

O Problema da Ordenação de Permutações por Reversões e Transposições

Andre Rodrigues Oliveira

Instituto de Computação
Universidade Estadual de Campinas
Orientador: Zaroni Dias
Dissertação de Mestrado

Agenda

- 1 Introdução
- 2 Conceitos
- 3 Heurísticas Básicas
- 4 Heurística de Grafo de Ciclos
- 5 Conclusões

Introdução

- Evolução
 - Mudança das características hereditárias.
 - Resultado de mutações ocorridas no material genético.
- Mutações pontuais
 - Afetam bases individuais do DNA.
 - Ocorrem com maior frequência.
- Rearranjos de genomas
 - Mutações que afetam grandes trechos do DNA.
 - Comuns em plantas, mamíferos, vírus e bactérias.
- Realizar a comparação entre genomas por meio de rearranjos.

Conceitos

- 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.

Conceitos

- A *permutação identidade* é a permutação ordenada e é denotada por $\iota = (1 \ 2 \ \dots \ n)$.
 - Para permutações com sinais, $\iota = (+1 \ +2 \ \dots \ +n)$.
- A *permutação estendida* é obtida a partir de π inserindo-se dois novos elementos: $\pi_0 = 0$ e $\pi_{n+1} = n + 1$.
- O número de elementos negativos em uma permutação com sinais é denotado por $neg(\pi)$.
- Denotamos por $u(\pi)$ o processo de remoção dos sinais dos elementos de uma permutação π .

Eventos de Mutação

Para efeito de comparação entre dois genomas, consideramos apenas eventos conservativos. Os principais eventos de mutação são:

- Reversão
- Transposição
- Fissão
- Fusão

Reversão

- Um *evento de reversão* $\rho_r(i, j)$, com $1 \leq i \leq j \leq n$, reverte a ordem e a orientação de $\pi[i..j]$, ou seja:

$$\pi_1 \quad \pi_2 \quad \dots \quad \pi_i \quad \pi_{i+1} \quad \dots \quad \pi_{j-1} \quad \pi_j \quad \dots \quad \pi_{n-1} \quad \pi_n$$

$$\pi_1 \quad \pi_2 \quad \dots \quad -\pi_j \quad -\pi_{j-1} \quad \dots \quad -\pi_{i+1} \quad -\pi_i \quad \dots \quad \pi_{n-1} \quad \pi_n$$

Reversão

- *Distância de reversão*: dadas as permutações π e σ , calcular a distância de reversão, $d_r(\pi, \sigma)$, consiste em obter uma sequência de k reversões ρ_1, \dots, ρ_k , de tamanho mínimo, tal que $d_r(\pi, \sigma) = k$ e $\pi \cdot \rho_1 \cdot \dots \cdot \rho_k = \sigma$.
- *Ordenação por reversões*: dada π , consiste em calcular a distância de ordenação por reversões, $d_r(\pi)$, entre π e ι , ou seja, $d_r(\pi) = d_r(\pi, \iota)$.

Transposição

- Um *evento de transposição* $\rho_t(i, j, k)$, com $1 \leq i < j < k \leq n + 1$, troca os blocos $\pi[i..j - 1]$ e $\pi[j..k - 1]$ de lugar, ou seja:

$$\pi_1 \quad \dots \quad \pi_{i-1} \quad \pi_i \quad \pi_{i+1} \quad \dots \quad \pi_{j-1} \quad \pi_j \quad \pi_{j+1} \quad \dots \quad \pi_{k-1} \quad \pi_k \quad \dots \quad \pi_n$$

$$\pi_1 \quad \dots \quad \pi_{i-1} \quad \pi_j \quad \pi_{j+1} \quad \dots \quad \pi_{k-1} \quad \pi_i \quad \pi_{i+1} \quad \dots \quad \pi_{j-1} \quad \pi_k \quad \dots \quad \pi_n$$

Transposição

- *Distância de transposição*: dadas π e σ , calcular a distância de transposição, $d_t(\pi, \sigma)$, consiste em obter uma sequência de k transposições ρ_1, \dots, ρ_k de tamanho mínimo, tal que $d_t(\pi, \sigma) = k$ e $\pi \cdot \rho_1 \cdot \dots \cdot \rho_k = \sigma$.
- *Ordenação por transposições*: dada π , consiste em calcular a distância de ordenação por transposições, $d_t(\pi)$, entre π e ι , ou seja, $d_t(\pi) = d_t(\pi, \iota)$.

Modelo de Rearranjo

- Um *modelo de rearranjo* β determina o conjunto de operações permitidas para transformar um genoma em outro.
- A distância de ordenação, $d_\beta(\pi)$, é específica para cada permutação.
- A maior distância de ordenação de uma permutação de tamanho n é o *diâmetro* da distância de rearranjo, denotada por $D_\beta(n)$.

O Problema de Ordenação por Reversões sem Sinais

- Em 1996, Bafna e Pevzner introduziram a idéia do *Grafo de Breakpoints*, e apresentaram um algoritmo de aproximação de fator 1.75.
- Em 1997, Caprara provou que este problema é NP-Difícil.
- Atualmente, o melhor algoritmo conhecido para o problema possui fator de aproximação 1.375, proposto por Berman e coautores em 2002.

