



Posgrado en
Optimización

Universidad
Autónoma
Metropolitana 
Casa abierta al tiempo **Azcapotzalco**

División de Ciencias Básicas e Ingeniería

Modelos combinatorios en ensamblamiento genético

Tesis para obtener el grado de Maestra
en Optimización

Elaborada por
Ing. Lidia Angélica García García

Asesora:
Dra. Guadalupe Rodríguez Sánchez
UAM-Azcapotzalco

25 de enero de 2018

Contenidos

Lista de Figuras	7
Índice de cuadros	9
1. Introducción	1
1.1. Ordenamiento por inversiones	2
1.1.1. Sistemas de isotropía	3
1.1.2. Δ -matroides	4
1.1.3. Obstáculos	4
1.2. Organización del trabajo	5
2. Conceptos matemáticos fundamentales	7
2.1. Elementos de teoría de gráficas	7
2.1.1. Paseos y circuitos	7
2.2. Campo $GF(2)$	8
2.3. Matroides binarios	8
2.4. Δ -matroides	11
2.4.1. Representación lineal de un Δ -matroide	12
2.4.2. Gráfica fundamental	13
2.4.3. Transformación de pivote principal	13
2.4.4. Menores	13
2.4.5. Conexidad de un Δ -matroide	14
2.5. Conceptos relacionados con el estudio de Δ -matroides	14
2.5.1. Alfabetos complementarios y palabras de doble ocurrencia	14
2.5.2. Gráfica de alternancia	15
2.5.3. Gráfica de alternancia signada	16
2.5.4. Complementación de lazo	16
2.5.5. Complementación local	17

3. Modelo genético	19
3.1. Construcción del modelo	19
3.1.1. Estructura del ADN	19
3.1.2. Problema del reordenamiento cromosómico	22
3.1.3. Ensamblamiento genético en ciliados	24
3.1.4. Operaciones de ADN en ciliados	26
3.1.5. Secuencias MDS	28
3.2. Permutaciones. Inversiones orientadas y no orientadas	29
3.2.1. Distancia de inversión	29
3.2.2. Representación de un cromosoma lineal	29
3.2.3. Inversiones orientadas	29
3.2.4. Obstáculos	30
3.2.5. Inversiones no orientadas	30
3.3. Gráfica de punto de rompimiento	31
3.3.1. Inversiones en $BG(\pi')$	32
3.3.2. Modelo de programación lineal entera	34
3.3.3. Obstáculos en la gráfica de punto de rompimiento	35
3.3.4. Eliminación de obstáculos en $BG(\pi')$	36
3.4. Gráfica $H_b(\pi')$	37
3.4.1. Complementación local e inversiones	38
3.4.2. Diagrama cordal	39
3.4.3. Matriz de intersección	41
3.4.4. Complemento local modificado de M	42
3.5. Paseos Eulerianos en multigráficas 4-regulares	44
3.5.1. Multigráfica 4-regular asociada a una permutación signada	44
3.5.2. Bitransiciones	46
3.5.3. Multigráfica codificada	48
3.6. Transformaciones en paseos Eulerianos	50
3.6.1. Separación	50
3.6.2. Conmutación	51
3.7. Gráficas fundamentales	51
3.7.1. Gráfica de alternancia relativa	52
3.8. Δ -matroides y sistemas de isotropía	53
4. Aplicación en el ordenamiento por inversiones	57
4.1. Inversiones orientadas y Δ -matroides	57
4.2. Inversión de corte	58

CONTENIDOS	5
4.3. Ecuación exacta para la distancia de inversión	59
4.4. Ciliados	60
5. Conclusiones y trabajo futuro	61
5.1. Conclusiones	61
5.2. Trabajo futuro	61
A. Sistemas de isotropía	63
.0.1. Presentaciones gráficas de sistemas de isotropía . . .	63
Bibliografía	65

Lista de Figuras

1.1.	Estructura y secuencia del ADN. Imagen tomada de [104].	1
2.1.	Diagrama cordal y gráfica de alternancia de una palabra de doble ocurrencia.	15
2.2.	Ejemplo de transposición sobre diagrama cordal.	16
3.1.	Bases nitrogenadas. Imagen tomada de [102].	19
3.2.	Enlaces de hidrógeno entre pares de bases complementarias en el ADN.	20
3.3.	Transcripción.	20
3.4.	Codones.	21
3.5.	Componentes estructurales del gen.	21
3.6.	Corte y empalme del ARN mensajero.	21
3.7.	Cromosoma lineal.	22
3.8.	Bloques de sintenia del cromosoma X en ratones y en humanos.	22
3.9.	Ejemplo del reordenamiento de inversión en un segmento de ADN.	23
3.10.	Escenario evolutivo entre el cromosoma X en ratones y en humanos.	23
3.11.	Escenario evolutivo con una inversión y una transposición.	24
3.12.	Oxytricha trifallax, Paramecium tetraurelia, Tetrahymena thermophila. Imagen tomada de [103].	24
3.13.	Ensamblamiento genético.	25
3.14.	Ejemplo de ensamblamiento genético en el gen actina I del ciliado Sterkiella nova.	26
3.15.	Segmento del gen β TTP en el ciliado Sterkiella histriomuscorum.	26
3.16.	Plegamiento molecular en la escisión de lazo.	27
3.17.	Recombinación homóloga.	27
3.18.	Eliminación de ADN no genético.	27

3.19. Recombinación de doble lazo.	27
3.20. Ejemplo de recombinación de horquilla en el gen Actina I del ciliado <i>Sterkiella Nova</i>	28
3.21. Gráfica de punto de rompimiento de la permutación no signada del del ejemplo 18.	32
3.22. Componentes en $BG(\pi'_{\text{ratón}})$	33
3.23. Ciclos en $BG(\pi'_{\text{ratón}})$	33
3.24. Inversión orientada en $BG(\pi'_{\text{ratón}})$	33
3.25. Obstáculos en la gráfica de punto de rompimiento.	35
3.26. Inversión de corte en la gráfica de punto de rompimiento.	36
3.27. Inversión de fusión.	37
3.28. Gráfica $H_b(\pi')$	38
3.29. Inversión en $H_b(\pi')$	38
3.30. Representación lineal de la gráfica en el ejemplo (19).	39
3.31. Ejemplo de aplicación de la transformación en (25).	40
3.32. $H(\pi'_{\text{ratón}})$	40
3.33. Ejemplo de secuencia óptima.	41
3.34. Gráfica de intersección asociada a la permutación π_1 del ejemplo 17.	42
3.35. Gráfica dirigida correspondiente a la permutación expandida π' del ejemplo 34.	45
3.36. Multigráfica 4-regular obtenida de la gráfica dirigida del ejemplo 35.	45
3.37. $G_\pi \Omega(TE(\Phi))$	47
3.38. $G_\pi \Omega(\Psi)$	47
3.39. Bitransiciones.	48
3.40. Gráfica de alternancia del paseo Euleriano $TE(\Phi)$ del ejemplo 37.	52
4.1. Gráfica de alternancia del paseo Euleriano $TE(\Phi)*v_4$ del ejemplo 41.	59
4.2. Multigráfica 4-regular G_ω de la permutación expandida ω' en (54).	60

Índice de cuadros

2.1. Suma y producto (mod 2)	8
3.1. Codificación de bitransiciones.	48
3.2. Suma \boxplus	54

Capítulo 1

Introducción

El ADN (ácido desoxirribonucleico) es una molécula con forma de doble hélice constituida por dos cadenas antiparalelas de nucleótidos. En el ADN los nucleótidos consisten en una base nitrogenada ligada a la desoxirribosa, la que a su vez está ligada al fosfato. Las bases nitrogenadas son: Adenina (A), Citosina (C), Guanina (G) y Timina (T) (figura 1.1). En un nivel abstracto, el ADN se representa como una secuencia sobre el alfabeto $\{A, C, G, T\}$.

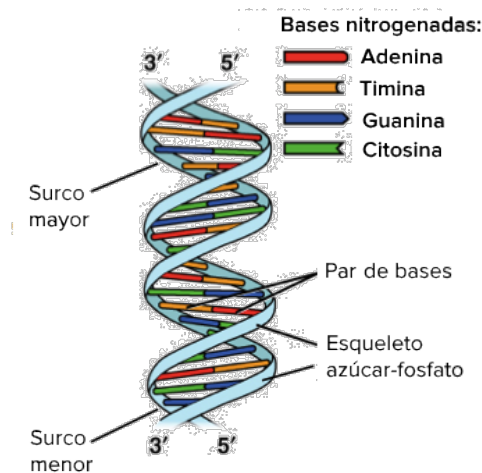


Figura 1.1: Estructura y secuencia del ADN. Imagen tomada de [104].

El genoma es la secuencia entera de ADN que almacena la información genética de un organismo. El genoma está particionado en subunidades llamadas cromosomas. Un cromosoma puede verse como una secuencia ordenada de genes. Un gen es un segmento de ADN involucrado en la producción de proteínas o en la codificación de ARN (ácido ribonucleico) funcional.

El reordenamiento cromosómico es una rama de la genómica comparada que estudia el orden genético entre diferentes especies. El objetivo es la inferencia de relaciones filogenéticas y la estimación del número de reordenamientos ocurridos durante la evolución de una especie a otra. El estudio de los reordenamientos cromosómicos también permite la construcción de bancos de datos que sirvan para atribuir funciones a los genes a partir de la comparación entre genomas.

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 del cromosoma. Tales mutaciones se conocen como reordenamientos cromosómicos. Asumiendo que la evolución es parsimoniosa y prefiere las secuencias con el mínimo de eventos [88, 89] y dados dos genomas, el problema general del reordenamiento cromosómico se puede plantear como la determinación de la secuencia más corta de reordenamientos que convierte a un genoma en el otro.

1.1. Ordenamiento por inversiones

En el modelo que consideramos, un genoma es un conjunto de cromosomas, donde cada cromosoma se representa como un conjunto de genes. Un gen se identifica con un entero sin signo. Cuando aparece en un cromosoma, un gen se asocia con un signo (" + " o " - ") que representa su orientación. Si Π es un genoma con N cromosomas y el k -ésimo cromosoma en Π contiene n_k genes, entonces:

$$\Pi = \{ \{ \pi_{11}, \pi_{12}, \dots, \pi_{1n_1} \}, \{ \pi_{21}, \pi_{22}, \dots, \pi_{2n_2} \}, \dots, \{ \pi_{N1}, \pi_{N2}, \dots, \pi_{Nn_N} \} \}$$

Los reordenamientos cromosómicos más comunes son la *inversión*, la *transposición* y la *transposición invertida* [72]. El presente trabajo se enfoca en la inversión la cuál corresponde al problema combinatorio de *ordenamiento por inversiones*. Sea $\pi = (\pi_1, \dots, \pi_{i-1}, \pi_i, \dots, \pi_j, \pi_{j+1}, \dots, \pi_n)$ un cromosoma en Π . Una inversión $\rho(i, j)$ reordena los genes dentro de π transformándolo en un cromosoma $(\pi_1, \dots, \pi_{i-1}, -\pi_j, \dots, -\pi_i, \pi_{j+1}, \dots, \pi_n)$. Consideremos un segundo genoma Γ y un cromosoma $\gamma \in \Gamma$ con el mismo número de genes que π . El *problema de la distancia de inversión* consiste en encontrar la secuencia más corta de inversiones que convierte a π en γ . La longitud de esta secuencia se conoce como la distancia de inversión entre π y γ y se denota con $d_r(\pi, \gamma)$. Las propiedades algebraicas de las permutaciones permiten formular el problema como la determinación del número mínimo de inversiones

que convierten a π en la permutación identidad, denotada como ι .

