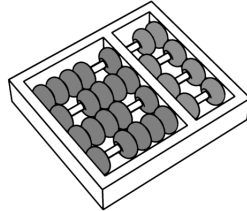


Universidade Estadual de Campinas

Instituto de Computação



PROPOSTA DE DISSERTAÇÃO DE MESTRADO

O Problema da Ordenação de Permutações
por Reversões e Transposições

Candidato:

Andre Rodrigues Oliveira

Orientador:

Prof. Dr. Zanoni Dias

24 de abril de 2014

Resumo

Rearranjos de Genomas, diferentemente das mutações mais comuns que afetam pontualmente o genoma, afetam grandes trechos do genoma e são um desafio para a teoria da computação, uma vez que, na maioria dos casos, computar a distância entre dois genomas considerando operações globais resultam em problemas NP-Difíceis. Além disso, temos o Problema da Ordenação de Genomas, onde buscamos calcular o número mínimo de eventos de rearranjos necessários para ordenar blocos conservados de genomas, onde os genomas são representados por permutações. Um modelo de rearranjo determina quais eventos de rearranjo são permitidos para ordenar uma permutação ou transformar uma permutação em outra. Dentre os eventos de rearranjo mais comuns, temos a reversão e a transposição. Modelos considerando os dois eventos citados acima, separadamente, já possuem vasta bibliografia. Entretanto, modelos considerando os dois eventos em conjunto ainda foram pouco explorados. Nosso trabalho de mestrado terá como meta estudar a versão do problema que permite o uso de reversões e transposições. Esta proposta apresenta alguns conceitos iniciais e resultados já existentes sobre este problema, bem como especifica o objetivo do nosso trabalho.

1 Introdução

O problema de rearranjo de genomas busca, dados dois genomas, encontrar uma sequência de rearranjos que transforma um genoma no outro. Esta sequência pode ser utilizada para estimar a distância evolutiva ou mesmo o grau de parentesco entre os dois genomas comparados. Diferente das mutações pontuais que afetam moléculas constituintes do genoma, os eventos de

rearranjo afetam grandes porções do genoma, tornando esta abordagem mais adequada para comparação de genomas completos.

O genoma de uma espécie é composto de cromossomos, representados como um conjunto ordenado de genes. Outra forma de representação dos cromossomos se dá por meio de blocos conservados, sendo estes blocos regiões de alta similaridade entre os dois genomas comparados. Estes blocos podem ser genes, ou qualquer subsequência conservada em ambos os genomas.

O objetivo em problemas de rearranjo de genomas é encontrar a menor sequência de eventos de rearranjo que transforma um genoma em outro. O tamanho desta sequência é denominado *distância evolucionária* entre indivíduos, e parte do princípio de que a natureza sempre utiliza o menor número de operações, com base em uma teoria conhecida como princípio da máxima parcimônia.

Na literatura foram propostos diversos eventos de rearranjo de genomas, também conhecidos como operações, dentre os quais podemos citar fusão, fissão, reversão, transposição e transreversão. Estes eventos de rearranjo formam um grupo de mutações que afetam um trecho do genoma, em contrapartida com as mutações pontuais que afetam apenas um nucleotídeo.

Um *modelo de rearranjo* determina o conjunto de operações permitidas para se transformar um genoma em outro. Nos estudos iniciais em rearranjos de genomas, o foco se deu em modelos de rearranjo que permitiam apenas um tipo de operação, capazes de explicar cenários evolutivos simples. Mais tarde, surgiram modelos que permitiam duas ou mais operações, possibilitando assim a análise de cenários evolutivos mais complexos.

As seções seguintes estão organizadas da seguinte forma: a Seção 2 apresenta conceitos teóricos importantes para este trabalho. A Seção 3 apresenta formalmente o Problema da Ordenação de Permutações por Reversões

e Transposições. A Seção 4 apresenta limitantes para problemas de rearranjo de genomas. A Seção 5 descreve os objetivos deste trabalho. A Seção 6 apresenta a metodologia que utilizaremos neste trabalho. A Seção 7 descreve o cronograma que seguiremos no decorrer do trabalho.

2 Fundamentação Teórica

Em Problemas de Rearranjos de Genomas, um genoma é representado como uma n -tupla cujos elementos representam os genes. Supondo que não haja repetição de genes, esta n -tupla é uma permutação $\pi = (\pi_1 \ \pi_2 \ \dots \ \pi_n)$, com $\pi_i \in \{-n, -(n-1), \dots, -2, -1, +1, +2, \dots, +(n-1), +n\}$ e $|\pi_i| \neq |\pi_j| \leftrightarrow i \neq j$. Cada elemento π_i possui um sinal, $+$ ou $-$, que indica a orientação do gene que ele representa. Quando não há informação sobre a orientação dos genes, este sinal é omitido.

Dadas duas permutações π e σ , e um modelo de rearranjo M , o problema de transformar a permutação π na permutação σ consiste em encontrar a menor sequência de operações $\rho_1, \rho_2, \dots, \rho_t$ pertencentes a M tal que $\pi\rho_1\rho_2\dots\rho_t = \sigma$. O tamanho desta sequência representa a distância entre as permutações π e σ com respeito ao modelo M , e é denotada por $d_M(\pi, \sigma)$ (neste caso, $d_M(\pi, \sigma) = t$). A maior distância entre duas permutações de tamanho n com respeito ao modelo M é chamada de *diâmetro* da distância de rearranjo, e é denotada por $D_M(n)$.

