{2k+1} \ \dots \ \overline{(\ell-1)k+1})^c \\
&= (0 \ \overline{\ell k+k+1} \ \dots \ \overline{nk+1} \ \overline{0+1} \ \overline{k+1} \ \overline{2k+1} \ \dots \ \overline{\ell k-k+1})^c \\
&= (0 \ \overline{k} \ \overline{2k} \ \dots \ \overline{nk})^c.
\end{aligned}$$

Como $m + (\overline{0} \ \overline{k} \ \overline{2k} \ \dots \ \overline{nk})^c = \underbrace{1 + \dots + 1}_{m-1} + 1 + (\overline{0} \ \overline{k} \ \overline{2k} \ \dots \ \overline{nk})^c =$
 $\underbrace{1 + \dots + 1}_{m-1} + (\overline{0} \ \overline{k} \ \overline{2k} \ \dots \ \overline{nk})^c = \dots = (\overline{0} \ \overline{k} \ \overline{2k} \ \dots \ \overline{nk})^c$, então a classe de equivalência tórica de $(\overline{0} \ \overline{k} \ \overline{2k} \ \dots \ \overline{nk})^c$ é unitária.

Por outro lado, considere que $\pi^c = (0\pi_1\pi_2 \dots \pi_n)^c$ é a única permutação que pertence a uma classe unitária; isto implica que $m + \pi^c$ para qualquer inteiro m , em particular para $m = 1$. Ou seja,

$$(0\pi_1\pi_2 \dots \pi_n)^c = (\overline{1} \ \overline{\pi_1+1} \ \overline{\pi_2+1} \ \dots \ \overline{\pi_n+1})^c.$$

Seja ℓ tal que $\overline{\pi_\ell+1} = 0$. Então $\overline{\pi_{\ell+1}+1} = \pi_1$, $\overline{\pi_{\ell+2}+1} = \pi_2$, \dots , $\overline{\pi_{\ell+n}+1} = \pi_n$, ou seja, $\pi_{\ell+i} + 1 \equiv \pi_i \pmod{n+1}$ para todo inteiro i entre 1 e n . Como $\pi_i \neq \pi_j$ se $i \neq j$, e $1 \leq \pi_1 \leq n$ para todo i entre 1 e n , o sistema de equações modulares só pode ser satisfeito se $\pi_i = ik$ onde $k = \overline{\pi_{\ell+1}+1}$. ■

O teorema 3.34 permite-nos determinar com precisão o número de vértices de $Tor(n)$, no caso em que $n + 1$ é primo. Segundo o teorema, se $n + 1$ é primo, há exatamente n classes de equivalência tórica que são unitárias. Dessa forma, os restantes $n! - n$ vértices estão distribuídos em classes de tamanho $n + 1$, logo há $\frac{n! - n}{n + 1}$ classes de tamanho $n + 1$. O teorema 3.35 sumariza este resultado.

Teorema 3.35. *Se $n + 1$ é primo, $Tor(n)$ possui exatamente $\frac{n! - n}{n + 1} + n$ vértices.*

Discutiremos prováveis desenvolvimentos futuros deste resultado no capítulo 5.

Capítulo 4

Uma aplicação: encontrando reversões e transposições entre os genomas humano e do chimpanzé

Este capítulo tem o propósito de apresentar os resultados encontrados na comparação entre os genomas do ser humano (*Homo sapiens*) e do chimpanzé (*Pan troglodytes*), fornecendo dados que indiquem a relevância biológica do estudo dos rearranjos de genomas por reversões e transposições. Este trabalho foi desenvolvido pelo aluno durante o período de doutorado sanduíche, quando esteve sob a orientação do prof. David Sankoff na Universidade de Ottawa, Canadá, com auxílio da CAPES através da bolsa BEX1182/05-6.

Os primeiros estudos que indicaram a presença de rearranjos entre genomas de dois organismos distintos datam da década de 1930, quando Sturtevant e Dobzhansky [23] relataram a existência de reversões no terceiro cromossomo de diferentes indivíduos da espécie *Drosophila pseudoobscura* (mosca das frutas), descobertos a partir da comparação dos padrões de faixas (ou bandas) encontrados nos cromossomos, que podem ser vistos pela observa-

ção em microscópio das células provenientes das glândulas salivares desses animais. Os mesmos autores, em um artigo seguinte [24], descreveram reversões entre diferentes espécies de *Drosophila*, fornecendo inclusive prováveis seqüências de reversões que transformam segmentos de cromossomos de um indivíduo nos segmentos correspondentes em outro – vale ressaltar, como curiosidade, que os pesquisadores definiram o conceito de pontos de quebra (não com esse nome) e a sua relação com a distância de reversão mais de 40 anos antes [26, 18] que esse conceito viesse a ser utilizado rotineiramente no estudo de rearranjos de genomas.

Em estudos mais recentes [19, 11], realizados após a publicação dos genomas completos de diversos organismos, houve uma mudança radical na forma de se detectarem rearranjos, uma vez que tornou-se possível compararem-se diretamente as cadeias de nucleotídeos dos cromossomos dos animais a serem estudados. Esta é a abordagem utilizada neste trabalho, a qual descreveremos em detalhe no decorrer deste capítulo. Destacamos que, nesta tese, pela primeira vez são consideradas transposições. Em [19], o genoma humano e o do camundongo foram comparados, levando-se em conta reversões, fusões (união de dois cromossomos A e B em um cromossomo AB), fissões (divisão de um cromossomo AB em dois cromossomos A e B) e translocações (onde ocorre uma troca de conteúdo entre dois cromossomos AB e CD, resultando em AD e CB). Em [11], os autores compararam os genomas humano e do chimpanzé, buscando apenas reversões.

4.1 Formalização do problema

Descreveremos aqui formalmente o problema de comparação entre dois genomas de organismos distintos. Em linhas gerais, desejamos comparar

dois conjuntos de cadeias, onde cada cadeia é composta por letras do alfabeto $\{a, t, c, g\}$, e descobrir onde há subcadeias comuns, ignorando algumas diferenças menores. A partir dessas subcadeias comuns, desejamos encontrar regiões onde houve reversões ou transposições.

Definição 4.1 (nucleotídeo). *Um nucleotídeo é um elemento do conjunto $\{a, t, c, g\}$.*

Definição 4.2 (cromossomo). *Um cromossomo é uma cadeia de nucleotídeos. Se T é um cromossomo, denotaremos os nucleotídeos de sua cadeia por $T_1T_2\dots T_{\ell_T}$, onde ℓ_T é o comprimento desse cromossomo.*

Definição 4.3 (genoma). *Um genoma é um conjunto de cromossomos.*

É comum, em diferentes organismos, haver cromossomos que desempenham funções similares. Por vezes, partes diferentes de um cromossomo em uma espécie correspondem a cromossomos inteiros em outra espécie. Um mapeamento entre cromossomos similares em dois genomas é dito uma *relação de sintenia*. A tabela 4.1 mostra a relação de sintenia entre os cromossomos do ser humano e do chimpanzé. Pode-se observar que parte do cromossomo humano número 2 é mapeada no cromossomo 12 do chimpanzé e outra parte é similar ao cromossomo 13. Para simplificar, consideraremos os cromossomos 12 e 13 do chimpanzé como um só, cuja cadeia é a concatenação das cadeias dos cromossomos 12 e 13, nesta ordem.

Como estamos interessados apenas em encontrar reversões e transposições, que são eventos de mutação que acontecem dentro de um cromossomo, iremos comparar apenas cromossomos similares segundo a relação de sintenia entre eles.

Dados dois cromossomos T e Q que possuam uma relação de sintenia entre eles, precisamos de um modo de associar subcadeias de caracteres de T

Tabela 4.1: Relação de sintenia entre os cromossomos do ser humano e do chimpanzé [28].

Humano	Chimpanzé	Humano	Chimpanzé	Humano	Chimpanzé
1	1	9	11	17	19
2	12 e 13(*)	10	8	18	17
3	2	11	9	19	20
4	3	12	10	20	21
5	4	13	14	21	22
6	5	14	15	22	23
7	6	15	16	X	X
8	7	16	18	Y	Y

(*) ver observação no texto.

a subcadeias de caracteres de Q , de tal forma que possamos identificar regiões similares entre elas. Daremos o nome de *alinhamento* a essa associação.

Definição 4.4 (alinhamento). *Sejam Q e T duas cadeias, com seus respectivos comprimentos ℓ_Q e ℓ_T . Um alinhamento é uma função parcial¹*

$$f : \{1, 2, \dots, \ell_Q\} \rightarrow \{1, 2, \dots, \ell_T\},$$

que mapeia uma posição i de Q na posição $f(i)$ de T .

Definição 4.5 (gráfico de pontos). *Sejam Q e T duas cadeias, com seus respectivos comprimentos ℓ_Q e ℓ_T , e f um alinhamento entre Q e T . Um gráfico de pontos (em inglês, dot plot) é uma matriz $[m_{i,j}]$, de ℓ_Q linhas por ℓ_T colunas, onde cada elemento da matriz é*

$$m_{i,j} = \begin{cases} 1, & \text{se } f(i) = j \\ 0, & \text{caso contrário.} \end{cases}$$

¹Uma função parcial é uma relação binária que associa elementos de um conjunto, denominado *domínio*, a no máximo 1 elemento de outro conjunto, denominado *contra-domínio*. Diferentemente de uma função comum, uma função parcial pode não estar definida para todo elemento do domínio.

O gráfico de pontos é uma representação gráfica de um alinhamento entre duas cadeias. No caso em que duas cadeias de mesmo comprimento alinharem-se perfeitamente, ou seja, a função de alinhamento é $f(i) = i$ para todo i , o gráfico de pontos é uma matriz diagonal.

Se duas cadeias Q, T têm o mesmo comprimento, e Q alinha-se com a cadeia reversa de T , ou seja, a função de alinhamento entre Q e T é $f(i) = \ell_T - i + 1$, o gráfico de pontos é a matriz

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ & \ddots & & \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

As figuras 4.1, 4.2 e 4.3 mostram exemplos de gráficos de pontos, onde estamos comparando duas cadeias no alfabeto $\{\mathbf{a}, \dots, \mathbf{i}\}$. Os rótulos $\mathbf{a}, \dots, \mathbf{i}$ das colunas e linhas dos gráficos de pontos foram colocados apenas para facilitar a visualização.

Em geral, um alinhamento entre duas cadeias pode conter espaços, ou “*gaps*” (fig. 4.1), reversões (fig. 4.2) e transposições (fig. 4.3) de subcadeias. A reversão de uma cadeia $T_1T_2 \dots T_{\ell_T-1}T_{\ell_T}$ é a cadeia $T_{\ell_T}T_{\ell_T-1} \dots T_2T_1$. Uma transposição de uma cadeia $T_1T_2 \dots T_{j-1}T_j \dots T_{\ell_T-1}T_{\ell_T}$ é uma troca na posição de dois blocos de caracteres: $T_j \dots T_{\ell_T-1}T_{\ell_T}T_1T_2 \dots T_{j-1}$. Portanto, o gráfico de pontos de um alinhamento poderá conter várias diagonais, e é a partir da análise dessas diagonais que iremos detectar onde ocorrem as reversões e transposições, como veremos a seguir.

Cadeias: $Q = abdefgi$, $T = abcd fghi$.

Alinhamento:

$$\begin{array}{cccccccc} a & b & & d & e & f & g & & i \\ \downarrow & \downarrow & & \downarrow & & \downarrow & \downarrow & & \downarrow \\ a & b & c & d & & f & g & h & i \end{array}$$

Gráfico de pontos:

	a	b	c	d	f	g	h	i
a	1	0	0	0	0	0	0	0
b	0	1	0	0	0	0	0	0
d	0	0	0	1	0	0	0	0
e	0	0	0	0	0	0	0	0
f	0	0	0	0	1	0	0	0
g	0	0	0	0	0	1	0	0
i	0	0	0	0	0	0	0	1

Figura 4.1: Exemplo de gráfico de pontos para um alinhamento com espaços (*gaps*).

Cadeias: $Q = abcdefghi$, $T = abfedcghi$.

Alinhamento:

$$\begin{array}{cccccccc} a & b & c & d & e & f & g & h & i \\ \downarrow & \downarrow & & \swarrow & \searrow & & \downarrow & \downarrow & \downarrow \\ a & b & f & e & d & c & g & h & i \end{array}$$

Gráfico de pontos:

	a	b	f	e	d	c	g	h	i
a	1	0	0	0	0	0	0	0	0
b	0	1	0	0	0	0	0	0	0
c	0	0	0	0	0	1	0	0	0
d	0	0	0	0	1	0	0	0	0
e	0	0	0	1	0	0	0	0	0
f	0	0	1	0	0	0	0	0	0
g	0	0	0	0	0	0	1	0	0
h	0	0	0	0	0	0	0	1	0
i	0	0	0	0	0	0	0	0	1

Figura 4.2: Exemplo de gráfico de pontos onde houve uma reversão na ordem dos elementos em uma das cadeias.

Cadeias: $Q = abcdefghi$, $T = abfcdehi$.

