

Cálculo de la distancia de inversión cromosómica usando estructuras matroidales

Lidia Angélica García García
Ma. Guadalupe Rodríguez Sánchez

18 de septiembre de 2019

Resumen

Diferentes especies comparten distintos porcentajes de información genética dependiendo de su proximidad evolutiva. Estos datos genéticos han sido heredados de un antecesor en común y no necesariamente tienen un orden similar en los genomas de cada especie debido a mutaciones a gran escala que afectan amplias regiones cromosómicas. Una de tales mutaciones es la inversión, la cuál produce una rotación de 180 grados en un segmento del cromosoma. Dados dos cromosomas sobre el mismo número de genes, el problema de la distancia de inversión consiste en determinar la secuencia más corta de inversiones que convierte a un cromosoma en el otro. El análisis del ordenamiento por inversiones involucra el manejo de un parámetro conocido como *obstáculo* que impide esta transformación. Las investigaciones sobre el problema de la distancia de inversión han permitido construir un modelo matemático cuyos elementos principales son las permutaciones y varias estructuras combinatorias como las multigráficas 4-regulares, las gráficas circulares y los sistemas de isotropía, éstos últimos fuertemente relacionadas con la teoría de delta-matroides. La aplicación de este modelo permite calcular la distancia de inversión aplicando dichos conceptos.

05E02

Palabras clave: Permutación, inversión, sistemas de isotropía, multigráficas 4-regulares.

1. Ordenamiento por inversiones

En el modelo aplicado, un genoma es un conjunto de cromosomas ¹, donde cada cromosoma es una secuencia ordenada de genes. Un gen es un segmento de ADN identificado por sus extremos 5' y 3'. El orden de los genes en el cromosoma es representado por una permutación la cuál es una función biyectiva que asocia a $i \in \{1, \dots, n\}$ un elemento $\pi_i \in \{1, \dots, n\}$ donde n es el número de genes. Una permutación signada es una permutación con un signo $+$ o $-$ en cada elemento, el cuál representa la orientación del gen correspondiente. Una *mutación* se define como un cambio en el material genético de la célula. Los reordenamientos cromosómicos son mutaciones a gran escala que afectan amplias regiones cromosómicas sin llegar a modificar el cromosoma en su totalidad. En el presente texto se abordará el reordenamiento cromosómico conocido como *inversión*.

Sea $\pi = (\pi_1, \dots, \pi_{i-1}, \pi_i, \dots, \pi_j, \pi_{j+1}, \dots, \pi_n)$ un cromosoma. Una inversión $\rho(i, j)$ reordena los genes dentro de π transformándolo en $(\pi_1, \dots, \pi_{i-1}, -\pi_j, \dots, -\pi_i, \pi_{j+1}, \dots, \pi_n)$, un nuevo cromosoma. Consideremos un segundo cromosoma γ con el mismo número de genes que π . La distancia de inversión se define como la secuencia más corta de inversiones que convierte a π en γ . El problema se puede formular como la determinación del número mínimo de inversiones que convierten a π en la permutación identidad ι , éste número se denota como $d_r(\pi, \iota)$ y se conoce como la *distancia de inversión de π* [13, 16]. Sea π un cromosoma lineal, sus telómeros se representan con los elementos 0 y $n+1$ como extremos en π , se obtiene una permutación $\pi = (0, \pi_1, \dots, \pi_n, n+1)$, así se dice que π es una permutación signada y *enmarcada*. De aquí en adelante todas las permutaciones serán enmarcadas y simplemente las denotaremos por π .

1.1. Inversiones orientadas

Un *par orientado* (π_i, π_j) con $i < j$, es un par de enteros consecutivos con signos opuestos que cumplen $|\pi_i| - |\pi_j| = \pm 1$. Un par (π_i, π_j) puede inducir dos tipos de inversión:

¹A lo largo del presente texto se consideran únicamente genomas eucariotas, es decir, conjuntos de cromosomas lineales.

$$\text{Si } \pi_i + \pi_j = +1 \text{ entonces se aplica } \rho(i, j - 1) \quad (1)$$

$$\text{Si } \pi_i + \pi_j = -1 \text{ entonces se aplica } \rho(i + 1, j) \quad (2)$$

Los pares orientados indican inversiones que producen elementos consecutivos que a su vez son enteros consecutivos. Una *inversión orientada* es una inversión inducida por un par orientado [2].

Ejemplo 1. Consideremos la permutación signada π en (3).

$$\pi = (\underset{\pi_0}{0}, \underset{\pi_1}{3}, \underset{\pi_2}{4}, \underset{\pi_3}{6}, \underset{\pi_4}{5}, \underset{\pi_5}{7}, \underset{\pi_6}{9}, \underset{\pi_7}{-2}, \underset{\pi_8}{1}, \underset{\pi_9}{8}, \underset{\pi_{10}}{10}) \quad (3)$$

El par orientado $(3, -2)$ induce una inversión del tipo (1) en π .

$$\rho(3, -2)(\pi) = (\underset{\pi_0}{0}, \underset{\pi_1}{-9}, \underset{\pi_2}{-7}, \underset{\pi_3}{-5}, \underset{\pi_4}{-6}, \underset{\pi_5}{-4}, \underset{\pi_6}{-3}, \underset{\pi_7}{-2}, \underset{\pi_8}{1}, \underset{\pi_9}{8}, \underset{\pi_{10}}{10})$$