Definindo-se a permutação identidade como $\iota_n = (1 \ 2 \ \dots \ n)$, a ordenação de uma permutação α na permutação identidade ι_n e a distância entre elas com respeito a um modelo M é denotada por $d_M(\alpha, \iota_n) = d_M(\alpha)$. Este processo é equivalente ao processo de transformar a permutação π na permutação σ se tomarmos $\alpha = \sigma^{-1} \cdot \pi$, onde σ^{-1} é a inversa de σ tal que $\sigma^{-1} \cdot \sigma = \iota_n$,

uma vez que $d_M(\pi, \sigma) = d_M(\sigma^{-1} \cdot \pi, \sigma^{-1} \cdot \sigma) = d_M(\alpha, \iota_n) = d_M(\alpha)$. Por exemplo, se $\pi = (1\ 3\ 2\ 5\ 4)$ e $\sigma = (2\ 4\ 5\ 1\ 3)$, então $\sigma^{-1} = (4\ 1\ 5\ 2\ 3)$, $\sigma^{-1} \cdot \sigma = (1\ 2\ 3\ 4\ 5) = \iota_5$, e $\sigma^{-1} \cdot \pi = (4\ 5\ 1\ 3\ 2)$.

2.1 Reversão

Uma reversão $\rho(i, j)$ é uma operação que inverte a ordem e a orientação dos genes entre os pontos i e j de um cromossomo. Formalmente, uma reversão $\rho(i, j)$, com $1 \leq i \leq j \leq n$, aplicada ao genoma $\pi = (\pi_1\ \pi_2\ \dots\ \pi_n)$, gera a permutação $\pi\rho(i, j) = (\pi_1\ \dots\ \pi_{i-1}\ \underline{-\pi_j\ \dots\ -\pi_i}\ \pi_{j+1}\ \dots\ \pi_n)$.

Quando não se conhece a orientação dos genes, tem-se o Problema de Reversão de Permutações sem Sinal, que foi provado ser NP-Completo por Caprara [7]. Bafna e Pevzner [2] apresentaram os primeiros estudos sobre este problema, o que resultou em um algoritmo de aproximação de fator 1.75. Posteriormente, Christie [8] apresentou um algoritmo de aproximação de fator 1.5. Atualmente, o melhor algoritmo conhecido para o problema possui fator de aproximação 1.375, proposto por Berman e coautores [4].

Uma variação deste problema, bastante estudada na literatura, consiste em um modelo de rearranjo em que as reversões sempre ocorrem no prefixo da permutação, conhecido como Problema de Ordenação de Panquecas [13, 17]. Atualmente, o melhor algoritmo conhecido para este problema, proposto por Fischer e Ginzinger [12], possui fator de aproximação 2.

Para o caso em que a orientação dos genes é conhecida, denominado Problema de Reversão de Permutações com Sinal, existem algoritmos polinomiais exatos, sendo que o primeiro foi apresentado por Hannenhalli e Pevzner [15]. Atualmente, o algoritmo mais eficiente disponível na literatura possui complexidade subquadrática [23]. Para o caso em que se deseja saber apenas o valor da distância de reversão de permutações, há também um algoritmo

linear [1].

2.2 Transposição

Uma transposição é um evento de rearranjo que tem a propriedade de trocar dois blocos adjacentes de genes de lugar. Formalmente, seja uma permutação $\pi = (\pi_1 \pi_2 \dots \pi_n)$. Uma transposição $\rho(i, j, k)$, onde $1 \leq i < j < k \leq n + 1$, remove o intervalo $[i, j - 1]$ de π e o insere entre π_{k-1} e π_k . O resultado é a permutação $\pi\rho(i, j, k) = (\pi_1 \dots \pi_{i-1} \underline{\pi_j \dots \pi_{k-1}} \underline{\pi_i \dots \pi_{j-1}} \pi_k \dots \pi_n)$. Definimos a distância de transposição entre duas permutações como um número mínimo de transposições necessárias para transformar uma permutação na outra.

O primeiro algoritmo de aproximação para este problema foi apresentado por Bafna e Pevzner [3], com complexidade de tempo de $O(n^2)$ e razão de aproximação 1.5. Dentre as contribuições do trabalho de Bafna e Pevzner, é possível citar uma estrutura em grafos chamada de Grafo de Ciclos. Esta estrutura permite a definição de limitantes inferiores e superiores fortes para o problema da distância de transposição.

Christie [9] propôs um algoritmo com fator de aproximação 1.5 mais simples que o de Bafna e Pevzner, mas com complexidade $O(n^4)$. Walter, Dias e Meidanis [25] desenvolveram um algoritmo com complexidade de tempo da ordem de $O(n^2)$, com fator de aproximação 2.25. Elias e Hartman [10] desenvolveram um algoritmo de aproximação na razão de 1.375 e complexidade $O(n^2)$, um avanço na razão de aproximação que não ocorria desde a publicação do trabalho de Bafna e Pevzner, oito anos antes.

Recentemente, Bulteau, Fertin e Rusu provaram que o problema da distância de transposição é NP-Difícil [6], resolvendo essa questão que permaneceu em aberto durante aproximadamente 15 anos.