Alinhamento:

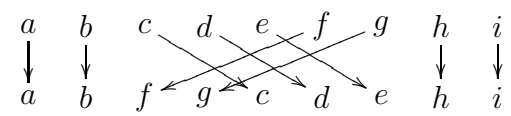


Gráfico de pontos:

	a	b	f	g	c	d	e	h	i
a	1	0	0	0	0	0	0	0	0
b	0	1	0	0	0	0	0	0	0
c	0	0	0	0	1	0	0	0	0
d	0	0	0	0	0	1	0	0	0
e	0	0	0	0	0	0	1	0	0
f	0	0	1	0	0	0	0	0	0
g	0	0	0	1	0	0	0	0	0
h	0	0	0	0	0	0	0	1	0
i	0	0	0	0	0	0	0	0	1

Figura 4.3: Exemplo de gráfico de pontos onde houve uma transposição na ordem dos elementos em uma das cadeias.

4.2 O método

Os alinhamentos no formato *net* entre as cadeias de DNA do ser humano (versão 17 de maio/2004) e do chimpanzé (versão 1, novembro/2003) foram obtidos do sítio *web* do projeto genoma da Universidade da Califórnia, Santa Cruz (<http://hgdownload.cse.ucsc.edu/goldenPath/hg17/vsPanTro1/>). O diagrama encontrado na figura 4.4 descreve de maneira sucinta como os dados são processados para a identificação das reversões e transposições.

O formato *net* descreve um alinhamento, obtido com o programa BLASTZ [20], entre uma cadeia alvo (*target*), neste caso, a do ser humano, e uma cadeia de busca (*query*), a do chimpanzé. Em resumo, o programa BLASTZ divide a cadeia alvo em um número arbitrário de segmentos contíguos de mais alto nível (ao qual chamaremos de nível 1), com comprimentos variáveis, de tal forma que cada segmento possua um alinhamento de pontuação máxima com algum segmento – possivelmente, com *gaps* – da cadeia de busca. Após isso, *gaps* são inseridos nos alinhamentos do nível anterior e outro conjunto de segmentos, cujo nível é 1 unidade maior do que o nível anterior, é alinhado entre esses *gaps*. Esse procedimento é repetido um número finito de vezes, por exemplo, 7 vezes, como é o caso dos dados mencionados anteriormente.

Os dados podem ser representados como uma coleção de tuplas:

$$S = (S.tipo, S.nível, S.t_{chr}, S.t_{início}, S.t_{fim}, S.q_{chr}, S.q_{início}, S.q_{fim}, S.fita),$$

onde $S.tipo \in \{fill, gap\}$ representa o tipo do segmento, se é um alinhamento (*fill*) ou um espaço (*gap*); $S.nível \in \mathbb{N}$ é o nível do segmento; $S.t_{chr}$ representa o cromossomo do genoma alvo (neste caso, ser humano) em que o alinhamento

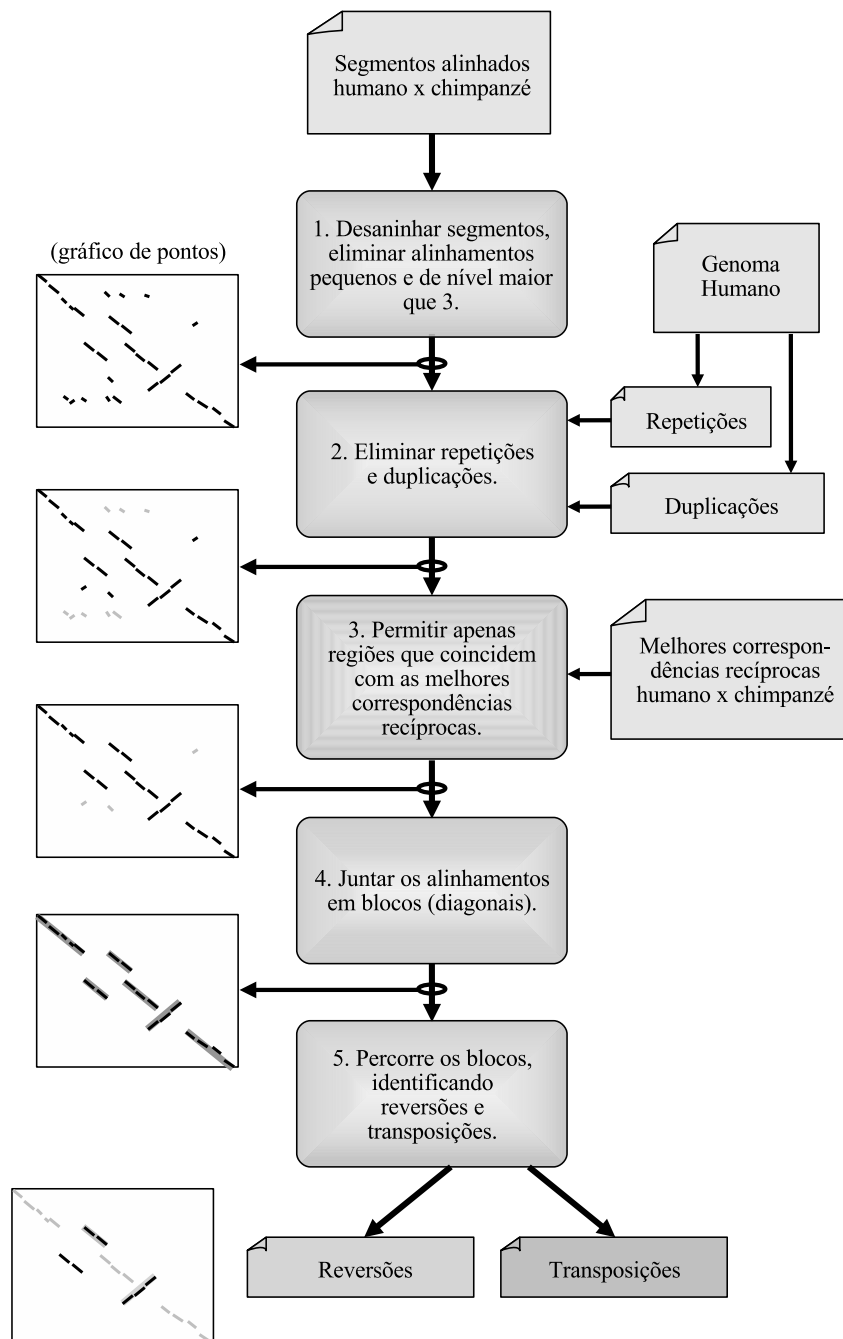


Figura 4.4: Diagrama esquemático do método adotado para identificar reversões e transposições entre os genomas humano e do chimpanzé.

é encontrado; $S.t_{início}$ e $S.t_{fim}$ são dois inteiros tais que $S.t_{início} < S.t_{fim}$, que indicam, respectivamente, as posições de início e fim² na cadeia do cromossomo $S.t_{chr}$ onde este alinhamento é encontrado; $S.q_{chr}$ é o cromossomo do genoma de busca (chimpanzé) no qual este segmento é encontrado; $S.q_{início}$ e $S.q_{fim}$ são as posições de início e fim do segmento na cadeia do cromossomo $S.q_{chr}$ onde este alinhamento é encontrado; e $S.fita \in \{+, -\}$ especifica se a subcadeia $T_{S.t_{início}} \dots T_{S.t_{fim}}$ é alinhada com a subcadeia $Q_{S.q_{início}} \dots Q_{S.q_{fim}}$ ($S.fita = "+"$) ou com a subcadeia reversa $Q_{S.q_{fim}} \dots Q_{S.q_{início}}$ ($S.fita = "-"$). As tuplas estão ordenadas em ordem crescente por $S.t_{início}$. Por exemplo, parte do alinhamento compreendido entre as posições 1025 e 245432976 no cromossomo 1 do ser humano é descrita desta forma:

$$\begin{aligned} & (fill, 1, chr1, 1025, 245432976, chr1, 1033, 229565299, +) \\ & (gap, 1, chr1, 52410, 54421, chr1, 60827, 62696, +) \\ & (fill, 2, 52820, 52926, chr1, 50599, 50705, +) \end{aligned}$$

Ou seja, a subcadeia entre as posições 1025 e 245432976 do cromossomo 1 do ser humano alinha-se com a subcadeia entre as posições 1033 e 229565299 do cromossomo 1 do chimpanzé. Entre as posições 52410 e 54421 do cromossomo 1 do ser humano e 60827 e 62696 do cromossomo 1 do chimpanzé, há um espaço (*gap*) no alinhamento. Parte desse espaço no alinhamento cromossomo 1 do ser humano é preenchido pelo alinhamento de nível 2 entre as posições 52820 e 52926 do cromossomo humano e as posições 50599 e 50705 do cromossomo do chimpanzé. O espaço no alinhamento divide o alinhamento de nível 1 em duas partes, a saber, uma parte à esquerda, com-

²Nota de implementação: os índices de início e fim nos dados obtidos do sítio *web* do projeto genoma da UCSC começam em 0, ou seja, o índice do primeiro caractere em uma cadeia é 0. Para lidarmos com índices começando em 1, adicionamos 1 unidade a cada índice.

preendida entre 1025 e 52409 (ser humano, cromossomo 1) e 1033 e 60826 (chimpanzé, cromossomo 1), e uma parte à direita, compreendida entre 54422 e 245432976 (ser humano, cromossomo 1) e 62697 e 229565299 (chimpanzé, cromossomo 1).

Descreveremos, a seguir, todos os passos do método, na ordem em que aparecem no diagrama da figura 4.4.

Passo 1: Neste passo, os segmentos alinhados estão aninhados, onde cada segmento no nível $i + 1$ está contido em um segmento do nível i . Para que o próximo passo do método seja executado, é necessário “desaninhar” os segmentos, ou seja, se um segmento X está contido em um segmento Y , devemos quebrar o segmento Y em duas partes, Y_1 e Y_2 , de tal forma que X não esteja contido em nenhuma das partes. Além disso, como desejamos descartar mudanças em escala muito pequena (que podem ser devidas a erros no seqüenciamento ou mutações de nucleotídeos que ocorrem comumente) devemos descartar alinhamentos pequenos ou cujo nível é muito grande. De modo a podermos comparar os resultados neste estudo àqueles encontrados em [11], descartamos segmentos em níveis superiores a 3 ou mais curtos do que 20 nucleotídeos. O procedimento para executar essa operação é descrito no algoritmo 3.

Faremos uma análise simples da complexidade temporal do algoritmo 3. As operações entre as linhas 5 e 18 são executadas 1 vez para cada segmento. Todas as operações compreendidas nesse laço são executadas em tempo $O(1)$; o laço entre as linhas 8 e 11 é executado um número fixo de vezes, uma vez que estamos limitando os níveis de aninhamento e o número de segmentos em D_{dir} depende da quantidade de níveis. Logo, o algoritmo é executado em tempo linear no número de segmentos.

Algoritmo 3 Desaninhamento dos segmentos.

- 1: seja D uma lista duplamente encadeada (*deque*) $\{D$ conterá os segmentos desaninhados $\}$
 - 2: **para todo** cromossomo alvo t_{chr} **faça**
 - 3: seja A o conjunto de segmentos alinhados ao cromossomo alvo t_{chr} , ordenados por seu índice t_{start}
 - 4: seja D_{dir} uma lista duplamente encadeada $\{D_{dir}$ conterá uma lista dos segmentos que estão à direita de S no cromossomo t_{chr} $\}$
 - 5: **para todo** segmento S em A **faça**
 - 6: **se** $S.nível \leq 3$ ou $comprimento_em_alvo(S) \geq 20$ ou $comprimento_em_busca(S) \geq 20$ **então**
 - 7: $S_{dir} \leftarrow primeiro_elemento(D_{dir})$
 - 8: **enquanto** $S_{dir}.t_{start} < S.t_{start}$ **faça** $\{$ se S_{dir} está à esquerda de S $\}$
 - 9: insira S_{dir} ao final de D
 - 10: remova S_{dir} de D_{dir}
 - 11: $S_{dir} \leftarrow primeiro_elemento(D_{dir})$
 - 12: **se** $S.tipo = fill$ **então**
 - 13: insira S no final de D
 - 14: **senão** $\{S.tipo = gap\}$
 - 15: $C \leftarrow remove_último(D)$ $\{C$ é o segmento que contém S $\}$
 - 16: sejam C_{esq}, C_{dir} as duas partes de C após inserir o *gap* representado por S , tais que $C_{esq}.t_{início} < C_{esq}.t_{fim} \leq S.t_{início} < S.t_{fim} \leq C_{dir}.t_{início} < C_{dir}.t_{fim}$
 - 17: insira C_{esq} ao final de D
 - 18: insira C_{dir} ao final de D_{dir}
 - 19: insira em D todos os segmentos que restaram em D_{dir}
 - 20: **retorna** D
-

Passo 2: após desaninhar os segmentos no formato *net*, são eliminados os segmentos cuja maior parte do conteúdo é de repetições e/ou duplicações. Uma *repetição* é uma subcadeia que ocorre mais de uma vez consecutiva em uma cadeia (exs.: aa...a e cagcag...cag). Uma duplicação é uma subcadeia de comprimento maior do que 1 que ocorre mais de uma vez em uma cadeia (por exemplo, em acgattagctcattac a subcadeia atta é uma duplicação). Grande parte do DNA humano é de repetições e/ou duplicações [22].