O Problema de Ordenação por Reversões com Sinais

- O primeiro algoritmo polinomial exato para este problema foi apresentado em 1995 por Hannenhalli e Pevzner, com complexidade $O(n^4)$.
- Atualmente, o algoritmo mais eficiente disponível possui complexidade $O(n\sqrt{n \log n})$, apresentado em 2004 por Tannier e Sagot.
- Há também um algoritmo linear que retorna apenas o valor da distância de reversão de permutações, apresentado em 2001 por Bader, Moret e Yan.

O Problema de Ordenação por Transposições

- O primeiro algoritmo de aproximação foi apresentado por Bafna e Pevzner em 1998, e possui fator de aproximação 1.5.
- Em 2006, Elias e Hartman desenvolveram um algoritmo de aproximação com fator de 1.375.
- Em 2011, Bultheau, Fertin e Rusu provaram que o problema da distância de transposição é NP-Difícil.

O Problema de Ordenação por Reversões e Transposições

- Este problema permite que os eventos de reversão e os eventos de transposição ocorram durante a ordenação de uma permutação qualquer π .
- Existem duas versões para este problema, quando considera-se permutações com sinais e quando considera-se permutações sem sinais.
- A complexidade de ambas as versões é desconhecida.

O Problema de Ordenação por Reversões e Transposições

Para o caso de permutações com sinais:

- Em 1998, Walter, Dias e Meidanis apresentaram um algoritmo de aproximação com razão 2 e limitantes para o diâmetro.

Para o caso de permutações sem sinais:

- Em 1998, Walter, Dias e Meidanis apresentaram um algoritmo de aproximação com razão 3.
- Em 2008, Rahman, Shatabda e Hasan apresentaram limitantes para a distância e um algoritmo 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, obtemos a aproximação 2.8334.
- Em 2014, Lintzmayer e Dias apresentaram limitantes para o diâmetro deste problema.

Breakpoints

- Um breakpoint para o problema de ordenação por reversões sem sinais ocorre entre um par de elementos π_i e π_{i+1} de π , com $0 \leq i \leq n$, se $|\pi_i - \pi_{i+1}| \neq 1$.
- O número de breakpoints de reversão de π é denotado por $b_r(\pi)$.
- $b_r(\pi) = 0 \Leftrightarrow \pi = \iota$.
- Exemplos:
 - $\pi = (0 \cdot 2 \ 3 \cdot 1 \cdot 4 \cdot 6 \ 5 \cdot 7)$ e $b_r(\pi) = 5$.
 - $\sigma = (0 \cdot 3 \ 4 \ 5 \cdot 2 \ 1 \cdot 6)$ e $b_r(\sigma) = 3$.

Breakpoints

- Um breakpoint para o problema de ordenação por reversões com sinais e 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 $\pi_{i+1} - \pi_i \neq 1$.
- O número de breakpoints de reversão com sinais e transposições de π é denotado por $b_{\bar{r}}(\pi)$ e $b_t(\pi)$, respectivamente.
- $b_{\bar{r}}(\pi) = b_t(\pi) = 0 \Leftrightarrow \pi = \iota$.
- Exemplos:
 - $\pi = (0 \cdot -1 \cdot +2 \ +3 \cdot -5 \ -4 \cdot +6 \ +7)$ e $b_{\bar{r}}(\pi) = 4$.
 - $\sigma = (0 \cdot 3 \ 4 \ 5 \cdot 2 \cdot 1 \cdot 6)$ e $b_t(\sigma) = 4$.

Grafo de Breakpoints

- O grafo de breakpoints de uma permutação π sem sinais é denotado por $G_b(\pi)$.
- O conjunto de vértices é dado por $V = \{\pi_0, \pi_1, \dots, \pi_n, \pi_{n+1}\}$.
- O conjunto de arestas é dividido em arestas pretas e arestas cinzas.
- Existe uma aresta preta entre π_i e π_{i+1} se existe um breakpoint de reversão entre estes elementos.
- Existe uma aresta cinza entre π_i e π_j se $\pi_j = \pi_i \pm 1$ e π_i e π_j não são adjacentes em π .

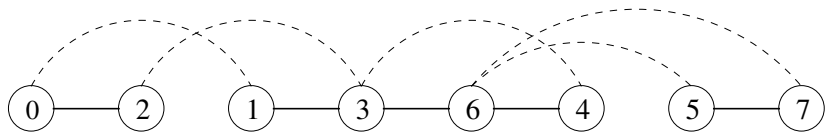


Figura: Grafo de breakpoints da permutação sem sinais $\pi = (2\ 1\ 3\ 6\ 4\ 5)$. Arestas cinzas estão representadas por arcos tracejados e arestas pretas por linhas retas.