El primer algoritmo de aproximación para el problema de la distancia de inversión fue propuesto por Kececioğlu y Sankoff en 1994 [70]. En 1996 Bafna y Pevzner [3] introdujeron la *gráfica de punto de rompimiento* (*breakpoint graph*) asociada a una permutación signada y revelaron importantes vínculos entre la distancia de inversión y la descomposición en ciclos de esta gráfica convirtiéndola en la base de los análisis subsecuentes del ordenamiento por inversiones. Sea $c(\pi)$ la cardinalidad de una partición de la gráfica de punto de rompimiento en ciclos de aristas disjuntas. Bafna y Pevzner demostraron que cualquier inversión puede cambiar el parámetro $c(\pi)$ a lo sumo en 1. Lo anterior define la cota inferior $d_r(\pi, \iota) \geq (n+1) - c(\pi)$ para una permutación signada π con n elementos.

En 1995 Kececioğlu y Sankoff [69] observaron que para permutaciones aleatorias la diferencia promedio entre la cota inferior definida por Bafna y Pevzner y la distancia exacta es menor que 1. En 1999 Hannenhalli y Pevzner [56] probaron la existencia de un nuevo parámetro que permitió eliminar la diferencia entre $d_r(\pi, \iota)$ y $(n+1) - c(\pi)$ y demostraron que el problema del ordenamiento por inversiones es polinomial. El parámetro introducido por Hannenhalli y Pevzner para la formulación exacta de la distancia de inversión es el *número de obstáculos* en la permutación signada π . Un *obstáculo* (*hurdle*) es una subsecuencia en π en la cuál no es aplicable la operación de inversión como se ha definido.

En 2000 Kaplan, Shamir, y Tarjan presentaron la *gráfica de superposición* (*overlap graph*) para una permutación signada [68], la cual se puede definir como una *gráfica circular* con lazos permitidos. La matriz de adyacencia de una gráfica circular es una $V \times V$ -matriz simétrica (antisimétrica) con entradas en el campo binario $GF(2)$. En esta tesis la gráfica de superposición será llamada *gráfica de intersección*.

Debido a un buen número de investigaciones sobre el reordenamiento cromosómico, se ha podido construir un modelo matemático que usa principalmente permutaciones y cuyo fundamento, son las estructuras matroidales que específicamente son las gráficas circulares, los Δ -matroides y los sistemas de isotropía.

1.1.1. Sistemas de isotropía

La *complementación local* de una gráfica simple $H = (V, E)$ en un vértice $v \in V$ invierte la relación de adyacencia de H sobre el conjunto de vecinos de H . El concepto de complementación local de gráficas simples fue

introducido por Kotzig en relación a las k -transformaciones de los paseos Eulerianos de multigráficas 4-regulares [71]. Las complementaciones locales se relacionan con las k -transformaciones solo cuando se aplican sobre gráficas circulares [15]. En el caso general existe un marco natural para tratar las complementaciones locales dado por los *sistemas de isotropía*.

Los sistemas de isotropía son estructuras algebraicas y combinatorias introducidas por Bouchet en [16] que unifican algunas propiedades comunes a las multigráficas 4-regulares y a pares de matroides binarios duales.

1.1.2. Δ -matroides

Los Δ -matroides [20] son estructuras combinatorias que constituyen una generalización de los matroides. Los Δ -matroides binarios normales representan de manera natural la organización de la molécula de ADN [31, 33, 38]. En 2017 Brijder propuso su aplicación en el problema de la distancia de inversión [26]. Sean π y ι los cromosomas de dos especies y sea H_b la gráfica de intersección del cromosoma π . En la presente tesis se demuestra que π y ι pueden representarse como vectores suplementarios del sistema de isotropía con gráfica fundamental H_b el cuál está determinado por el Δ -matroide \mathcal{D}_{H_b} .

La *transformación de pivote principal* (*principal pivot transform*) introducida por Tucker [98], invierte parcialmente una matriz dada. Sea $A(H_b)$ la matriz de adyacencia de la gráfica H_b y sea un subconjunto $X \subseteq V$ con $\det(A(H_b)[X]) = 1$. Brijder probó [26] que la transformación de pivote principal sobre X seguida de la eliminación de los elementos de X corresponde a una inversión en π . Bouchet demostró [23] que en el caso de las matrices simétricas sobre $GF(2)$ la transformación de pivote principal satisface una definición equivalente en términos de Δ -matroides debido a que cumplen el axioma de intercambio de Δ -matroides.

1.1.3. Obstáculos

Como se mencionó anteriormente un obstáculo es una subsecuencia en la permutación signada π donde la inversión no es aplicable. Hannenhalli y Pevzner propusieron dos operaciones denominadas *inversión de corte* e *inversión de fusión* para la eliminación de estas estructuras. La inversión de corte y la inversión de fusión no están definidas de manera directa en la gráfica de intersección y su definición se hace por medio de la gráfica de punto de rompimiento o de la *gráfica de cubierta* (*cover graph*) de π^1 . En el presente trabajo un obstáculo se define como una submatriz principal

¹El concepto de gráfica de cubierta no será utilizado en esta tesis.

singular de $A(H_b)$. La inversión de corte se presenta como el *complemento local modificado* postulado por Traldi en [94], el cuál está directamente relacionado con las transformaciones de Kotzig.

1.2. Organización del trabajo

La presente tesis se estructura como sigue. En el capítulo 2, se revisan los conceptos matemáticos fundamentales empleados a lo largo de este trabajo. Las secciones 2.3 y 2.4 presentan respectivamente las definiciones de matroide binario y Δ -matroide.

En la sección 3.1 del capítulo 3 se da una introducción al problema del reordenamiento cromosómico. La subsección 3.1.3 describe la estructura de ciertos organismos unicelulares llamados ciliados que son empleados como modelo en la genómica comparativa. Recientes investigaciones en teoría de inversiones han propuesto la aplicación de las tres operaciones irreversibles con las que estos organismos ordenan su ADN (definidas en la subsección 3.1.4) en la deducción de potenciales relaciones evolutivas relativas al fenómeno del reordenamiento cromosómico [61, 63].

En la sección 3.2 se presentan las definiciones de inversión, distancia de inversión, inversiones orientadas e inversiones no orientadas en términos de la permutación signada con la que se representa el orden y la orientación de los genes en un cromosoma lineal. La gráfica de punto de rompimiento asociada a dicha permutación es definida en la sección 3.3. La subsección 3.3.2 describe un modelo de programación lineal entera para el problema de la distancia de inversión. El concepto de gráfica de intersección está dado en la sección 3.4. La matriz de adyacencia de esta gráfica se presenta en la subsección 3.4.3. En la subsección 3.4.4 se define la inversión de corte de las subsecciones 3.2.5 y 3.3.4 en términos de la operación sobre matrices antisimétricas conocida como *complemento local modificado*.

Los paseos Eulerianos en la multigráfica 4-regular conexa asociada a una permutación signada se estudian en la sección 3.5. En la subsección 3.5.3 se da la definición de la multigráfica codificada (G, τ) asociada a un sistema de isotropía S con gráfica fundamental H^2 .

Las transformaciones aplicables sobre un paseo Euleriano en una multigráfica 4-regular, descritas por Kotzig y posteriormente extendidas por Bouchet, se definen en la sección 3.6. El concepto de gráficas fundamentales es presentado en la sección 3.7. La sección 3.8 define el sistema de

²El concepto de un sistema de isotropía S se revisa en el apéndice A.

isotropía con gráfica fundamental $H_b(\Phi, \Psi)$ descrita en la subsección 3.7.1 y su relación con los Δ -matroides.

El capítulo 4 presenta la aplicación de los conceptos expuestos en los contextos del ordenamiento por inversiones y del ensamblamiento genético en ciliados. En la sección 4.1 las inversiones en un cromosoma lineal π son descritas en términos del Δ -matroide binario normal $\mathcal{D}_{H_b(\Phi, \Psi)}$. La definición de la distancia de inversión $d_r(\pi, \iota)$ en el caso en que π sea una permutación signada sin obstáculos se muestra en la ecuación (52) de esta sección. La relación entre el complemento local modificado y las transformaciones sobre paseos Eulerianos en multigráficas 4-regulares es presentada en la sección 4.2 mientras que la fórmula exacta para la distancia de inversión está dada por la ecuación (53) de la sección 4.3. Finalmente la descripción de como el modelo aplicado se generaliza para genes y cromosomas circulares se da en la sección 4.4.

Conceptos matemáticos fundamentales

2.1. Elementos de teoría de gráficas

Una gráfica G se define como un par ordenado (V, E) , donde V es un conjunto finito y E es un conjunto de multiconjuntos de dos elementos de V . Los elementos de V son llamados *vértices*, y los elementos de E son llamados *aristas* de G . Se dice que una arista es *incidente* con los dos vértices que ella une. Por ejemplo, la arista $\{v, w\}$ es incidente con los vértices v y w y los vértices v y w son llamados *vértices adyacentes*. En adelante una arista $\{v, w\}$ se denotará simplemente por vw . Un *lazo* es una arista vv incidente con un mismo vértice v [75].

Una *digráfica* o *gráfica dirigida* D es un par ordenado (V, E) , donde V es el conjunto de vértices y E es el conjunto de *arcos* o *aristas dirigidas*. Cada arco (v, w) es un *par ordenado* de vértices distintos v y w .

Una *multigráfica* G es una gráfica con lazos permitidos y donde más de una arista puede ser incidente con el mismo par de vértices. Se dice que multigráfica es una *gráfica simple con lazos permitidos* si sus aristas son incidentes a lo sumo con un mismo par de vértices. La *vecindad* de un vértice v de una gráfica G es el subconjunto de vértices definido en (1):

$$n(v) = \{w : vw \in E(G)\}. \quad (1)$$

El *grado* de un vértice v se define como el número de aristas incidentes en v , $deg(v) = |n(v)|$. Cada lazo contribuye en dos al grado de v .

2.1.1. Paseos y circuitos

Un paseo es una sucesión de aristas $e_{i_1}, e_{i_2}, \dots, e_{i_k}$ tal que el vértice terminal de e_{i_j} coincide con el vértice inicial de $e_{i_{(j+1)}}$ para $j \in \{1, \dots, k-1\}$. Un *circuito*

es un paseo $e_{i_1}, e_{i_2}, \dots, e_{i_k}$ en el cuál el vértice terminal e_{i_k} coincide con el vértice inicial e_{i_1} .

Un *paseo Euleriano* en una gráfica G , se define como un paseo que pasa a través de cada arista en G una y sólo una vez. De modo similar, un *circuito Euleriano* en G , se define como un circuito que pasa a través de cada arista de G una y sólo una vez. Una gráfica G es llamada *Euleriana* si tiene un circuito Euleriano.

Teorema 2.1 (Teorema de Euler). *Sea $G = (V, E)$ una gráfica o multigráfica conexa, son equivalentes para G :*

- (i) G es Euleriana.
- (ii) Todo vértice de G es de grado par.
- (iii) El conjunto $E(G)$ puede descomponerse en circuitos.

Multigráficas 4-regulares

Sea $G = (V, E)$ una *multigráfica Euleriana*, se dice que G es 4-regular si $\deg(v) = 4 \forall v \in V$. Una *partición Euleriana*, (también llamada partición de circuitos o ξ -descomposición [71]) en G es una partición de E en circuitos de aristas disjuntas.

2.2. Campo $GF(2)$

Un campo es un conjunto de elementos que tiene definidas dos operaciones sobre sus elementos, denominadas *suma* y *producto*, las cuáles cumplen ciertas propiedades. Un ejemplo de un campo es el conjunto de los números reales con las operaciones de suma y producto habituales. Un campo definido sobre un conjunto finito es llamado campo finito. Al campo definido sobre el conjunto $\{0, 1\}$, se le conoce como $GF(2)$. El campo binario $GF(2)$ consta de los elementos 0 y 1 , con las operaciones suma y producto módulo 2 (cuadro 2.1).

\oplus	0	1	\wedge	0	1
0	0	1	0	0	0
1	1	0	1	0	1

Cuadro 2.1: Suma y producto (mod 2)

2.3. Matroides binarios

Los *matroides* fueron introducidos por Whitney [86] en 1935. En esta tesis consideraremos algunos conceptos de teoría de matroides. Para un estudio más profundo sugerimos los textos de referencia [53, 87, 97]. Un *matroide*

M es un par ordenado (V, \mathcal{I}) , donde V es un conjunto finito e \mathcal{I} es una familia de subconjuntos de V que cumplen las siguientes condiciones:

- I1** $\emptyset \in \mathcal{I}$.
- I2** Si $I \in \mathcal{I}$ y $I' \subseteq I$, entonces $I' \in \mathcal{I}$.
- I3** Si I_1 y I_2 están en \mathcal{I} y $|I_1| < |I_2|$, entonces $\exists e \in (I_2 - I_1)$ tal que $I_1 \cup e \in \mathcal{I}$.