As informações sobre repetições foram obtidas do sítio *web* do grupo de bioinformática da UCSC³, e as informações sobre duplicações foram retiradas do banco de dados “Human Genome Segmental Duplication Database”, do Centro de Genômica Aplicada do Canadá⁴. Os dados foram transformados em uma coleção de tuplas $D = (D.t_{chr}, D.t_{início}, D.t_{fim})$ descrevendo regiões no genoma alvo que são repetições/duplicações, onde $D.t_{chr}$ é o cromossomo do genoma alvo e $D.t_{início}, D.t_{fim}$ são respectivamente os índices de início e fim na cadeia do cromossomo alvo. A interseção entre cada segmento alinhado e os segmentos repetidos/duplicados é calculada e, se o tamanho da interseção dividido pelo comprimento do alinhamento está acima de um determinado limiar, o alinhamento é descartado. No nosso caso, usamos 0.2 como limite (assim como [11]); ou seja, se mais de 20% do conteúdo de um segmento alinhado é de repetições ou duplicações, ele é descartado. No nosso estudo, é importante descartar repetições e duplicações pois eles podem gerar alinhamentos espúrios, apesar de estarmos perdendo alguma informação.

Passo 3: neste passo, os segmentos alinhados são comparados com as melhores correspondências recíprocas (*reciprocal best matches*) entre as cadeias alvo e de busca. As melhores corres-

³<http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/>, arquivos *_rnsk.txt.gz

⁴<http://projects.tcag.ca/humandup>

pondências recíprocas são representadas por tuplas na forma $R = (R.t_{chr}, R.t_{início}, R.t_{fim}, R.q_{chr}, R.q_{início}, R.q_{fim}, R.fita)$, de tal forma que o alinhamento de pontuação máxima para a subcadeia $T_{R.t_{início}} \dots T_{R.t_{fim}}$ no cromossomo alvo $R.t_{chr}$ é aquele que se alinha à subcadeia $Q_{R.q_{início}} \dots Q_{R.q_{fim}}$ no cromossomo de busca $R.q_{chr}$, e *vice-versa*. Cada segmento alinhado é verificado e, caso não esteja contido em uma melhor correspondência recíproca, ele é descartado.

Passo 4: os segmentos restantes são agrupados em blocos, de tal forma que segmentos no mesmo bloco estão próximos da mesma diagonal no gráfico de pontos (fig. 4.5e). Um bloco é representado por uma tupla

$$B = (B.t_{chr}, B.t_{início}, B.t_{fim}, B.q_{chr}, B.q_{início}, B.q_{fim}, B.fita).$$

Como os segmentos nem sempre seguem perfeitamente uma diagonal, pois podem haver *gaps* no alinhamento (como na fig. 4.1), usamos uma abordagem heurística para montar os blocos. Esta abordagem, descrita no parágrafo a seguir, tenta imitar o modo como alguém reconhece, intuitivamente, uma diagonal em um grupo de segmentos próximos no gráfico de pontos.

Para cada cromossomo alvo, um novo bloco é criado e o segmento S_1 mais à esquerda no gráfico de pontos (ou seja, aquele em que $S_1.t_{início}$ possui o menor valor) é adicionado a esse bloco. Neste momento, o bloco é descrito por

$$B = (S_1.t_{chr}, S_1.t_{início}, S_1.t_{fim}, S_1.q_{chr}, S_1.q_{início}, S_1.q_{fim}, S_1.fita).$$

Consideramos, agora, uma reta que passa pelos pontos cujas coordenadas são $(B.t_{início}, B.q_{início})$ e $(B.t_{fim}, B.q_{fim})$, e o lugar geométrico dos pontos cuja distância à reta é menor do que $w/2$ e que estão a, no máximo, uma

distância⁵ d do ponto $(B.t_{fim}, B.q_{fim})$ – veja fig. 4.5a. Se o próximo segmento S_2 é tal que $S_2.fita = S_1.fita$, e $S_2.q_{chr} = S_1.q_{chr}$, e S_2 está dentro da região descrita, ele é adicionado ao bloco B , que torna-se, então

$$B = (S_1.t_{chr}, S_1.t_{início}, \mathbf{S_2.t_{fim}}, S_1.q_{chr}, S_1.q_{início}, \mathbf{S_2.q_{fim}}, S_1.fita),$$

(fig. 4.5b); caso contrário, outro bloco é criado e adicionamos S_2 a ele (figs. 4.5c e 4.5d). Este procedimento é repetido para todos os segmentos alinhados no cromossomo alvo, por ordem crescente de seus índices $t_{início}$ (fig. 4.5e).

Passo 5: O último passo no nosso método é identificar as reversões e transposições a partir dos blocos previamente montados. Isso é facilmente realizado, para cada cromossomo alvo, percorrendo-se os blocos por ordem crescente de seus índices $t_{início}$. Sejam B_a , B_b e B_c três blocos consecutivos no mesmo cromossomo alvo; se $B_a.fita = B_c.fita = +$, $B_c.fita = -$, e $B_a.q_{início} < B_b.q_{início} < B_c.q_{início}$, então B_b é um bloco onde houve uma reversão; se $B_a.fita = B_b.fita = B_c.fita = +$, $B_a.q_{início} < B_c.q_{início}$, mas $B_b.q_{início} < B_a.q_{fim}$, então houve uma transposição que moveu o bloco B_b (veja fig. 4.5f).

4.3 Resultados

Foram encontradas 542 reversões (contrastando com as 1576 reversões em [11]) e 255 transposições. Os gráficos 4.7 e 4.6 mostram a distribuição dos tamanhos dos segmentos que sofreram reversão e dos que sofreram trans-

⁵Nota de implementação: os valores w e d são parâmetros pré-fixados. Usamos, para este trabalho, $d = 200000$ e $w = 150000$.

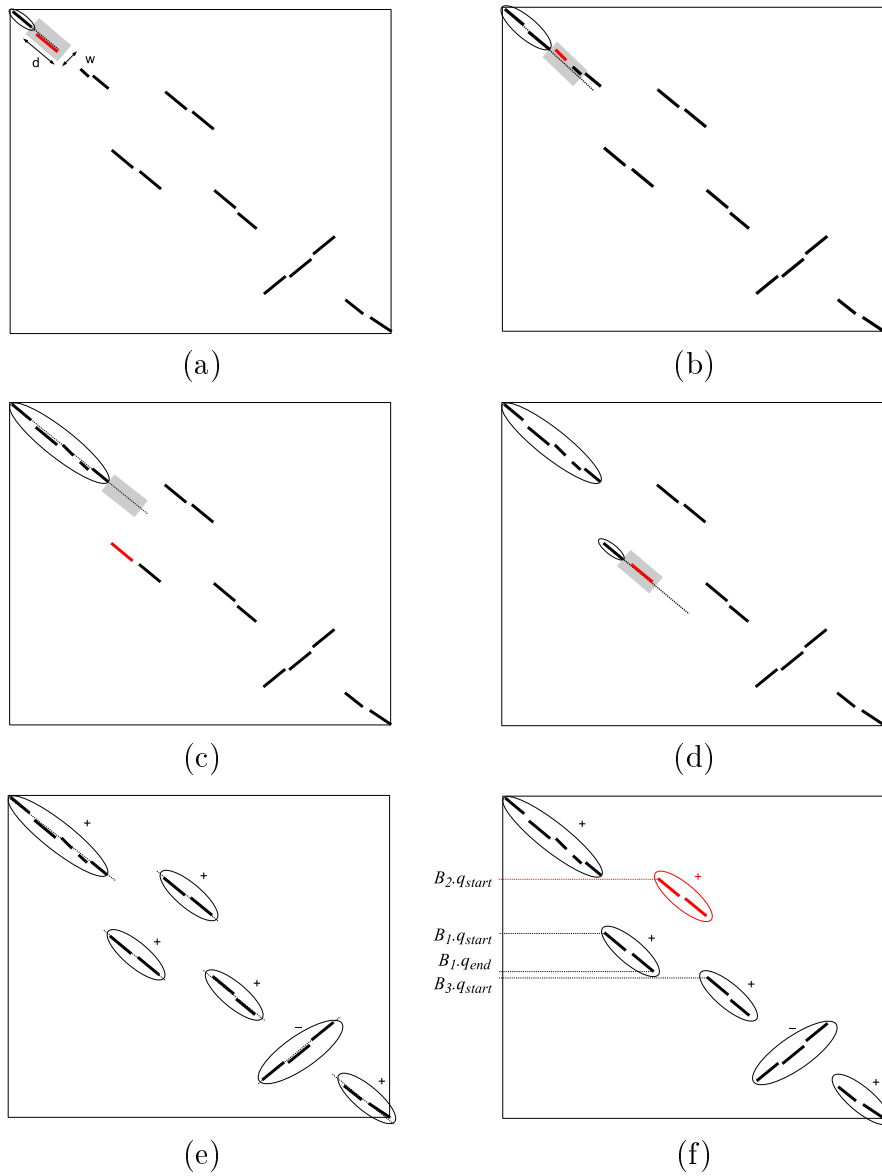


Figura 4.5: Agrupando segmentos alinhados em blocos (diagonais): (a) inicie com o segmento mais à esquerda; (b) adicionando segmentos ao bloco; (c) próximo alinhamento está fora da região; (d) um novo bloco é criado; (e) blocos encontrados após a execução do algoritmo; (f) identificação de reversões e transposições.

posição, respectivamente. A listagem com todas as reversões e transposições encontradas está no apêndice B.

Vemos que as reversões encontradas neste trabalho, em muitos casos, coincidem com as encontradas em [11]. Por exemplo, a reversão entre as posições 18313193 e 26198378 do cromossomo humano 15 (posições 20589043 a 28367385 do cromossomo 16 do chimpanzé) aparece no trabalho de Feuk *et al.* como entre as posições 18313193 e 26244831 do mesmo cromossomo humano (posições 20542252 a 28367385 do cromossomo 16 do chimpanzé). Observa-se, em alguns casos, algumas reversões encontram-se “divididas” em várias partes, como as 9 reversões entre as posições 34441312 e 95939805 do cromossomo 5 do ser humano (posições 19227253 a 81490988 do cromossomo 4 do chimpanzé), e que em [11] está listada como apenas uma reversão. É importante notar que, no trabalho de Feuk *et al.*, há uma fase de validação manual dos resultados, em que os autores uniram vários segmentos em um só, enquanto que os resultados obtidos nesta tese são obtidos puramente pela execução computacional do método.

Ressaltamos também que a distribuição do tamanho das reversões encontradas neste trabalho difere de [11]. No nosso trabalho, grande parte das reversões encontradas correspondem a segmentos alinhados com comprimento entre 1000 e 10000 nucleotídeos no ser humano, enquanto que Feuk *et al.* encontraram grande parte das reversões com comprimento de até 250 nucleotídeos. Essa diferença é devida aos parâmetros d e w usados por nós no passo 4 do método. Valores menores para d e w fariam com que a distribuição das reversões e transposições tendesse mais para os comprimentos menores, porém isso causaria um número ainda maior de reversões “quebradas” em pequenos segmentos. Os parâmetros d e w foram ajustados de forma que as reversões encontradas coincidissem da melhor forma possível com as do

trabalho de Feuk *et al.* [11], sem termos que recorrer a ajustes manuais nos resultados após a execução do método, de maneira que os resultados entre os dois trabalhos fossem comparáveis.

O método descrito nesta tese segue as mesmas linhas gerais do método de Feuk *et al.*, com a diferença que estamos levando em conta transposições no passo 5. Procuramos também descrever de maneira mais detalhada todos os passos da execução. O passo 4, em especial, praticamente não é descrito em [11], o que obrigou-nos a desenvolver um novo procedimento. Acreditamos que este procedimento seja ligeiramente melhor do que o original, uma vez que o fato de que encontramos menos reversões, com distribuição de tamanho tendendo a segmentos maiores, indica que muitos segmentos que eram descritos separadamente no artigo original foram unidos em um único segmento neste trabalho, de maneira automática. Um exemplo disso é a reversão compreendida entre as posições 41673868 e 55003655 do cromossomo humano 17 (posições 50169262 a 63465787 do cromossomo 19 do chimpanzé), que em [11] está separada em duas partes, de 41673868 a 41733269 e de 54956779 a 55003655 no cromossomo humano (respectivamente, posições 63397185 a 63465787 e posições 50169262 a 50216602 no cromossomo correspondente do chimpanzé).

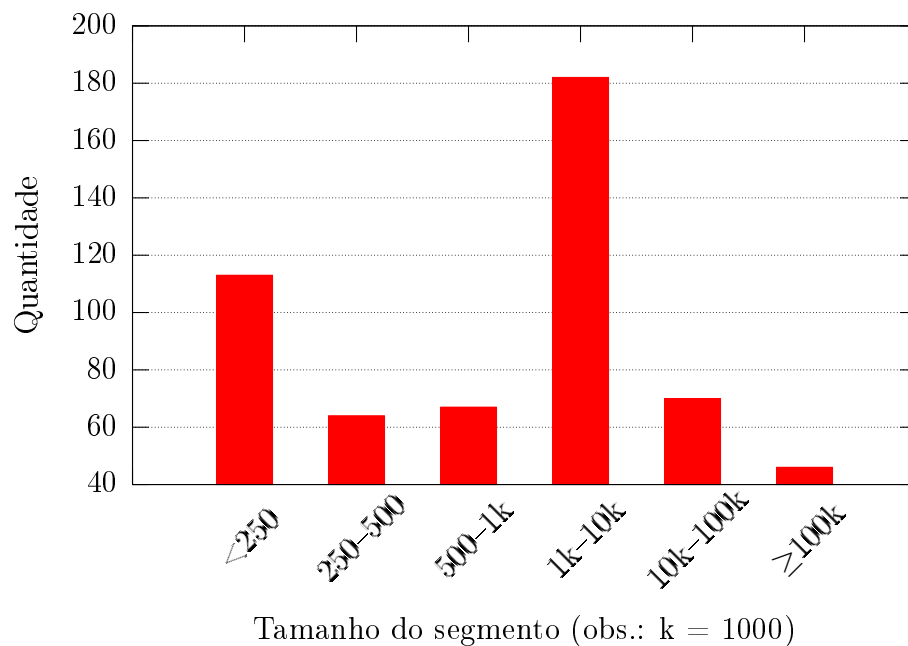


Figura 4.6: Distribuição dos tamanhos dos segmentos que sofreram reversão.