El par orientado $(-9, 8)$ induce una inversión del tipo (2) en $\rho(3, 2)(\pi)$:

$$\rho(-9, 8)\rho(3, -2)(\pi) = (\underset{\pi_0}{0}, \underset{\pi_1}{-9}, \underset{\pi_2}{-8}, \underset{\pi_3}{-1}, \underset{\pi_4}{2}, \underset{\pi_5}{3}, \underset{\pi_6}{4}, \underset{\pi_7}{6}, \underset{\pi_8}{5}, \underset{\pi_9}{7}, \underset{\pi_{10}}{10})$$

Al aplicar la inversión inducida por $(-8, 7)$ sobre el resultado anterior, se tiene:

$$\rho(-8, 7)\rho(-9, 8)\rho(3, -2)(\pi) = (0, -9, -8, -7, -5, -6, -4, -3, -2, 1, 10)$$

La inversión inducida por el par $(-2, 1)$ en la permutación anterior, produce:

$$\rho(-2, 1)\rho(-8, 7)\rho(-9, 8)\rho(3, -2)(\pi) = (0, -9, -8, -7, -5, -6, -4, -3, -2, -1, 10)$$

Al aplicar la inversión inducida por $(-9, 10)$ en el resultado anterior, se tiene:

$$\rho(-9, 10)\rho(-2, 1)\rho(-8, 7)\rho(-9, 8)\rho(3, -2)(\pi) = (0, 1, 2, 3, 4, 6, 5, 7, 8, 9, 10)$$

Podemos observar que la permutación en $\rho(-9, 10)\rho(-2, 1)\rho(-8, 7)\rho(-9, 8)\rho(3, -2)(\pi)$ del ejemplo 1 no está totalmente ordenada sin embargo no es posible seguir aplicando inversiones orientadas debido a la presencia de una estructura conocida como obstáculo.

1.2. Inversiones no orientadas

Supongamos que π es una permutación signada y enmarcada cuyos elementos son todos positivos y no consecutivos. Consideremos un orden circular (inducido al fijar 0 como el sucesor de $n + 1$) en π . Un intervalo enmarcado en π , es un intervalo de la forma $i, \pi_{j+1}, \pi_{j+2}, \dots, \pi_{j+k-1}, i + k$. Tal que todos los enteros entre i e $i + k$ pertenecen al intervalo $[i, \dots, i + k]$. Un *obstáculo* en π es un intervalo enmarcado que no contiene un intervalo enmarcado más corto [2]. Los obstáculos en π se eliminan con dos tipos de inversiones no orientadas: la *inversión de fusión* (*merge reversal*) y la *inversión de corte* (*cut reversal*). El presente artículo se enfoca en la inversión de corte la cual invierte el segmento entre i e $i + 1$ de un obstáculo $i, \pi_{j+1}, \pi_{j+2}, \dots, i + 1, \dots, \pi_{j+k-1}, i + k$ [15], produciendo una permutación con el mismo número de elementos. Ésta última característica nos permite tener un control mayor en el modelo. Dejamos la inversión de fusión para un estudio posterior.

Ejemplo 2. Denotemos la última permutación obtenida en el ejemplo 1 como $\pi_{(-9,10)} = (0, 1, 2, 3, 4, 6, 5, 7, 8, 9, 10)$. La permutación en (4) muestra la inversión de corte aplicada en la permutación $\pi_{(-9,10)}$:

$$\pi = (0, 1, 2, 3, 4, \underset{i}{-6}, \underset{i+1}{5}, \underset{i+k}{7}, 8, 9, 10). \quad (4)$$

Después de la inversión de corte la permutación en (4) se ordena con dos inversiones orientadas:

$$\pi_{(-6,5)} = (0, 1, 2, 3, 4, -6, -5, 7, 8, 9, 10),$$

$$\pi_{(4,-5)} = (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10).$$

De acuerdo a lo mostrado en los ejemplos 1 y 2 la permutación signada en (3) queda ordenada con 7 inversiones orientadas y 1 inversión no orientada, por lo tanto $d(\pi, \iota) = 8$.

Un súper obstáculo es un obstáculo cuya eliminación crea nuevos obstáculos. Sea π una permutación signada. Se dice que π es una *fortaleza* (*fortress*), si tiene un número impar de obstáculos y todos ellos son súper obstáculos. Así podemos definir la variable booleana $f(\pi)$,

$$f(\pi): \begin{cases} 1 & \text{si } \pi \text{ es una fortaleza} \\ 0 & \text{en otro caso} \end{cases}$$

2. Palabras de doble ocurrencia

Sea λ un alfabeto finito y m una palabra sobre λ , donde cada letra ocurre exactamente dos veces. Se dice que m es una *palabra de doble ocurrencia* [4]. Sea π un cromosoma si cada gen $i \in \pi$ es representado por sus extremos $5'$ y $3'$ el resultado es una palabra de doble ocurrencia que denotaremos con π' . La palabra de doble ocurrencia π' se obtiene al tomar el valor absoluto de cada gen $i \in \pi$ y aplicar las asignaciones en (5) y (6):

$$i \rightarrow ((i-1)_b, i_a) \quad \text{si } i > 0 \quad (5)$$

$$i \rightarrow (i_a, (i-1)_b) \quad \text{si } i < 0 \quad (6)$$

Los elementos i_a y $(i-1)_b$ corresponden a los extremos $5'$ y $3'$ del gen. El orden en estos elementos modela la orientación del gen.

