Algoritmos e Combinatória em Rearranjo de Genomas

Luís Felipe Ignácio Cunha - IC/UFF

lfignacio@ic.uff.br

Seminário de Combinatória do IME-UFF
26 de Junho de 2017

Conteúdo

Princípio da evolução

''Durante a evolução, moléculas de hereditariedade são costuradas, modificadas, cortadas, alongadas, encurtadas e revertidas``

Princípio da evolução

''Durante a evolução, moléculas de hereditariedade são costuradas, modificadas, cortadas, alongadas, encurtadas e revertidas``

(François Jacob, 1984)

Princípio da evolução

''Durante a evolução, moléculas de hereditariedade são costuradas, modificadas, cortadas, alongadas, encurtadas e revertidas``

(François Jacob, 1984)

- Moléculas de DNA são responsáveis por toda informação genética dos seres.
- De uma célula para outra, uma proteína para outra, conteúdos de DNA são quase similares, porém suas organizações se diferem drasticamente.
- **Rearranjo de Genomas** são mutações que afetam a organização do DNA.

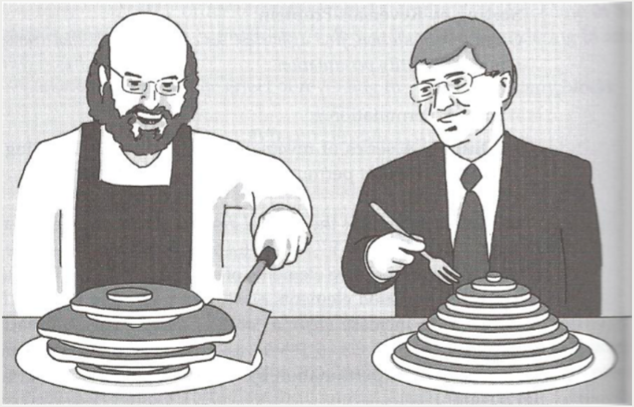
Rearranjo de genomas

Rearranjo de genomas

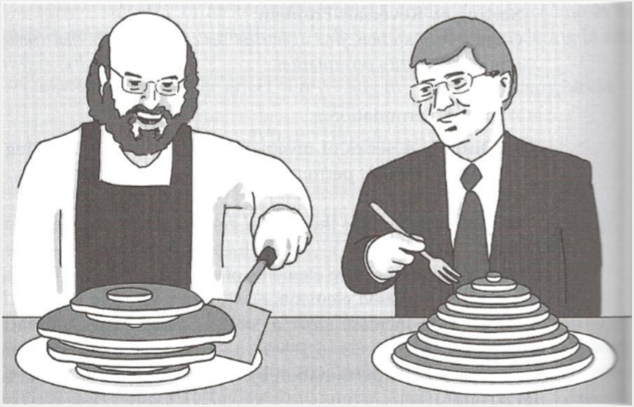
Rearranjo de genomas

Hannenhalli e Pevzner
Journal of the ACM, 1999

Flipagem de panquecas



Flipagem de panquecas



Papadimitriou e Gates
Discrete Mathematics, 1979

Modelo matemático

- Genomas vs. Permutações:

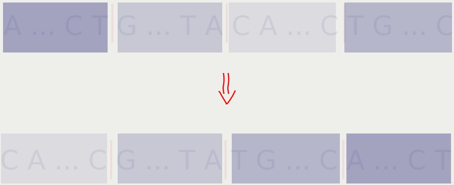
A ... C T | G ... T A | C A ... C | T G ... C

$\Downarrow$

C A ... C | G ... T A | T G ... C | A ... C T

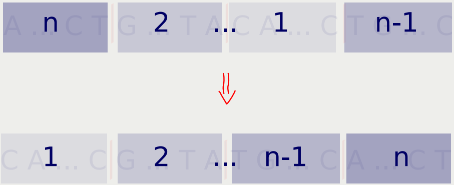
Modelo matemático

- Genomas vs. Permutações:



Modelo matemático

- Genomas vs. Permutações:



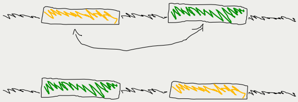
Conteúdo

Operações em permutações

- **Movimento de blocos**:
- **Movimento de blocos curtos**:
- **Transposição**:
- k**-multi corte restrito**:

Operações em permutações

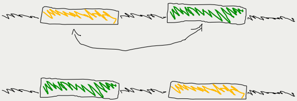
- **Movimento de blocos**:



- **Transposição**:
- **Movimento de blocos curtos**:
- *k*-**multi corte restrito**:

Operações em permutações

- **Movimento de blocos**:



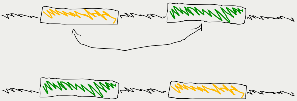
- **Transposição**:

- **Movimento de blocos curtos**:

- ***k*-multi corte restrito**:

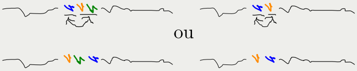
Operações em permutações

- **Movimento de blocos**:



- **Transposição**:

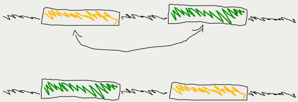
- **Movimento de blocos curtos**:



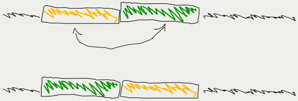
- ***k*-multi corte restrito**:

Operações em permutações

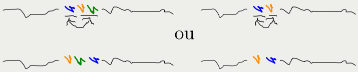
- **Movimento de blocos:**



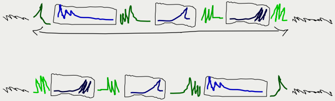
- **Transposição:**



- **Movimento de blocos curtos:**

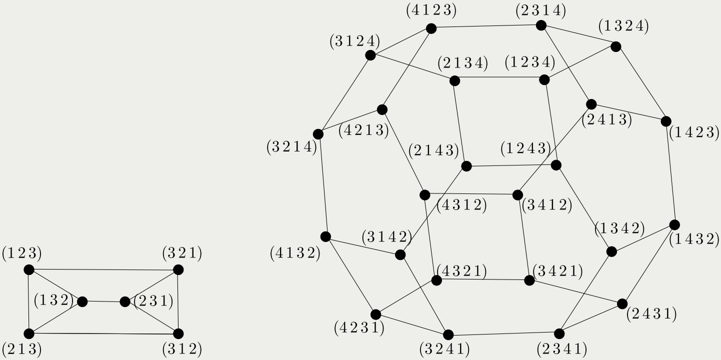


- ***k*-multi corte restrito:**



Grafos associados

- **Ex.:** Trocas de elementos consecutivos:



Distância

Premissa:
Distância evolucionária é o menor número de eventos de rearranjo necessários para transformar um genoma em outro.

Distância $d_M(\pi)$:

$d_M(\pi) =$ menor número de operações pela métrica **M** para ordenar $\boldsymbol{\pi}$.

Distância

Premissa:
Distância evolucionária é o menor número de eventos de rearranjo necessários para transformar um genoma em outro.

Distância $d_M(\pi)$:

$d_M(\pi)$ = menor número de operações pela métrica **M** para ordenar $\boldsymbol{\pi}$.

9 3 6 7 4 2 1 5 8

$\Downarrow$

?

$\Downarrow$

1 2 3 4 5 6 7 8 9

Distância

Premissa:
Distância evolucionária é o menor número de eventos de rearranjo necessários para transformar um genoma em outro.

Distância $d_M(\pi)$:

$d_M(\pi) =$ menor número de operações pela métrica **M** para ordenar $\boldsymbol{\pi}$.

	Movimento de blocos	Transposição	Movimento de blocos curtos	*k*-Multi corte restrito
Distância	Polinomial (Christie, 99)	*NP*-completo (Bulteau et al., 10)	Em aberto	Em aberto

Diâmetro

Diâmetro $D_M(n)$:

$D_M(n)$ = máxima distância pela métrica **M** para permutações de tamanho ***n***.

Diâmetro

Diâmetro $D_M(n)$:

$D_M(n)$ = máxima distância pela métrica **M** para permutações de tamanho ***n***.

	Movimento de blocos	Transposição	Movimento de blocos curtos	k-Multi corte restrito
Distância	Polinomial (Christie, 99)	*NP*-completo (Bulteau et al., 10)	Em aberto	Em aberto
Diâmetro	Polinomial (Christie, 99)	Em aberto	Polinomial (Heath e Vergara, 01)	Em aberto

Diâmetro

Diâmetro $D_M(n)$:

$D_M(n)$ = máxima distância pela métrica **M** para permutações de tamanho ***n***.

	Movimento de blocos	Transposição	Movimento de blocos curtos	*k*-Multi corte restrito
Distância	Polinomial (Christie, 99)	*NP*-completo (Bulteau et al., 10)	Em aberto	Em aberto
Diâmetro	Polinomial (Christie, 99)	Em aberto	Polinomial (Heath e Vergara, 01)	Em aberto

- Diâmetro de Transposição: $D_t(n) \geq \lfloor \frac{n+1}{2} \rfloor + 1$
  - Proposta de limite $D_t(n) \geq \frac{17n+1}{33}$ (Lu e Yang, *SIAM J. Disc. Math.* 2010).

Diâmetro

Diâmetro $D_M(n)$:

$D_M(n)$ = máxima distância pela métrica **M** para permutações de tamanho ***n***.

	Movimento de blocos	Transposição	Movimento de blocos curtos	k-Multi corte restrito
Distância	Polinomial (Christie, 99)	*NP*-completo (Bulteau et al., 10)	Em aberto	Em aberto
Diâmetro	Polinomial (Christie, 99)	Em aberto	Polinomial (Heath e Vergara, 01)	Em aberto

- Diâmetro de Transposição: $D_t(n) \geq \lfloor \frac{n+1}{2} \rfloor + 1$
  - Proposta de limite $D_t(n) \geq \frac{17n+1}{33}$ (Lu e Yang, *SIAM J. Disc. Math.* 2010).
  1. Invalidamos esta proposta;
  2. Mostramos novas famílias infinitas com distâncias iguais ao limite corrente;
  3. Mostramos estratégia para obter $D_t(n) \geq \lfloor \frac{n+1}{2} \rfloor + 2$.

Diâmetro

Diâmetro $D_M(n)$:

$D_M(n)$ = máxima distância pela métrica **M** para permutações de tamanho ***n***.