Diferentemente do problema da distância de reversão, o problema da distância de transposição não leva em consideração a orientação dos genes, uma vez que transposições apenas trocam elementos de posição.

2.3 Breakpoints

Seja a *permutação estendida* que pode ser obtida a partir de π inserindo-se dois novos elementos: $\pi_0 = 0$ e $\pi_{n+1} = n + 1$.

Um breakpoint para o problema de ordenação por reversões sem sinal ocorre entre um par de elementos π_i e π_{i+1} de π , com $0 \leq i \leq n$, se $|\pi_i - \pi_{i+1}| \neq 1$. A permutação identidade é a única permutação que não possui breakpoints. Por exemplo, $\pi = (0 \cdot 3 \ 4 \cdot 1 \ 2 \cdot 5)$, onde “.” representa cada breakpoint desta permutação. O número de breakpoints de reversões em uma permutação sem sinal π é denotado por $b_r(\pi)$. No exemplo acima, temos que $b_r(\pi) = 3$.

Um breakpoint para o problema de ordenação por reversões com sinal ou para o problema de ordenação por transposições ocorre entre um par de elementos π_i e π_{i+1} , com $0 \leq i \leq n$ se $\pi_{i+1} - \pi_i \neq 1$. A permutação identidade é a única permutação que não possui breakpoints. O número de breakpoints de reversões em uma permutação com sinal π é denotado por $b_{\bar{r}}(\pi)$. O número de breakpoints de transposições em uma permutação π é denotado por $b_t(\pi)$. Por exemplo, seja $\pi = (0 \ +1 \cdot -2 \cdot +3 \ +4 \cdot -5 \cdot +6$ e $\sigma = (0 \ 1 \ 2 \cdot 4 \cdot 3 \cdot 5 \ 6)$. Logo, $b_{\bar{r}}(\pi) = 4$ e $b_t(\sigma) = 3$.

2.4 Grafo de Breakpoints

O grafo de breakpoints de uma permutação π sem sinal é um grafo de arestas coloridas $G_b(\pi)$, com conjunto de vértices dado por $V = \{\pi_0, \pi_1, \dots, \pi_n, \pi_{n+1}\}$, que correspondem a cada elemento da notação estendida de π , e conjunto de

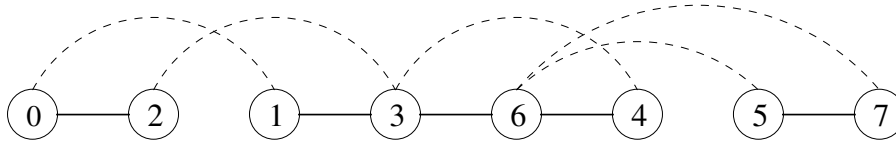


Figura 1: Grafo de breakpoints para a permutação $\pi = (2\ 1\ 3\ 6\ 4\ 5)$. Arestas cinzas estão representadas por arcos pontilhados e arestas pretas por linhas retas.

arestas dado por $E = E_P \cup E_C$, onde E_P é o conjunto de arestas pretas e E_C é o conjunto de arestas cinzas. Existe uma aresta preta entre π_i e π_{i+1} se existe um breakpoint entre π_i e π_{i+1} , $0 \leq i \leq n$. Existe uma aresta cinza entre dois vértices π_i e π_j , para algum $0 \leq i < j \leq n + 1$ se $\pi_j = \pi_i \pm 1$ e π_i e π_j não são consecutivos em π .

Um ciclo no grafo $G_b(\pi)$ é dito alternante se as cores de cada duas arestas consecutivas deste ciclo são sempre distintas. O tamanho de um ciclo alternante C , denotado por $l(C)$, é o número de arestas pretas neste ciclo. Um ciclo é pequeno se $l(C) = 2$ e longo se $l(C) > 2$. Uma permutação π é simples se $G_b(\pi)$ não possui ciclos longos.

Denota-se por $c(\pi)$ o número de ciclos de uma decomposição máxima de $G_b(\pi)$ em ciclos alternantes com arestas disjuntas. Encontrar uma decomposição máxima no número de ciclos alternantes é um problema NP-Difícil [7], cujo melhor algoritmo de aproximação, proposto por Lin e Jiang [19], possui fator $1,4193 + \epsilon$, para algum $\epsilon > 0$. A Figura 1 mostra o grafo de breakpoints para a permutação $\pi = (2\ 1\ 3\ 6\ 4\ 5)$.

Uma extensão do grafo de breakpoints pode ser utilizada para permutações com sinal. Seja a transformação de uma permutação com sinal π de tamanho n para uma permutação sem sinal π' de tamanho $2n$ da seguinte maneira: substitua elementos positivos π_i de π por $(2x - 1, 2x)$ em π' e elementos negativos π_i de π por $(2x, 2x - 1)$ em π' , onde $x = |\pi_i|$. Dizemos que

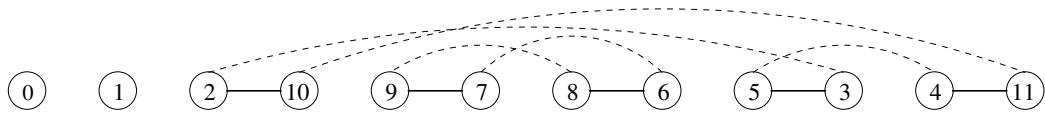


Figura 2: Grafo de breakpoints para a permutação $\pi = (+1 -5 +4 -3 +2)$. Arestas cinzas estão representadas por arcos pontilhados e arestas pretas por linhas retas.

a permutação π' é a imagem da permutação com sinal π . No grafo de breakpoints de π' todo vértice possui grau 2 ou 0, cada componente conexo é um ciclo alternante e cada ciclo alternante possui, no mínimo, 2 arestas pretas, o que nos leva a uma decomposição única em ciclos alternantes. A Figura 2 exibe o grafo de breakpoints para a permutação $\pi = (+1 -5 +4 -3 +2)$.