Ejemplo 3. La palabra de doble ocurrencia π' asociada a la permutación signada π en (3) del ejemplo 1 se muestra en (7).

$$\pi' = (0_a, 2_b, 3_a, 3_b, 4_a, 5_b, 6_a, 4_b, 5_a, 6_b, 7_a, 8_b, 9_a, 2_a, 1_b, 0_b, 1_a, 7_b, 8_a, 9_b) \quad (7)$$

3. Gráfica de intersección de π'

Sea V el conjunto de todos los elementos en π' . El diagrama $BG(\pi')$ consta de una circunferencia con un conjunto finito de cuerdas. La circunferencia tiene las etiquetas de la palabra π' , en el orden en que aparecen en π' ; los extremos de las cuerdas están conectados por aristas que representan el orden en la permutación identidad. Si asignamos un color a las aristas que se encuentran en la circunferencia y un color distinto a las cuerdas podremos considerar el diagrama como un conjunto de ciclos de longitud par que alternan aristas de cada color [14]. El diagrama $BG(\pi')$ es la representación circular de la *gráfica de punto de rompimiento* de π' . Con base en $BG(\pi')$ se puede construir $H_b(\pi')$, como la gráfica de intersección del diagrama $BG(\pi')$, donde cada cuerda es representada por un vértice y dos vértices son adyacentes si las correspondientes cuerdas se intersectan en el

diagrama; $v, w \in V$. Si se construye una palabra con las etiquetas que se encuentran colocadas sobre la circunferencia del diagrama $BG(\pi')$, empezando por 0_a y recorriendo la circunferencia siguiendo el orden de las manecillas de un reloj hasta encontrar la etiqueta de partida se tiene una palabra de doble ocurrencia. Adicionalmente la orientación de los genes en π' se codifica añadiendo lazos a los vértices de $H_b(\pi')$, un vértice orientado es un vértice con lazo.

En el contexto del estudio sobre las estructuras combinatorias que se derivan del estudio de gráficas 4-regulares, comenzado por Kotzig [18] y desarrollado por Bouchet [5], a la gráfica de punto de rompimiento se le llama *diagrama cordal* y a la gráfica de intersección se le conoce como *gráfica circular*.

Ejemplo 4. Diagrama $BG(\pi')$ (figura 1) y gráfica de intersección $H_b(\pi')$ (figura 2) de la palabra de doble ocurrencia π' del ejemplo 3.

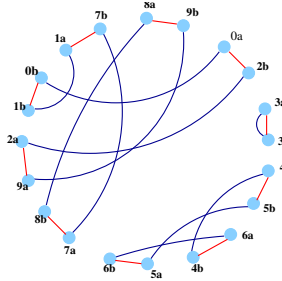


Figura 1: $BG(\pi')$

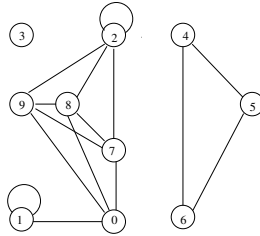


Figura 2: $H_b(\pi')$

Sea H_b una gráfica con lazos permitidos, pero sin aristas múltiples. La *matriz de adyacencia* $A(H_b)$ de H_b es una $V \times V$ -matriz sobre $\mathbf{GF}(2)$ ² donde para los vértices v, v' de H_b , la entrada etiquetada por (v, v') es 1 si y sólo si v y v' son adyacentes. Se considera que un vértice v de H_b es adyacente a si mismo si v es un vértice con lazo.

Ejemplo 5. Consideremos la gráfica $H_b(\pi')$ en 2 del ejemplo 4 cuya matriz de adyacencia $A(H_b(\pi'))$ se ilustra en (8).

$$A(H_b(\pi')) = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{matrix} \quad (8)$$

Sea $r_A(A(H_b(\pi')))$ el rango de la matriz $A(H_b(\pi'))$.

Teorema 1. (Brijder 2017, [11]). Para toda permutación signada π , se tiene que:

$$d(\pi, \iota) \geq r_A(A(H_b)) \quad (9)$$

En el caso de permutaciones sin obstáculos la desigualdad en la ecuación (9) del teorema (1) se convierte en la igualdad $d(\pi, \iota) = r(A(H_b))$. Sea $h(\pi)$ el número de obstáculos en la permutación signada π y sea $f(\pi)$ la variable booleana que nos indica si π es o no una fortaleza. La ecuación exacta para la distancia de inversión en términos de matrices de tipo simétrico esta dada por:

²En el caso binario, una matriz antisimétrica es una matriz simétrica con su diagonal principal nula [20].

$$d(\pi, \iota) = r(A(H_b)) + h(\pi) + f(\pi) \quad (10)$$

Denotamos por $A[X]$ a la submatriz de $A(H_b)$ indexada en renglones y columnas por $X \subseteq V$. Las matrices $A[X]$ son llamadas *submatrices principales*. Sea $A(H_b(\pi'))$ la matriz de adyacencia de la gráfica $H_b(\pi')$. Las submatrices principales correspondientes a obstáculos en la gráfica de intersección son matrices antisimétricas, no invertibles de tamaño impar [10].

Consideremos las transformaciones definidas en (5) y (6) que nos permiten construir la palabra de doble ocurrencia π' . Con base en π' se obtendrá una multigráfica 4-regular con un sistema de isotropía asociado, el cual introduciremos en la siguiente sección.