La condición **I3** se conoce como el *axioma de incremento de independencia*. Se dice que $M = (V, \mathcal{I})$ es un matroide sobre V . Los miembros de \mathcal{I} son los *conjuntos independientes* de M y V es el *conjunto base* de M . Un subconjunto de V que no está en \mathcal{I} es llamado *dependiente* [87]. Una *base* de M , es un conjunto maximal independiente de M .

Lema 2.1. Si B_1 y B_2 son bases de un matroide M , entonces $|B_1| = |B_2|$.

Si \mathcal{B} es una familia de bases del matroide M , entonces por **I1**:

- B1** \mathcal{B} es no vacío.

Lema 2.2. \mathcal{B} satisface la siguiente condición:

- B2** Si $B_1, B_2 \in \mathcal{B}$ y $x \in (B_1 - B_2)$, entonces $\exists y \in (B_2 - B_1)$ tal que $(B_1 - x) \cup y \in \mathcal{B}$.

Teorema 2.2. Sea V un conjunto y \mathcal{B} una colección de subconjuntos de V que satisface **B1** y **B2**. Sea \mathcal{I} la familia de subconjuntos de V que están contenidos en algún miembro de \mathcal{B} . Entonces (V, \mathcal{I}) es un matroide que tiene la familia de bases \mathcal{B} .

Corolario 2.1. Sea B una base del matroide M . Si $e \in V(M) - B$ entonces $B \cup e$ contiene un único circuito $C(e, B)$. Además $e \in C(e, B)$.

Llamamos a $C(e, B)$ el *circuito fundamental* de e con respecto a B .

Definición 2.3.1 (Matroide lineal). Sea \mathbb{F} un campo, $A \in \mathbb{F}^{m \times n}$ una $m \times n$ -matriz sobre \mathbb{F} . Sea $V = \{1, \dots, n\}$ el conjunto de etiquetas de las columnas de A . Un subconjunto $I \subseteq V$ es independiente si las columnas etiquetadas por I son linealmente independientes en el espacio vectorial (m, \mathbb{F}) .

Un matroide binario es un matroide lineal que puede ser representado sobre $GF(2)$. Un matroide binario no se modifica si una fila de su matriz de representación es sumada a otra, si las filas son permutadas, o si una fila de ceros se añade o se elimina. El resultado de permutar las columnas de una

matriz es una nueva matriz que representa a un matroide binario *isomorfo*.

Definición 2.3.2. Sea I la matriz identidad $r \times r$. Una representación estándar de un matroide binario \mathcal{M} de rango r es una matriz de la forma $(I|A)$ [92].

Si A es una matriz con entradas en $\mathbf{GF}(2)$, entonces $\mathcal{M}(IA)$ denota al matroide con representación estándar $(I|A)$. Esencialmente, la representación $(I|A)$ se puede describir de la siguiente manera: los elementos del matroide correspondientes a las columnas de I son una base B y para cada elemento $v \notin B$, el circuito fundamental $C(v, B)$ incluye a v junto con los elementos de B correspondientes a las entradas distintas de cero de la columna v de A [91].

Teorema 2.3 (Traldi 2005, [91]). Sean A_1 y A_2 $r \times \{n - r\}$ -matrices con entradas en $\mathbf{GF}(2)$. Entonces $\mathcal{M}(IA_1) \cong \mathcal{M}(IA_2)$ si y sólo si $(I|A_2)$ puede ser obtenido de $(I|A_1)$ por medio de los siguientes tipos de operaciones sobre matrices de la forma $(I|A)$:

- (a) Permutación de las columnas de A .
- (b) Permutación de las columnas de I y de las filas de $(I|A)$, usando la misma permutación.
- (c) Supongamos que la entrada jk de A es $a_{jk} = 1$. Entonces reemplazamos a_{bc} con $1 + a_{bc}$ siempre que $b \neq j$, $c \neq k$, $a_{jc} = 1$ y $a_{bk} = 1$.

Se dice que la operación de tipo (c) es un *intercambio de base* que involucra a la j -ésima columna de I y a la k -ésima columna de A .

Si H_b es una gráfica simple con lazos permitidos. La *matriz de adyacencia* $A(H_b)$, es una $v \times v$ -matriz sobre $\mathbf{GF}(2)$ 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 sí mismo si v es un vértice con lazo).

Sea $(I|A(H_b)|I + A(H_b))$, donde I es la $v \times v$ -matriz identidad, $A(H_b)$ es la $v \times v$ -matriz de adyacencia de H_b e $I + A(H_b)$ es la $v \times v$ -matriz suma de las matrices anteriores. Sea $IAS(H_b) = (A(H_b)|I + A(H_b))$, denotamos con $\mathcal{M}(IAS(H_b))$ al matroide binario representado por la matriz:

$$IAS(H_b) = (I|A(H_b)|I + A(H_b)) \tag{2}$$

Para todo vértice $v \in V(H_b)$ existen tres columnas de $IAS(H_b)$ asociadas a v : v_ϕ en I , v_χ en $A(H_b)$ y v_ψ en $I + A(H_b)$. El conjunto que contiene a todos los elementos que corresponden a las columnas de $IAS(H_b)$ es el conjunto base del matroide binario $\mathcal{M}(IAS(H_b))$ y se denotará con $W(H_b)$. El conjunto de vértices $\{v_\phi, v_\chi, v_\psi\}$ es la *tripleta de vértices* correspondientes a v . El grupo de permutaciones de los símbolos ϕ , χ y ψ se denotará con S_3 [91].

Supongamos que H_b^1 y H_b^2 son gráficas simples con lazos permitidos, y que existe un isomorfismo compatible $\beta: \mathcal{M}(IAS(H_b^1)) \rightarrow \mathcal{M}(IAS(H_b^2))$. Entonces este isomorfismo consta de dos partes:

- 1 Existe una biyección inducida $V(H_b^1) \rightarrow V(H_b^2)$ (la cuál también se denotará con β). Sin embargo, con respecto a isomorfismo, asumiremos que $V(H_b^1) = V(H_b^2)$.
- 2 Existe una función $f_\beta: V(H_b^1) \rightarrow S_3$ tal que $\beta(v_\iota) = \beta(v)_{f_\beta(v)(\iota)} \forall v \in V(H_b^1) \forall \iota \in \{\phi, \chi, \psi\}$. En este caso se dice que β es un isomorfismo compatible determinado por f_β .

Sea H_b una gráfica simple con lazos permitidos. El matroide binario restringido $\mathcal{M}(IA(H_b))$ se define como el matroide representado por la matriz (3):

$$IA(H_b) = (I|A(H_b)) \quad (3)$$

2.4. Δ -matroides

Definición 2.4.1. Un sistema de conjuntos (sobre V) es un par ordenado $\mathcal{D} = (V, \mathcal{F})$ donde V es un conjunto finito y \mathcal{F} es una familia de subconjuntos de V . Un sistema de conjuntos es llamado propio si $\mathcal{F} \neq \emptyset$.

Axioma 2.1. Axioma de Intercambio Simétrico

Para $F_1, F_2 \in \mathcal{F}$ y $x \in F_1 \Delta F_2$, $\exists y \in F_1 \Delta F_2$, tal que $F_1 \Delta \{x, y\} \in \mathcal{F}$.

Donde Δ es el operador diferencia simétrica entre conjuntos esto es, $X \Delta Y = (X \cup Y) \setminus (X \cap Y)$.

Definición 2.4.2. Un Δ -matroide es un sistema propio de conjuntos $\mathcal{D} = (V, \mathcal{F})$, que satisface el axioma de intercambio simétrico. A los elementos de \mathcal{F} se les llama conjuntos factibles o bases.

Ejemplo 1

Sea \mathcal{D} un Δ -matroide con el conjunto base $v = \{2, 3, 4\}$ y con la familia de conjuntos factibles,

$$\mathcal{F} = \{\{\emptyset\}, \{2\}, \{3\}, \{2, 4\}, \{3, 4\}, \{2, 3, 4\}\}$$

Si $F_1 = \{2, 4\}$ y $F_2 = \{3\}$, entonces $F_1 \Delta F_2 = \{2, 3, 4\}$. Para $x \in F_1 \Delta F_2$ y $y \in F_1 \Delta F_2$:

$$\text{Si } x = 2 \text{ y } y = 2$$

$$\begin{aligned} F_1 \Delta \{2\} &= \{2, 4\} \cup \{2\} \setminus \{2, 4\} \cap \{2\} \\ &= \{4\}, \{4\} \notin \mathcal{F} \end{aligned}$$

$$\text{Si } x = 2 \text{ y } y = 3$$

$$\begin{aligned} F_1 \Delta \{2, 3\} &= \{2, 4\} \cup \{2, 3\} \setminus \{2, 4\} \cap \{2, 3\} \\ &= \{3, 4\}, \{3, 4\} \in \mathcal{F} \end{aligned}$$

Para $X \subseteq V$ se consideran $\mathcal{F} \Delta X = \{F \Delta X : F \in \mathcal{F}\}$ y $\mathcal{D} \Delta X = (V, \mathcal{F} \Delta X)$. La transformación $\mathcal{D} \rightarrow \mathcal{D} \Delta X$ es llamada *torcimiento* (*twisting*). Se dice que \mathcal{D} y $\mathcal{D} \Delta X$ son Δ -equivalentes.

2.4.1. Representación lineal de un Δ -matroide

Una matriz de tipo simétrico es una matriz que es simétrica o antisimétrica. Sea \mathbb{F} un campo y A una $V \times V$ -matriz de tipo simétrico con entradas en \mathbb{F} . Denotamos por $A[X]$ a la submatriz de A indexada por $X \subseteq A$. Si $\mathcal{F} = \{X \subseteq V : A[X] \text{ es no singular}\}$, entonces $\mathcal{D} = (V, \mathcal{F})$ es un Δ -matroide. Las matrices $A[X]$ son llamadas *submatrices principales*. Por convención, se asume que $A[X]$ es no singular si $A[X] = \emptyset$ [23, 84]. Se dice que un Δ -matroide es *normal* si cumple que $\emptyset \in \mathcal{F}$. Un Δ -matroide \mathcal{D} tal que para todo $F_1, F_2 \in \mathcal{F}$, $|F_1|$ y $|F_2|$ tienen la misma paridad se dice que es *par*.

Definición 2.4.3. La matriz A es una representación lineal de un Δ -matroide \mathcal{D} con respecto a una base X si,

$$\mathcal{D} = \mathcal{D}(A) \Delta X \tag{4}$$

Un Δ -matroide *binario* es un Δ -matroide equivalente a $\mathcal{D} = (V, \mathcal{F}(A))$ para alguna matriz simétrica (antisimétrica) A sobre $GF(2)$. La representación lineal de un Δ -matroide binario \mathcal{D} se define por medio de la matriz de adyacencia de una gráfica fundamental de \mathcal{D} [6].

Ejemplo 2

Sea A la matriz simétrica con entradas en $\mathbf{GF}(2)$ en (5). \mathcal{F} en (6) es el conjunto de etiquetas correspondientes a las submatrices principales invertibles de A .

$$A = \begin{matrix} & \begin{matrix} 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{matrix} \quad \mathcal{F} = \{\emptyset, 2, 3, 24, 34, 234\} \quad (6)$$

(5)

2.4.2. Gráfica fundamental

Sea $\mathcal{D} = (V, \mathcal{F})$ un Δ -matroide y sea X una base de \mathcal{D} . Se define la gráfica simple $H = H_{\mathcal{D}}(X)$, con $E(H) = \{vw : X \Delta \{v, w\} \in \mathcal{F}\}$. A $H_{\mathcal{D}}(X)$ se la llama *gráfica fundamental* del Δ -matroide \mathcal{D} con respecto a X .