2.5 Grafo de Ciclos

A estrutura Grafo de Ciclos, introduzida por Bafna e Pevzner [3], serve como base de uma série de limitantes e de algoritmos para o Problema da Ordenação por Transposições.

O grafo de ciclos $G_c(\pi)$, de uma permutação $\pi = (\pi_1 \pi_2 \dots \pi_n)$, é formado pelo conjunto de vértices $V(\pi)$, o conjunto de arestas pretas $E_p(\pi)$ e o conjunto de arestas cinzas $E_c(\pi)$ tais que:

- $V(\pi) = \{-\pi_1, +\pi_1, -\pi_2, +\pi_2, \dots, -\pi_n, +\pi_n\} \cup \{0, -(n+1)\}$
- $E_c(\pi) = \bigcup_{i=1}^{n+1} \{+(i-1), -i\}$
- $E_p(\pi) = \bigcup_{i=1}^{n+1} \{-\pi_i, +\pi_{i-1}\}$

As arestas pretas são numeradas de 1 a $n+1$, sendo que a aresta $(-\pi_i, +\pi_{i-1})$ recebe o rótulo i . Todo vértice de $G(\pi)$ possui uma aresta cinza pareada com uma aresta preta. Isto implica na existência de uma decomposição

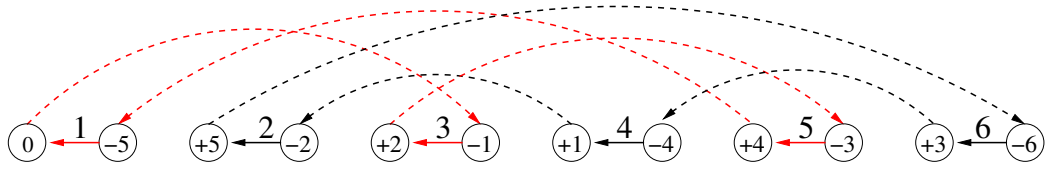


Figura 3: Grafo de ciclos $G(\pi)$. Arestas cinzas estão representadas por arcos pontilhados e arestas pretas por linhas retas. O grafo contém dois ciclos: $C_1 = (5, 1, 3)$ e $C_2 = (6, 4, 2)$. C_1 é um ciclo orientado e C_2 é um ciclo não orientado.

única das arestas de $G(\pi)$ em ciclos com arestas de cores alternantes, chamados de ciclos alternantes. Seja $c(\pi)$ o número de ciclos alternantes de $G(\pi)$. Seja $c_{\text{impar}}(\pi)$ o número de ciclos alternantes com número ímpar de arestas pretas. A permutação ι é a única cujo grafo de ciclos alternantes possui $n+1$ ciclos, sendo todos ímpares. Um ciclo alternantes é chamado de k -ciclo se ele possui $2k$ arestas. Um ciclo C é representado pela listagem dos rótulos das arestas pretas na ordem em que aparecem no ciclo, $C = (i_1, \dots, i_k)$. Assume-se que i_1 é a aresta preta rotulada com o maior número em C . Um k -ciclo é chamado não orientado se as arestas pretas formam uma sequência decrescente. Caso contrário, o ciclo é chamado de orientado. Uma propriedade importante dos ciclos orientados é que permitem a aplicação de uma transposição capaz de gerar um novo ciclo de tamanho ímpar, o que é necessário durante a ordenação de uma permutação, uma vez que a permutação identidade contém apenas ciclos de tamanho ímpar.

A Figura 3 mostra o grafo de ciclos para a permutação $\pi = (5\ 2\ 1\ 4\ 3)$. O grafo possui dois ciclos, $C_1 = (5, 1, 3)$ e $C_2 = (6, 4, 2)$, dos quais C_1 é um ciclo orientado e C_2 é um ciclo não orientado.

3 O Problema da Ordenação de Permutações por Reversões e Transposições

Este problema permite que tanto os eventos de reversão, definidos na Seção 2.1, quanto os eventos de transposição, definidos na Seção 2.2, ocorram durante a ordenação de uma permutação qualquer π . Uma investigação preliminar desse problema foi realizada por Blanchette, Kunisawa e Sankoff [5]. Existem duas versões para este problema, quando considera-se permutações com sinal e quando considera-se permutações sem sinal. Ambas as versões possuem complexidade desconhecida.