Grafo de Ciclos

- O grafo de ciclos de uma permutação π com sinais é um grafo de arestas coloridas $G(\pi)$, onde:
 - O conjunto de vértices é dado por $V = \{+0, -\pi_1, +\pi_1, \dots, -\pi_n, +\pi_n, -(n+1)\}$.
 - O conjunto de arestas é 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.
 - $E_P = \{(-\pi_i, +\pi_{i-1}), \text{ para } 1 \leq i \leq n+1\}$.
 - $E_C = \{+(i-1), -i\}, \text{ para } 1 \leq i \leq n+1\}$.

Grafo de Ciclos

- Denota-se por $c(\pi)$ o número de ciclos alternados de $G(\pi)$.
- Seja $c_{\text{impar}}(\pi)$ o número de ciclos alternados com número ímpar de arestas pretas de $G(\pi)$.
- A permutação ι é a única cujo grafo de ciclos possui $n + 1$ ciclos, sendo todos ímpares e com apenas uma aresta preta cada ($c(\iota) = c_{\text{impar}}(\iota) = n + 1$).

Limitantes

Para o Problema da Ordenação por Reversões sem sinais:

- $d_r(\pi) \geq \frac{b_r(\pi)}{2}$ (Bafna e Pevzner, 1996).
- $d_r(\pi) \leq b_r(\pi) - 1$ (Kececioglu e Sankoff, 1995).
- $D_r(n) = n - 1$ (Bafna e Pevzner, 1996).

Limitantes

Para o Problema da Ordenação por Transposições:

- $\frac{n+1 - c_{\text{impar}}(G(\pi))}{2} \leq d_t(\pi) \leq \frac{3(n+1 - c_{\text{impar}}(G(\pi)))}{4}$ (Bafna e Pevzner, 1998).
- $\lfloor \frac{n+1}{2} \rfloor \leq D_t(n)$ (Bafna e Pevzner, 1998).
- $D_t(n) \leq \lfloor \frac{2n-2}{3} \rfloor$ (Eriksson e coautores, 2001).

Limitantes

Para o Problema da Ordenação por Reversões e Transposições com sinais:

- $d_{\bar{r}t}(\pi) \geq \frac{n+1-c_{\text{impar}}(\pi)}{2}$.
- $\lfloor \frac{n}{2} \rfloor + 2 \leq D_{\bar{r}t}(n)$ (Walter, Dias e Meidanis, 1998).

Para o Problema da Ordenação por Reversões e Transposições sem sinais:

- $d_{rt}(\pi) \geq \frac{b_r(\pi)}{3}$.
- $\lceil \frac{n}{2} \rceil \leq D_{rt}(n)$, para $n \geq 4$ (Lintzmayer e Dias, 2014).

Heurísticas Básicas

- Foi definido um score específico para cada heurística.
- A cada passo, uma heurística aplica todas as operações possíveis e calcula o score de cada permutação resultante
- É aplicado então a operação que resulta no melhor score possível.

Heurísticas Básicas

- Códigos Esquerdo e Direito.
- Breakpoints.
- LIS.
- LIS+LDS.
- LIS+Breakpoints.
- Entropia.
- Inversões.

Códigos Esquerdo e Direito

- Código esquerdo de um elemento: $lc(\pi_i) = |\{\pi_j | \pi_j > \pi_i \text{ e } 0 \leq j \leq i - 1\}|$, para $1 \leq i \leq n$.
- Código esquerdo de uma permutação: $lc(\pi) = (lc(\pi_1) \dots lc(\pi_n))$
- Código direito de um elemento: $rc(\pi_i) = |\{\pi_j | \pi_j < \pi_i \text{ e } i + 1 \leq j \leq n + 1\}|$, para $1 \leq i \leq n$.
- Código direito de uma permutação: $rc(\pi) = (rc(\pi_1) \dots rc(\pi_n))$

Códigos Esquerdo e Direito

- *Platô*: sequência maximal de elementos do código que possuem o mesmo valor não-nulo.
- Dados os códigos esquerdo e direito de uma permutação π , denotamos por $pl(\pi)$ o mínimo entre os platôs $pl(lc(\pi))$ e $pl(rc(\pi))$.
- $score_{lc+rc}(\pi) = pl(u(\pi)) + neg(\pi)$.

Lema

Dada uma permutação π tal que $\pi \neq \iota$, sempre é possível diminuir o valor de $score_{lc+rc}(\pi)$ com uma operação.

- Complexidade da heurística: $O(n^6)$

Códigos Esquerdo e Direito

- Exemplo:

$$\begin{aligned}\alpha &= (6 \ 3 \ 2 \ 1 \ 4 \ 5) \\ lc(\alpha) &= (0 \ 1 \ 2 \ 3 \ 1 \ 1) \\ rc(\alpha) &= (5 \ 2 \ 1 \ 0 \ 0 \ 0)\end{aligned}$$

- Neste caso, $pl(lc(\alpha)) = 4$, $pl(rc(\alpha)) = 3$ e $pl(\alpha) = 3$.
- $score_{lc+rc}(\alpha) = pl(u(\alpha)) + neg(\alpha) = 3 + 0 = 3$.

Breakpoints

- Dada uma permutação π sem sinais, $score_{BP}(\pi) = b_r(\pi)$.
- Dada uma permutação π com sinais, $score_{BP}(\pi) = b_t(\pi)$.