2.4.3. Transformación de pivote principal

Sea \mathcal{D} un Δ -matroide representado por una matriz de tipo simétrico A , con respecto a la base X y sea vw una arista de $H = H_{\mathcal{D}}(X)$. Entonces, el conjunto $X' = X \Delta \{v, w\}$ es una base de \mathcal{D} y $\alpha = [\{v, w\}]$ es una matriz no singular. En 1987 Bouchet demostró que al hacer la descomposición por bloques en (7), se obtiene una matriz de tipo simétrico A' en (8), además $\mathcal{D}(A') = \mathcal{D}(A) \Delta \{v, w\}$.

$$A = \left(\begin{array}{c|c} \alpha & \beta \\ \gamma & \delta \end{array} \right) \quad (7) \quad A' = \left(\begin{array}{c|c} \alpha^{-1} & -\alpha^{-1}\beta \\ \gamma\alpha^{-1} & \delta - \gamma\alpha^{-1}\beta \end{array} \right) \quad (8)$$

La igualdad (4) implica que $\mathcal{D} = \mathcal{D}(A') \Delta X'$, por lo tanto A' es una representación de \mathcal{D} con respecto a X [7]. La transformación $A \mapsto A'$ es llamada *transformación de pivote principal* o simplemente *pivote*.

2.4.4. Menores

Sea $\mathcal{D} = (V, \mathcal{F})$ un Δ -matroide. Para $x \in V$ se definen dos Δ -matroides con conjunto base $V \setminus x$ y cuyas familias de conjuntos factibles están dadas por:

$$\begin{aligned} \mathcal{F} \setminus x &= \{F : F \subseteq V \setminus x, F \in \mathcal{F}\} \\ \mathcal{F} \circ x &= \{F : F \subseteq V \setminus x, F \cup \{x\} \in \mathcal{F}\} \end{aligned}$$

Se conoce a $\mathcal{D} \setminus x = (V \setminus x, \mathcal{F} \setminus x)$ como un menor elemental de D , obtenido por el **borrado** del elemento x y a $D \circ x = (V \setminus x, \mathcal{F} \circ x)$ como un menor elemental de D , obtenido por la *contracción* del elemento x [84].

Propiedad 1 (Bouchet 1991, [13]). *Para cualquier Δ -matroide $\mathcal{D} = (V, \mathcal{F})$, $x \in V$ y $F \subseteq V$:*

$$(D \Delta F) \setminus x = (D \circ x) \Delta (F - x) \quad \text{si } x \in F$$

Teorema 2.4. Sea $\mathcal{D} = (V, \mathcal{F})$ un Δ -matroide con una matriz de representación A sobre \mathcal{F} . Cualquier menor elemental D_μ de D tiene una matriz de representación A_μ sobre \mathcal{F} . Además A_μ , es antisimétrica (simétrica) si A es antisimétrica (simétrica).

2.4.5. Conexidad de un Δ -matroide

Para un par de Δ -matroides $\mathcal{D}_1 = (V_1, \mathcal{F}_1)$ y $\mathcal{D}_2 = (V_2, \mathcal{F}_2)$ con $V_1 \cap V_2 = \emptyset$, la *suma directa* $\mathcal{D}_1 \oplus \mathcal{D}_2$ designa al Δ -matroide (V, \mathcal{F}) definido por $V = V_1 \cup V_2$ y $\mathcal{F} = \{F_1 \cup F_2 : F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2\}$ [42, 51]. En este caso se dice que \mathcal{D}_1 y \mathcal{D}_2 son *sumandos* de \mathcal{D} . Se dice que un Δ -matroide es *conexo* si no es la suma directa de dos Δ -matroides con conjunto base no vacío.

2.5. Conceptos relacionados con el estudio de Δ -matroides

2.5.1. Alfabetos complementarios y palabras de doble ocurrencia

Una *permutación* es una función biyectiva que asocia a $i \in \{1, \dots, n\}$ un elemento $\pi_i \in \{1, \dots, n\}$. Sea $\beta : X \mapsto X$ una permutación sobre un conjunto finito X . β es una *involución* si $\beta(\beta(x)) = x, \forall x \in X$.

Sea λ un alfabeto, esto es, un conjunto finito de símbolos. Las secuencias a_1, a_2, \dots, a_n con $a_i \in \lambda$ son palabras sobre λ . El conjunto de todas las palabras sobre λ , incluyendo a la palabra vacía, se denota con λ^* .

Ejemplo 3

Tomemos, como ejemplo $\lambda = \{a, b, c\}$. Si λ_i es el conjunto de palabras formadas por i símbolos.

$$\begin{aligned} \lambda^0 &= \{\emptyset\} \\ \lambda^1 &= \{a, b, c\} \\ \lambda^2 &= \{ab, ac, bc, ba, ca, cb\} \\ \lambda^* &= \bigcup_{j=0}^{\infty} \lambda^j \end{aligned}$$

Denotamos como $L(a)$, al conjunto de letras en una palabra a sobre un alfabeto λ . La palabra a , es una palabra de doble ocurrencia si cada $x \in L(a)$ ocurre exactamente dos veces en a . Un alfabeto λ junto con una involución $\beta : \lambda \mapsto \lambda$ es llamado *alfabeto complementario* y una palabra sobre λ es una *palabra de doble ocurrencia signada* [45]. Sea (λ, β) un alfabeto

complementario, $\beta(a)$ se denotará como $\bar{a} \forall a \in \lambda$. Una *permutación signada* es una palabra sobre λ donde cada símbolo de λ aparece exactamente una vez.

2.5.2. Gráfica de alternancia

Sea m una palabra de doble ocurrencia sobre el alfabeto finito ν . Un *par no ordenado* νw de elementos distintos de ν es una *alternancia* de m , si al leer m encontramos alternativamente $\nu \dots w \dots \nu \dots w$ o $w \dots \nu \dots w \dots \nu$ [10].

La *gráfica de alternancia* de m es la gráfica simple H_m definida sobre el conjunto de vértices ν , tal que existe una arista de ν a w si νw es una alternancia de m . [10]. Otra manera de definir la gráfica H_m es considerando un diagrama D que consta de una circunferencia con un conjunto finito de cuerdas etiquetadas con los elementos de ν . La gráfica H_m es la gráfica de intersección del diagrama D , es decir H_m tiene como vértices a los elementos de ν y existe una arista del vértice ν al vértice w si las correspondientes cuerdas se intersecan en el diagrama D . En este contexto a D se le conoce como el *diagrama cordal* y a H_m como la *gráfica circular* [23]. En el diagrama D , se puede asumir sin pérdida de generalidad que no existen dos cuerdas que compartan un extremo en común. Se puede recorrer el círculo y construir la palabra m si se consideran las etiquetas de las intersecciones de las cuerdas con la circunferencia, recorriendo la circunferencia a partir de la intersección de una cuerda con la circunferencia tomando una dirección fija sobre la circunferencia dando una vuelta completa hasta llegar al punto inicial.

Ejemplo 4

La figura 2.1 muestra el diagrama cordal y la gráfica de alternancia y el diagrama cordal de la palabra de doble ocurrencia $m = 1321545234$.

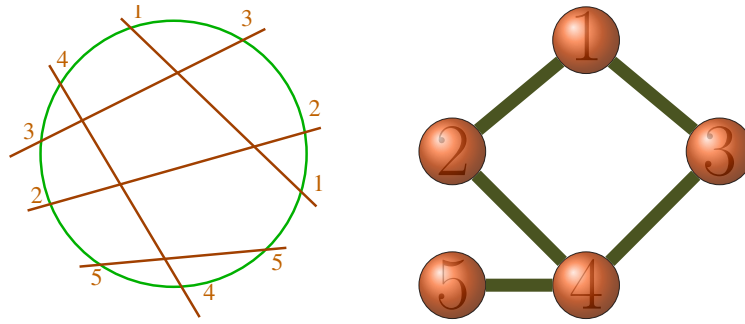


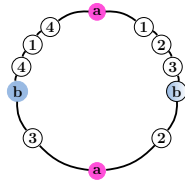
Figura 2.1: Diagrama cordal y gráfica de alternancia de una palabra de doble ocurrencia.

Definición 2.5.1. *Transposición [2]*

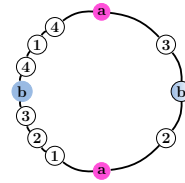
Sea H el diagrama con la palabra m colocada sobre la circunferencia, como en la figura ?? sin las cuerdas y sean v y w un par de letras que forman una alternancia en m . Una transposición sobre vw es la gráfica resultante de intercambiar una de las secuencias de letras de v a w con la otra secuencia de v a w .

Ejemplo 5

Consideremos la palabra $m = a123b2a3b414$.



(a) Diagrama cordal de m .



(b) Transposición sobre ab .

Figura 2.2: Ejemplo de transposición sobre diagrama cordal.

2.5.3. Gráfica de alternancia signada

Considérese un alfabeto de la forma $(V, \bar{\cdot})$ y una palabra de doble ocurrencia $a \in V$. La gráfica de alternancia de a , es la gráfica simple $H_a = (V, E, v)$ consistente en:

- * Un conjunto finito V junto con una *involución* $v \mapsto \bar{v}$, tal que para $\forall v \in V$, $v \neq \bar{v}$ y si $v \in V$ entonces $\bar{\bar{v}} \in V$.
- * El conjunto finito de aristas E definido en la subsección 2.5.2.

Para cada $v \in V$, \bar{v} se representará con un lazo sobre v .

2.5.4. Complementación de lazo

Sea H_b^1 una gráfica simple con lazos permitidos y sea H_b^2 la gráfica resultante de H_b^1 por el complemento del estatus de lazo de un vértice $v \in H_b^1$. Entonces $IAS(H_b^2)$ (ver sección 2.3) es la matriz obtenida de $IAS(H_b^1)$ por el intercambio de las columnas v_χ y v_ψ . Este intercambio es un ejemplo de la operación tipo (a) del teorema 2.3, de acuerdo con el cuál existe un isomorfismo compatible $\mathcal{M}(IAS(H_b^1)) \mapsto \mathcal{M}(IAS(H_b^2))$ determinado por la función $f : \mathcal{V}(H_b^1) \mapsto S_3$ dada en (9):

$$f(w) = \begin{cases} \text{La transposición } \chi\phi, & \text{si } w = v \\ 1 & \text{si } w \neq v \end{cases} \quad (9)$$

Teorema 2.5 (Traldi 2015, [91]). *Sean H_b^1 y H_b^2 gráficas simples con lazos permitidos, y supongamos $v \in V(H_b^1)$. Entonces las siguientes condiciones son equivalentes:*

- (1) *Con respecto a isomorfismo, H_b^2 es la gráfica obtenida de H_b^1 al complementar el estatus de lazo de v .*
- (2) *Existe un isomorfismo compatible $\beta : \mathcal{M}(IAS(H_b^1)) \mapsto \mathcal{M}(IAS(H_b^2))$ tal que $f\beta(v) = (\chi\psi)$ y $f\beta(w) = 1$, $\forall w \neq v$.*

2.5.5. Complementación local

Existen dos distintas versiones de la complementación local: la *complementación local simple* y la *complementación local no simple*. La primera fue introducida por Bouchet [24] y por Rosenstiehl y Read [85], como parte de la teoría de alternancia en multigráficas 4-regulares. La complementación local simple no contempla la creación de lazos. La complementación local no simple es parte de la teoría de Transformaciones de Pivote Principal sobre $GF(2)$ (ver subsección 2.4.3).

Sea H_b una gráfica simple con lazos permitidos. Para un vértice con lazo $v \in H_b$:

- (1) El complemento local simple de H_b en v es la complementación del estatus de adyacencia de cada par de vecinos de v .
- (2) El complemento local no simple de H_b con respecto a v , $A(H_b^{ns} * v)$, es la complementación del estatus de adyacencia de todo par de vecinos de v y la complementación del estatus de lazo de cada vecino de v [27, 91].
- (4) $H_b^{inv} * v$ denota a la gráfica obtenida de $H_b^{ns} * v$ al eliminar todas las aristas incidentes con v (incluyendo al lazo en v). De forma equivalente, $H_b^{inv} * v$ es la complementación local no simple de la subgráfica inducida por la vecindad cerrada de v , $\{w : vw \in E(H_b)\}$.