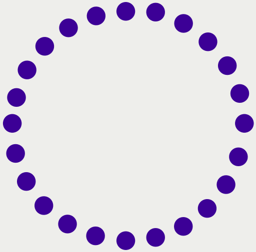
	Movimento de blocos	Transposição	Movimento de blocos curtos	*k*-Multi corte restrito
Distância	Polinomial (Christie, 99)	*NP*-completo (Bulteau et al., 10)	Em aberto	Em aberto
Diâmetro	Polinomial (Christie, 99)	Em aberto	Polinomial (Heath e Vergara, 01)	Em aberto

- Diâmetro de Transposição: $D_t(n) \geq \lfloor \frac{n+1}{2} \rfloor + 1$
  - Proposta de limite $D_t(n) \geq \frac{17n+1}{33}$ (Lu e Yang, *SIAM J. Disc. Math.* 2010).
  1. Invalidamos esta proposta;
  2. Mostramos novas famílias infinitas com distâncias iguais ao limite corrente;
  3. Mostramos estratégia para obter $D_t(n) \geq \lfloor \frac{n+1}{2} \rfloor + 2$.

  Cunha, Hausen, Kowada, de Figueiredo, *SIAM J. Disc. Math.* 2013.

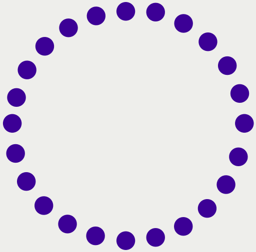
Problemas de centralidade

Permutação mais próxima



Problemas de centralidade

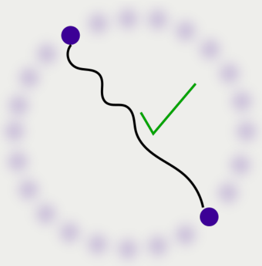
Permutação mais próxima



1. Entrada: Conjunto de objetos e um inteiro f.
2. É conhecida a ``distância'' entre qualquer par de objetos.
3. Existe um objeto com distância para cada objeto da entrada $\leq f$?

Problemas de centralidade

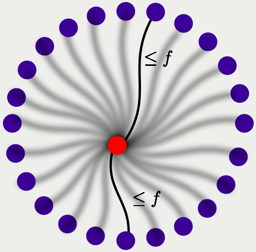
Permutação mais próxima



1. Entrada: Conjunto de objetos e um inteiro f.
2. É conhecida a ``distância'' entre qualquer par de objetos.
3. Existe um objeto com distância para cada objeto da entrada $\leq f$?

Problemas de centralidade

Permutação mais próxima



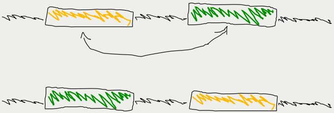
1. Entrada: Conjunto de objetos e um inteiro f.
2. É conhecida a ``distância'' entre qualquer par de objetos.
3. Existe um objeto com distância para cada objeto da entrada $\leq f$?

Conteúdo

Conteúdo

Distância de movimentos de blocos

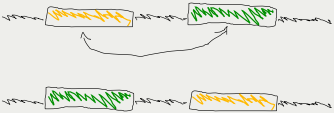
- **Polinomia**l: Christie (1996)
- Como obter $d_{mb}(\pi)$?



Distância de movimentos de blocos

- **Polinomial**: Christie (1996)
- Como obter $d_{mb}(\pi)$?

  Diagrama de realidade e desejo



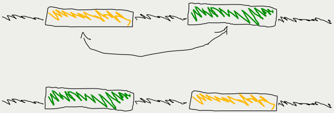
$0 \quad -10 \quad +10 \quad -9 \quad +9 \quad -8 \quad +8 \quad -7 \quad +7 \quad -1 \quad +1 \quad -6 \quad +6 \quad -11 \quad +11 \quad -5 \quad +5 \quad -4 \quad +4 \quad -3 \quad +3 \quad -2 \quad +2 \quad -12$

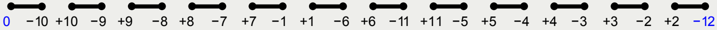
$G([10\ 9\ 8\ 7\ 1\ 6\ 11\ 5\ 4\ 3\ 2])$

Distância de movimentos de blocos

- **Polinomial**: Christie (1996)
- Como obter $d_{mb}(\pi)$?

  Diagrama de realidade e desejo



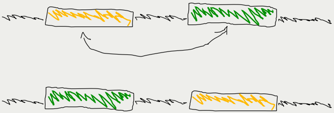


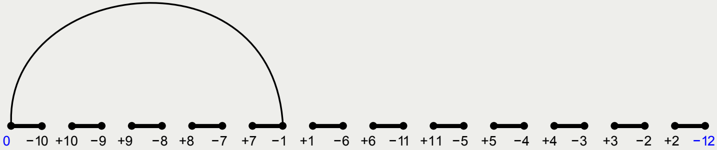
$G([10\ 9\ 8\ 7\ 1\ 6\ 11\ 5\ 4\ 3\ 2])$

Distância de movimentos de blocos

- **Polinomial**: Christie (1996)
- Como obter $d_{mb}(\pi)$?