Para o caso de permutações com sinal, Walter, Dias e Meidanis [24] apresentaram um algoritmo de aproximação com razão 2 e limitantes para o diâmetro. Meidanis, Walter e Dias [21] apresentaram limitantes para a distância, apresentados na Seção 4. Gu, Peng e Sudborough [14] acrescentaram o evento de transversão ao problema. No evento de transversão, um bloco é destacado do genoma e inserido em outra posição, mas com a ordem e a orientação invertidas. Formalmente, seja $\rho(i, j, k)$ uma transreversão. Então, $\pi\rho(i, j, k) = (\pi_1 \dots \pi_{i-1} - \pi_{k-1} \dots - \pi_j \pi_i \dots \pi_{j-1} \pi_k \dots \pi_n)$. Gu, Peng e Sudborough [14] apresentaram um algoritmo de aproximação na razão 2 quando as três operações são permitidas. Lin e Xue [20] apresentaram um algoritmo com fator de aproximação 1.75 quando, além das três operações citadas anteriormente, é adicionada também a operação *RevRev*, que aplica reversões em dois blocos consecutivos da permutação em apenas uma operação. Mais tarde, Hartman e Sharan [16] melhoraram esta razão para 1.5.

Para o caso de permutações sem sinal, Walter, Dias e Meidanis [24] apresentaram um algoritmo de aproximação com razão 3. Em 2008, Rahman, Shatabda e Hasan [22] apresentaram limitantes para a distância e um al-

goritmo de aproximação com fator $2k$, onde k é o fator de aproximação do algoritmo utilizado para a decomposição de ciclos. Dado o melhor valor de k conhecido [19], o fator de aproximação deste algoritmo torna-se $2.8386 + \epsilon$, para qualquer $\epsilon > 0$.

4 Limitantes para Problemas de Rearranjo de Genomas

No Problema da Ordenação por Reversões sem sinal, uma reversão $\rho(i, j)$ corta dois pontos da permutação: entre (π_{i-1}, π_i) e (π_j, π_{j+1}) . Sendo assim, uma reversão pode criar, manter ou remover até 2 breakpoints e, portanto, $\Delta b_r(\pi, \rho) = b_r(\pi \cdot \rho) - b_r(\pi) \in \{-2, -1, 0, 1, 2\}$. Um limitante inferior para a distância, proposto por Bafna e Pevzner [2], pode ser derivado desta propriedade: $d_r(\pi) \geq \frac{b_r(\pi)}{2}$. Além disso, sempre é possível encontrar uma reversão que diminua em pelo menos 1 o número de breakpoints se $\pi \neq \iota$, e a última reversão para a ordenação de uma permutação sempre remove dois breakpoints. Logo, temos como limitante superior $d_r(\pi) \leq b_r(\pi) - 1$, proposto por Kececioglu e Sankoff [18]. Já em relação ao diâmetro, Bafna e Pevzner [2] provaram que $D_r(n) = n - 1$.

Para o Problema da Ordenação por Transposições, uma transposição $\rho(i, j, k)$ corta três pontos da permutação: entre (π_{i-1}, π_i) , (π_{j-1}, π_j) e (π_{k-1}, π_k) . Deste modo, uma transposição pode criar, manter ou remover até 3 breakpoints e, portanto, $\Delta b_t(\pi, \rho) = b_t(\pi \cdot \rho) - b_t(\pi) \in \{-3, -2, -1, 0, 1, 2, 3\}$. Assim, o limitante inferior para a distância é $d_t(\pi) \geq \frac{b_t(\pi)}{3}$, proposto por Bafna e Pevzner [3]. Além disso, sempre é possível encontrar uma transposição que diminua em pelo menos 1 o número de breakpoints se $\pi \neq \iota$. Logo, temos o limite superior para a distância: $d_t(\pi) \leq b_t(\pi)$. Estes limitantes

utilizam apenas o conceito de breakpoints. Limitantes mais precisos, propostos por Bafna e Pevzner [3], consideram o número de ciclos ímpares no grafo de ciclos da permutação π : $\frac{n+1-c_{\text{ímpar}}(G(\pi))}{2} \leq d_t(\pi) \leq \frac{3(n+1-c_{\text{ímpar}}(G(\pi)))}{4}$. Em relação ao diâmetro, Bafna e Pevzner [3] provaram o limitante inferior, $\lfloor \frac{n+1}{2} \rfloor \leq D_t(n)$, e Eriksson e coautores [11] provaram o limitante superior, $D_t(n) \leq \lfloor \frac{2n-2}{3} \rfloor$.

Já para o problema da Ordenação por Reversões e Transposições, considerando permutações com sinal e sem sinal, Walter, Dias e Meidanis [24] provaram um limitante inferior para a distância, $d_{rt}(\pi), d_{\bar{rt}}(\pi) \geq \frac{n+1-c_b(\pi)}{2}$, e um limitante inferior para o diâmetro, $\lfloor \frac{n}{2} \rfloor + 2 \leq D_{rt}(n), D_{\bar{rt}}(n)$.

Uma *família* é um grupo de permutações que compartilham uma mesma regra de definição. Para estudar melhor o Problema da Ordenação de Permutações por Reversões e Transposições, definimos uma família de permutações com sinal e uma família de permutações sem sinal e, para cada família, conjecturamos a distância de ordenação exata correspondente. Para isso, elaboramos um algoritmo de ordenação específico para cada família.