Sea $\mathcal{M}(IAS(H_b))$ el matroide binario representado por la matriz $IAS(H_b)$ (ver sección 2.3). Reemplazar $A(H_b)$ con $A(H_b^{ns} * v)$ tiene exactamente el mismo efecto sobre $IAS(H_b)$ que la operación del tipo (c) del teorema 2.3. Esta

operación es equivalente a un cambio de base haciendo un intercambio entre v_ϕ y v_χ (si v es un vértice con lazo) o entre v_ϕ y v_ψ (si v es un vértice sin lazo) [91].

Capítulo 3

Modelo genético

3.1. Construcción del modelo

3.1.1. Estructura del ADN

Existen dos tipos de ácidos nucleicos: el *desoxirribonucleico*, ADN, y el *ribonucleico*, ARN. Los *ácidos nucleicos* son las moléculas responsables del *almacenamiento*, *expresión*, y *replicación* de la información genética. Una cadena de ácido nucleico es un polímero de nucleótidos. Un nucleótido está formado por tres componentes: una *pentosa*, (2'-desoxirribosa en el ADN, ribosa en el ARN), un *grupo fosfato* (el cuál une a dos pentosas sucesivas mediante enlaces fosfodiéster entre el grupo 5'-hidroxilo de la primera y el grupo 3'-hidroxilo de la siguiente) y una *base nitrogenada* (púrica o pirimidínica) (figura 3.1).

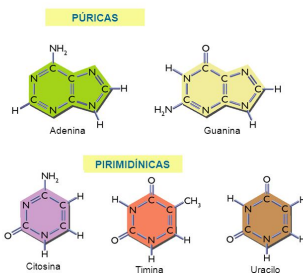


Figura 3.1: Bases nitrogenadas.
Imagen tomada de [102].

Las *bases púricas* que componen los ácidos nucleicos son: Adenina (A) y Guanina (G). Las *bases pirimidínicas* son Citosina (C), Timina (T) y Úracilo (U). La Citosina está presente en ambos ácidos nucleicos, La Timina únicamente se encuentra en el ADN y el Úracilo en el ARN.

La molécula de ADN consta de dos cadenas antiparalelas enrolladas alrededor de un mismo eje formando una doble hélice unida por *enlaces de hidrógeno entre pares de bases púricas y pirimidínicas complementarias* y por interacciones hidrofóbicas. En los patrones de enlaces de hidrógeno más frecuentes (definidos por Watson y Crick [74]) A se une con T y G se une con C (figura 3.2).

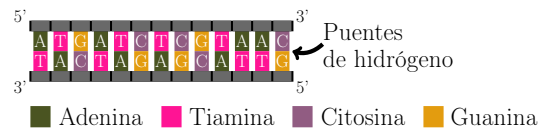


Figura 3.2: Enlaces de hidrógeno entre pares de bases complementarias en el ADN.

Durante la replicación del ADN una enzima rompe los enlaces de hidrógeno que unen a las cadenas sencillas de ADN y cada una de ellas puede actuar como modelo para la reproducción de la cadena complementaria. La polaridad de una hebra es positiva si una versión del *transcrito* (ARN) de la misma secuencia es traducida a proteína (figura 3.3).

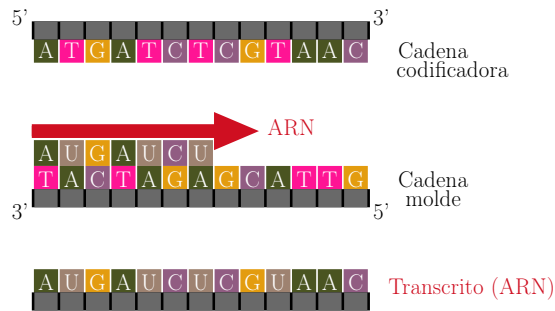


Figura 3.3: Transcripción.

El código genético se guarda en las secuencias de bases de cada cadena sencilla de la doble hélice. Estas secuencias se representan como palabras sobre el alfabeto de nucleótidos $\mathcal{N} = \{A, C, G, T\}$. El alfabeto \mathcal{N} y la involución $\beta : \mathcal{N} \rightarrow \mathcal{N}$ forman un *alfabeto complementario* (subsección 2.5.1). La función β modela la relación complementaria **A – T**, **C – G** de Watson y Crick [45, 66, 67].

$$\beta(A) = T, \quad \beta(T) = A, \quad \text{y} \quad \beta(C) = G, \quad \beta(G) = C$$

Sea \mathcal{N}^i el conjunto de palabras con i símbolos. Los *codones* son los elementos del conjunto \mathcal{N}^3 y contienen la codificación para aminoácidos específicos (ejemplo 6).

Ejemplo 6



Figura 3.4: Codones.

El codón ACG codifica para el aminoácido serina y el codón ACC codifica para el aminoácido treonina (figura 3.4).

Un *gen* es un fragmento de cadena sencilla de ADN que codifica para una proteína (en algunos casos para una molécula de ARN), identificado por un *codón de inicio* y un *codón de término* (figura 3.5). La orientación de un gen depende de la polaridad de la cadena sencilla a la que pertenece y de su posición dentro de la misma.

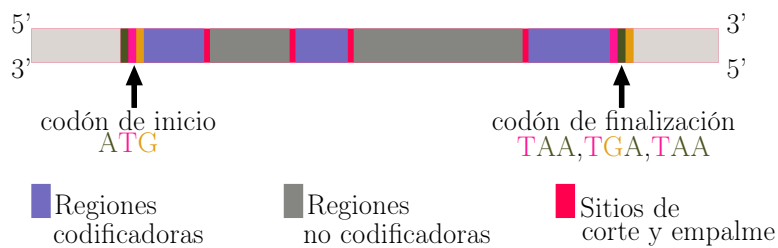


Figura 3.5: Componentes estructurales del gen.

El ARN mensajero (*mARN*) codifica la secuencia de aminoácidos especificada por un gen. Para que esta codificación se realice es necesario eliminar las regiones que no especifican aminoácidos y ensamblar las regiones que codifican para proteína. El proceso descrito es conocido como *corte y empalme* (*splicing*) del mARN (figura 3.6).

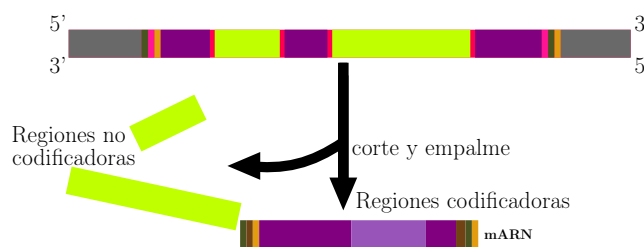


Figura 3.6: Corte y empalme del ARN mensajero.

El genoma eucariota es un conjunto de cromosomas lineales (figura 3.7). Dado un conjunto de genes, un *cromosoma* se define como un conjunto de adyacencias formado por pares de extremos de genes, donde cada gen está contenido en exactamente una adyacencia.

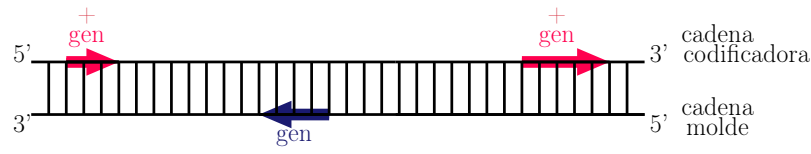


Figura 3.7: Cromosoma lineal.

Los *telómeros* son "cubiertas" de ADN especializado que protegen al cromosoma lineal del deterioro causado por la replicación. Los telómeros son regiones compuestas por repeticiones de la misma secuencia corta de ADN, la cuál varía entre organismos, pero en seres humanos y otros mamíferos es $5' - TTAGGG - 3'$.

3.1.2. Problema del reordenamiento cromosómico

Sean Π y Γ los genomas de dos especies con un genoma ancestral común γ . Un *bloque de sintenia* es un segmento de información genética conservado de γ en Π y Γ (ejemplo 7) [76, 78, 81].

Ejemplo 7

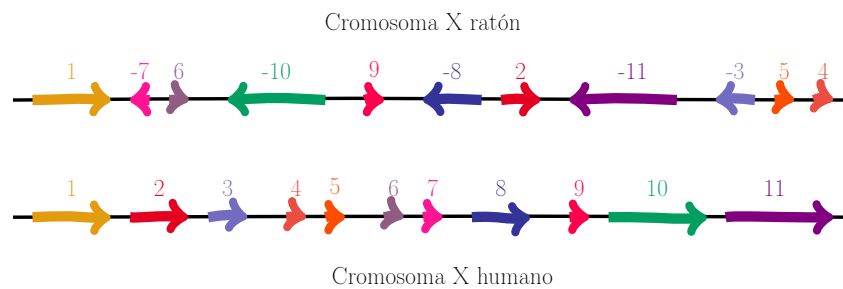


Figura 3.8: Bloques de sintenia del cromosoma X en ratones y en humanos.

Existe un conjunto de eventos evolutivos conocidos como *reordenamientos cromosómicos* (*chromosomal rearrangements*) que pueden convertir a Π en γ y luego en Γ [72, 79]. Los reordenamientos cromosómicos tienen lugar por un mecanismo conocido como *recombinación homóloga*, el cuál ocurre entre secuencias de ADN con segmentos idénticos de pares de bases [77].

- (1) Los segmentos idénticos se alinean marcando los sitios de corte, llamados *puntos de rompimiento*, en el cromosoma.
- (2) Una enzima corta la estructura de doble hélice en los puntos de rompimiento.
- (3) El segmento separado se reordena.
- (4) Una enzima repara la ruptura.

Los reordenamientos cromosómicos preponderantes son la *inversión*, la *transposición* y la *transposición invertida* [72]. La presente tesis se enfoca en la inversión, la cuál produce una rotación de 180 grados en una región del cromosoma (ejemplo 8).

Ejemplo 8



Figura 3.9: Ejemplo del reordenamiento de inversión en un segmento de ADN.

Sean π y γ un par de cromosomas con el mismo número de genes en los genomas Π y Γ . La secuencia de reordenamientos cromosómicos que transforma a π en γ se conoce como *escenario evolutivo*. Un escenario evolutivo puede constar de uno o más tipos de reordenamientos (ejemplos 9 y 10).

Ejemplo 9

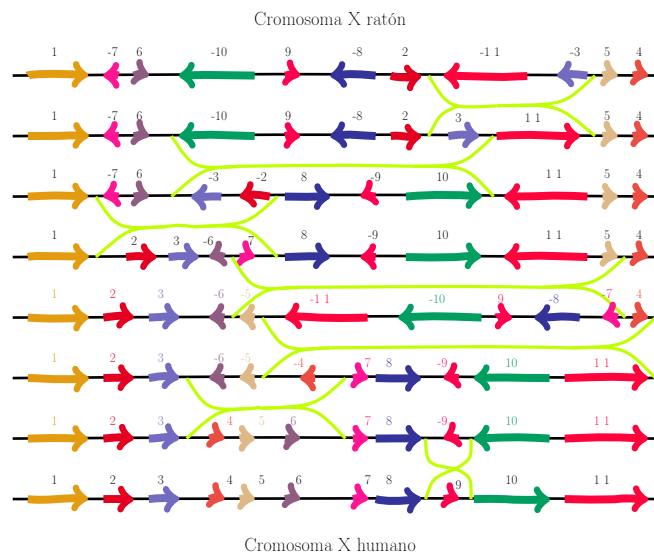


Figura 3.10: Escenario evolutivo entre el cromosoma X en ratones y en humanos.

La figura 3.10 muestra la secuencia de inversiones (marcadas en verde) que transforman el cromosoma X de ratones en el

cromosoma X humano.