  Diagrama de realidade e desejo



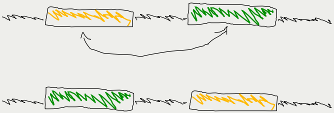


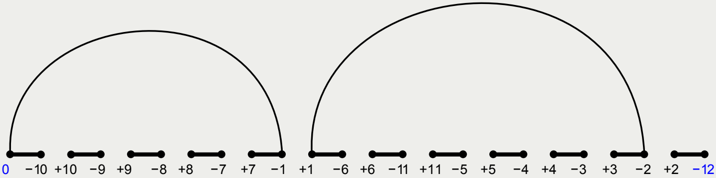
$G([10\ 9\ 8\ 7\ 1\ 6\ 11\ 5\ 4\ 3\ 2])$

Distância de movimentos de blocos

- **Polinomial**: Christie (1996)
- Como obter $d_{mb}(\pi)$?

  Diagrama de realidade e desejo



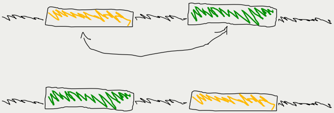


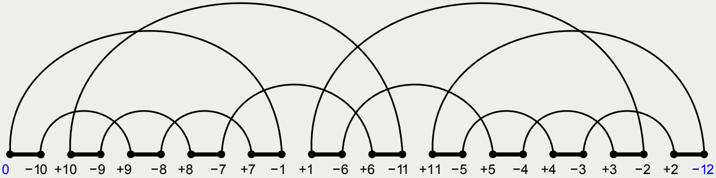
$G([10\ 9\ 8\ 7\ 1\ 6\ 11\ 5\ 4\ 3\ 2])$

Distância de movimentos de blocos

- **Polinomial**: Christie (1996)
- Como obter $d_{mb}(\pi)$?

  Diagrama de realidade e desejo



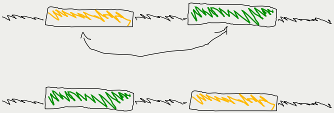


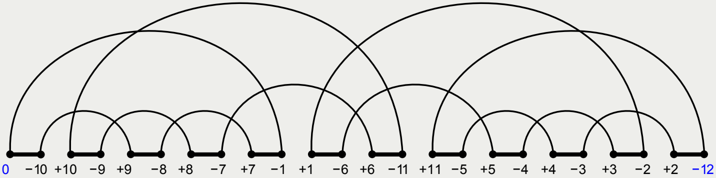
$G([10\ 9\ 8\ 7\ 1\ 6\ 11\ 5\ 4\ 3\ 2])$

Distância de movimentos de blocos

- **Polinomial**: Christie (1996)
- Como obter $d_{mb}(\pi)$?

  Diagrama de realidade e desejo





$G([10\ 9\ 8\ 7\ 1\ 6\ 11\ 5\ 4\ 3\ 2])$

$$d_{mb}(\pi) = \frac{n+1-c(\pi)}{2}.$$

Conteúdo

Ordenação por Transposições

- *NP*-**completo** Bulteau *et al.* (2010)
  - Esteve em aberto por mais de 15 anos.
- Como obter limites para $d_t(\pi)$?

Ordenação por Transposições

- *NP*-**completo** Bulteau *et al.* (2010)
  - Esteve em aberto por mais de 15 anos.
- Como obter limites para $d_t(\pi)$?

  Diagrama de realidade e desejo

Limite inferior

$$d_t(\pi) \geq \frac{n+1-c_{impar}(\pi)}{2}.$$

Ordenação por Transposições

- *NP*-**completo** Bulteau *et al.* (2010)
  - Esteve em aberto por mais de 15 anos.
- Como obter limites para $d_t(\pi)$?

  Diagrama de realidade e desejo

Limite inferior

$$d_t(\pi) \geq \frac{n+1-c_{impar}(\pi)}{2}.$$

Limite superior

$$d_t(\pi) \leq \frac{2n}{3}.$$

Algoritmos aproximativos

- Melhores algoritmos, até agora:

Razão de aproximação

- Elias e Hartman, 2006: 1,375-aproximativo
  - complexidade: $O(n^2)$

Complexidade

- Feng e Zhu, 2007: Árvore de permutação
- Hartman e Shamir, 2005: 1,5-aproximativo
  - complexidade: $O(n \log n)$

- Por que não usar **árvore de permutação** no **1,375-aproximativo**?
  - Proposta de Firoz *et al.* (*Journal of Comput. Biol.*, 2011).

Algoritmos aproximativos

- Melhores algoritmos, até agora:

Razão de aproximação

- Elias e Hartman, 2006: 1,375-aproximativo
  - complexidade: $O(n^2)$

Complexidade

- Feng e Zhu, 2007: Árvore de permutação
- Hartman e Shamir, 2005: 1,5-aproximativo
  - complexidade: $O(n \log n)$

- Por que não usar **árvore de permutação** no **1,375-aproximativo**?
  - Proposta de Firoz *et al.* (*Journal of Comput. Biol.*, 2011).