4. Sistemas de isotropía

Sea K el espacio vectorial de dimensión 2 sobre $GF(2)$ con la forma bilineal, antisimétrica no degenerada, $(a \wedge b)$, $a \wedge b = 1 \iff 0 \neq a \neq b \neq 0$. Para cada conjunto finito V consideramos a K^V como un espacio vectorial de dimensión $2|V|$ sobre $GF(2)$ con la forma bilineal, antisimétrica no degenerada, $\Phi \wedge \Psi = \sum (\Phi_v \wedge \Psi_v : v \in V)$. Si L es un subespacio totalmente isotrópico de K^V (esto es, $\Phi, \Psi \in L \implies \Phi \wedge \Psi = 0$) entonces $\dim(L) \leq |V|$. Un *sistema de isotropía* es un par $s = (L, V)$, donde V es un conjunto finito y L es un subespacio $L \in K^V$ tal que $\dim(L) = |V|$. Para $\Phi \in K^V$ y un subconjunto $X \subseteq V$, el vector $\gamma \in K^V$ definido por $\gamma_v = \Phi_v$ si $v \in X$ y $\gamma_v = 0$ si $v \notin X$, se denotará con ΦX . Sea $\hat{\Phi} = \{\Phi X : X \subseteq V\}$, $\hat{\Phi}$ es un subespacio de K^V y la dimensión de $\hat{\Phi}$ es igual al número de elementos v tales que $\Phi_v \neq 0$. Sea $K' = K \setminus \{0\}$, se dice que Φ es completo si $\Phi \in K'^V$, tal que $\dim(\hat{\Phi}) = |V|$. Dos vectores completos Φ y Ψ son *suplementarios* si $\Phi_v \neq \Psi_v \quad \forall v \in V$, esto es si $\hat{\Phi}$ y $\hat{\Psi}$ son subespacios suplementarios de K^V . El *rango* de un vector $\Phi \in K^V$ con respecto a un sistema de isotropía $s = (L, V)$ se define como $r(\Phi) = \dim(\hat{\Phi} \cap L)$. Un *vector Euleriano* de s es un vector $\Phi \in K'^V$ tal que $r(\Phi) = 0$ [5, 7].

5. Multigráfica 4-regular asociada a π

Para construir una multigráfica 4-regular G . Empezamos por expandir la transformación definida en (5) y (6) para π' por medio de la inclusión de un gen (denotado como M_i), M_0 como primer elemento y M_{n+1} como último elemento y por la inclusión de ADN no codificante (denotado como I_i) entre cada gen, de manera que se alternan en la expansión como $M_0 \dots I_0 \dots M_{i_1} \dots I_j \dots M_{i_2} \dots I_{j+1} \dots$, con $i = 1, \dots, n$; $j \in \{1, \dots, n\}$, donde en los lugares que aparece \dots van los elementos ordenados de π' .

Ejemplo 6. La expansión de π' mostrada en (7) del ejemplo 3, se ilustra en (11).

$$\pi' = (M_0 0_a I_0 2_b M_3 3_a I_1 3_b M_4 4_a I_2 5_b M_6 6_a I_3 4_b M_5 5_a I_4 6_b M_7 7_a I_5 8_b M_9 9_a I_6 2_a M_2 1_b I_7 0_b M_1 1_a I_8 7_b M_8 8_a I_9 9_b M_{10}) \quad (11)$$

5.1. Gráfica dirigida

La gráfica $D_{\pi'}$ es una gráfica auxiliar que se define como una gráfica dirigida donde cada vértice se asocia con un único extremo del gen y cada arco representa un segmento de ADN (codificante o no) entre dos extremos, ver el ejemplo 7.

Ejemplo 7. La gráfica $D_{\pi'}$ muestra la gráfica dirigida correspondiente a la permutación expandida π' del ejemplo 6.

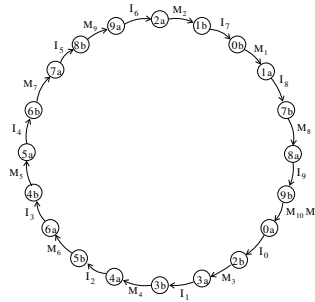


Figura 3: $D(\pi')$

Una multigráfica 4-regular G_π asociada a un cromosoma π se obtiene a partir de D_π modificando la transformación en (5) y (6) como se expresa en (12):

$$(i_b, i_a)) \mapsto i \quad (12)$$

Cada vértice v de G_π se relaciona con su elemento par en la transformación en (12) y cada arco de D_π se convierte en una arista no dirigida de G_π . La multigráfica G_π es conexa para toda permutación signada π , [11, 12].

Ejemplo 8. La multigráfica 4-regular en la figura 4 se obtiene de la gráfica dirigida D_π del ejemplo 7.

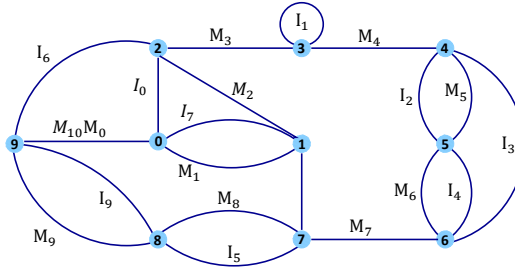


Figura 4: G_π

5.2. Bitransiciones

Consideraremos que cada arista e de una multigráfica 4-regular G está compuesta por un par de *medias aristas* $\{h', h''\}$ con solo un vértice final en G , tal que cada media arista incide en un extremo de e [3, 5, 8, 17]. Para cada vértice v de G , el conjunto de medias aristas incidentes en v , se denota con h_v . Ya que G es 4-regular $|h_v| = 4, \forall v \in V$. Una *transición* en v es un subconjunto de tamaño dos de h_v . Una *bitransición* en v es un par $\omega = \{t', t''\}$ de transiciones disjuntas incidentes en un mismo vértice v (ver figura 7). Un conjunto de bitransiciones $\Omega = \omega_1, \omega_2, \dots, \omega_k$ es *disjunto* si no contiene bitransiciones incidentes en un mismo vértice. Se define $\hat{\Omega}(G) = \{\Omega : \Omega \text{ es un conjunto disjunto de bitransiciones en } G\}$. Todo conjunto $\Omega \in \hat{\Omega}$ induce una *partición Euleriana* de G , es decir un conjunto de circuitos tales que

cada arista de G pertenece a un único circuito. *Cortar* una bitransición incidente en v consiste en eliminar v de G y reemplazarlo con dos vértices v' incidente a las medias aristas de t' y v'' incidente a las medias aristas de t'' ; así se obtiene una nueva gráfica denotada con $G|_{\omega}$. *Cortar* un conjunto $\Omega \in \hat{\Omega}$ implica construir en la gráfica $G|\Omega = G|_{\omega_1|\omega_2|\dots|\omega_k}$ (la cuál se denotará con $G|\Omega$ ya que el orden $\omega_1, \omega_2, \dots, \omega_k$ es irrelevante [9]).