Ejemplo 10

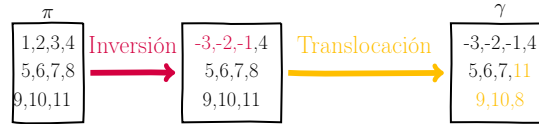


Figura 3.11: Escenario evolutivo con una inversión y una transposición.

La longitud de un escenario evolutivo se correlaciona con la distancia evolutiva entre dos especies. El *problema general del reordenamiento cromosómico* consiste en la determinación de escenarios evolutivos de longitud mínima (parsimonia máxima).

Una *permutación signada* es una permutación donde cada elemento tiene asociado un signo (+, -). Las permutaciones signadas modelan el orden relativo en los bloques de sintenia, ya que reflejan la estructura de doble hélice del ADN [80]. Sean $\pi = (\pi_1, \dots, \pi_n)$ y $\gamma = (\gamma_1, \dots, \gamma_n)$ las permutaciones signadas que representan el orden y la orientación de los genes en los cromosomas π y γ . Al reasignar la etiqueta de cada gen $i \in \pi$ podemos asumir que $\gamma = \iota = (1, 2, \dots, n)$, donde ι denota a la *permutación identidad*. Así el problema general del reordenamiento cromosómico es equivalente a ordenar π con el número mínimo de reordenamientos permitidos.

3.1.3. Ensamblamiento genético en ciliados

Los ciliados (*ciliophora*) son organismos eucariotas unicelulares que ordenan su material genético mediante reordenamientos genómicos a gran escala [41, 46], lo que los convierte en organismos modelo para el análisis de la evolución molecular [61, 63]. Los ciliados se caracterizan por la presencia de dos núcleos, un *macronúcleo* (MAC) y un *micronúcleo* (MIC). El MAC contiene al genoma transcripcionalmente activo y el genoma transcripcionalmente silente está contenido en el MIC.

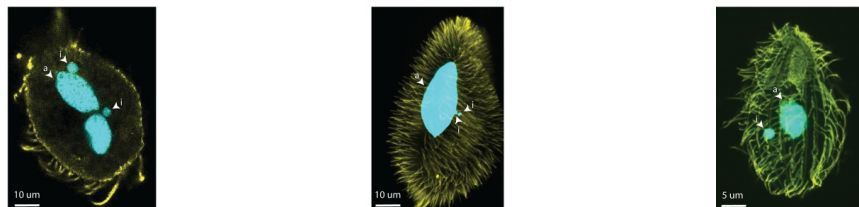


Figura 3.12: Oxytricha trifallax, Paramecium tetraurelia, Tetrahymena thermophila. Imagen tomada de [103].

El micronúcleo está fragmentado en alrededor de 200,000¹ segmentos codificadores conocidos como *secuencias destinadas al macronúcleo* (*macromolecular destined sequences, MDSs*). Las *secuencias internamente eliminadas* (*internally eliminated sequences, IESs*) son regiones no codificadoras que interrumpen a los MDSs. El macronúcleo se constituye de nanocromosomas del tamaño de un gen. El *ensamblamiento genético* consiste en el ordenamiento de los segmentos codificadores mediante cortes y empalmes lo que produce genes transcripcionalmente competentes que se expresan en un nuevo MAC (figura 3.13).

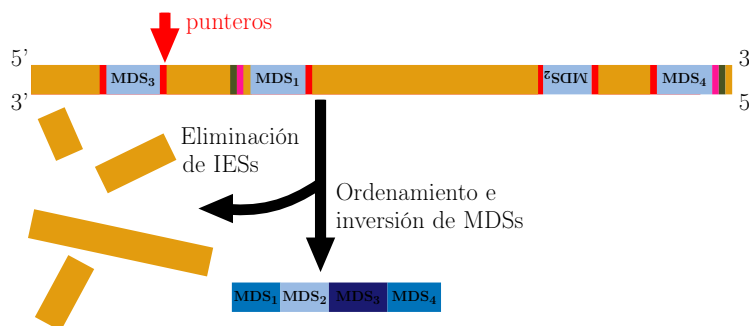


Figura 3.13: Ensamblamiento genético.

Las secuencias destinadas al macronúcleo pueden ser elementales o compuestas. Un *MDS elemental* es un segmento codificador presente en la forma MIC. Los *MDSs compuestos* se producen durante el ensamblamiento genético como resultado de cortes y empalmes de la molécula de ADN. Este proceso de rupturas y pegados se lleva a cabo por recombinación homóloga entre pares de secuencias repetidas llamadas *punteros*. Los punteros son secuencias cortas de nucleótidos localizadas en los extremos de los MDSs elementales (exceptuando a los que se ubican en los extremos del gen). La secuencia en el borde izquierdo se conoce como *puntero de entrada* y la secuencia en el borde derecho es llamada *puntero de salida* (ejemplo 11).

¹Dependiendo de la especie [41].

Ejemplo 11

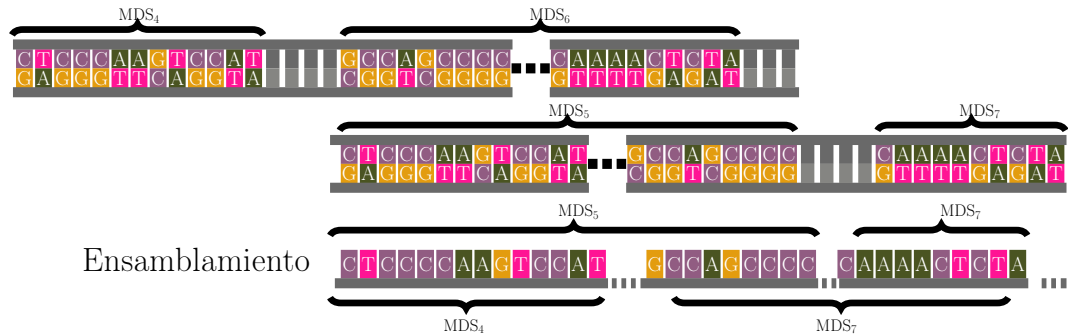


Figura 3.14: Ejemplo de ensamblamiento genético en el gen actina I del ciliado *Sterkiella nova*.

La figura 3.14 muestra los punteros en los MDSs 4, 6, 5 y 7 del gen actina I. Notar los emparejamientos entre el puntero de salida del MDS_4 y el puntero de entrada del MDS_5 , entre el puntero de salida del MDS_5 y el puntero de entrada del MDS_6 y entre el puntero de salida del MDS_6 y el puntero de entrada del MDS_7 .

3.1.4. Operaciones de ADN en ciliados

De acuerdo con el *modelo intramolecular de ensamblamiento genético* postulado en [46, 82, 83], los ciliados ordenan su ADN mediante tres operaciones *irreversibles* conocidas como: *escisión de lazo (loop excision)*, *recombinación de doble lazo (double loop recombination)* y *recombinación de horquilla (hairpin recombination)*.

Escisión de lazo

La escisión de lazo, elimina un segmento no codificador ubicado entre un par de punteros consecutivos con la misma orientación.

Ejemplo 12



Figura 3.15: Segmento del gen β TTP en el ciliado *Sterkiella histriomuscorum*.

La figura 3.15 muestra una región del gen β TTP con el puntero de salida del MDS_2 y el puntero de entrada del MDS_3 en el

Sterkiella histriomuscorum. La molécula de ADN se pliega alineando el par de punteros (figura 3.16) y se lleva a cabo la recombinación homóloga (figura 3.17). Como resultado el *IES* es eliminado como una molécula circular y los fragmentos *MDS*₂ y *MDS*₃ quedan unidos (figura 3.18).

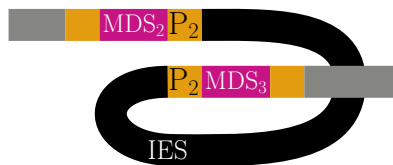


Figura 3.16: Plegamiento molecular en la escisión de lazo.

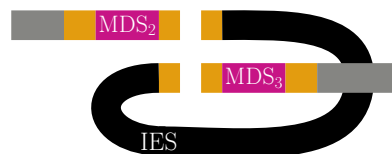


Figura 3.17: Recombinación homóloga.

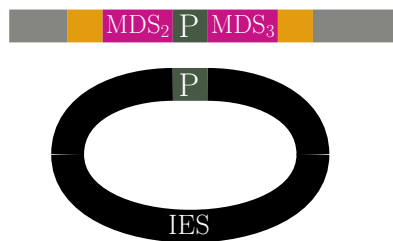


Figura 3.18: Eliminación de ADN no genético.

Recombinación de doble lazo

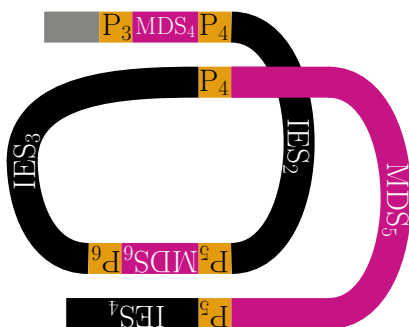


Figura 3.19: Recombinación de doble lazo.

La *recombinación de doble lazo*, (*double loop recombination*), es aplicable a una molécula de ADN con dos pares de punteros que delimitan regiones superpuestas. La molécula se pliega en dos lazos alineando los dos pares de punteros y este doble plegamiento intercambia las posiciones de las regiones trasladadas (figura 3.19).

Recombinación de horquilla

La *recombinación de horquilla* (*hairpin recombination*) invierte el orden y la orientación de una región del gen y ocurre entre pares de repeticiones con orientación opuesta (ejemplo 13).

Ejemplo 13

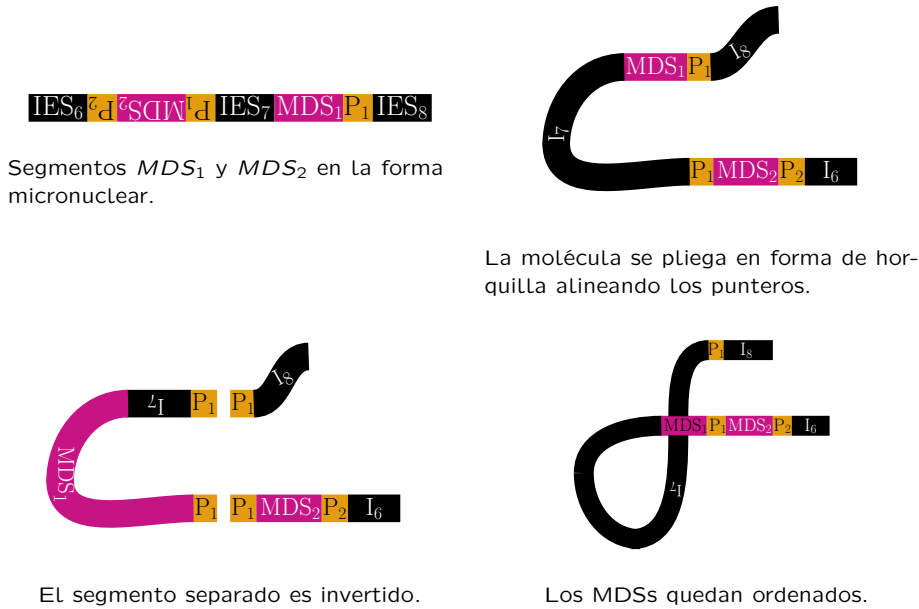


Figura 3.20: Ejemplo de recombinación de horquilla en el gen Actina I del ciliado *Sterkiella Nova*.

La recombinación de horquilla y el reordenamiento de inversión se pueden ver como operaciones análogas que actúan en distintos niveles moleculares [61, 63]. Sin embargo es importante recordar que la recombinación de horquilla es un evento irreversible.

3.1.5. Secuencias MDS

Un gen compuesto por $k \geq 1$ segmentos codificadores es representable como una palabra sobre el *alfabeto de MDSs* $\mu_k = \{M_{i,j} : 1 \leq i \leq j \leq k\}$. Donde el símbolo $M_{i,i}$ (o simplemente M_i) abstrae a un MDS elemental y los símbolos $M_{i,j}$ para $j > i$ abstraen MDSs compuestos. Las palabras sobre μ_k son llamadas *secuencias MDS* y se denotarán con la letra ϑ . Toda $\vartheta \in \mu_k^*$ puede ser modelada por una *permutación signada* ϖ .