Lema

Dada uma permutação π , com $\pi \neq \iota$, sempre é possível diminuir o valor de $score_{BP}(\pi)$ com uma operação.

- Complexidade da heurística: $O(n^2)$.

LIS

- Subsequência crescente de elementos de π tal que o tamanho desta subsequência seja o maior possível.
- Para $\alpha = (4\ 2\ 3\ 5\ 1\ 7\ 6\ 8)$, temos que $|LIS(\alpha)| = |\{2, 3, 5, 6, 8\}| = 5$
 - Não é única: $\{2, 3, 5, 7, 8\}$ também é uma LIS de α .

LIS

- $score_{LIS}(\pi) = |LIS(u(\pi))| - neg(\pi)$.

Lema

Dada uma permutação π tal que $\pi \neq \iota$, sempre é possível aumentar o valor de $score_{LIS}(\pi)$ com uma operação.

- Complexidade da heurística: $O(n^5 \log n)$.

LIS+LDS

- Além da subsequência crescente máxima, considera também a subsequência decrescente máxima de π .
- Por exemplo, para $\alpha = (4\ 2\ 3\ 5\ 1\ 7\ 6\ 8)$, temos que $|LDS(\alpha)| = |\{4, 2, 1\}| = 3$
 - Não é única: $\{4, 3, 1\}$ também é uma *LDS* de α .

LIS+LDS

- $score_{LIS+LDS}(\pi) = \max(|LIS(u(\pi))|, |LDS(u(\pi))|) - neg(\pi)$.

Lema

Dada uma permutação π tal que $\pi \neq \iota$, sempre é possível aumentar o valor de $score_{LIS+LDS}(\pi)$ com uma operação.

- Complexidade da heurística: $O(n^5 \log n)$.

LIS+Breakpoints

- Utiliza as métricas LIS e Breakpoints para decidir qual operação aplicar.
- $score_{LIS+BP}(\pi) = score_{LIS}(\pi) - b(\pi)$.

Lema

Dada uma permutação π tal que $\pi \neq \iota$, sempre é possível aumentar o valor de $score_{LIS+BP}(\pi)$ com uma operação.

- Complexidade da heurística: $O(n^5 \log n)$.

Entropia

- Nível de “desordem” dos elementos da permutação.
- $entropia(\pi_i) = ||\pi_i| - i|$.
- $entropia(\pi) = \sum_{i=1}^n entropia(\pi_i)$.
- Por exemplo, para $\sigma = (1\ 4\ 3\ 5\ 2\ 6)$, temos $entropia(\sigma) = (0\ 2\ 0\ 1\ 3\ 0) = 2 + 1 + 3 = 6$.

Entropia

- $score_{ENTRO}(\pi) = entropia(\pi) + neg(\pi)$.

Lema

Dada uma permutação π tal que $\pi \neq \iota$, sempre é possível diminuir o valor de $score_{ENTRO}(\pi)$ com uma operação.

- Complexidade da heurística: $O(n^5)$.

Inversões

- Ocorre uma *inversão* entre dois elementos π_i e π_j da permutação π se $|\pi_i| > |\pi_j|$ e $j > i$
- O número de inversões em uma permutação π é denotado por $inv(\pi)$.
- Por exemplo, para $\sigma = (2\ 5\ 3\ 4\ 1\ 6)$, $inv(\sigma) = |\{(2, 1), (5, 3), (5, 4), (5, 1), (3, 1), (4, 1)\}| = 6$.

Inversões

- $score_{INV}(\pi) = inv(\pi) + neg(\pi)$.

Lema

Dada uma permutação π tal que $\pi \neq \iota$, sempre é possível diminuir o valor de $score_{INV}(\pi)$ com uma operação.

- Complexidade da heurística: $O(n^5 \log n)$.

Banco de permutações

- Com o objetivo de comparar a performance das heurísticas, foram criados dois bancos de permutações, um com permutações com sinais e outro com permutações sem sinais.
- As permutações possuem tamanhos variando entre 10 e 100, com intervalos de tamanho 5.
- Para cada tamanho, 10000 permutações foram obtidas a partir da permutação ι , aplicando 10 reversões e 10 transposições de forma aleatória.
 - É possível ordenar qualquer permutação gerada com no máximo 20 operações.

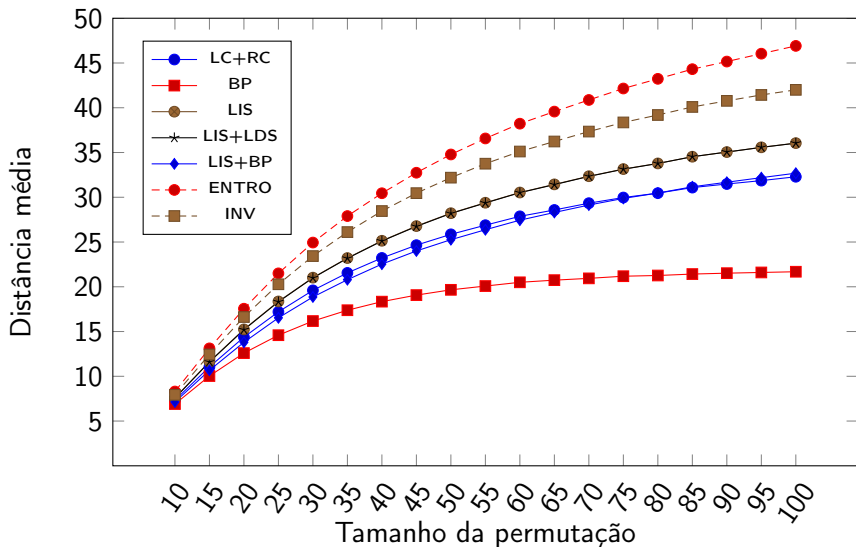


Figura: Distância média para permutações com sinais.