Algoritmos aproximativos

- Melhores algoritmos, até agora:

Razão de aproximação

- Elias e Hartman, 2006: 1,375-aproximativo
  - complexidade: $O(n^2)$

Complexidade

- Feng e Zhu, 2007: Árvore de permutação
- Hartman e Shamir, 2005: 1,5-aproximativo
  - complexidade: $O(n \log n)$

- Por que não usar **árvore de permutação** no **1,375-aproximativo**?
  - Proposta de Firoz *et al.* (*Journal of Comput. Biol.*, 2011).

Algoritmos aproximativos

- Melhores algoritmos, até agora:

Razão de aproximação

- Elias e Hartman, 2006: 1,375-aproximativo
  - complexidade: $O(n^2)$

Complexidade

- Feng e Zhu, 2007: Árvore de permutação
- Hartman e Shamir, 2005: 1,5-aproximativo
  - complexidade: $O(n \log n)$

- Por que não usar **árvore de permutação** no **1,375-aproximativo**?
  - **Mostramos contra exemplos para a corretude.**

Novo algoritmo 1,375-aproximativo

- Propomos um novo algoritmo 1,375-aproximativo para obter complexidade $O(n \log n)$.
  - Adaptamos o algoritmo de Elias e Hartman e utilizamos a árvore de permutação.

Novo algoritmo 1,375-aproximativo

- Propomos um novo algoritmo 1,375-aproximativo para obter complexidade $O(n \log n)$.
  - Adaptamos o algoritmo de Elias e Hartman e utilizamos a árvore de permutação.

- Sempre encontramos uma 11/8-sequência.
  - 11 operações tais que pelo menos 8 são ´´boas``.
  - Sequências encontradas por um algoritmo de força bruta. `http://compscinet.org/research/sbt1375`

Novo algoritmo 1,375-aproximativo

- Propomos um novo algoritmo 1,375-aproximativo para obter complexidade $O(n \log n)$.
  - Adaptamos o algoritmo de Elias e Hartman e utilizamos a árvore de permutação.

- Sempre encontramos uma 11/8-sequência.
  - 11 operações tais que pelo menos 8 são ´´boas``.
  - Sequências encontradas por um algoritmo de força bruta. `http://compscinet.org/research/sbt1375`

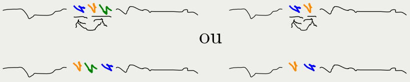
Cunha, Hausen, Kowada, de Figueiredo, *Journal of Comput. Biol.*, 2015.

Conteúdo

Distância de movimentos de blocos curtos

- **Em aberto**

  Quando restrito a soma dos blocos $= 2$, resolvido pelo *Bubble sort*.

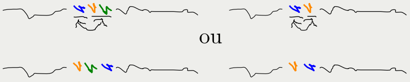


- Como obter limites para $d_{mbc}(\pi)$?

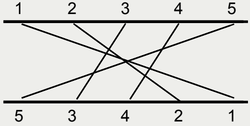
Distância de movimentos de blocos curtos

- **Em aberto**

  Quando restrito a soma dos blocos $= 2$,
  resolvido pelo *Bubble sort*.



- Como obter limites para $d_{mbc}(\pi)$? Grafo de permutação.

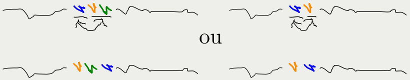


$PG([5\ 3\ 4\ 2\ 1])$.

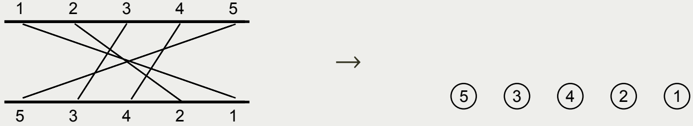
Distância de movimentos de blocos curtos

- **Em aberto**

  Quando restrito a soma dos blocos $= 2$, resolvido pelo *Bubble sort*.



- Como obter limites para $d_{mbc}(\pi)$? Grafo de permutação.

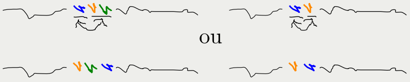


$PG([5\ 3\ 4\ 2\ 1])$.

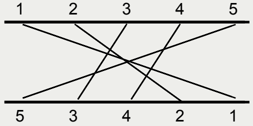
Distância de movimentos de blocos curtos

- **Em aberto**

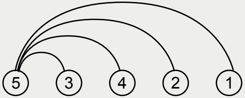
  Quando restrito a soma dos blocos $= 2$, resolvido pelo *Bubble sort*.



- Como obter limites para $d_{mbc}(\pi)$? Grafo de permutação.



$\rightarrow$

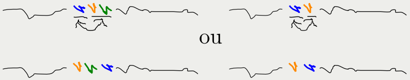


$PG([5\ 3\ 4\ 2\ 1])$.

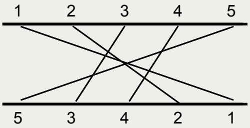
Distância de movimentos de blocos curtos

- **Em aberto**

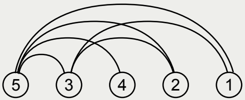
  Quando restrito a soma dos blocos $= 2$, resolvido pelo *Bubble sort*.



- Como obter limites para $d_{mbc}(\pi)$? Grafo de permutação.



$\rightarrow$

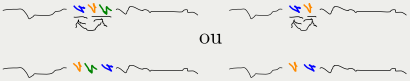


$PG([5\ 3\ 4\ 2\ 1])$.