Considere as seguintes famílias de permutações:

$$\begin{aligned}
- \pi_n^1 &= (+n \ + (n-1) \ - (n-2) \ [trinca_1] \ [trinca_2] \ \dots \ [trinca_k]), \text{ com:} \\
- k &= \lceil \frac{n-3}{3} \rceil; \\
- trinca_i &= (+ (3i-2) \ + 3i \ - (3i-1)), \text{ para } 1 \leq i < k; \\
- trinca_k &= \begin{cases} (+ (3k-2) \ + 3k \ - (3k-1)), & \text{se } n \equiv 3 \pmod{3}; \\ (+ (3k-2) \ + (3k-1)), & \text{se } n \equiv 2 \pmod{3}; \\ (+ (3k-2)), & \text{se } n \equiv 1 \pmod{3}. \end{cases} \\
- \pi_n^2 &= \begin{cases} (n \ (n-2) \ \dots \ 2 \ (n-1) \ (n-3) \ \dots \ 1), & \text{se } n \text{ for par}; \\ (n \ (n-2) \ \dots \ 1 \ (n-1) \ (n-3) \ \dots \ 2) & \text{se } n \text{ for ímpar.} \end{cases}
\end{aligned}$$

Lema 1. *Para o Problema da Ordenação por Reversões e Transposições, para permutações com sinal, $d_{\bar{r}t}(\pi_n^1) \leq n - \lfloor \frac{n-2}{3} \rfloor$, para $n \geq 3$.*

Demonstração. Considere a família π_n^1 . É possível ordenar utilizando o número conjecturado de operações. Aplique duas transposições $\tau(1, 2, n+1)$ e $\tau(1, 2, n)$, colocando os dois últimos elementos da permutação nas posições corretas. Faça as seguintes operações, dependendo do valor de n :

1. Se $n \not\equiv 0 \pmod{3}$: Aplique uma reversão $\rho(2, n-2)$.
2. Se $n \equiv 0 \pmod{3}$: Aplique uma reversão $\rho(2, n-3)$ e uma transposição $\tau(3, n-2, n-1)$.

Repita $k-1$ vezes as seguintes operações: aplique uma transposição $\tau(n-4-3i, n-3-3i, n-2-3i)$ e uma reversão $\rho(n-3-3i, n-3-3i)$, com $0 \leq i \leq k-2$. Aplique uma reversão $\rho(1, n-2)$ para ordenar a permutação. Assim, temos $d_{\bar{r}t}(\pi_n^1) \leq 2 + 1 + 2(k-1) + 1 = 2k + 2 = 2\lceil \frac{n-3}{3} \rceil + 2 = 2\lceil \frac{n}{3} \rceil = n - \lfloor \frac{n-2}{3} \rfloor$, se $n \not\equiv 0 \pmod{3}$ e $d_{\bar{r}t}(\pi_n^1) \leq 2 + 2 + 2(k-1) + 1 = 2k + 3 = 2\lceil \frac{n-3}{3} \rceil + 3 = 2\lceil \frac{n}{3} \rceil + 1 = n - \lfloor \frac{n-2}{3} \rfloor$, se $n \equiv 0 \pmod{3}$. \square

Conjectura 1. *Para o Problema da Ordenação por Reversões e Transposições, para permutações com sinal, $d_{\bar{r}t}(\pi_n^1) = n - \lfloor \frac{n-2}{3} \rfloor$, para $n \geq 3$.*

A conjectura acima é verdadeira para $n \leq 11$.

Conjectura 2. *Para o Problema da Ordenação por Reversões e Transposições, para permutações com sinal, $D(n) = d_{\bar{r}t}(\pi_n^1) = n - \lfloor \frac{n-2}{3} \rfloor$, para $n \geq 3$.*

A conjectura acima é verdadeira para $n \leq 10$.

Lema 2. *Para o Problema da Ordenação por Reversões e Transposições, considerando permutações sem sinal, $d_{rt}(\pi_n^2) \leq \lceil \frac{n}{2} \rceil$, para $n \geq 4$.*

Demonstração. Considere a família π_n^2 . É possível ordenar usando o número conjecturado de operações:

1. Faça as seguintes operações, dependendo da paridade de n :

1.1 Se n for ímpar, use $\lfloor \frac{n}{2} \rfloor$ transposições para posicionar os elementos da segunda sequência $((n-1), (n-3), \dots, 2)$ entre os elementos adequados na primeira sequência. Ou seja, posicione o elemento 2 entre o 3 e o 1, o 4 entre o 5 e o 3, e assim por diante.

1.2 Se n for par, faça as seguintes operações, dependendo da paridade de $\frac{n}{2}$:

1.2.1 Se $\frac{n}{2}$ for par, aplique $(\frac{n}{4}) - 1$ transposições da forma $\tau(3, \frac{n}{2} + 1, n - 1), \tau(5, \frac{n}{2} + 1, n - 3), \tau(7, \frac{n}{2} + 1, n - 5), \dots, \tau(\frac{n}{2} - 1, \frac{n}{2} + 1, \frac{n}{2} + 3)$. A seguir, aplique $\frac{n}{4}$ transposições ou reversões da forma $\tau(2, 3, 4) = \rho(2, 3), \tau(6, 7, 8) = \rho(6, 7), \tau(10, 11, 12) = \rho(10, 11), \dots, \tau(n - 2, n - 1, n) = \rho(n - 2, n - 1)$.