Ejemplo 14

Consideremos la secuencia $\vartheta = M_3 M_4 M_6 M_5 M_7 M_9 \overline{M_2} M_1 M_8$ correspondiente al gen Actina I del ciliado *Sterkiella Nova*. La inversión del segmento MDS_2 se indica con $\overline{M_2}$. ϑ es representada por la permutación $\varpi = (3, 4, 6, 5, 7, 9, -2, 1, 8)$.

Sea $\pi = (\pi_1, \dots, \pi_n)$ un cromosoma con n genes y sea $\varpi = (\varpi(i)_1, \dots, \varpi(i)_{k_i})$. k_i el número de segmentos codificadores en el i ésimo gen de π . El ensamblamiento genético definido en la subsección 3.1.3 es la transformación $\varpi \rightarrow \iota$.

3.2. Permutaciones. Inversiones orientadas y no orientadas

3.2.1. Distancia de inversión

Sea $\pi = (\pi_1, \dots, \pi_n)$ un cromosoma $1 \leq i \leq j \leq n$. Una *inversión* $\rho(i, j)$ reordena los genes dentro del cromosoma $\pi = (\pi_1, \dots, \pi_{i-1}, \pi_i, \dots, \pi_j, \pi_{j+1}, \dots, \pi_n)$ y transforma a π en el cromosoma $(\pi_1, \dots, \pi_{i-1}, -\pi_j, \dots, -\pi_i, \pi_{j+1}, \dots, \pi_n)$. El *problema de la distancia de inversión* consiste en determinar la secuencia de inversiones $\rho_1, \dots, \rho_{d_r}$ de longitud mínima que que convierta a π en ι . El valor óptimo de la solución d_r se conoce como la de la distancia de inversión de π y se denota como $d_r(\pi, \iota)$ [40, 59].

3.2.2. Representación de un cromosoma lineal

En teoría de inversiones, los telómeros de un cromosoma lineal son representados por los elementos 0 y $n+1$, tal que $\pi = (0, \pi_1, \dots, \pi_n, n+1)$. Se dice que π es una permutación signada y *enmarcada*.

3.2.3. Inversiones orientadas

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

$$\text{Si } \pi_i + \pi_j = +1 \text{ entonces } \rho(i, j-1) \quad (10)$$

$$\text{Si } \pi_i + \pi_j = -1 \text{ entonces } \rho(i+1, j) \quad (11)$$

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 [5, 80].

Ejemplo 15

Sea $\pi_{\text{ratón}} = (0, 1, -7, 6, -10, 9, -8, 2, -11, -3, 5, 4, 12)$ la permutación que describe el orden relativo de los genes en el cromosoma X del ratón. Inicialmente, los pares orientados en $\pi_{\text{ratón}}$ son $(-7, 6)$, $(-10, 9)$, $(9, -8)$, $(2, -3)$ y $(-3, 4)$. El par $(2, -3)$ induce una inversión del tipo (11).

$$\pi_{\text{ratón}_{\rho(2,-3)}} = (0, 1, -7, 6, -10, 9, -8, 2, 3, 11, 5, 4, 12)$$

La cuál produce un nuevo par orientado $(-10, 11)$, las inversiones orientadas no necesariamente crean nuevos pares orientados como se verificará más adelante.

3.2.4. Obstáculos

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 [5, 96].

3.2.5. Inversiones no orientadas

Los obstáculos en π se eliminan con dos tipos de inversión no orientada: la *inversión de corte* (*cut reversal*) y la *inversión de fusión* (*merge reversal*). Ambas operaciones son inducidas por pares no orientados en π . La *inversión de corte* invierte el segmento entre i e $i+1$ de un obstáculo $i, \pi_{j+1}, \pi_{j+2}, \dots, i+1, \dots, \pi_{j+k-1}, i+k$.

Ejemplo 16

Sea $\pi = (0, 3, 4, 6, 5, 7, 9, -2, 1, 8, 10)$ la representación como cromosoma lineal del gen MIC en el ejemplo (14). El único obstáculo de π se marca con azul en (12).

$$\pi = (0, 3, 4, 6, 5, 7, 9, -2, 1, 8, 10) \tag{12}$$

Al invertir el segmento entre $i = 4$ e $i + 1 = 5$ se obtiene la secuencia en (13) ,

$$\pi = (0, 3, 4, -6, 5, 7, 9, -2, 1, 8, 10) \tag{13}$$

La *inversión de fusión* invierte el intervalo cuyos extremos $i+k$ e i' , son los puntos finales de dos obstáculos diferentes $i, \dots, i+k, \dots, i', \dots, i'+k'$.

Ejemplo 17

Consideremos la permutación $\pi_1 = (0, 2, 4, 6, 5, 7, 9, 8, 10, 3, 11, 1, 12)$. La fusión de los obstáculos $7, 9, 8, 10$ y $11, 1, 12, 0, 2$ se muestra en (14).

$$\pi_{1\rho(10,11)} = (0, 2, 4, 6, 5, 7, 9, 8, -11, -3, -10, 1, 12) \quad (14)$$

π_1 contiene además el obstáculo 4,6,5,7 el cuál se elimina con la inversión de corte ilustrada en el ejemplo 16.

3.3. Gráfica de punto de rompimiento

Los puntos de rompimiento descritos en la subsección 3.1.2 permiten comparar la posición de los bloques de sintenia (genes) entre pares de cromosomas. El primer paso para identificar tales puntos es aplicar la transformación $T_1: \pi \mapsto \pi'$, la cuál convierte a la permutación signada π de orden n en la permutación no signada π' en (15) conocida como la *imagen* de π [3].

$$\pi' = (\pi'_0, \dots, \pi'_{2n-1}) \quad (15)$$

Cada gen $i \in \pi$ es representado en π' por sus extremos $5'$ y $3'$ (ver subsección 3.1.1) de acuerdo a las asignaciones en (16) y (17):

$$i \rightarrow (2i - 1, 2i) \quad \text{si } i > 0 \quad (16)$$

$$i \rightarrow (2i, 2i - 1) \quad \text{si } i < 0 \quad (17)$$

La permutación no signada π' modela la estructura de doble hélice de la molécula de ADN ya que los elementos $2i$ y $2i - 1$ corresponden a los extremos $5'$ y $3'$ del gen. Los elementos 0 y $n + 1$ en π asociados con los telómeros del cromosoma lineal (ver subsección 3.2.2) son representados por $\pi'_0 = 0$ y $\pi'_{2n-1} = 2n - 1$.

Ejemplo 18

$$\pi' = (0, 5, 6, 7, 8, 11, 12, 9, 10, 13, 14, 17, 18, 4, 3, 1, 2, 15, 16, 19)$$

$\pi'_0 \pi'_1 \pi'_2 \pi'_3 \pi'_4 \pi'_5 \pi'_6 \pi'_7 \pi'_8 \pi'_9 \pi'_{10} \pi'_{11} \pi'_{12} \pi'_{13} \pi'_{14} \pi'_{15} \pi'_{16} \pi'_{17} \pi'_{18} \pi'_{19}$

Imagen de la permutación signada π del ejemplo 16.

La *gráfica de punto de rompimiento* de π' es una gráfica bicoloreada, $BG(\pi') = (V, E_R, E_A)$, donde el conjunto V está compuesto por los elementos de π' . El conjunto de aristas rojas, E_R , representa las adyacencias en π y el conjunto de aristas azules, E_A , codifica las adyacencias en ι . Cada arista en E_R es de la forma (π'_{2i}, π'_{2i+1}) , mientras que las aristas en E_A conectan a los pares $(2i, 2i + 1) \forall i \in \{0, \dots, n\}$. Todos los vértices de $BG(\pi')$ tienen el mismo número de aristas incidentes de E_R y de E_A , por lo tanto $|E_R| = |E_A|$.

Un *ciclo alternante* en $BG(\pi')$ es una secuencia de aristas $r_1, a_1, r_2, a_2, \dots, r_m, a_m$ donde $r_i \in E_R$ y $a_i \in E_A$ para $i \in \{1, \dots, m\}$; r_i y a_j inciden en un vértice común

para $i = j \in \{1, \dots, m\}$ y para $i = j + 1 \in \{1, \dots, m\}$ [40]. Una *partición* de $BG(\pi)$ en ciclos alternantes, es una partición tal que cada arista de $BG(\pi')$ está contenida en exactamente un ciclo de la partición. La cardinalidad de una partición es el número de ciclos que aparecen en $BG(\pi')$, se denotará con $c(\pi)$. Una partición de máxima cardinalidad es óptima. Los *ciclos triviales* contienen exactamente una arista roja y una arista azul, mientras que los *ciclos largos* se componen de al menos cuatro aristas alternantes. Se dice que π está ordenada si todos los ciclos en $BG(\pi')$ son triviales.

Ejemplo 19

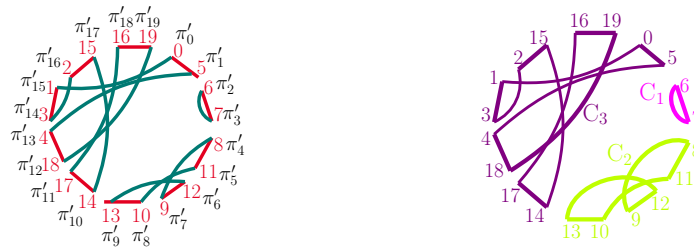


Figura 3.21: Gráfica de punto de rompimiento de la permutación no signada del ejemplo 18.

Las figuras en 18 ilustran la gráfica $BG(\pi')$. La figura de la derecha está coloreada de manera que los ciclos se distingan más claramente. La partición de $BG(\pi')$ está compuesta por un ciclo trivial $C_1 = (6, 7), (7, 6)$ y dos ciclos largos:

$$C_2 = (8, 11), (11, 10), (10, 13), (13, 12), (12, 9), (9, 8) \quad \text{y}$$

$$C_3 = (14, 17), (17, 16), (16, 19), (19, 18), (18, 4), (4, 5), (5, 0), (0, 1), (1, 3), (3, 2), (2, 15), (15, 14)$$

3.3.1. Inversiones en $BG(\pi')$

Toda inversión en la permutación signada π tiene una correspondencia única con una inversión en la permutación imagen π' [4, 56]. Las inversiones $\rho(i, j)$ en π se simulan con inversiones no signadas $\rho(2i - 1, 2j)$ en π' . Se dice que una inversión $\rho(i, j)$ actúa sobre una arista azul a , si a incide con dos aristas rojas (π'_{i-1}, π'_i) y (π'_j, π'_{j+1}) . Si las dos aristas rojas pertenecen al mismo ciclo C se dice que ρ actúa sobre C . Es importante notar que si una inversión $\rho(i, j)$ actúa sobre una arista azul $a = (\pi'_k, \pi'_l)$, entonces cualquier elemento (distinto de $\pi'_i, \pi'_j, \pi'_k, \pi'_l$) está contenido en el intervalo $[i, j]$, si y sólo si está contenido en el intervalo $[k, l]$. Una *inversión propia* ρ en π' se define como una inversión que incrementa en 1 el parámetro $c(\pi)$ [25, 56–59]. Las inversiones propias en π' corresponden a las inversiones orientadas sobre π definidas en la subsección 3.2.3.

Una arista azul es *orientada* si la inversión que actúa sobre ella es una inversión propia y no orientada en otro caso. Dos aristas azules (π'_a, π'_b) y (π'_c, π'_d) se *intersecan* si $a < c < b < d$. Dos ciclos se intersecan si contienen aristas azules intersecantes. Un ciclo es *orientado* si contiene al menos una arista orientada y *no orientado* en otro caso. Una *componente conexa* de $BG(\pi')$ se define como un conjunto de ciclos intersecantes. Una componente es *orientada* si contiene al menos un ciclo orientado y *no orientada* en otro caso [56].