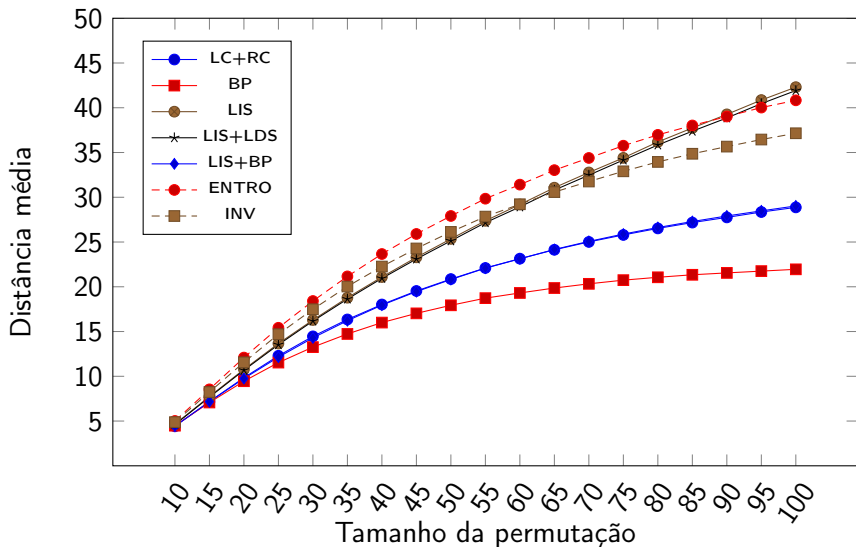


Figura: Distância média para permutações sem sinais.

Heurística de Grafo de Ciclos

- Tem como idéia básica a utilização do Grafo de Ciclos para ordenar permutações com sinais ou sem sinais.
- Criado banco de configurações que auxiliam a heurística durante a ordenação.
- Dividida em 3 etapas:
 - 1 Remoção de Breakpoints.
 - 2 Análise dos ciclos.
 - 3 Busca de uma sequência de operações em bancos de configurações.

Grafo de Ciclos (permutações sem sinais)

- O grafo de ciclos $G(\pi)$ de uma permutação π sem sinais é obtido a partir da decomposição de $G_b(\pi)$ em ciclos.
 - Uma decomposição máxima de $G_b(\pi)$ em ciclos alternantes é um problema NP-Difícil.
 - O melhor fator de aproximação conhecido é $1.4167 + \epsilon$, para $\epsilon > 0$, proposto por Chen em 2013.
- Os ciclos unitários, que não fazem parte de $G_b(\pi)$ são adicionados a $G(\pi)$.

Grafo de Ciclos

- Um ciclo do grafo de ciclos é dito *curto* caso contenha no máximo 3 arestas pretas, e *longo* caso contrário.
- Uma permutação é dita *simples* se seu grafo de ciclos contém apenas ciclos curtos.
- Arestas pretas de $G(\pi)$ são rotuladas de 1 a $n + 1$ e possuem sinal $+$ ou $-$, dependendo da ordem em que aparecem no ciclo.
- Um ciclo é dito *não orientado* se suas arestas pretas formam uma sequência decrescente, e *orientado* caso contrário.
- Um ciclo é dito *convergente* se possui apenas arestas pretas positivas, e *divergente* caso contrário.

Grafo de Ciclos

Dado $G(\pi)$ de uma permutação qualquer:

- A distância $dist(r, t)$ entre duas arestas pretas r e t no mesmo ciclo C é o número de arestas cinzas que devem ser percorridas para ir de r para t .
- $\Delta c(\pi, \rho_t), \Delta c(\pi, \rho_r) \in \{2, 0, -2\}$.
- $\Delta_{C_{\text{impar}}}(\pi, \rho_t), \Delta_{C_{\text{impar}}}(\pi, \rho_r) \in \{2, 0, -2\}$.

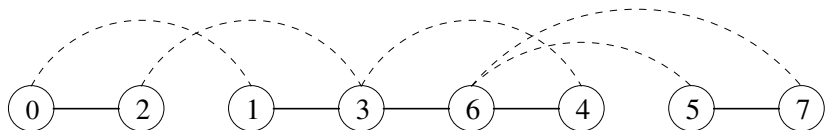


Figura: Grafo de breakpoints da permutação sem sinais $\pi = (2\ 1\ 3\ 6\ 4\ 5)$.