1.2.2 Se $\frac{n}{2}$ for ímpar, aplique $\lfloor \frac{n}{4} \rfloor - 1$ transposições da forma $\tau(3, \frac{n}{2} + 1, n - 1), \tau(5, \frac{n}{2} + 1, n - 3), \tau(7, \frac{n}{2} + 1, n - 5), \dots, \tau(\frac{n}{2} - 2, \frac{n}{2} + 1, \frac{n}{2} + 4)$. A seguir, aplique a transposição $\tau(\frac{n}{2}, \frac{n}{2} + 1, \frac{n}{2} + 3)$. Depois, aplique $\lfloor \frac{n+2}{8} \rfloor$ transposições ou reversões da forma $\tau(2, 3, 4) = \rho(2, 3), \tau(6, 7, 8) = \rho(6, 7), \tau(10, 11, 12) = \rho(10, 11), \dots, \tau(\frac{n}{2} - 3, \frac{n}{2} - 2, \frac{n}{2} - 1) = \rho(\frac{n}{2} - 3, \frac{n}{2} - 2)$. Então, aplique a operação $\tau(\frac{n}{2} + 1, \frac{n}{2} + 2, \frac{n}{2} + 3) = \rho(\frac{n}{2} + 1, \frac{n}{2} + 2)$ seguida de $\lfloor \frac{n}{8} \rfloor$ transposições ou reversões da forma $\tau(\frac{n}{2} + 3, \frac{n}{2} + 4, \frac{n}{2} + 5) = \rho(\frac{n}{2} + 3, \frac{n}{2} + 4), \tau(\frac{n}{2} + 7, \frac{n}{2} + 8, \frac{n}{2} + 8) = \rho(\frac{n}{2} + 7, \frac{n}{2} + 8), \dots, \tau(\frac{n}{2} - 3, \frac{n}{2} - 2, \frac{n}{2} - 1) = \rho(\frac{n}{2} - 3, \frac{n}{2} - 2)$.

2. Use uma reversão $\rho(1, n)$ para reverter a sequência completa.

Note que $d_{rt}(\pi_n^2) \leq \lfloor \frac{n}{2} \rfloor + 1 = \lceil \frac{n}{2} \rceil$, $d_{rt}(\pi_n^2) \leq \frac{n}{4} - 1 + \frac{n}{4} + 1 = \frac{n}{2}$, se $\frac{n}{2}$ for par, e $d_{rt}(\pi_n^2) \leq (\lfloor \frac{n}{4} \rfloor - 1) + 1 + \lfloor \frac{n+2}{8} \rfloor + 1 + \lfloor \frac{n}{8} \rfloor = \lfloor \frac{n}{4} \rfloor + \lfloor \frac{n+2}{8} \rfloor + \lfloor \frac{n}{8} \rfloor + 1 = \frac{n}{2}$, se $\frac{n}{2}$ for ímpar. \square

Conjectura 3. *Para o Problema da Ordenação por Reversões e Transposições, considerando permutações sem sinal, $d_{rt}(\pi_n^2) = \lceil \frac{n}{2} \rceil$, para $n \geq 4$.*

Conjectura 4. *Para o Problema da Ordenação por Reversões e Transposições, considerando permutações sem sinal, $D(n) = d_{rt}(\pi_n^2) = \lceil \frac{n}{2} \rceil$, para $n \geq 4$.*

As Conjecturas 3 e 4 são verdadeiras para $n \leq 13$.

5 Objetivos

Para o Problema da Ordenação de Permutações por Reversões e Transposições apresentado na Seção 3, considerando permutações com sinal e permutações sem sinal, pretendemos melhorar os limitantes existentes, obter algoritmos de aproximação com fatores melhores que os existentes assim como estudar algumas classes de permutações, que podem ajudar a entender melhor o problema. Acreditamos que os fatores de aproximação atual, 2 para permutações com sinal e $2k$ para permutações sem sinal, onde k é o fator de aproximação do algoritmo utilizado para a decomposição de ciclos, possam ser melhorados, tendo em vista que os melhores algoritmos de aproximação para o problema de transposição e o problema de reversão sem sinal possuem fator de aproximação 1.375 [4, 10]. Além disso, trabalharemos na prova das conjecturas propostas na Seção 4.

6 Metodologia

O trabalho a ser realizado é, em grande parte, teórico. O estudo de teoremas e provas dos trabalhos já existentes na literatura são a principal base para o estabelecimento dos limitantes e das classes de permutações específicas. Sempre que possível e necessário, no entanto, programas serão desenvolvidos para comprovar a validade dos resultados obtidos.

7 Cronograma

	2013					2014										2015									
	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	
1	*	*	*	*				*	*	*	*														
2		*	*	*	*	*	*	*	*				*	*				*	*	*					
3					*	*	*																		
4									*																
5								*	*	*	*	*	*												
6													*	*	*	*	*	*							
7																		*	*	*	*	*	*	*	
8												*	*				*	*			*	*	*		
9																							*		
10																									*

1. Obtenção dos créditos obrigatórios em disciplinas do programa de mestrado;
2. Revisão Bibliográfica;
3. Escrita da proposta de mestrado;
4. Exame de qualificação do mestrado;
5. Investigação sobre limitantes, considerando permutações com sinal e sem sinal, para o valor do diâmetro;

6. Definição das classes de permutações com sinal e permutações sem sinal para as quais é possível determinar a distância exata;
7. Investigação de algoritmo com um melhor fator de aproximação, para as versões do problema que consideram permutações com sinal e sem sinal;
8. Escrita da dissertação;
9. Revisão da dissertação;
10. Defesa da dissertação.