Ejemplo 9. El orden en la permutación expandida π' del ejemplo 6 induce un paseo Euleriano en la multigráfica 4-regular conexa de la figura 4. Este paseo, que denotaremos como $TE(\Phi)$, se muestra en (13).

$$TE(\Phi) = M'_1 M''_1, I'_8 I''_8, M'_8 M''_8, I'_9 I''_9, M_{10} M'_0 M_{10} M''_0, I'_0 I''_0, M'_3 M''_3, I'_1 I''_1, M'_4 M''_4, I'_2 I''_2, M'_6 M''_6, I'_3 I''_3, M'_5 M''_5, I'_4 I''_4, M'_7 M''_7, I'_5 I''_5, M'_9 M''_9, I'_6 I''_6, M'_2 M''_2, I'_7 I''_7 \quad (13)$$

$TE(\Phi)$ define al conjunto de bitransiciones $\Omega(TE(\Phi))$.

- $\Omega(TE(\Phi))$
- $v_0 \quad \{\{I''_7, M'_1\}, \{M_{10} M''_0, I'_0\}\}$
 - $v_1 \quad \{\{M''_1, I'_8\}, \{M'_2, I'_7\}\}$
 - $v_2 \quad \{\{I''_6, M'_2\}, \{I'_0, M'_3\}\}$
 - $v_3 \quad \{\{M''_3, I'_1\}, \{I''_1, M'_4\}\}$
 - $v_4 \quad \{\{M''_4, I'_2\}, \{I'_3, M'_5\}\}$
 - $v_5 \quad \{\{I''_2, M'_6\}, \{M'_5, I'_4\}\}$
 - $v_6 \quad \{\{M''_6, I'_3\}, \{I''_4, M'_7\}\}$
 - $v_7 \quad \{\{I''_8, M'_8\}, \{M''_9, I'_5\}\}$
 - $v_8 \quad \{\{M'_8, I'_9\}, \{I''_5, M'_9\}\}$
 - $v_9 \quad \{\{I''_9, M_{10} M'_0\}, \{M'_9, I'_6\}\}$

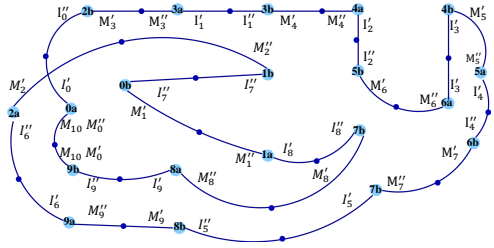


Figura 5: $G_{\pi}|\Omega(TE(\Phi))$

Consideremos un segundo conjunto de bitransiciones $\Omega \in \hat{\Omega}(G_{\pi})$ denotado con $\Omega(\Psi)$. $\Omega(\Psi)$ induce la partición Euleriana $P(\Omega(\Psi))$ en (14) ilustrada en la gráfica de la figura 6.

$$P(\Omega(\Psi)) = \{\{M'_1 M''_1, M''_2 M'_2, M'_3 M''_3, M'_4 M''_4, M'_5 M''_5, M'_6 M''_6, M'_7 M''_7, M'_8 M''_8, M'_9 M''_9, M_{10} M'_0 M_{10} M''_0\}, \{I'_3 I''_3, I'_2 I''_2, I'_4 I''_4\}, \{I'_1 I''_1\}, \{I'_7 I''_7, I'_8 I''_8, I'_5 I''_5, I'_9 I''_9, I'_6 I''_6, I'_0 I''_0\}\} \quad (14)$$

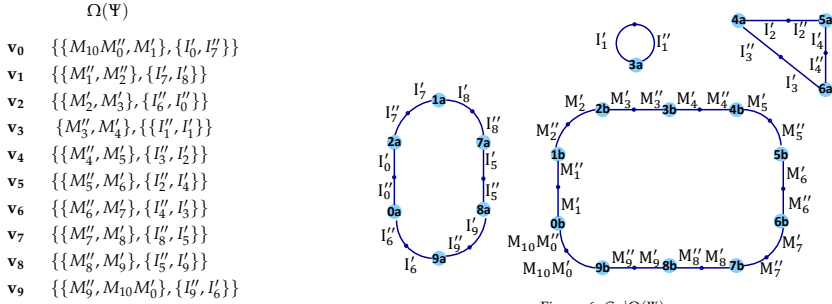


Figura 6: $G_\pi|\Omega(\Psi)$

En las figuras 5 y 6 aparece cada arista dividida en sus dos medias aristas mediante un vértice más pequeño sin etiqueta.