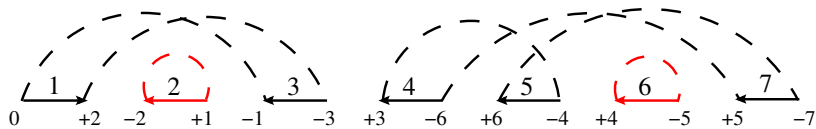


Figura: Um grafo de ciclos $G(\pi)$ obtido através de uma decomposição de $G_b(\pi)$.

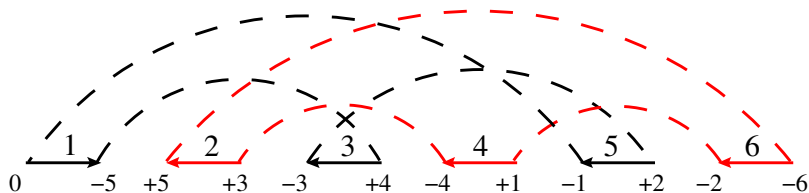


Figura: Grafo de ciclos da permutação com sinais $\pi = (+5 \ -3 \ -4 \ -1 \ -2)$, com os ciclos ímpares $C_1 = (6, 4, 2)$, representado pela cor vermelha, e $C_2 = (5, -1, 3)$, representado pela cor preta.

Remoção de Breakpoints

- A primeira etapa da heurística possui três passos:
 - 1 Procura uma transposição ρ_t que aplicada em π remove 3 breakpoints.
 - 2 Procura uma reversão ρ_r que aplicada em π remove 2 breakpoints.
 - 3 Procura uma transposição ρ_t que aplicada em π remove 2 breakpoints.
- Pode não encontrar uma operação que aplicada em π remova pelo menos 2 breakpoints.
- Uma operação encontrada em um dos passos acima garante o fator de aproximação de 1.5.

Análise dos ciclos existentes

- A segunda etapa da heurística possui cinco passos:
 - 1 Ciclo orientado convergente com duas distâncias ímpares.
 - 2 Ciclos pares divergentes.
 - 3 Ciclos pares convergentes que se cruzam.
 - 4 Ciclos pares convergentes.
 - 5 Ciclo orientado convergente sem duas distâncias ímpares.
- Uma operação encontrada em um dos passos acima garante o fator de aproximação de 1.5.

Criação dos bancos de configurações

- Considera apenas permutações simples.
- As configurações podem conter no máximo um ciclo par e deve ser convergente.
- Configurações geradas a partir de três ciclos: $(2, 1)$, $(3, -2, 1)$ e $(3, -2, -1)$.
- As configurações foram estendidas com um dos ciclos a seguir: $(3, 2, 1)$, $(3, -2, 1)$, $(3, -1, 2)$, $(3, 2, -1)$, $(3, 1, -2)$, $(3, -2, -1)$ e $(3, -1, -2)$.
- As permutações são estendidas com o objetivo de encontrar uma sequência válida com fator de no máximo 1.8.

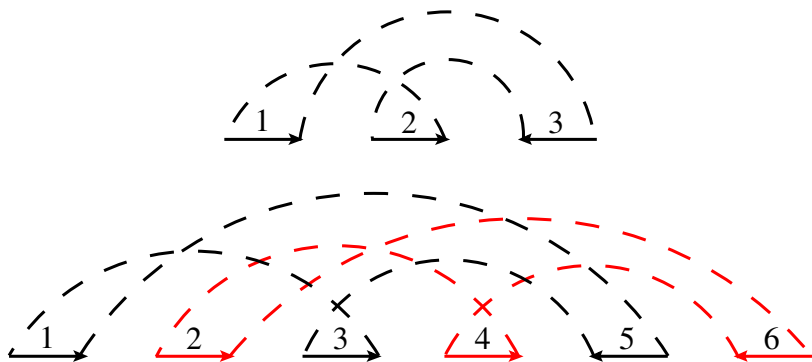


Figura: A configuração $(3, -2, -1)$, que não possui uma sequência válida com fator menor ou igual a 1.8. Uma extensão da configuração $(3, -2, -1)$ com um ciclo $(3, -2, -1)$, gerando a configuração $(6, -4, -2)$, $(5, -3, -1)$, cuja sequência válida possui as operações $\rho_r(1, 4)$, $\rho_t(2, 4, 6)$, $\rho_r(1, 4)$ e fator 1.5.

Criação dos bancos de configurações

- As configurações geradas a partir do ciclo par $(2, 1)$ foram agrupadas em um banco chamado *banco par*.
- Existe uma sequência válida para toda configuração com 4 ciclos.
- Aproximação máxima da sequência válida igual a 1.667.

Criação dos bancos de configurações

- As configurações geradas pelos ciclos ímpares foram agrupadas em um banco chamado *banco ímpar*.
- No *banco ímpar* existem configurações com até 4 ciclos com uma sequência válida com fator 1.8.
- Nem toda configuração com 4 ciclos do *banco ímpar* possui uma sequência válida.