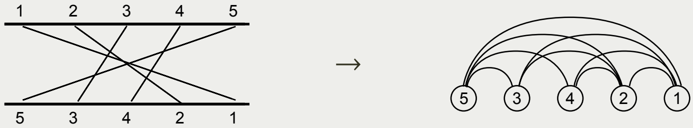
Distância de movimentos de blocos curtos

- **Em aberto**

  Quando restrito a soma dos blocos $= 2$,
  resolvido pelo *Bubble sort*.



- Como obter limites para $d_{mbc}(\pi)$? Grafo de permutação.

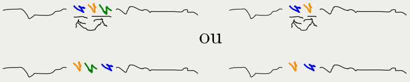


$PG([5\ 3\ 4\ 2\ 1])$.

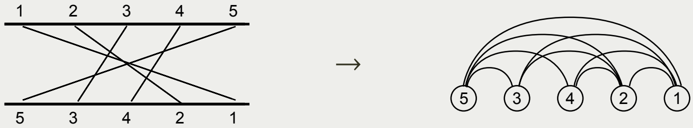
Distância de movimentos de blocos curtos

- **Em aberto**

  Quando restrito a soma dos blocos $= 2$, resolvido pelo *Bubble sort*.



- Como obter limites para $d_{mbc}(\pi)$? Grafo de permutação.

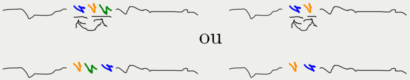


$$5 = \lceil \frac{\#Arestas\ de\ PG(\pi)}{2} \rceil \leq d_{mbc}(\pi) \leq \#Arestas\ de\ PG(\pi) = 9.$$

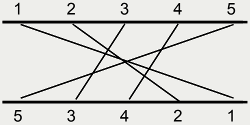
Distância de movimentos de blocos curtos

- **Em aberto**

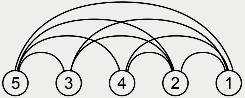
  Quando restrito a soma dos blocos $= 2$,
  resolvido pelo *Bubble sort*.



- Como obter limites para $d_{mbc}(\pi)$? Grafo de permutação.



$\longrightarrow$

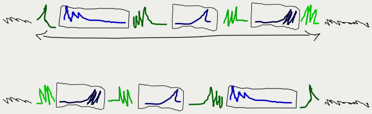


$$5 = \lceil \frac{\#Arestas\ de\ PG(\pi)}{2} \rceil \leq d_{mbc}(\pi) \leq \#Arestas\ de\ PG(\pi) = 9.$$

- Ordenação ótima obtida por ordenar cada componente conexa.
  - Não vale para movimento de blocos limitados por valores > 3.

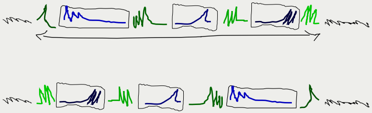
Distância de *k*-multi corte restrito

- $k = 0$: *NP*-**completo** (Caprara, 1997)
- $k \geq 1$: **Em aberto**.



- test

Distância de *k*-multi corte restrito



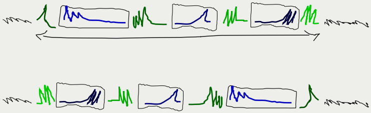
- $k = 0$: *NP*-**completo** (Caprara, 1997)
- $k \geq 1$: **Em aberto**.

Limites para Distância:

- test

Limites para Diâmetro:

Distância de *k*-multi corte restrito



- $k = 0$: *NP*-**completo** (Caprara, 1997)
- $k \geq 1$: **Em aberto**.

Limites para Distância:

Pontos de quebra $b(\pi)$,

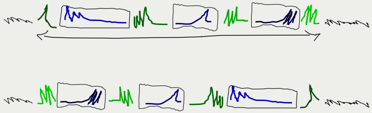
Diagrama de realidade e desejo,

Ciclos algébricos $pc(\pi)$,

Limites para Diâmetro:

- test

Distância de *k*-multi corte restrito



- $k = 0$: *NP*-**completo** (Caprara, 1997)
- $k \geq 1$: **Em aberto**.

Limites para Distância:

- test

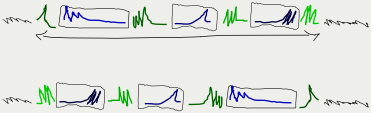
Pontos de quebra $b(\pi)$, $d_{Kmcr}(\pi) \geq \frac{b(\pi)}{2(k+1)}$.

Diagrama de realidade e desejo,

Ciclos algébricos $pc(\pi)$,

Limites para Diâmetro:

Distância de k-multi corte restrito



- $k = 0$: *NP*-**completo** (Caprara, 1997)
- $k \geq 1$: **Em aberto**.

Limites para Distância:

- test

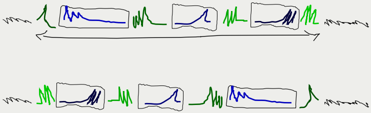
Pontos de quebra $b(\pi)$, $d_{Kmcr}(\pi) \geq \frac{b(\pi)}{2(k+1)}$.

Diagrama de realidade e desejo, $d_{Kmcr}(\pi) \geq \frac{b(\pi)-c(\pi)}{k+1}$.