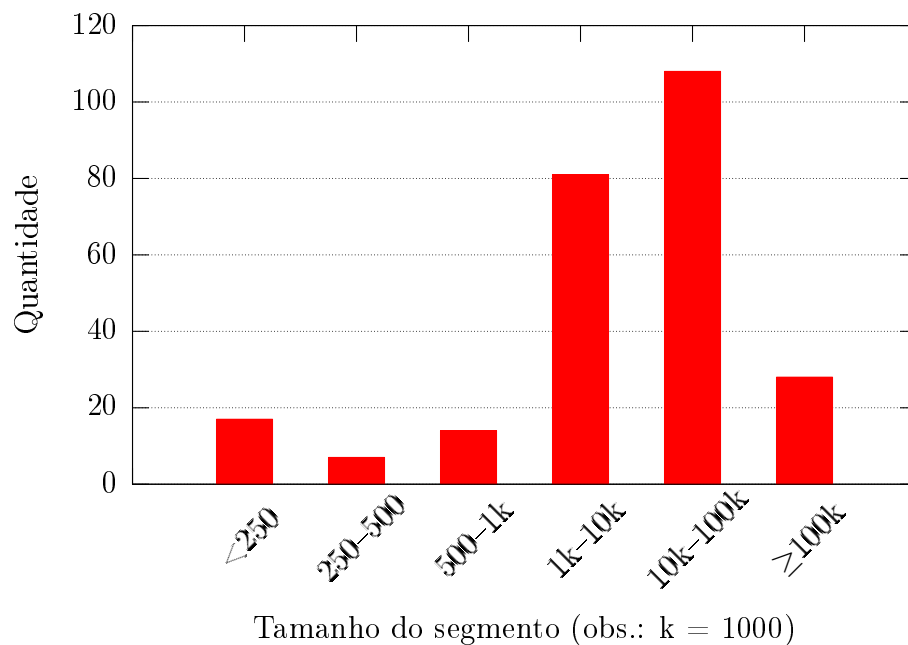


Figura 4.7: Distribuição dos tamanhos dos segmentos que sofreram transposição.

Capítulo 5

Conclusão

Delinearemos, nesta parte final, as realizações alcançadas nesta tese e os caminhos que poderão ser trilhados a partir dos desenvolvimentos deste trabalho.

No capítulo 2, revisamos dois limites inferiores para a distância de transposição encontrados na literatura, um baseado em pontos de quebra e o outro baseado na estrutura do diagrama de realidade e desejo. Na demonstração do limite inferior usando o diagrama de realidade e desejo, descrevemos formalmente a estrutura dos ciclos desse diagrama e como uma transposição pode alterar esses ciclos, usando as figuras apenas como meio de auxílio ao entendimento, e não como principal meio para a demonstração. Ressaltamos que não encontramos referências anteriores onde essa análise dos ciclos do diagrama de realidade e desejo para o problema de rearranjos por transposições é feita com tal rigor.

Apresentamos no capítulo 3 o problema de rearranjos por transposições como um problema de teoria dos grafos. Somente encontramos uma única referência na literatura, devida a Eriksson *et al.* [10], em que tal abordagem é utilizada. Nessa referência, a abordagem mostrou-se frutífera, uma vez que

os autores conseguiram uma demonstração alternativa para o fato de que a distância de transposição entre a permutação reversa $\rho_{[n]}$ e a identidade $\iota_{[n]}$ é de $\lfloor n/2 \rfloor + 1$ através do uso da equivalência tórica, além de terem determinado o diâmetro de transposição para o conjunto S_n das permutações de n elementos para $13 \leq n \leq 15$, através de uma busca no grafo tórico $Tor(n)$, que reduziu o número de permutações que deveriam ser examinadas para determinar-se o diâmetro. No processo de determinação do diâmetro de S_{13} e S_{15} , eles também encontraram uma permutação que está a uma distância maior da identidade do que a reversa, invalidando então uma conjectura de Meidanis, Walter e Dias [16].

Curiosamente, no artigo de Eriksson *et al.*, os autores descrevem muito sucintamente o grafo tórico, e não apresentam nenhuma estimativa do número de vértices desse grafo. Nesta tese, damos uma definição detalhada do grafo tórico e estimamos que ele possui ao menos $\frac{n!-2}{n+1} + 2$ vértices; mais ainda, para o caso em que $n + 1$ é primo, determinamos que $Tor(n)$ possui exatamente $\frac{n!-n}{n+1} + n$ vértices.

Ainda restam muitas perguntas em aberto sobre o problema de rearranjos por transposições. Até o presente momento, não se encontrou nenhum algoritmo polinomial para ordenar uma permutação usando apenas transposições, nem se demonstrou que o problema é *NP*-completo.

A determinação do diâmetro de transposição é outro problema que ainda não foi completamente solucionado, uma vez que não se sabe nenhuma fórmula geral para $D_t(n)$ (caso ela exista). Sabemos, devido ao fato de que $d_t(\rho_{[n]}, \iota_{[n]}) = \lfloor n/2 \rfloor + 1$, que o diâmetro de transposição não é inferior a $\lfloor n/2 \rfloor + 1$. Os resultados de Eriksson *et al.* [10] e Elias e Hartman [9] indicam que $D_t(n)$ é estritamente maior do que $\lfloor n/2 \rfloor + 1$, nos casos em que

$n \geq 13$ é ímpar. Para valores pares de $n \geq 16$, não se sabe se o diâmetro de transposição é igual ou maior do que o limite inferior $\lfloor n/2 \rfloor + 1$.

Creemos estar próximos de determinar o diâmetro de transposição $D_t(16)$, uma vez que, pelo fato de 17 ser primo, sabemos um pouco mais sobre a estrutura do grafo tórico $Tor(16)$: ele possui exatos 16 vértices que correspondem a classes de equivalência tórica com apenas 1 elemento, e $\frac{16!-16}{17}$ vértices que correspondem a classes com 17 elementos. Em nossos estudos, para todos os casos testados em que $n+1$ é primo, no grafo de rearranjos por transposições $TRG(n)$ os caminhos mínimos entre o vértice correspondente à identidade $\iota_{[n]}$ e a um vértice diametral σ não passam por vértices que estão em classes de equivalência tórica que contém apenas 1 elemento, a não ser, é claro, o primeiro vértice $\iota_{[n]}$ e o último vértice σ . Por exemplo, no $TRG(4)$, qualquer caminho mínimo entre o vértice $\iota_{[4]}$ e o vértice $\rho_{[4]}$, que faz par diametral com $\iota_{[4]}$, passa por vértices que pertencem a classes de equivalência tórica contendo 5 elementos, exceto, claro, pelo primeiro vértice $\iota_{[4]}$ e pelo último vértice $\rho_{[4]}$.

Outra característica que pudemos observar, para todos os pares de vértices π, σ que nós testamos, se π e σ estão em dois $TRGs$ distintos $TRG_{\pi_1}(n)$ e $TRG_{\sigma_1}(n)$, respectivamente, então sempre há um caminho mínimo entre esses dois vértices que passa apenas por $TRG_{\pi_1}(n)$ e $TRG_{\sigma_1}(n)$. Em termos de permutações, isso significa que, dadas duas permutações π e σ tais que $\pi_1 \neq \sigma_1$, existe uma seqüência de transposições t_1, t_2, \dots, t_q que transforma π em σ e somente troca o primeiro elemento uma única vez, como é mostrado na figura 5.1.

Infelizmente, não conseguimos desenvolver uma demonstração formal para essa característica, mas esperamos no futuro poder demonstrá-la. Essa propriedade é interessante, pois, dados dois vértices correspondendo a duas

$$\begin{aligned}
\pi &= (\pi_1 \pi_2 \dots \pi_n) \\
\pi \cdot t_1 &= (\pi_1 \dots \dots \dots) \\
&\vdots \\
\pi \cdot t_1 \cdot t_p &= (\pi_1 \dots \dots \dots) \\
\pi \cdot t_1 \cdot \dots \cdot t_p \cdot t_{p+1} &= (\sigma_1 \dots \dots \dots) \\
&\vdots \\
\pi \cdot t_1 \cdot \dots \cdot t_q = \sigma &= (\sigma_1 \sigma_2 \dots \sigma_n)
\end{aligned}$$

Figura 5.1: Uma seqüência de transposições que troca o primeiro elemento uma única vez.

permutações π e σ em $TRG(n)$, poderíamos restringir a busca de caminhos mínimos entre esses dois vértices apenas aos subgrafos $TRG_{\pi_1}(n)$ e $TRG_{\sigma_1}(n)$.

Com relação ao capítulo 4, apresentamos um método para comparação dos genomas humano e do chimpanzé que procura reversões e transposições que possam ter ocorrido entre esses dois genomas, sendo que não encontramos referências anteriores de trabalhos que procurassem transposições entre dois genomas nessa escala e de maneira puramente computacional.

O trabalho de Feuk *et al.* [11], que nos inspirou o estudo feito no capítulo 4, procurava apenas por reversões, e essas passaram por uma validação manual após a etapa computacional. Nesta tese usamos uma abordagem puramente computacional, baseada na original, a qual descrevemos em maior detalhe do que a descrição feita em [11], e levando em conta transposições. Ressalta-se que encontramos resultados diferentes, com menos reversões, os quais atribuímos a um melhor agrupamento dos segmentos alinhados em blocos executado pelo nosso método.

Esperamos em breve fazer uma revisão dos resultados do capítulo 4, utilizando a comparação entre as versões mais recentes *hg18* do genoma humano e *panTro2* do genoma do chimpanzé. Esses dados tornaram-se disponíveis apenas recentemente, o que não permitiu a sua inclusão neste texto. Além disso, planejamos fazer um estudo mais detalhado do efeito dos parâmetros d e w , usados no passo 4 do método, nas reversões e transposições encontradas.

Apêndice A

Alguns parâmetros conhecidos

A tabela a seguir sumariza valores conhecidos para o diâmetro de transposição e a distância entre a permutação reversa e a identidade, para até 15 elementos. Apresentamos também, para conveniência do leitor, a quantidade de vértices, o grau e a quantidade de arestas de $TRG(n)$.

Os valores para a distância de transposição entre a permutação reversa e a identidade já eram conhecidos para n até 10 desde o artigo de Bafna e Pevzner [4], e já se havia descoberto uma seqüência com $\lfloor \frac{n}{2} \rfloor + 1$ transposições para ordenar a permutação reversa, o que garantia um limite superior para a sua distância. Meidanis, Walter e Dias [16] demonstraram que $d(\rho_{[n]}, \iota_{[n]}) \geq \lfloor \frac{n}{2} \rfloor + 1$, estabelecendo assim a igualdade entre esse limite e a distância; eles conjecturaram que esse era o diâmetro de transposição para S_n , uma vez que, para qualquer permutação com $n \leq 10$ elementos, $d(\pi, \iota_{[n]}) \leq d(\rho_{[n]}, \iota_{[n]})$.

Eriksson et al. [10] encontraram, por meio de busca exaustiva, permutações com 13 e 15 elementos que estavam a uma distância maior da identidade que $\rho_{[13]}$ e $\rho_{[15]}$, respectivamente. Eles também determinaram o diâmetro para $13 \leq n \leq 15$.

Tabela A.1: Parâmetros conhecidos relacionados ao problema de ordenação por transposições.

n	$ V $	Δ	$ E $	$d(\rho_{[n]}, \iota_{[n]})$	$D_t(n)$
1	1	0	0	0	0
2	2	1	1	1	1
3	6	4	12	2	2
4	24	10	120	3	3
5	120	20	1200	3	3
6	720	35	12600	4	4
7	5040	56	141120	4	4
8	40320	84	1693440	5	5
9	362880	120	21772800	5	5
10	3628800	165	299376000	6	6
11	39916800	220	4390848000	6	6
12	479001600	286	68497228800	7	7
13	6227020800	364	1133317785600	7	8
14	87178291200	455	19833061248000	8	8
15	1307674368000	560	366148823040000	8	9

Apêndice B

Lista de reversões e transposições encontradas entre o ser humano e o chimpanzé

Dados obtidos pela comparação do genoma humano (*hg17*) e do chimpanzé (*PanTro1*).