Criação dos bancos de configurações

- Criado o *banco 2.0*, com configurações geradas a partir do ciclo $(3, -2, 1)$.
- As configurações deste banco foram estendidas com um dos ciclos a seguir: $(3, 2, 1)$, $(3, -2, 1)$, $(3, 2, -1)$ e $(3, -1, -2)$.
- No *banco 2.0* existem configurações com até 3 ciclos e sequência válida com fator 2.0.
- Toda configuração com 3 ciclos do *banco 2.0* possui uma sequência válida.

Criação dos bancos de configurações

- Configurações contendo apenas combinações do ciclo não orientado convergente (3, 2, 1).
- Enumeração realizada por Elias e Hartman em 2006.
- Garante o fator de aproximação 1.375 quando a configuração possui mais de 8 ciclos.
- Para configurações com menor quantidade de ciclos, o fator de aproximação é de 1.5.
- Estão agrupadas em um banco chamado *banco EH*.

Banco	Número de ciclos	Número de configurações	Total
par	2	2	1806
	3	161	
	4	1643	
ímpar	2	6	443748
	3	1468	
	4	442274	
2.0	1	1	257
	2	6	
	3	250	
EH	3	1	4002165
	4	22	
	5	5696	
	6	53898	
	7	377877	
	8	1450662	
	9	1077521	
	10	1034940	
	11	1140	
	12	264	
	13	144	

Tabela: Número de configurações com uma sequência válida para cada um dos bancos de configurações criados.

Busca em bancos de configurações

- A terceira etapa da heurística busca por uma sequência válida em quatro bancos criados, na seguinte ordem:
 - 1 Configurações do *banco EH*.
 - 2 Configurações do *banco par*.
 - 3 Configurações do *banco ímpar*.
 - 4 Configurações do *banco 2.0*.
- Sempre encontra uma sequência de operações e garante o fator de aproximação de 2.0.

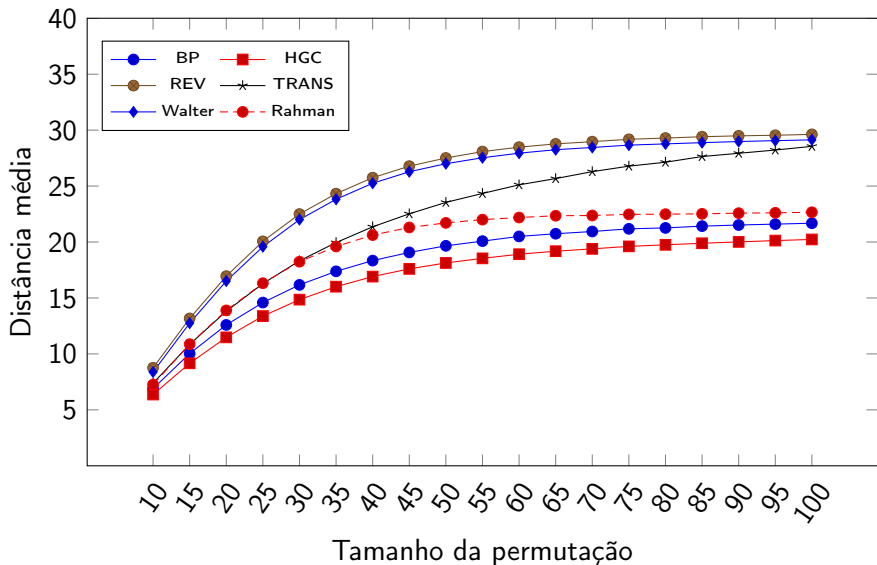


Figura: Distância média para permutações com sinais.

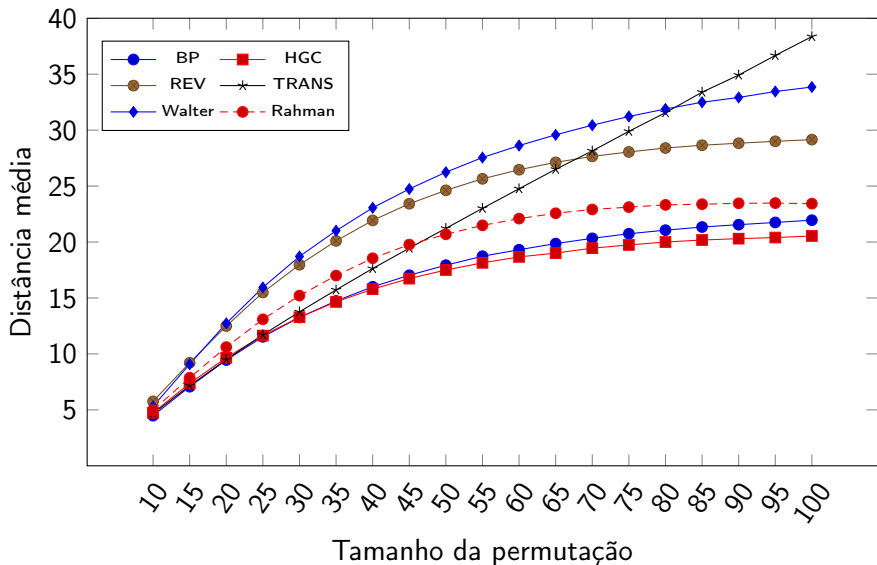


Figura: Distância média para permutações sem sinais.