Ciclos algébricos $pc(\pi)$,

Limites para Diâmetro:

Distância de k-multi corte restrito



- $k = 0$: *NP*-**completo** (Caprara, 1997)
- $k \geq 1$: **Em aberto**.

Limites para Distância:

- test

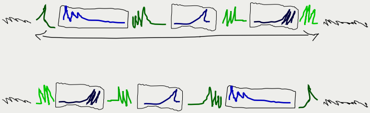
Pontos de quebra $b(\pi)$, $d_{Kmcr}(\pi) \geq \frac{b(\pi)}{2(k+1)}$.

Diagrama de realidade e desejo, $d_{Kmcr}(\pi) \geq \frac{b(\pi)-c(\pi)}{k+1}$.

Ciclos algébricos $pc(\pi)$, $d_{Krmb} \leq n - pc(\pi)$.

Limites para Diâmetro:

Distância de *k*-multi corte restrito



- $k = 0$: *NP*-**completo** (Caprara, 1997)
- $k \geq 1$: **Em aberto**.

Limites para Distância:

- test

Pontos de quebra $b(\pi)$, $d_{Kmcr}(\pi) \geq \frac{b(\pi)}{2(k+1)}$.

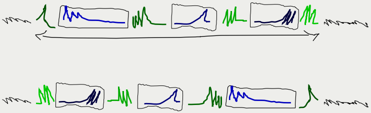
Diagrama de realidade e desejo, $d_{Kmcr}(\pi) \geq \frac{b(\pi)-c(\pi)}{k+1}$.

Ciclos algébricos $pc(\pi)$, $d_{Krmb} \leq n - pc(\pi)$.

Limites para Diâmetro:

$\frac{n}{2} \leq D_1(n) \leq \frac{2n}{3}$.

Distância de *k*-multi corte restrito



- $k = 0$: *NP*-**completo** (Caprara, 1997)
- $k \geq 1$: **Em aberto**.

Limites para Distância:

- test

Pontos de quebra $b(\pi)$, $d_{Kmcr}(\pi) \geq \frac{b(\pi)}{2(k+1)}$.

Diagrama de realidade e desejo, $d_{Kmcr}(\pi) \geq \frac{b(\pi)-c(\pi)}{k+1}$.

Ciclos algébricos $pc(\pi)$, $d_{Krmb} \leq n - pc(\pi)$.

Limites para Diâmetro:

$\frac{n}{2} \leq D_1(n) \leq \frac{2n}{3}$.

$\log n \leq D_n(n) \leq \log^2 n$.

Conteúdo

Cadeia mais próxima pela distância de Hamming

CADEIA MAIS PRÓXIMA PELA DISTÂNCIA DE HAMMING ($H-CMP$)
ENTRADA: Conjunto de cadeias $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_\ell\}$ de tamanho $\boldsymbol{m}$ cada e $\mathbf{f} \geq \mathbf{0}$.
PERGUNTA: Existe uma cadeia $\boldsymbol{\sigma}$ de tamanho $\boldsymbol{m}$ tal que $\max_{i=1,\ldots,\ell} \boldsymbol{d_H(s_i, \sigma) \leq f}$?

- Distância de Hamming: Número de posições com elementos distintos entre duas cadeias.

$$d_H(010111, \underline{1}10\underline{0}11) = 2$$

- H-CMP é *NP*-completo para alfabeto binário (Lanctot *et al.*, 03).

Cadeia mais próxima pela distância de Hamming

CADEIA MAIS PRÓXIMA PELA DISTÂNCIA DE HAMMING ($H-CMP$)
ENTRADA: Conjunto de cadeias $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_\ell\}$ de tamanho $\boldsymbol{m}$ cada e $\mathbf{f} \geq \mathbf{0}$.
PERGUNTA: Existe uma cadeia $\boldsymbol{\sigma}$ de tamanho $\boldsymbol{m}$ tal que $\max_{i=1,\ldots,\ell} \boldsymbol{d_H(s_i, \sigma)} \leq \boldsymbol{f}$?

- Distância de Hamming: Número de posições com elementos distintos entre duas cadeias.

$$d_H(010111, \underline{1}10\underline{0}11) = 2$$

- H-CMP é *NP*-completo para alfabeto binário (Lanctot *et al.*, 03).

Cadeia mais próxima pela distância de Hamming

CADEIA MAIS PRÓXIMA PELA DISTÂNCIA DE HAMMING ($H-CMP$)
ENTRADA: Conjunto de cadeias $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_\ell\}$ de tamanho $\boldsymbol{m}$ cada e $\mathbf{f} \geq \mathbf{0}$.
PERGUNTA: Existe uma cadeia $\boldsymbol{\sigma}$ de tamanho $\boldsymbol{m}$ tal que $\max_{i=1,\ldots,\ell} \boldsymbol{d_H(s_i, \sigma) \leq f}$?

- Distância de Hamming: Número de posições com elementos distintos entre duas cadeias.

$$d_H(010111, \underline{1}10\underline{0}11) = 2$$

- H-CMP é *NP*-completo para alfabeto binário (Lanctot *et al.*, 03).