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr1	2517646	2518621	chr1	2454344	2455299
chr1	2519272	2519822	chr1	2453143	2453693
chr1	2520571	2523007	chr1	2449903	2452393
chr1	2523902	2527785	chr1	2445136	2449022
chr1	3712839	3712998	chr1	3566667	3566826
chr1	13148309	13200117	chr1	13286514	13337904
chr1	13464365	13466487	chr1	13240584	13242708
chr1	13485310	13485716	chr1	13221364	13221769
chr1	13489275	13490451	chr1	13216662	13217799
chr1	16976992	16977037	chr1	17138801	17138846

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr1	20593108	20594513	chr1	20910458	20911857
chr1	20597128	20609738	chr1	20894566	20907019
chr1	20611728	20613661	chr1	20890637	20892572
chr1	20619586	20621685	chr1	20881595	20883691
chr1	29879461	29879822	chr1	30479398	30479768
chr1	32551140	32565896	chr1	33301125	33316277
chr1	76335600	76335953	chr1	74893099	74893452
chr1	77940100	77940163	chr1	115379286	115379349
chr1	78762874	78763055	chr1	77355275	77355456
chr1	101529434	101529500	chr1	100578437	100578503
chr1	105075590	105083136	chr1	104138745	104146363
chr1	109923787	111948702	chr1	110317888	112511796
chr1	111949406	113371599	chr1	108881401	110316525
chr1	113376859	113399189	chr1	112897292	112921391
chr1	113406945	113655554	chr1	112590603	112889536
chr1	113669596	113678640	chr1	112567480	112576558
chr1	113689399	113696451	chr1	112549219	112556726
chr1	120750113	120818517	chr1	123893067	123961143
chr1	142346417	142436732	chr1	113013969	113106265
chr1	142441814	142468082	chr1	112982531	113008920
chr1	142468186	142468216	chr1	112982397	112982427
chr1	142482538	142483844	chr1	112966767	112968075
chr1	142486281	142525080	chr1	112925060	112964330
chr1	143746122	143746419	chr1	127189070	127189367
chr1	143761475	143765090	chr1	127171526	127175137
chr1	146602496	146603996	chr1	128380958	128382897
chr1	194489414	194489643	chr1	177143436	177143665
chr1	202704365	202875288	chr1	186160249	186330610
chr1	244791936	244792170	chr1	228802491	228802726
chr1	244804651	244804813	chr1	228795397	228795562
chr1	245161343	245181641	chr1	229037924	229055883
chr2	1589507	3020329	chr12	1618753	3079857
chr2	8047586	8051501	chr12	8699866	8703776
chr2	34257653	34257734	chr12	35653559	35653640
chr2	35422484	35423968	chr12	36837336	36838807

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr2	41106315	41107062	chr12	42653964	42654711
chr2	47311482	47311555	chr12	49345033	49345108
chr2	64721050	64721103	chr12	66953510	66953563
chr2	75227541	75227685	chr12	77633749	77633893
chr2	82061054	82062191	chr12	84887797	84888936
chr2	82523243	82523454	chr12	85356478	85356686
chr2	87050058	87052404	chr12	91015149	91017500
chr2	87281131	87282300	chr12	115425210	115426385
chr2	89928470	89934904	chr12	92088699	92095126
chr2	95584615	95612009	chr12	98625570	98652750
chr2	95612830	95641389	chr12	98601718	98624749
chr2	95642238	95667415	chr12	98575492	98600867
chr2	95944236	95945192	chr12	91843937	91844890
chr2	95947451	95947919	chr12	91841418	91841886
chr2	96058959	96060357	chr12	114194388	114195766
chr2	97639594	97652197	chr12	114173863	114184806
chr2	97653641	97654925	chr12	114171257	114172405
chr2	108336113	108336154	chr12	116087478	116087519
chr2	108336369	108336392	chr12	116087240	116087263
chr2	110588149	110592231	chr12	113372794	113376883
chr2	111152313	111152668	chr12	114164407	114164763
chr2	119671893	119672106	chr13	5691849	5692061
chr2	130639888	130640808	chr13	18029495	18030418
chr2	130653778	130655585	chr13	18014606	18016425
chr2	132016651	132021355	chr13	17913453	17918131
chr2	138838732	138842320	chr13	27851068	27854662
chr2	156445101	156445419	chr13	45633794	45634112
chr2	175410035	175411087	chr13	41945350	41946403
chr2	176498500	176498678	chr13	66135344	66135522
chr2	181457577	181476067	chr13	71255064	71273715
chr2	192911245	192912333	chr13	82807288	82808378
chr2	205309948	205310033	chr13	95323076	95323160
chr2	226241151	226241794	chr13	116825868	116826511
chr2	233255894	233256377	chr13	123847311	123847794
chr2	234256246	234262638	chr13	124994291	125000743

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr2	234262677	234265587	chr13	124992592	124994252
chr2	235304784	235306111	chr13	126028805	126030130
chr2	242226139	242227030	chr13	133702783	133703688
chr3	850688	857030	chr2	870711	877143
chr3	12698978	12699055	chr2	13153577	13153653
chr3	30797191	30797214	chr2	31779002	31779025
chr3	32329909	32330398	chr2	33322984	33323473
chr3	50902683	50908872	chr2	52487419	52493604
chr3	54167662	54169999	chr2	55817738	55820077
chr3	96204791	96204880	chr2	97282180	97282269
chr3	96465140	96465739	chr2	97543169	97544452
chr3	125892323	125892475	chr2	127718878	127719030
chr3	129357522	129357582	chr2	131249275	131249335
chr3	160103702	160103817	chr2	163284067	163284183
chr3	170088600	170091341	chr2	173577309	173580049
chr3	178384355	178385788	chr2	182107531	182108971
chr3	195542642	195542939	chr2	199663960	199664257
chr4	7729208	7729262	chr3	7804002	7804056
chr4	9186711	9188235	chr3	4293520	4295045
chr4	31377772	31383231	chr3	31972192	31977650
chr4	44671174	57204059	chr3	89426696	102294272
chr4	57204090	65134026	chr3	81507847	89426100
chr4	65134679	65134719	chr3	81507154	81507194
chr4	65136339	65136527	chr3	81505328	81505516
chr4	65139017	69814463	chr3	77032076	81502838
chr4	69877633	69879425	chr3	76976450	76978240
chr4	69881467	69921783	chr3	76967336	76974414
chr4	69926314	69926348	chr3	76962774	76962808
chr4	69927072	69955942	chr3	76930351	76962050
chr4	69958067	69958107	chr3	76928181	76928221
chr4	69960105	69960722	chr3	76925546	76926162
chr4	69961745	69962082	chr3	76924165	76924514
chr4	69963561	69964749	chr3	76921498	76922688
chr4	69965997	69966133	chr3	76920130	76920266
chr4	69968658	69969220	chr3	76917041	76917604

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr4	69970449	69971035	chr3	76915217	76915803
chr4	69971450	69972157	chr3	76914425	76914807
chr4	69973706	69973782	chr3	76912821	76912897
chr4	69978953	69979046	chr3	76907557	76907650
chr4	69980747	69996651	chr3	76890067	76905858
chr4	69999340	69999392	chr3	76887328	76887380
chr4	70001224	70001530	chr3	76885193	76885499
chr4	70002776	70014956	chr3	76871668	76883944
chr4	70049084	70049775	chr3	76836587	76837279
chr4	70067000	70067039	chr3	76819295	76819334
chr4	70067550	70142281	chr3	76744159	76818784
chr4	70142789	70142829	chr3	76743611	76743651
chr4	70160778	70163700	chr3	76722619	76725575
chr4	70194502	70195711	chr3	76705455	76706658
chr4	70199000	70200236	chr3	76696390	76697632
chr4	70202031	70202121	chr3	76694454	76694534
chr4	70202617	70221269	chr3	76675176	76693958
chr4	70222529	70225224	chr3	76671222	76673908
chr4	70302808	70305686	chr3	76650039	76652973
chr4	70306193	70306259	chr3	76649474	76649540
chr4	70308899	70309129	chr3	76646705	76646935
chr4	70311042	70311076	chr3	76644372	76644402
chr4	70332809	70332870	chr3	76641131	76641193
chr4	70332951	70332975	chr3	76641026	76641050
chr4	70337956	70340751	chr3	76634884	76636014
chr4	70340814	70340903	chr3	76634732	76634821
chr4	70422628	70435659	chr3	76621260	76634279
chr4	70436227	70436271	chr3	76620644	76620688
chr4	70436312	70436344	chr3	76620571	76620603
chr4	70449208	70454713	chr3	76614524	76619981
chr4	70454763	70454807	chr3	76614433	76614474
chr4	70455569	70455611	chr3	76613628	76613670
chr4	70456189	70457302	chr3	76611935	76613049
chr4	70458162	70460866	chr3	76608616	76611332
chr4	70477456	70527585	chr3	76539420	76590995

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr4	70542722	70542807	chr3	76524193	76524278
chr4	70542959	70822121	chr3	76243550	76524041
chr4	70822450	73403379	chr3	60332218	76241736
chr4	73403961	73404368	chr3	60331225	60331632
chr4	73407826	86315437	chr3	47358490	60327796
chr4	89204343	89212421	chr3	105199787	105207803
chr4	95237011	95237106	chr3	111451872	111451967
chr4	96649338	96649460	chr3	112943838	112943960
chr4	110986047	110986151	chr3	127482961	127483065
chr4	117391316	117392194	chr3	134058500	134058973
chr4	133402241	133425660	chr3	150399225	150399111
chr4	134282344	134286773	chr3	151263251	151267665
chr4	136313622	136314106	chr3	153357216	153357699
chr4	136671147	136671176	chr3	153716686	153716715
chr4	138000890	138007037	chr3	155178372	155190586
chr4	139261019	139261141	chr3	156436842	156436964
chr4	147242947	147260093	chr3	164494747	164511914
chr4	187377194	187377450	chr3	205596591	205596847
chr4	190983692	190988265	chr3	209200281	209203151
chr5	271217	274101	chr4	1705326	1708235
chr5	18589078	20727255	chr4	99547730	101743160
chr5	20850617	20851856	chr4	96574003	96575247
chr5	20875403	20877453	chr4	96550190	96552241
chr5	20887990	20922374	chr4	96500509	96539653
chr5	20923688	20925463	chr4	96497414	96499196
chr5	20928974	20929027	chr4	96493780	96493833
chr5	20931269	20932838	chr4	96489957	96491531
chr5	20935159	20935255	chr4	96487526	96487622
chr5	20940092	20948477	chr4	96474269	96482670
chr5	20949927	20951017	chr4	96471731	96472819
chr5	21324017	21406595	chr4	96015778	96098860
chr5	21410252	21429294	chr4	95992994	96012082
chr5	21497080	21498165	chr4	95924119	95925204
chr5	21509063	21538023	chr4	95884595	95913212
chr5	21558880	26652011	chr4	90802394	45509920

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr5	26675757	29130801	chr4	88312408	90786579
chr5	29149433	29466532	chr4	87991277	88294259
chr5	29468590	29475693	chr4	46133448	87989221
chr5	29492470	32038443	chr4	85333796	87885851
chr5	34441312	35330976	chr4	81499890	82417810
chr5	35338145	37428245	chr4	79330797	81490988
chr5	37428293	69137573	chr4	46435085	79330689
chr5	69142826	69143999	chr4	46428678	46429841
chr5	69157807	70202026	chr4	45016863	46414653
chr5	70204319	70206240	chr4	45012671	45014586
chr5	70222561	79212162	chr4	35976338	44997358
chr5	79213749	94703190	chr4	20474573	35974761
chr5	94704723	95939805	chr4	19227253	20473021
chr5	98949085	99410215	chr4	104683812	105147879
chr5	101051689	101056216	chr4	119354501	119359034
chr5	105928549	105934652	chr4	111817053	111823176
chr5	105934680	105938726	chr4	111812991	111817025
chr5	117519850	117520222	chr4	123723971	123724342
chr5	125750968	125751243	chr4	30007891	30008166
chr5	144513021	144526373	chr4	151189848	151203168
chr5	168469975	168470808	chr4	175713416	175714249
chr5	170193335	170193672	chr4	185621451	185621789
chr5	171817898	171818115	chr4	179139711	179139890
chr5	177317719	177320282	chr4	184821613	184824244
chr6	8929968	8930071	chr5	9301907	9302010
chr6	9854774	9854855	chr5	10069236	10069317
chr6	31117202	31118074	chr5	31581562	31582432
chr6	62022373	62071854	chr5	62678876	62734478
chr6	62072461	62159925	chr5	62585162	62676997
chr6	69459242	69459330	chr5	70569697	70569785
chr6	69906942	69908737	chr5	71020457	71022258
chr6	72090380	72090993	chr5	73222667	73223297
chr6	84072796	84072887	chr5	85709245	85709336
chr6	93505644	93513309	chr5	95250727	95258373
chr6	93513346	93515235	chr5	95248807	95250690