Regresando al problema del cálculo de la distancia de inversión entre dos cromosomas pertenecientes a dos especies que comparten cierto porcentaje de información genética, notemos que dada la construcción de la gráfica $\Omega(TE(\Phi))$, ésta corresponde al primer cromosoma y la gráfica $\Omega(\Psi)$ al segundo, que en el modelo presentado corresponde a la permutación identidad con el mismo número de genes que el primer cromosoma, ver [11]. Además, los conjuntos $\Omega(TE(\Phi))$ y $\Omega(\Psi)$ no tienen bitransiciones en común, en este caso se dice que los conjuntos son *suplementarios*.

6. Multigráfica codificada

Sea G una multigráfica conexa 4-regular y sea $TE = h'_0h''_0, h'_1h''_1, \dots, h'_{n-1}h''_{n-1}$ un paseo Euleriano de G , expresado mediante las medias aristas que lo forman. Consideremos un vértice v de G tal que $v = v_i = v_j$, ($i \neq j$). Existen exactamente tres pares de bitransiciones en v : $\{\{h'_{j-1}, h'_j\}, \{h''_{i-1}, h''_i\}\}$, $\{\{h'_{j-1}, h'_i\}, \{h''_{i-1}, h''_j\}\}$ y $\{\{h'_{j-1}, h''_{i-1}\}, \{h'_i, h''_j\}\}$ los cuáles se ilustran en la figura 7.

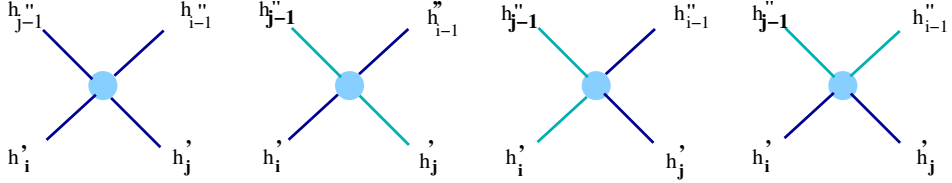


Figura 7: Bitransiciones.

Asociamos un elemento $\tau_v(\omega) \in K'^V$, a cada una de las tres posibles bitransiciones ω incidentes en v . Entonces existe una correspondencia biyectiva entre las tres bitransiciones en v y los elementos en $K'^V = \{\phi, \chi, \psi\}K'^V$, [3, 5, 6]. Denotamos estas bitransiciones por:

$$\begin{aligned}\tau_v(\{\{h''_{j-1}, h'_j\}, \{h''_{i-1}, h'_i\}\}) &= \phi \\ \tau_v(\{\{h''_{j-1}, h'_i\}, \{h''_{i-1}, h'_j\}\}) &= \chi \\ \tau_v(\{\{h''_{j-1}, h''_{i-1}\}, \{h'_i, h'_j\}\}) &= \psi\end{aligned}$$

Cuadro 1: Codificación de bitransiciones.

Una *codificación de bitransiciones* es una familia $\tau = (\tau_v : v \in V)$ [3, 8].

Ejemplo 10. *Asignación de etiquetas para los conjuntos disjuntos de bitransiciones $\Omega(TE(\Phi))$ y $\Omega(\Psi)$ del ejemplo 9.*

	$v0$	$v1$	$v2$	$v3$	$v4$	$v5$	$v6$	$v7$	$v8$	$v9$
$\tau(\Omega(TE(\Phi)))$	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
$\tau(\Omega(\Psi))$	χ	ψ	ψ	χ	χ	χ	χ	χ	χ	χ

Una *multigráfica codificada* es un par (G, τ) , donde G es una multigráfica 4-regular y τ una codificación de bitransición [5, 6]. Para todo conjunto $\Omega \in \hat{\Omega}(G)$ se define un vector $T(\Omega) \in K^V$ tal que $T(\Omega)_v$ es la codificación de la bitransición de Ω incidente en v , $\forall v \in V$ [3]. Si no existe una bitransición de Ω incidente en v entonces $T(\Omega)_v = 0$. En otro caso $T(\Omega)_v$ es la etiqueta de la bitransición de Ω incidente en v . Sea $S = (L, V)$ el sistema de isotropía asociado a una multigráfica

codificada (G, τ) . Una bitransición inducida por un vector $\Phi \in K^V$ en un vértice v de G es una bitransición tal que $\tau_v(\omega) = \Phi_v$. Denotamos con $\Omega(\Phi)$ al conjunto de todas las bitransiciones inducidas por Φ . Todo conjunto de bitransiciones (partición Euleriana) es determinado por un único vector de K^V [8]. Consideremos el conjunto de bitransiciones $\Omega(TE(\Phi))$ que induce el paseo Euleriano $TE(\Phi)$ en (13) del ejemplo 9. Sea Φ un vector en K^V tal que para cada $v \in V$ $\tau(\Omega(TE(\Phi)))_v = \Phi_v$, sea $P(\Omega(TE(\Phi)))$ la partición Euleriana de G_π inducida por Φ (en la figura 5) y sea $I(G_\pi)$ el número de componentes conexas en G_π . $P(\Omega(TE(\Phi)))$ satisface la igualdad:

$$|P(\Omega(TE(\Phi)))| - I(G_\pi) = 0 \quad (15)$$

Un sistema de isotropía s que se puede asociar de la manera descrita con una multigráfica codificada se conoce como un *sistema gráfico* [6, 7].