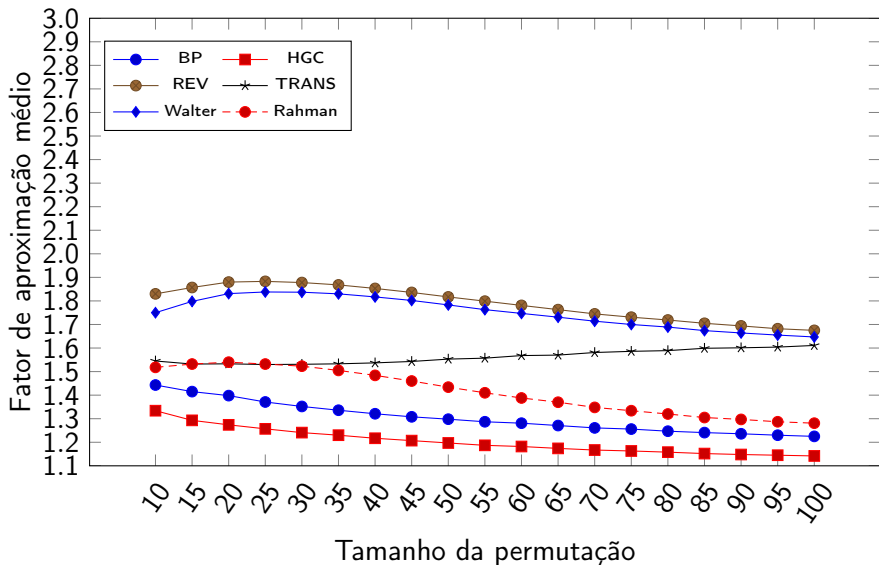


Figura: Fator de aproximação médio para permutações com sinais.

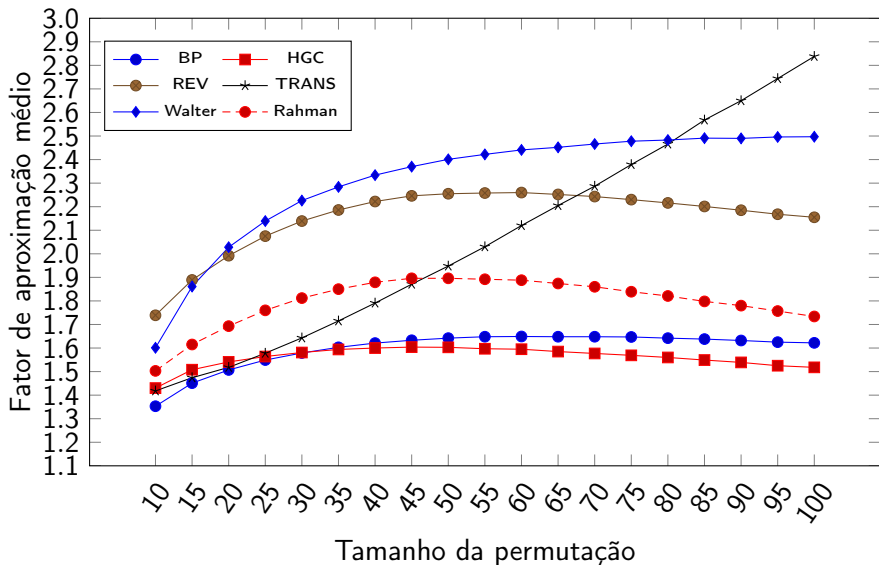


Figura: Fator de aproximação médio para permutações sem sinais.

Conclusões

- Para permutações com sinais, a Heurística de Grafo de Ciclos retornou em média as menores distâncias e os menores fatores de aproximação médios para todos os tamanhos de permutações testadas.
 - O maior fator de aproximação obtido foi de 2.0, sendo este o fator teórico garantido.
- Para permutações sem sinais, a Heurística de Grafo de Ciclos retornou em média as menores distâncias e os melhores fatores de aproximação médios para as permutações testadas com tamanho maior que 30.
 - O maior fator de aproximação obtido foi 2.142, abaixo do fator teórico garantido de 2.8386 (2k).

Publicações

U. Dias, A. R. Oliveira, and Z. Dias. An improved algorithm for the sorting by reversals and transpositions problem. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB'2014)*, pages 400–409, Newport Beach, California, USA, 2014.

- Versão preliminar da Heurística de Grafo de Ciclos
- Não considerava a etapa de remoção de breakpoints.
- Os bancos de configurações foram gerados através de outras regras.

Trabalhos Futuros

- Investigação sobre as configurações que não conseguiram uma sequência válida com aproximação 1.8.
- Análise do limitante inferior do problema.
- Investigação da complexidade do Problema de Ordenação de Permutações por Reversões e Transposições, para permutações com sinais e permutações sem sinais.

O Problema da Ordenação de Permutações por Reversões e Transposições

Andre Rodrigues Oliveira

Instituto de Computação
Universidade Estadual de Campinas
Orientador: Zaroni Dias
Dissertação de Mestrado