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr6	141661682	141662115	chr5	144560012	144559999
chr6	161133096	161133360	chr5	164951297	164951561
chr6	161156373	161156621	chr5	164795771	164796016
chr6	168005703	168006083	chr5	172433421	172433801
chr6	168911261	168912284	chr5	173379136	173380147
chr7	5055820	5056938	chr6	5477080	5478198
chr7	5059608	5060400	chr6	5471696	5472488
chr7	5802995	5804555	chr6	6998563	7000122
chr7	5806557	5806761	chr6	6996358	6996561
chr7	5807329	5807638	chr6	6995477	6995786
chr7	5808986	5809246	chr6	6993874	6994134
chr7	5810563	6518007	chr6	6281602	6992559
chr7	6518064	6518796	chr6	6280820	6281545
chr7	6520014	6542455	chr6	6255758	6279604
chr7	6547108	6547275	chr6	6242255	6242424
chr7	8665460	8668362	chr6	9112677	9115597
chr7	35266178	35266490	chr6	36438161	36438477
chr7	35735584	35736380	chr6	47137232	47138012
chr7	39199429	39203939	chr6	45164706	45169239
chr7	39356320	39356647	chr6	45163346	45163670
chr7	39356827	39356853	chr6	45163135	45163163
chr7	39358321	40652615	chr6	43599928	45161662
chr7	40653710	43664983	chr6	40575221	43598946
chr7	53030844	53704127	chr6	54721836	55400709
chr7	56865010	56865460	chr6	63977238	63977688
chr7	62995497	62995843	chr6	66542426	66542770
chr7	63238470	63238636	chr6	58193586	58193752
chr7	63279472	63288663	chr6	58201400	58207900
chr7	63698718	63715865	chr6	65496122	65522817
chr7	71641035	71678158	chr6	67690928	67735494
chr7	71680620	71681516	chr6	67687571	67688465
chr7	71684002	71684527	chr6	67684565	67685085
chr7	71686722	71689051	chr6	67680043	67682374
chr7	71691335	71692411	chr6	67676682	67677759
chr7	71694654	71697119	chr6	67671948	67674444

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr7	71698388	71700329	chr6	67668317	67670583
chr7	71717423	71719267	chr6	67649351	67651204
chr7	73930376	73932608	chr6	75480638	75482859
chr7	74648295	74649150	chr6	73985150	73986006
chr7	75478985	75479051	chr6	76578106	76578172
chr7	80553522	80588840	chr6	81781455	81816825
chr7	94925418	94927419	chr6	96537024	96539015
chr7	122719914	122719958	chr6	125068740	125068784
chr7	123758232	123758821	chr6	126119037	126119626
chr7	142859016	142873533	chr6	145622204	145636849
chr7	143894744	143895647	chr6	146520226	146521133
chr7	143895670	143895940	chr6	146519934	146520203
chr8	6142792	6144697	chr7	6356679	6358566
chr8	7133431	7143502	chr7	12762192	12764642
chr8	16188127	16205890	chr7	16940697	16961590
chr8	33960604	33960688	chr7	35144690	35144778
chr8	48378817	48388148	chr7	49931351	49940672
chr8	50533585	50537114	chr7	52115939	52119468
chr8	54573512	54573776	chr7	56244720	56244983
chr8	65411772	65411858	chr7	67414353	67414439
chr8	75000285	75000373	chr7	77039920	77040007
chr8	106889337	106889470	chr7	109325452	109325591
chr8	118990567	118991046	chr7	121716096	121716577
chr8	138210629	138210986	chr7	141203488	141203846
chr8	145115791	145116751	chr7	148393731	148394687
chr9	8188514	8188570	chr11	8114750	8114806
chr9	17472936	17473208	chr11	17709896	17710168
chr9	30298832	30299643	chr11	30769125	30769933
chr9	31418405	31419607	chr11	31890619	31891827
chr9	39216440	39217454	chr11	40306483	40307484
chr9	40264618	40264846	chr11	40342380	40342608
chr9	40268671	40269446	chr11	40337605	40338589
chr9	40303706	40306273	chr11	40300206	40302757
chr9	40307691	40326864	chr11	40280144	40298960
chr9	42021637	42025889	chr11	42694575	42698829

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr9	43443244	43444954	chr11	40035040	40036757
chr9	45310284	45443087	chr11	48890266	49027748
chr9	45456030	45456358	chr11	48876990	48877318
chr9	65274191	65275776	chr11	50242296	50243868
chr9	65347271	65348544	chr11	50176595	50177770
chr9	65390752	65422486	chr11	50102749	50134554
chr9	66360398	66364447	chr11	50083717	50087843
chr9	67226640	67231153	chr11	52043645	52046530
chr9	74127695	74127943	chr11	57958754	57959002
chr9	85570782	85572536	chr11	67702214	67703968
chr9	85629056	85629384	chr11	67642700	67643029
chr9	85643439	85643872	chr11	67614057	67614490
chr9	87815543	87878496	chr11	71829384	71888651
chr9	87983061	87983777	chr11	71955686	71956405
chr9	101684793	101685039	chr11	85942590	85942836
chr9	103014613	103015658	chr11	87275614	87276660
chr9	107780931	107781972	chr11	92305187	92306229
chr9	112932144	112939927	chr11	97487694	97495489
chr9	114294256	114294714	chr11	98836036	98836494
chr9	114294735	114298706	chr11	98832031	98836015
chr9	133143245	133144592	chr11	118010582	118012032
chr9	137374371	137375141	chr11	112748695	112749465
chr10	4978653	4985937	chr8	5110466	5117754
chr10	11532239	11532846	chr8	11672638	11673245
chr10	26910848	26921757	chr8	27945554	27953365
chr10	28304028	28304486	chr8	27877016	27877473
chr10	29798980	29813588	chr8	31401845	31416734
chr10	46456036	46478424	chr8	47442764	47467450
chr10	47242108	47246284	chr8	47759429	47763621
chr10	47491219	47818295	chr8	76352134	53052081
chr10	51298068	51298866	chr8	51157839	51158638
chr10	51427050	51427854	chr8	42810356	42811160
chr10	51636570	51753561	chr8	52912725	53025666
chr10	51806257	52095219	chr8	52569883	52860029
chr10	52098397	52099973	chr8	52566554	52568137

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr10	52101904	52103621	chr8	52562902	52564623
chr10	58276526	58278360	chr8	59191962	59193795
chr10	80960960	80965932	chr8	83359412	83364373
chr10	81274348	81274464	chr8	83359193	83359311
chr10	81576151	81576548	chr8	82480075	82480472
chr10	93193325	93195766	chr8	94916473	94918921
chr10	117877464	117877546	chr8	120049634	120049716
chr10	133302133	133303009	chr8	136046554	136047425
chr11	1649113	1659255	chr9	1788012	1798234
chr11	1882620	1892054	chr9	2087334	2096796
chr11	7755655	7756268	chr9	7874772	7875385
chr11	42337395	42337539	chr9	42811330	42811495
chr11	51275554	51276084	chr9	55539188	55539721
chr11	85523229	85523331	chr9	87170029	87170131
chr11	87850908	87851661	chr9	89685364	89686117
chr11	104607594	104608398	chr9	106545199	106545996
chr11	113877739	113878055	chr9	115882091	115882407
chr11	123763149	123767999	chr9	125842835	125847683
chr11	131430121	131430184	chr9	133580180	133580243
chr11	131993156	132013281	chr9	134163135	134184084
chr12	41131550	41134133	chr10	41724125	41726731
chr12	46134049	46188190	chr10	47107288	47158728
chr12	65800790	65805965	chr10	67367044	67372220
chr12	73652434	73652740	chr10	75376587	75376893
chr12	85744858	85754030	chr10	87868698	87877955
chr12	90305218	90305416	chr10	92472745	92472943
chr12	121450986	121453340	chr10	124522273	124524645
chr12	130297741	130297852	chr10	133863643	133863755
chr12	130530820	130530894	chr10	133698826	133698896
chr13	43030632	43030710	chr14	42741732	42741811
chr13	64757212	64758128	chr14	47119465	47120381
chr13	66222497	66222575	chr14	48618702	48618780
chr13	66783979	66784135	chr14	49179917	49180073
chr13	81825546	81830859	chr14	64534970	64540282
chr13	82171545	82171869	chr14	64879482	64879789

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr13	103550518	103551071	chr14	86946211	86946772
chr13	111949900	111950403	chr14	95583094	95583598
chr14	23327740	23336793	chr15	22270862	22280363
chr14	24035952	24038656	chr15	22933309	22936028
chr14	28831277	28831999	chr15	27783038	27783761
chr14	35454226	35461892	chr15	34781017	34788613
chr14	39050969	39056010	chr15	38411245	38416290
chr14	48784628	48785040	chr15	48171749	48172162
chr14	50209204	50209393	chr15	49624803	49624995
chr14	63404199	63406567	chr15	62971561	62973918
chr14	66935373	66935652	chr15	66553926	66554204
chr14	67673916	67674158	chr15	67298025	67298267
chr14	83504719	83504993	chr15	83635449	83635719
chr14	94477163	94478182	chr15	94914132	94915213
chr14	94557546	94558514	chr15	94995463	94996432
chr15	18313193	26198378	chr16	20589043	28367385
chr15	26517076	26517128	chr16	20517959	20518011
chr15	26996736	28489097	chr16	18428715	20042977
chr15	28537891	28539884	chr16	18377909	18379898
chr15	28619971	28657889	chr16	18260001	18297864
chr15	32488648	32488844	chr16	20043428	20043621
chr15	43031921	43032209	chr16	42952460	42952748
chr15	54303233	54303316	chr16	54650726	54650810
chr15	75993598	75995105	chr16	77521029	77522540
chr15	76009462	76020690	chr16	77495349	77506574
chr15	80747259	80768499	chr16	76642228	84369603
chr15	80831752	80832545	chr16	84470569	84471360
chr15	86187948	86191937	chr16	87032683	87036699
chr15	100152676	100157947	chr16	99231735	99237010
chr16	3170240	3170337	chr18	3054237	3054333
chr16	11926201	11927014	chr18	2725256	2726069
chr16	15351393	15351472	chr18	15162748	15162827
chr16	20352897	20354866	chr18	20953406	20955375
chr16	28390711	28545637	chr18	29110008	29273443
chr16	29403156	29404068	chr18	2722888	2723113

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr16	29798667	29798848	chr18	30560215	30560396
chr16	32566759	32571784	chr18	34023931	34028923
chr16	32631298	32631591	chr18	33964047	33964340
chr16	32634466	32635635	chr18	33960174	33961346
chr16	32638178	32652853	chr18	33941902	33957568
chr16	32654079	32654167	chr18	33940590	33940678
chr16	32678731	32680380	chr18	33914356	33916015
chr16	32717043	32718704	chr18	33876325	33877994
chr16	32733849	32734073	chr18	33861532	33861756
chr16	32739522	32740886	chr18	33854722	33856086
chr16	32744017	32744700	chr18	33850885	33851569
chr16	32976588	32977674	chr18	33628260	33629346
chr16	33044811	33383509	chr18	35022835	33561625
chr16	33424422	33424969	chr18	34979897	34980436
chr16	33460481	33469775	chr18	34937127	34946276
chr16	33471481	33472677	chr18	34934181	34935382
chr16	33490132	33491455	chr18	34915223	34916543
chr16	54195836	54199127	chr18	48257868	48261286
chr16	66087808	66089081	chr18	60342437	60343711
chr16	72921348	72922230	chr18	63092101	63092982
chr16	73797610	73798480	chr18	68210625	68211494
chr16	73801975	73810399	chr18	68198650	68207133
chr16	73810467	73813062	chr18	68196102	68198582
chr16	83746447	83747108	chr18	78663196	78663857
chr17	1083532	1083824	chr19	559823	560109
chr17	11790717	11791814	chr19	12143443	12144547
chr17	21736344	21737295	chr19	21828675	21829630
chr17	22580597	22580779	chr19	29739389	29739570
chr17	27406044	27419428	chr19	29758203	29772494
chr17	41673868	55003655	chr19	50169262	63465787
chr17	55560245	57660951	chr19	47477836	49592706
chr17	57667301	57667610	chr19	47471343	47471663
chr17	57667961	57668264	chr19	47470685	47470988
chr17	57693605	60076862	chr19	45056525	47470565
chr17	69643821	69644236	chr19	72974571	72974985

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr18	123124	10585975	chr17	2076623	12677269
chr18	60799148	60799218	chr17	57660612	57660682
chr18	61351670	61351904	chr17	58215144	58215376
chr18	63800683	63800738	chr17	60746817	60746872
chr18	64219658	64219739	chr17	61180664	61180745
chr18	64706235	64706474	chr17	61678306	61678545
chr19	7743266	7743291	chr20	8051718	8051753
chr19	11998329	12000934	chr20	12412138	12414745
chr19	16750522	17003347	chr20	17983388	18240445
chr19	19779932	19780343	chr20	21676134	21676525
chr19	20312197	20416758	chr20	21134608	21690020
chr19	20581576	20581881	chr20	21569070	21569375
chr19	20583565	20585228	chr20	21565703	21567382
chr19	20601946	20602824	chr20	21547786	21548666
chr19	20623018	20623316	chr20	21675352	21675650
chr19	20640153	20665417	chr20	21485215	21510456
chr19	20666654	20670128	chr20	21480510	21483978
chr19	20679787	20680094	chr20	21470355	21470663
chr19	20682802	20693034	chr20	21457302	21467653
chr19	23090917	23092520	chr20	22583710	22586827
chr19	23108868	23111530	chr20	22570306	22571880
chr19	23119367	23120168	chr20	22560799	22561614
chr19	23254271	23254624	chr20	24894865	24895219
chr19	38699394	38699416	chr20	35215146	35215168
chr19	41500388	41501320	chr20	39276094	39277034
chr19	41504868	41505162	chr20	39275241	39275529
chr19	41506278	41513992	chr20	39266380	39274125
chr19	41515494	41523182	chr20	39262787	39264898
chr19	41524055	41931651	chr20	38847173	39261914
chr19	41932387	42208004	chr20	38571492	38846437
chr19	42284262	42425485	chr20	38338869	38431451
chr19	42426358	42434425	chr20	38329947	38337996
chr19	42435131	42435635	chr20	38328766	38329272
chr19	42443363	42443759	chr20	38320501	38320896
chr19	43960157	43967092	chr20	40889975	40899409