7. Conmutación

Consideremos una multigráfica codificada (G, τ) , su sistema gráfico asociado S , y un vector Euleriano Φ de S . Sea $TE(\Phi) = h'_0 h''_0, h'_1 h''_1, \dots, h'_{n-1} h''_{n-1}$ una sucesión de medias aristas que forman un paseo Euleriano en G inducido por Φ . La secuencia de vértices de $TE(\Phi)$ es $V(\Phi) = v_0, v_1, \dots, v_{n-1}$ donde v_i es el vértice incidente a $TE(\Phi) = h'_{i-1} h''_{i-1}, h'_i$ para $i \in \{0, \dots, n-1\}$. $V(\Phi)$ es una *palabra de doble ocurrencia* sobre $V(G)$. Sea $v \in V(G)$ y supongamos $v = v_i = v_j$ ($i \neq 0$). El resultado de reemplazar la bitransición codificada por Φ en v con la bitransición con la etiquetada con ψ en v , es el conjunto de bitransiciones de un nuevo paseo Euleriano denotado con $TE(\Psi)^*v$. Se dice que $TE(\Psi)^*v$ es obtenido por la *conmutación* de $TE(\Psi)$ en v [6, 7].

Para cada vértice $v \in V(G)$, dado que $TE(\Phi)$ es un paseo Euleriano, asociamos con la bitransición en v el elemento $\phi \in K^V$, denotémoslo como ϕ_v . Éste corresponde a una forma de despegar el vértice v de grado 4 en dos vértices v_1 y v_2 de grado 2 que coincide en v con $TE(\Phi)$. Hay otras dos formas de despegar v que corresponden a los elementos χ_v y ψ_v correspondientes a χ y ψ respectivamente.

Si se cambia en v la forma de despegar por χ_v o ψ_v , y se conservan las demás ϕ_w para $w \neq v$ en $TE(\Phi)$, solamente una de las dos elecciones para v , nos lleva a

un nuevo paseo Euleriano [18], para todo vértice de G .

8. Otra fórmula para calcular $d(\pi, \iota)$

Sea π una permutación signada, a partir de ella y siguiendo el procedimiento explicado, podemos construir una multigráfica 4-regular G_π . Si G_π es conexa, le corresponde un paseo Euleriano $TE(\Phi)$ con su codificación asociada, es decir $G_\pi|(\Phi)$; esta corresponde al cromosoma de la primera especie, sea esta s_1 . Cuando se obtiene $G_\pi|(\Psi)$ que corresponde a la permutación identidad asociada al cromosoma de la segunda especie, a la cual denominaremos s_0 . Observamos que $G_\pi|(\Psi)$ es una partición de circuitos, suplementaria a $G_\pi|(\Phi)$ y que cumple con tener un circuito asociado a la permutación identidad ι , cuyas aristas están etiquetadas por las M_i , $i = 0, \dots, n + 1$. Éstas representan el material genético codificante y otros circuitos con I_j como etiquetas de sus aristas.

Definimos $c(\pi)$ como el número de ciclos en la partición $G_\pi|\Omega(\Psi)$ obtenida a partir de la permutación signada π , excluyendo el ciclo con las etiquetas M_i . En [1] se demuestra el siguiente teorema.

Teorema 2. (Bafna-Pezner, [1]). Sea π una permutación signada con n elementos. Entonces:

$$d(\pi, \iota) \geq n + 1 - c(\pi) \quad (16)$$

La desigualdad (16) está expresada como igualdad en (17). En [1] encontramos la igualdad expresada en el siguiente teorema.

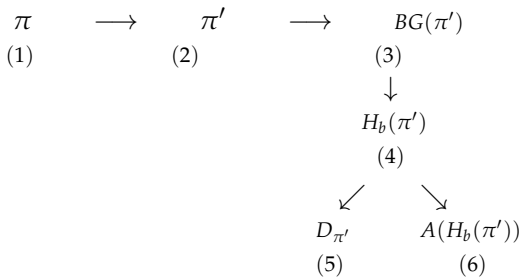
Teorema 3. (Hannenhalli, Pezner, [15]). Sea π una permutación signada. Si cada componente conexa de H_π tiene al menos un vértice con lazo (loop) o consiste de un vértice aislado, entonces:

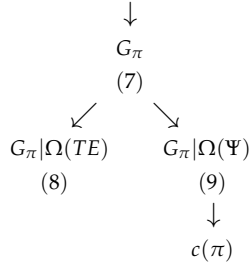
$$d(\pi, \iota) = n + 1 - c(\pi) \quad (17)$$

La fórmula (17) es muy simple y si hemos construido la partición $G_\pi | (\Psi)$ y ésta cumple las condiciones del teorema 3, el cálculo de $d(\pi, \iota)$ resulta sencillo. En la mayoría de los casos solo tendremos la desigualdad (16) como una aproximación. En 1993, Kececioglu y Sankoff conjeturaron que el ordenamiento por inversiones es un problema *NP-Duro*. En [19] Kececioglu y Sankoff presentan dos algoritmos para calcular la distancia de inversión, uno que es una aproximación usando el algoritmo Greedy y que encuentra una solución cercana a la óptima con complejidad computacional del orden de $O(n^2)$ y un algoritmo exacto que encuentra la solución usando el método de ramificación y acotamiento en $O(mL(n, n))$, donde m el tamaño del árbol construido para resolver el problema y $L(n, n)$ es el tiempo para resolver un programa lineal con n variables y n restricciones.

Ejemplo 11. Consideremos los conjuntos de bitrasciones $\Omega(TE(\Phi))$ y $\Omega(\Psi)$ cuya codificación se muestra en el ejemplo 10. Sean Φ y Ψ vectores en K^V tal que para cada $v \in V$, $\tau(\Omega(TE(\Phi)))_v = \Phi_v$ y $\tau(\Omega(\Psi))_v = \Psi_v$. El vector Φ induce el paseo Euleriano $TE(\Phi)$ descrito en (13), mientras que Ψ induce la partición P_Ψ descrita en (14), ilustradas en el ejemplo 9. La gráfica circular de P_Ψ con respecto al paseo Euleriano $TE(\Phi)$ es equivalente a la gráfica 2 en el ejemplo 4. Además, podemos observar que $c(\pi) = 3$, para la partición P_Ψ .