É importante ressaltar que o cronograma apresentado pode ser adaptado à medida que os avanços ocorrerem, uma vez que tais avanços poderão nos levar a resultados mais promissores que outros, o que nos faria dedicar mais tempo em uma ou mais atividades em detrimento de outras.

Referências

- [1] D. A. Bader, B. M. E. Moret, and M. Yan. A linear-time algorithm for computing inversion distance between signed permutations with an experimental study. *Journal of Computational Biology*, 8:483–491, 2001.
- [2] V. Bafna and P. A. Pevzner. Genome rearrangements and sorting by reversals. *SIAM Journal of Computing*, 25(2):272–289, 1996.
- [3] V. Bafna and P. A. Pevzner. Sorting by transpositions. *SIAM Journal on Discrete Mathematics*, 11(2):224–240, 1998.
- [4] P. Berman, S. Hannenhalli, and M. Karpinski. 1.375-approximation algorithm for sorting by reversals. In *Proceedings of the 10th Annual*

- European Symposium on Algorithms (ESA '2002)*, pages 200–210, Rome, Italy, 2002.
- [5] M. Blanchette, T. Kunisawa, and D. Sankoff. Parametric genome rearrangement. *Journal of Computational Biology*, 172:11–17, 1996.
- [6] L. Bulteau, G. Fertin, and I. Rusu. Sorting by transpositions is difficult. *SIAM Journal on Discrete Mathematics*, 26(3):1148–1180, 2012.
- [7] A. Caprara. Sorting by reversals is difficult. In *Proceedings of the 1st Annual International Conference on Computational Molecular Biology (RECOMB'1997)*, pages 75–83, Santa Fe, New Mexico, USA, 1997.
- [8] D. A. Christie. A $3/2$ -approximation algorithm for sorting by reversals. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'1998)*, pages 244–252, San Francisco, California, USA, 1998.
- [9] D. A. Christie. *Genome rearrangement problems*. PhD thesis, Glasgow University, 1998.
- [10] I. Elias and T. Hartman. A 1.375 -approximation algorithm for sorting by transpositions. *Transactions on Computational Biology and Bioinformatics*, 3(4):369–379, 2006.
- [11] H. Eriksson, K. Eriksson, J. Karlander, L. Svensson, and J. Wastlund. Sorting a bridge hand. *Discrete Mathematics*, 241(1-3):289–300, 2001.
- [12] J. Fischer and S. W. Ginzinger. A 2 -approximation algorithm for sorting by prefix reversals. In *Proceedings of the 13th Annual European Conference on Algorithms (ESA'2005)*, pages 415–425, Palma de Mallorca, Spain, 2005.

- [13] W. H. Gates and C. H. Papadimitriou. Bounds for sorting by prefix reversal. *Discrete Mathematics*, 27:47–57, 1979.
- [14] Q.-P. Gu, S. Peng, and H. Sudborough. A 2-approximation algorithm for genome rearrangements by reversals and transpositions. *Theoretical Computer Science*, 210(2):327–339, 1999.
- [15] S. Hannenhalli and P. A. Pevzner. Transforming men into mice (polynomial algorithm for genomic distance problem). In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS'1995)*, pages 581–592, Milwaukee, Wisconsin, USA, 1995.
- [16] T. Hartman and R. Sharan. A 1.5-approximation algorithm for sorting by transpositions and transreversals. *Journal of Computer and System Sciences*, 70(3):300–320, 2005.
- [17] M. H. Heydari and I. H. Sudborough. On the diameter of the pancake network. *Journal of Algorithms*, 25(1):67–94, 1997.
- [18] J. Kececioglu and D. Sankoff. Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. *Algorithmica*, 13:180–210, 1995.
- [19] G. Lin and T. Jiang. A further improved approximation algorithm for breakpoint graph decomposition. *Journal of Combinatorial Optimization*, 8(2):183–194, 2004.
- [20] G.-H. Lin and G. Xue. Signed genome rearrangement by reversals and transpositions: models and approximations. *Theoretical Computer Science*, 259(1-2):513–531, 2001.

- [21] J. Meidanis, M. E. M. T. Walter, and Z. Dias. A lower bound on the reversal and transposition diameter. *Journal of Computational Biology*, 9(5):743–745, 2002.
- [22] A. Rahman, S. Shatabda, and M. Hasan. An approximation algorithm for sorting by reversals and transpositions. *Journal of Discrete Algorithms*, 6(3):449–457, 2008.
- [23] E. Tannier, A. Bergeron, and M.-F. Sagot. Advances on sorting by reversals. *Discrete Applied Mathematics*, 155(6-7):881–888, 2007.
- [24] M. E. M. T. Walter, Z. Dias, and J. Meidanis. Reversal and transposition distance of linear chromosomes. In *Proceedings of the 5th International Symposium on String Processing Information Retrieval (SPIRE'1998)*, pages 96–102, Santa Cruz, Bolivia, 1998.
- [25] M. E. M. T. Walter, Z. Dias, and J. Meidanis. A new approach for approximating the transposition distance. In *Proceedings of the 7th International Symposium on String Processing Information Retrieval (SPIRE'2000)*, pages 199–208, A Coruña, Spain, 2000.