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr19	43967231	43968946	chr20	40888114	40889836
chr19	44744970	44783796	chr20	41790122	41825382
chr19	44811377	44856311	chr20	41728626	41773696
chr19	44858816	44859412	chr20	41725530	41726126
chr19	44859685	44859722	chr20	41725220	41725257
chr19	46084464	46090486	chr20	43246086	43257402
chr19	50320133	50320566	chr20	47462106	47462544
chr19	55132132	55133579	chr20	53231566	53233002
chr19	55135022	55260077	chr20	53105203	53230106
chr19	56405878	56407909	chr20	54051621	54053608
chr19	58564789	58565088	chr20	55557435	55557738
chr20	5502021	5502167	chr21	5487022	5487168
chr20	10736465	10737082	chr21	10614467	10615084
chr20	20993589	20994495	chr21	20914931	20915834
chr20	29430595	29430624	chr21	31128228	31128257
chr20	29430662	29433167	chr21	31125688	31128190
chr20	34682601	34682679	chr21	36920267	36920345
chr20	45914002	45915219	chr21	49188064	49189294
chr20	45919641	45920431	chr21	49185704	49186492
chr21	17267126	17268064	chr22	17345197	17346140
chr21	17343291	17345441	chr22	17425394	17449830
chr21	40320324	40329742	chr22	40364021	40373719
chr21	44712843	44713230	chr22	44886454	44886837
chr22	16133235	16133912	chr23	15992708	15993362
chr22	16986339	16988050	chr23	16912130	16913845
chr22	20133826	20146044	chr23	20128104	20141186
chr22	22912447	22913559	chr23	21100845	21101996
chr22	22937288	22937472	chr23	21092736	21092921
chr22	22944282	22944404	chr23	21090216	21090341
chr22	22974522	22974625	chr23	17293542	17293652
chr22	28159879	28160174	chr23	28249865	28250159
chr22	34532842	34533087	chr23	34657429	34657673
chrX	11522793	11522861	chrX	11846551	11846619
chrX	42859196	42860277	chrX	70002508	70003589
chrX	44238224	44239382	chrX	45684191	45685354

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chrX	46567858	46586760	chrX	48143503	48162481
chrX	47712647	47714346	chrX	49282130	49283829
chrX	47953906	47954488	chrX	53921356	53921938
chrX	48772251	48775268	chrX	50094857	50097857
chrX	50616684	50618452	chrX	51836086	51837205
chrX	50740106	50740896	chrX	51709502	51710292
chrX	51230041	51243942	chrX	52500113	52513970
chrX	51306799	51319066	chrX	52420189	52432725
chrX	51340855	51342167	chrX	52397088	52398400
chrX	51344190	51410012	chrX	52327911	52395065
chrX	52734884	52748605	chrX	53774910	53786587
chrX	54951041	54957226	chrX	56217249	56223146
chrX	55389764	55392909	chrX	56756131	56759249
chrX	55394598	55400449	chrX	56747050	56752905
chrX	55400755	55403244	chrX	56744261	56746744
chrX	62187579	62190522	chrX	63935505	63938445
chrX	70357459	70376725	chrX	72625218	72645640
chrX	70402015	70416450	chrX	72557555	72572875
chrX	72079036	72080198	chrX	74279680	74280847
chrX	100678976	100681250	chrX	103919573	103921828
chrX	101556657	101561388	chrX	104549605	104554465
chrX	101585728	101595375	chrX	104515227	104524877
chrX	103075575	103075883	chrX	106442722	106443033
chrX	105328722	105329830	chrX	108893936	108895043
chrX	117129791	117130280	chrX	121101131	121101620
chrX	136705671	136711436	chrX	141372842	141378601
chrX	138810865	138863111	chrX	143818338	143870605
chrX	138877264	138950857	chrX	143729418	143804185
chrX	140489952	140490769	chrX	144461907	144462727
chrX	140491772	140492097	chrX	144460220	144460544
chrX	140492337	140500207	chrX	144452094	144459974
chrX	145586302	145587551	chrX	150760747	150761988
chrX	145589216	145589771	chrX	150788001	150788554
chrX	145611452	145612689	chrX	150779596	150780818
chrX	148392597	148402585	chrX	153781671	153791516

Tabela B.1: Reversões entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chrX	148410169	148411021	chrX	153773074	153773922
chrX	148413013	148437797	chrX	153751620	153771082
chrX	148469758	148509463	chrX	153671767	153711814
chrX	149242145	149254719	chrX	154471926	154484543
chrX	153100837	153119930	chrX	158420864	158440049
chrX	153122170	153127182	chrX	158413497	158418587
chrX	153127822	153132095	chrX	158408556	158412857

Tabela B.2: Transposições entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr1	9851327	9970838	chr1	9875303	9999704
chr1	25424037	25428336	chr1	25758168	25762464
chr1	25432124	25433979	chr1	25764659	25766509
chr1	25436447	25437484	chr1	25768769	25769809
chr1	25439743	25443296	chr1	25772084	25775874
chr1	25447427	25455824	chr1	25780144	25794766
chr1	25457935	25458137	chr1	25796890	25797092
chr1	25479183	25479364	chr1	26022026	26022207
chr1	67917157	67937683	chr1	66284773	66305290
chr1	141890250	141969666	chr1	124420263	108870098
chr1	144747150	144748193	chr1	125915935	125916979
chr1	144982743	144982921	chr1	124240788	124240965
chr1	146434560	146494320	chr1	127879570	127938533
chr1	149883423	149911710	chr1	131664187	131692150
chr1	179376792	179402329	chr1	161784569	161810753
chr1	179449860	179474849	chr1	161858252	161883298
chr1	193552136	193553951	chr1	176172089	176173905
chr2	4613317	4982777	chr12	3180610	3549749
chr2	5003805	5066463	chr12	3570777	3640741
chr2	34180557	34207908	chr12	35275837	35303322
chr2	34255541	34257274	chr12	35350954	35352673
chr2	34265134	34409254	chr12	35504748	35504390
chr2	46008211	46055398	chr12	47760774	47808586

Tabela B.2: Transposições entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr2	46063033	46066016	chr12	47816359	47819346
chr2	46068248	46072933	chr12	47821579	47826264
chr2	76579588	76616025	chr12	79148543	79184834
chr2	76630861	76658664	chr12	79199670	79228292
chr2	106445187	106445250	chr12	89906760	89906819
chr2	108337621	108754738	chr12	111379777	111786334
chr2	159124113	159177814	chr13	48260032	48339923
chr2	170205482	170210231	chr13	59659917	59664640
chr2	170214475	170216210	chr13	59668883	59670637
chr2	170217754	170220189	chr13	59672181	59674612
chr2	170231287	170242937	chr13	59681938	59693592
chr2	170333321	170348335	chr13	59783976	59803217
chr2	178292901	178317565	chr13	67939970	67939209
chr2	222213555	222335218	chr13	112431038	112553072
chr2	236295277	236302542	chr13	127022709	127030865
chr2	236303074	236399512	chr13	127031397	127132814
chr3	6980687	7079006	chr2	7193524	7292493
chr3	13119194	13146342	chr2	13573483	13600656
chr3	13158287	13178033	chr2	13612601	13629031
chr3	95060856	95060905	chr2	92723029	92723078
chr3	102254502	102256726	chr2	103504763	103503641
chr3	102257638	102257749	chr2	103504548	103504658
chr3	102275490	102408209	chr2	103525558	103665458
chr3	113671175	113701976	chr2	115265912	115296982
chr3	164697551	164838498	chr2	167977417	168108262
chr4	19219840	19237725	chr3	19572666	19589167
chr4	37245156	37262320	chr3	37621168	37638357
chr4	37277729	37303225	chr3	37653771	37679492
chr4	120198532	120254361	chr3	136673236	136729364
chr4	120275541	120315429	chr3	136749838	136790962
chr4	120652715	122111733	chr3	137253041	138771026
chr4	137783463	137887112	chr3	154832462	154937168
chr4	172238345	172313480	chr3	189911034	189988088
chr4	172356469	172381455	chr3	190031077	190056271
chr4	182120133	182139501	chr3	200249047	200268420

Tabela B.2: Transposições entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr5	10284365	10343708	chr4	10506134	10565301
chr5	16803928	16865682	chr4	17232221	17293772
chr5	16867998	16869490	chr4	17296083	17297616
chr5	16872037	16872621	chr4	17300141	17300725
chr5	16873774	16874040	chr4	17301872	17302138
chr5	134831543	134835164	chr4	141187271	141190888
chr5	134851490	134987927	chr4	141207214	141338766
chr5	146452796	146543289	chr4	153166463	153256499
chr5	156744792	156783004	chr4	163745632	163781695
chr5	156803030	156819442	chr4	163801725	163818156
chr5	156821586	156823220	chr4	163820300	163821933
chr5	156837126	156859713	chr4	163833992	163856689
chr6	3949364	3974556	chr5	3905988	3931229
chr6	4918016	5026444	chr5	5132370	5238974
chr6	45903141	45923027	chr5	46503732	46523714
chr6	55239932	55241822	chr5	56229941	56231828
chr6	55243589	55263295	chr5	56233595	56253680
chr6	55289379	55303171	chr5	56276812	56290630
chr6	74545013	74570547	chr5	75749649	75775138
chr6	102540096	102544682	chr5	104524254	104528825
chr6	134507456	134587571	chr5	137149741	137230406
chr6	134587873	134587901	chr5	137230710	137230738
chr6	134590010	134593110	chr5	137233501	137232582
chr6	147636219	147663774	chr5	150937768	150965177
chr6	150352899	150363793	chr5	153760033	153770768
chr6	150368161	150387045	chr5	153774419	153793160
chr6	157211185	157227258	chr5	160778344	160794303
chr6	157246943	157313237	chr5	160813988	160882144
chr6	158242781	158243965	chr5	161936135	161937320
chr6	158251818	158277452	chr5	161945203	161970631
chr6	161703512	161766204	chr5	165554038	165616569
chr6	161803103	161811063	chr5	165651447	165659014
chr6	162001662	162122232	chr5	166090101	166210572
chr7	4801023	4802017	chr6	5100200	5101184
chr7	9248001	9261270	chr6	9659322	9670950

Tabela B.2: Transposições entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr7	9277211	9293394	chr6	9694785	9711005
chr7	9330288	9345386	chr6	9747899	9763212
chr7	55516579	55518134	chr6	44574111	44575658
chr7	81773965	81854595	chr6	83047089	83132005
chr7	81872395	81884928	chr6	83159684	83173721
chr7	114713371	114772407	chr6	116795624	116853865
chr8	8310382	8324021	chr7	8785724	8799337
chr8	8379900	8476201	chr7	8799443	8904556
chr8	58356827	58365983	chr7	60097097	60106266
chr8	133703270	133762353	chr7	136477892	136537059
chr8	133787517	133797170	chr7	136575187	136584858
chr9	11778544	11780717	chr11	11701595	11703769
chr9	24049148	24068073	chr11	24380145	24399227
chr9	64653325	64654475	chr11	48398556	48399700
chr9	64660635	64683829	chr11	48405957	48429428
chr9	92727341	92730399	chr11	71781836	71784870
chr9	104884620	104886798	chr11	89202355	89204529
chr9	104888397	104890617	chr11	89206119	89209088
chr9	104891778	104896498	chr11	89210249	89214950
chr9	104902996	104905004	chr11	89221392	89223393
chr9	104910897	104923850	chr11	89229294	89242281
chr9	133045960	133080185	chr11	117930362	117965910
chr9	133111939	133112736	chr11	117997664	117998461
chr10	16944408	16953580	chr8	17464849	17478686
chr10	16953862	17895446	chr8	17478969	18421196
chr10	45055192	45066059	chr8	17422018	17446867
chr10	47051406	47069758	chr8	47751246	47765534
chr10	60679009	60694270	chr8	61363922	61379340
chr10	64670166	64765559	chr8	65508296	65604850
chr10	81407537	81408342	chr8	82489370	82490172
chr10	114484675	114577212	chr8	116412783	116506452
chr11	54929353	54929803	chr9	55473260	55473709
chr11	54941871	54942119	chr9	55498545	55504157
chr11	55002338	55008364	chr9	55580534	55616195
chr11	64364955	64368615	chr9	65357708	65361373