9. Esquema del proceso para el cálculo de $d(\pi, \iota)$





- (1) permutación signada
- (2) palabra de doble ocurrencia
- (3) diagrama cordal o gráfica de punto de rompimiento
- (4) gráfica circular o gráfica de intersección
- (5) gráfica dirigida auxiliar
- (6) matriz de adyacencia de $H_b(\pi')$
- (7) multigráfica 4-regular
- (8) $G_\pi| \Omega(TE)$ gráfica Euleriana
- (9) $G_\pi| \Omega(\Psi)$ partición Euleriana
- (10) número de ciclos no cidificadores

Agradecimientos

Las autoras agradecemos a nuestros colegas Diego Frausto Tamayo y Jesús Rodríguez Martínez su apoyo en la elaboración de algunas figuras que forman este documento.

Referencias

- [1] Vineet Bafna and Pavel A. Pevzner.
Genome rearrangements and sorting by reversals.
 SIAM Journal on Computing, (1996). 25(2):272–289.

- [2] Anne Bergeron.
A very elementary presentation of the Hannenhalli–Pevzner theory.
Discrete Applied Mathematics, 146(2), (2005). pp 134-145.
- [3] André Bouchet and Laurence Ghier.
Connectivity and β -invariants of isotropic systems and 4-regular graphs.
Discrete Mathematics, 161(1-3), (1996). pp 25-44.
- [4] André Bouchet.
Circle graph obstructions.
Journal of Combinatorial Theory, Series B, 60(1), (1994). pp 107-144.
- [5] André Bouchet.
Compatible Euler tours and supplementary Eulerian vectors.
European journal of combinatorics, vol. 14, no 6, (1993). pp 513-520.
- [6] André Bouchet.
k-transformations, local complementations and switching.
In Cycles and Rays Springer Netherlands. (1990). pp. 41-50.
- [7] André Bouchet.
Graphic presentations of isotropic systems.
Journal of Combinatorial Theory, Series B, 45(1), (1988). pp 58-76.
- [8] André Bouchet.
Isotropic systems.
European J. Combin. 8 (1987). pp 231-244.
- [9] André Bouchet.
Greedy Algorithm and Symmetric Matroids.
Math. Program. 38, 2 (September 1987). pp 147-159.
- [10] André Bouchet.
Representability of Δ -matroids.
Colloquia Mathematica Societatis János Bolyai 52. Combinatorics, Eger (Hungary), (1987) pp 167-182.

- [11] Robert Brijder.
Sorting by Reversals and the Theory of 4-Regular Graphs.
arXiv preprint arXiv:1701.07463. (2017).
- [12] Robert Brijder.
Recombination faults in gene assembly in ciliates modeled using multimatroids.
Theoretical Computer Science, 608. (2015). pp 27-35.
- [13] Alberto Caprara.
Sorting permutations by reversals and Eulerian cycle decompositions.
SIAM journal on discrete mathematics, 12(1), (1999). pp 91-110.
- [14] Simona Grusea, and Anthony Labarre.
The distribution of cycles in breakpoint graphs of signed permutations.
Discrete Applied Mathematics, 161(10). (2013) pp 1448-1466.
- [15] Sridhar Hannenhalli, and Pavel A. Pevzner.
Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals.
Journal of the ACM (JACM), 46(1), (1999) pp 1-27.
- [16] Sridhar Hannenhalli, and Pavel A. Pevzner.
Transforming men into mice (polynomial algorithm for genomic distance problem).
In Foundations of Computer Science, Proceedings., 36th Annual Symposium. (1995) pp. 581-592 IEEE.
- [17] Bill Jackson.
A characterisation of graphs having three pairwise compatible Euler tours.
Journal of Combinatorial Theory, Series B, 53(1), (1991). pp 80-92.
- [18] A. Kotzig.
Eulerian lines in finite 4-valent graphs and their transformations.
Theory of graphs, Proceedings of the Colloquium, Tihany, Hungary. (1966)
pp 219-230.
Academic Press, New York (1968)

- [19] John Kececioglu, and David Sankoff.
Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement.
Algorithmica 13: 180. <https://doi.org/10.1007/BF01188586> (1995)
- [20] Maria Guadalupe Rodríguez Sánchez.
Delta-matroides rueda ternarios.
Morfismos, México, Vól. 6 Núm. 1. (2002) pp. 31-55
- [21] Lorenzo Traldi.
Binary matroids and local complementation.
European Journal of Combinatorics, 45, (2015) pp 21-40
- [22] Lorenzo Traldi.
Interlacement in 4-regular graphs: a new approach using nonsymmetric matrices.
arXiv preprint arXiv:1204.0482. (2012).

20.

LIDIA ANGÉLICA GARCÍA GARCÍA
MA. GUADALUPE RODRÍGUEZ SÁNCHEZ

Dra. Ma. Guadalupe Rodríguez Sánchez
Profesora de tiempo completo e investigadora del
Departamento de Ciencias Básicas, UAM Azcapotzalco.
Av. San Pablo Xalpa 180, Reynosa Tamaulipas, 02200,
Ciudad de México, CDMX

rsmg@azc.uam.mx

Mtra. Lidia Angélica García García
Maestra en optimización,
UAM Azcapotzalco Av. San Pablo Xalpa 180, Reynosa
Tamaulipas, 02200, Ciudad de México, CDMX

stokes_1560@hotmail.com