Cadeia mais próxima pela distância de Hamming

CADEIA MAIS PRÓXIMA PELA DISTÂNCIA DE HAMMING ($H - CMP$)
ENTRADA: Conjunto de cadeias $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_\ell\}$ de tamanho $\boldsymbol{m}$ cada e $\mathbf{f} \geq \mathbf{0}$.
PERGUNTA: Existe uma cadeia $\boldsymbol{\sigma}$ de tamanho $\boldsymbol{m}$ tal que $\max_{i=1,\ldots,\ell} \boldsymbol{d_H(s_i, \sigma) \leq f}$?

- Distância de Hamming: Número de posições com elementos distintos entre duas cadeias.

$$d_H(010111, \underline{1}10\underline{0}11) = 2$$

- H-CMP é *NP*-completo para alfabeto binário (Lanctot *et al.*, 03).
  - E quando as cadeia da entrada e saída são permutações?

NP-completudes

- *M*-PMP é *NP*-completo para *M*:
  - Movimentos de blocos;
  - Movimentos de blocos curtos;
  - Pontos de quebra.
- Quando consideramos genomas ´´distintos`` de permutações:
  - *single-cut-or-join*-GMP é *NP*-completo.

NP-completudes

- *M*-PMP é *NP*-completo para *M*:
  - Movimentos de blocos;
  - Movimentos de blocos curtos;
  - Pontos de quebra.
- Quando consideramos genomas ''distintos`` de permutações:
  - *single-cut-or-join*-GMP é *NP*-completo.

- Hamming-CMP $\propto$ M-PMP
  1. Transforme uma cadeia binária s em uma permutação particular π_s.

NP-completudes

- *M*-PMP é *NP*-completo para *M*:
  - Movimentos de blocos;
  - Movimentos de blocos curtos;
  - Pontos de quebra.
- Quando consideramos genomas ´´distintos`` de permutações:
  - *single-cut-or-join*-GMP é *NP*-completo.

- Hamming-CMP $\propto$ M-PMP
  1. Transforme uma cadeia binária s em uma permutação particular π_s.
  2. Mostre que $d_M(\pi_s) = f(d_H(s))$.

NP-completudes

- *M*-PMP é *NP*-completo para *M*:
  - Movimentos de blocos;
  - Movimentos de blocos curtos;
  - Pontos de quebra.
- Quando consideramos genomas ''distintos`` de permutações:
  - *single-cut-or-join*-GMP é *NP*-completo.

- Hamming-CMP $\propto$ M-PMP
  1. Transforme uma cadeia binária s em uma permutação particular π_s.
  2. Mostre que $d_M(\pi_s) = f(d_H(s))$.
  3. Mostre que uma solução para Hamming-CMP implica numa solução para Movimento de blocos-PMP, e vice e versa.

Conteúdo

Conclusões e trabalhos futuros

- Apresentamos:
    - Problemas envolvendo permutações (Distância, diâmetro e centralidade).
    - Propriedades, algoritmos e *NP*-completudes.

Conclusões e trabalhos futuros

- Apresentamos:
  - Problemas envolvendo permutações (Distância, diâmetro e centralidade).
  - Propriedades, algoritmos e *NP*-completudes.
- Alguns trabalhos futuros:

Conclusões e trabalhos futuros

- Apresentamos:
  - Problemas envolvendo permutações (Distância, diâmetro e centralidade).
  - Propriedades, algoritmos e *NP*-completudes.
- Alguns trabalhos futuros:
  - Distância:

    Algoritmos para demais k-multi corte restritos.

Conclusões e trabalhos futuros

- Apresentamos:
  - Problemas envolvendo permutações (Distância, diâmetro e centralidade).
  - Propriedades, algoritmos e *NP*-completudes.

- Alguns trabalhos futuros:
  - Distância:

    Algoritmos para demais k-multi corte restritos.
  - Diâmetro:

    Provar diâmetro para 1-multi corte restrito $D_1(n) = \frac{n}{2}$.

n	2	3	4	5	6	7	8	9	10	11
$\mathbf{D_0(n)}$	**1**	2	**3**	4	**5**	6	**7**	8	**9**	10
$\mathbf{D_1(n)}$	**1**	**1**	2	2	**3**	**3**	4	4	**5**	**5**
$\mathbf{D_2(n)}$	**1**	**1**	2	2	**3**	**3**	**3**	**3**	4	4

Conclusões e trabalhos futuros

- Apresentamos:
  - Problemas envolvendo permutações (Distância, diâmetro e centralidade).
  - Propriedades, algoritmos e *NP*-completudes.
- Alguns trabalhos futuros:
  - Distância:

    Algoritmos para demais k-multi corte restritos.
  - Diâmetro:

    Provar diâmetro para 1-multi corte restrito $D_1(n) = \frac{n}{2}$.

n	2	3	4	5	6	7	8	9	10	11
$\mathbf{D_0(n)}$	**1**	2	**3**	4	**5**	6	**7**	8	**9**	10
$\mathbf{D_1(n)}$	**1**	**1**	2	2	**3**	**3**	4	4	**5**	**5**
$\mathbf{D_2(n)}$	**1**	**1**	2	2	**3**	**3**	**3**	**3**	4	4

  - Permutação mais próxima:

    Relação com problema de mediana.

    Centralidade é **min-max** e mediana é **min-sum**.

Muito obrigado!