Tabela B.2: Transposições entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr11	84790239	84816079	chr9	86250004	86276040
chr11	84829825	84873463	chr9	86291945	86336219
chr11	84896968	84936072	chr9	86359724	86399658
chr11	86028205	86064056	chr9	87685351	87721361
chr11	86077267	86101462	chr9	87735070	87759242
chr11	122623494	122624860	chr9	124661010	124662383
chr11	132877375	132886132	chr9	135079039	135087791
chr12	3980370	4148040	chr10	4061334	4231922
chr12	5155440	5191698	chr10	5546251	5582556
chr12	5195070	5197718	chr10	5585926	5588574
chr12	5252365	5253559	chr10	5524434	5525628
chr12	5255431	5265910	chr10	5527496	5538862
chr12	45028111	45029617	chr10	45651203	45652713
chr12	45459386	45471754	chr10	46185409	46198141
chr12	45618730	45658351	chr10	46371333	46409820
chr12	68619395	68628855	chr10	70163967	70173148
chr12	79656211	79679143	chr10	81298869	81321782
chr12	79743995	79751655	chr10	81386636	81394301
chr12	79785716	79812100	chr10	81428362	81456538
chr12	84382291	84421854	chr10	86295681	86335351
chr12	95355993	95545451	chr10	97684133	97870095
chr12	110319960	110320158	chr10	112817128	112817326
chr12	110328720	110331172	chr10	112825887	112828316
chr12	114388862	114558092	chr10	117076099	117246056
chr13	20274243	20291110	chr14	19244061	19261598
chr13	20291287	20292483	chr14	19261773	19262964
chr13	20306685	20386269	chr14	19277166	19357609
chr13	39386344	39407546	chr14	38761623	38783223
chr13	39431747	39449987	chr14	38806549	38825813
chr13	89119202	89138052	chr14	71905654	71924390
chr13	96585327	96585364	chr14	79644082	79644119
chr13	113117913	113120139	chr14	96740555	96742710
chr13	113121307	113125816	chr14	96743877	96748490
chr14	35333606	35362705	chr15	34498624	34527483
chr14	35396125	35410155	chr15	34549055	34563118

Tabela B.2: Transposições entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr14	46325229	46342062	chr15	45569127	45586030
chr15	26631487	26631637	chr16	18777486	18777636
chr15	26674044	26674703	chr16	18793701	18794366
chr15	26694777	26695182	chr16	18815505	18815910
chr15	26697276	26697534	chr16	18818158	18818410
chr15	26698735	26699265	chr16	18819600	18820129
chr15	26699906	26700151	chr16	18820944	18821189
chr15	26701548	26718360	chr16	18822587	18840093
chr15	26720040	26720123	chr16	18841783	18841866
chr15	32358588	32455984	chr16	32004962	32105187
chr15	50881712	50905768	chr16	50932033	50960839
chr15	50914178	50917279	chr16	50969249	50972326
chr15	50925960	51060470	chr16	50981007	51137011
chr16	5193917	5351439	chr18	5140243	5298664
chr16	10120541	10126274	chr18	10254677	10260393
chr16	10161663	10275637	chr18	10295782	10409941
chr16	11844319	11847341	chr18	12173922	12176977
chr16	11861063	11874872	chr18	12190699	12204931
chr16	11875651	11877577	chr18	12205709	12207623
chr16	11880258	11881659	chr18	12210304	12211707
chr16	11890893	11901093	chr18	12220941	12231136
chr16	11904386	11919269	chr18	12233406	12248212
chr16	11919484	11920857	chr18	12248427	12249794
chr16	29037472	29045593	chr18	21765993	21776814
chr16	29218048	29257293	chr18	12360687	12403486
chr16	32311370	32321996	chr18	33211687	33222290
chr16	32348559	32350215	chr18	33248888	33250520
chr16	32355384	32356725	chr18	33255693	33257032
chr16	32379405	32380153	chr18	33280005	33280754
chr16	32390868	32392204	chr18	33291484	33292990
chr16	32751426	32753799	chr18	33648591	33651070
chr16	32776891	32786329	chr18	33674167	33683517
chr16	32791983	32793149	chr18	33689174	33690340
chr16	32796158	32798164	chr18	33693353	33695341
chr16	32808333	32808461	chr18	33705558	33705682

Tabela B.2: Transposições entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr16	32813632	32814613	chr18	33712488	33713471
chr16	32830609	32839251	chr18	33731253	33740621
chr16	32854772	32857126	chr18	33756149	33758510
chr16	32869130	32870272	chr18	33770521	33771667
chr16	32883041	32883410	chr18	33784439	33784810
chr16	32887109	32888093	chr18	33788513	33789594
chr16	32889739	32897766	chr18	33791242	33799294
chr16	75407217	75408388	chr18	69875237	69876406
chr16	75430310	75435846	chr18	69898328	69903855
chr16	75436628	75574258	chr18	69904637	70045215
chr16	86450685	86451956	chr18	21978596	21979812
chr16	86587392	86637860	chr18	21778393	12417068
chr16	86864217	86866412	chr18	12363063	12364963
chr16	86872495	86872806	chr18	12367790	12368101
chr17	20241281	20241988	chr19	16977719	16978425
chr17	33685370	33815809	chr19	36872555	36925549
chr17	60101984	60102176	chr19	44345935	44346127
chr17	60194424	60194466	chr19	44150669	44150711
chr17	60345569	60347022	chr19	63411517	63413011
chr17	73346686	73364733	chr19	76840832	76863598
chr18	52473402	52542892	chr17	49156893	49227015
chr19	12207468	12225451	chr20	12549515	12567582
chr19	12230661	12235238	chr20	12572786	12577361
chr19	33510380	33533333	chr20	29686996	29709976
chr19	33550017	33715506	chr20	29726660	29892842
chr19	62121428	62122013	chr20	59635709	59636295
chr19	62247336	62259964	chr20	59834320	59846963
chr20	23074561	23083559	chr21	23026293	23035264
chr20	23083590	23084498	chr21	23035265	23036166
chr20	23115481	23236592	chr21	23068270	23188450
chr20	23375650	23382667	chr21	23662417	23669441
chr20	23403981	23404470	chr21	23701805	23702301
chr20	23404982	23405592	chr21	23702663	23703347
chr20	23405671	23405782	chr21	23703418	23703528
chr20	23407282	23408758	chr21	23704044	23705529

Tabela B.2: Transposições entre os genomas humano e do chimpanzé.

cromoss. humano	início humano	fim humano	cromoss. chimp.	início chimp.	fim chimp.
chr20	33954954	33956310	chr21	35987090	35988428
chr20	34023191	34033728	chr21	36055307	36065554
chr20	34036506	34089748	chr21	36068332	36120187
chr20	34103747	34142128	chr21	36134185	36172908
chr21	45650247	45651673	chr22	45834153	45835591
chr21	45777671	45834558	chr22	45961589	46020168
chr22	16856407	16858438	chr23	16781739	16783772
chrX	3525594	3538781	chrX	3293852	3307034
chrX	3541077	3693092	chrX	3309328	3484457
chrX	6934272	6935664	chrX	6893653	6895060
chrX	6938320	6939220	chrX	6897722	6898617
chrX	27742330	27744043	chrX	28622232	28623942
chrX	30607470	30643683	chrX	31620120	31667206
chrX	52808468	53667754	chrX	54053335	54917774
chrX	55659542	58189747	chrX	57020366	59650496
chrX	98095369	98216277	chrX	101185329	101305388
chrX	101754373	101767315	chrX	105094079	105107098
chrX	101769874	101771252	chrX	105109658	105111036
chrX	101845358	101866034	chrX	105185142	105205874
chrX	118775787	118776918	chrX	122811297	122812436
chrX	118779876	131822452	chrX	122815394	136173087
chrX	139868449	139882065	chrX	144717749	144731002

Referências bibliográficas

- [1] BAFNA, V., PEVZNER, P. A., “Genome rearrangements and sorting by reversals”, In: *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pp. 148–157, 1993.
- [2] BAFNA, V., PEVZNER, P. A., “Sorting by reversals: genome rearrangements in plant organelles and evolutionary history of X chromosome”, *Mol. Biol. Evol.*, v. 12, n. 2, pp. 239–246, março/1995.
- [3] BAFNA, V., PEVZNER, P. A., “Genome rearrangements and sorting by reversals”, *SIAM J. Comput.*, v. 25, n. 2, pp. 272–289, 1996.
- [4] BAFNA, V., PEVZNER, P. A., “Sorting by transpositions”, *SIAM J. Discrete Math.*, v. 11, n. 2, pp. 224–240, maio/1998.
- [5] CAPRARA, A., “Sorting permutations by reversals and eulerian cycle decompositions”, *SIAM J. Discrete Math.*, v. 12, n. 1, pp. 91–110, 1999.
- [6] CHRISTIE, D. A., “Sorting permutations by block-interchanges”, *Inform. Process. Lett.*, v. 60, n. 4, pp. 165–169, 1996.
- [7] CRANSTON, D. W., SUDBOROUGH, H. I., WEST, D. B., “Short proofs for cut-and-paste sorting of permutations”. Aceito para publicação

- em *Discrete Math.*, versão online em <http://dx.doi.org/10.1016/j.disc.2007.01.011>, fevereiro/2007.
- [8] DIAS, Z., *Rearranjo de genomas: uma coletânea de artigos*. Tese de doutorado, Unicamp, Campinas, São Paulo, 2002.
- [9] ELIAS, I., HARTMAN, T., “A 1.375-approximation algorithm for sorting by transpositions”, *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, v. 3, n. 4, pp. 369–379, 2006.
- [10] ERIKSSON, H., ERIKSSON, K., KARLANDER, J., SVENSSON, L., WÄSTLUND, J., “Sorting a bridge hand”, *Discrete Math.*, v. 241, n. 1, pp. 289–300, outubro/2001.
- [11] FEUK, L., MACDONALD, J. R., TANG, T., CARSON, A. R., LI, M., RAO, G., KHAJA, R., SCHERER, S. W., “Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies”, *PLoS Genet.*, v. 1, n. 4, p. e56, outubro/2005.
- [12] HANNENHALLI, S., PEVZNER, P., “Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals”, In: *STOC '95: Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, (New York, NY, USA), pp. 178–189, ACM Press, 1995.
- [13] KECECIOGLU, J., SANKOFF, D., “Exact and approximation algorithms for the inversion distance between two permutations”, *Algorithmica*, n. 13, pp. 180–210, 1995.

- [14] LABARRE, A., “New bounds and tractable instances for the transposition distance”, *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, v. 3, n. 4, pp. 380–394, dezembro/2006.
- [15] MEIDANIS, J., DIAS, Z., *An Alternative Algebraic Formalism for Genome Rearrangements*, pp. 213–223. Kluwer Academic Publishers, 2000.
- [16] MEIDANIS, J., WALTER, M. E. M. T., DIAS, Z., “Transposition distance between a permutation and its reverse”, In: *Proceedings of the 4th South American Workshop on String Processing* (BAEZA-YATES, R., ed.), (Valparaíso, Chile), pp. 70–79, Carleton University Press, 1997. Uma versão deste trabalho foi publicada em [8].
- [17] MIRA, C. V. G., *Análise Algébrica de Problemas de Rearranjo em Genomas: Algoritmos e Complexidade*. Tese de doutorado, Unicamp, Campinas, São Paulo, 2007.
- [18] NADEAU, J. H., TAYLOR, B. A., “Lengths of chromosomal segments conserved since divergence of man and mouse”, *Proc. Natl. Acad. Sci. USA*, v. 81, n. 3, pp. 814–818, 1984.
- [19] PEVZNER, P., TESLER, G., “Genome rearrangements in mammalian evolution: lessons from human and mouse genomes”, *Genome Res.*, v. 13, n. 1, pp. 37–45, janeiro/2003.
- [20] SCHWARTZ, S., KENT, W. J., SMIT, A., ZHANG, Z., BAERTSCH, R., HARDISON, R. C., DAVIDHAUSSLER, MILLER, W., “Human–Mouse Alignments with BLASTZ”, *Genome Res.*, v. 13, n. 1, pp. 103–107, 2002.

- [21] SETUBAL, C., MEIDANIS, J., *Introduction to Computational Molecular Biology*. PWS Publishing, janeiro/1997.
- [22] SMIT, A. F. A., “The origin of interspersed repeats in the human genome”, *Current Opin. Genet. Dev.*, v. 6, n. 6, pp. 743–748, dezembro/1996.
- [23] STURTEVANT, A. H., DOBZHANSKY, T., “Inversions in the third chromosome of wild races of *Drosophila pseudoobscura*, and their use in the study of the history of the species”, *Proc. Natl. Acad. Sci. USA*, v. 22, n. 7, pp. 448–450, 1936.
- [24] STURTEVANT, A. H., DOBZHANSKY, T., “Inversions in the chromosomes of *Drosophila pseudoobscura*”, *Genetics*, v. 23, n. 1, pp. 28–64, 1938.
- [25] WALTER, M. E. T., FREIRE, A. L., MOURA, C., XAVIER, D., “An approximation algorithm for the problem of sorting by transversals”, In: *Third Brazilian Workshop on Bioinformatics – WOB*, pp. 191–195, 2004.
- [26] WATTERSON, G., EWENS, W., HALL, T., MORGAN, A., “The chromosome inversion problem”, *J. Theoret. Biol.*, v. 99, n. 1, pp. 1–7, 1982.
- [27] YANCOPOULOS, S., ATTIE, O., FRIEDBERG, R., “Efficient sorting of genomic permutations by translocation, inversion and block interchange”, *Bioinformatics*, v. 21, n. 16, pp. 3340–3346, agosto/2005.

- [28] YUNIS, J. J., PRAKASH, O., “The origin of man: A chromosomal pictorial legacy”, *Science*, v. 215, n. 4539, pp. 1525–1530, 1982.