Ejemplo 20

$$\pi'_{\text{ratón}} = (0, 1, 2, 14, 13, 11, 12, 20, 19, 17, 18, 16, 15, 3, 4, 22, 21, 6, 5, 9, 10, 7, 8, 23) \\ \pi'_{21} \pi'_{22} \pi'_{23}$$

Imagen de la permutación signada $\pi'_{\text{ratón}}$ del ejemplo 15.

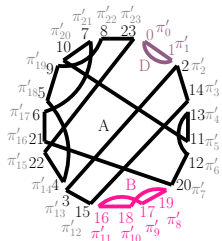


Figura 3.22: Componentes en $BG(\pi'_{\text{ratón}})$.

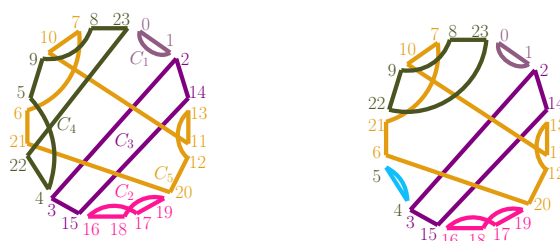


Figura 3.23: Ciclos en $BG(\pi'_{\text{ratón}})$.

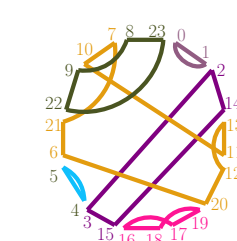


Figura 3.24: Inversión orientada en $BG(\pi'_{\text{ratón}})$.

Las componentes A y B de $BG(\pi'_{\text{ratón}})$ son componentes orientadas ya que contienen a los ciclos C_2, C_4 y C_5 . La gráfica en 3.24 muestra el efecto de la inversión sobre la arista orientada $\{4,5\}$.

Teorema 3.1 (Bafna y Pevzner 1996, [3]). *Sea π una permutación signada con n elementos:*

$$d_r(\pi, \iota) \geq |E_R| - c(\pi)$$

La cota inferior $|E_R| - c(\pi)$ proporciona el número de inversiones orientadas que ordenan una permutación π . Las permutaciones que alcanzan la cota inferior definida en 3.1 son permutaciones sin obstáculos.

Ejemplo 21

Retomando la permutación signada $\pi_{\text{ratón}}$ del ejemplo 15 y la permutación no signada $\pi'_{\text{ratón}}$ en el ejemplo 20, se tiene que $d_{\text{ratón}}(\pi_{\text{ratón}}, \iota) = 12 - 5 = 7$.

3.3.2. Modelo de programación lineal entera

En la modelación que se describe a continuación ni la permutación ni la secuencia de inversiones de la solución al problema de la distancia de inversión son vistas de forma directa como variables y/o restricciones del modelo. La programación lineal entera es usada para determinar la solución del problema consistente en determinar el número de ciclos en una partición de máxima cardinalidad de $BG(\pi')$. La solución de este problema produce una cota ajustada sobre $d_r(\pi, \iota)$ [40, 73]. Sea \mathcal{C} el conjunto de ciclos alternantes de $BG(\pi')$. Para cada $C \in \mathcal{C}$ se define la variable de decisión x_C .

$$\text{máx} \quad \sum_{C \in \mathcal{C}} x_C \quad (18)$$

$$\text{s.a.} \quad \sum_{C \ni e} x_C \leq 1, \quad e \in E_R \cup E_A \quad (19)$$

$$x_C \in \{0, 1\}, \quad C \in \mathcal{C} \quad (20)$$

La restricción (19) asegura que cada arista $e \in E_R \cup E_A$ este contenida como máximo un en un ciclo C tal que $x_C = 1$. En una solución óptima x^* de (18)-(20), el conjunto de ciclos C tal que $x_C^* = 1$ forma una descomposición de cardinalidad máxima.

Se puede obtener una cota inferior válida $c^*(\pi)$ sobre $c(\pi)$ resolviendo la relajación del problema de programación lineal (18)-(20), obtenida reemplazando la restricción (20) con:

$$x_C \geq 0, \quad C \in \mathcal{C} \quad (21)$$

En el modelo definido por (18), (19) y (21), $x_C \leq 1$ se implica trivialmente. El dual del problema de programación lineal (18),(19) y (21) se lee

$$\text{mín} \quad \sum_{e \in E_R \cup E_A} y_e \quad (22)$$

$$\text{s.a.} \quad \sum_{e \in C} y_e \geq 1, \quad C \in \mathcal{C} \quad (23)$$

Donde cada variable y_e , $e \in E_R \cup E_A$ no tiene restricción de signo. Resolver la relajación del problema de programación lineal (18),(19) y (21) implica la resolución de un problema de programación lineal con $|E_R \cup E_A| = \mathcal{O}(n)$

restricciones y $|\mathcal{C}| = \mathcal{O}(2^n)$ variables. En [73] Lancia et. al., resuelven el problema en tiempo polinomial mediante el uso de un algoritmo de ramificación y acotamiento.

3.3.3. Obstáculos en La gráfica de punto de rompimiento

Sea L una componente conexa de $BG(\pi')$, Las posiciones en los extremos izquierdo y derecho de L se definen en (24):

$$L_{\min} = \min_{C \in L} \min_{\pi_i \in C} \quad y \quad L_{\max} = \max_{C \in L} \max_{\pi_i \in C} \tag{24}$$

Una componente L *separa* a las componentes L' y L'' si existe una arista $(\pi_i, \pi_j) \in L$ tal que el intervalo $[L'_{\min}, L'_{\max}]$ esté contenido en el intervalo $[i, j]$ pero el intervalo $[L''_{\min}, L''_{\max}]$ no esté contenido en el intervalo $[i, j]$. Un *obstáculo* es una componente no orientada que no separa a otras componentes no orientadas.

Ejemplo 22

La figura 3.25 ilustra la gráfica de punto de rompimiento asociada a la permutación signada del ejemplo 17.

Consideremos la arista $\{\pi'_2, \pi'_{17}\} = (4, 5)$ en la componente **A**, Las posiciones en los extremos de **B** y **D** son respectivamente (π'_{20}, π'_1) y (π'_{10}, π'_{15}) , tal que $B_{\min} = 20$, $B_{\max} = 1$ y $D_{\min} = 10$, $D_{\max} = 15$. La componente **A** separa a las componentes **B** y **D** ya que $[10, 15]$ está contenido en $[12, 17]$ en tanto que $[20, 1]$ no lo está.

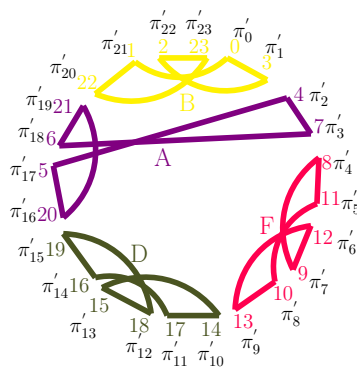


Figura 3.25: Obstáculos en la gráfica de punto de rompimiento.

En la subsección 3.2.4 un obstáculo se definió como un intervalo enmarcado que no contiene un intervalo enmarcado más pequeño.

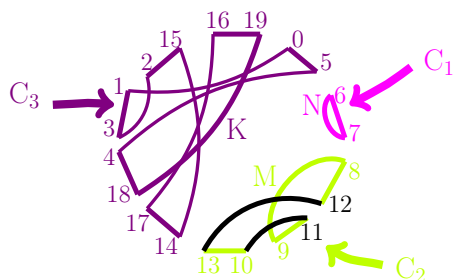
Lema 3.1 (Bergeron 2005, [5]). *Los intervalos enmarcados de la forma $[i, i+k]$ en una permutación signada π tienen una correspondencia uno a uno con los intervalos enmarcados de la forma $[2i, 2(i+k)-1]$ en la permutación imagen π' .*

Los extremos de un intervalo enmarcado $[i, i+k]$ corresponden respectivamente a los pares $2i-1, 2i$ y $2(i+k)-1, 2(i+k)$. Todos los elementos entre i e $i+k$ ocurren en el intervalo $[i, i+k]$ si y sólo si todos los elementos entre $2i$ y $2(i+k)-1$ aparecen en el intervalo $[2i, 2(i+k)-1]$. Si $[2i, 2(i+k)-1]$ es un *obstáculo* entonces contiene exactamente los enteros entre $2i$ y $2(i+k)-1$, tal que las únicas aristas azules en este intervalo son: $(2i, 2i+1), (2i+2, 2i+3), \dots, (2(i+k)-2, 2(i+k)-1)$ y ninguna otra arista azul interseca este conjunto.

3.3.4. Eliminación de obstáculos en $BG(\pi')$

La *inversión de corte* definida en la sección 3.2.5 invierte el segmento entre $2i$ y $2i+1$ contenido en el intervalo $[2i, 2(i+k)-1]$, esto es, actúa dentro de un mismo ciclo c . Esta operación no altera el parámetro $c(\pi)$ (ejemplo 23).

Ejemplo 23

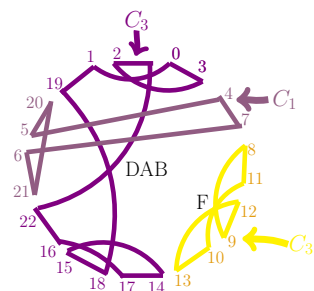


La inversión de corte sobre $BG(\pi')$ mostrada en la figura 3.26 corresponde al corte de obstáculo en del ejemplo 16.

Figura 3.26: Inversión de corte en la gráfica de punto de rompimiento.

La *inversión de fusión*, $\rho(i+k, i')$ une los extremos de dos obstáculos separados por una componente no orientada produciendo una nueva componente orientada y disminuyendo en 1 el parámetro $c(\pi)$ (ejemplo 24).

Ejemplo 24



La inversión $\rho(19, 22)$ fusiona las componentes no orientadas **B**, **D** y **A** de la gráfica en el ejemplo 22.

Figura 3.27: Inversión de fusión.

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 $f(\pi)$:

$$f(\pi) = \begin{cases} 1 & \text{Si } \pi \text{ es una fortaleza} \\ 0 & \text{En otro caso} \end{cases}$$

Sea $h(\pi)$ el número de obstáculos en la permutación signada π . El teorema 3.2 proporciona una ecuación exacta para la distancia de inversión.

Teorema 3.2 (Hannenhalli y Pevzner, 1999 [56]). *Para cualquier permutación signada π :*

$$d_r(\pi, \iota) = |E_R| - c(\pi) + h(\pi) + f(\pi)$$

Ejemplo 25

De acuerdo con el teorema 3.2, la distancia de inversión para las permutaciones signadas π y π_1 de los ejemplos 16 y 17 es:

$$d_r(\pi, \iota) = 10 - 3 + 1 = 8$$

$$d_r(\pi_1, \iota) = 12 - 4 + 3 = 11$$

3.4. Gráfica $H_b(\pi')$

Sea π' la imagen de una permutación signada π bajo τ_1 . La gráfica $H_b(\pi')$ se define como la gráfica donde cada arista de $BG(\pi')$ es representada con un vértice y dos vértices son adyacentes si las cuerdas correspondientes se intersecan en la gráfica de punto de rompimiento. Los vértices orientados de $H_b(\pi')$, corresponden a las aristas orientadas de $BG(\pi')$. En esta tesis se define $H_b(\pi')$ como una gráfica simple (V, E) con lazos permitidos, cuya matriz de adyacencia es una matriz simétrica con entradas en $GF(2)$. En [5, 43, 56, 57],

$H_b(\pi')$, se define como una gráfica simple *bicoloreada* $H = (V, E, c)$, donde el 2-coloreo $c: V \mapsto \{0, 1\}$ de sus vértices, codifica su orientación [52].

Ejemplo 26

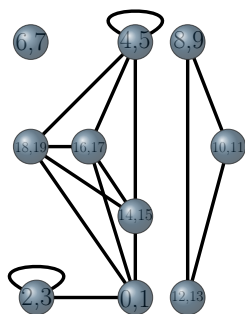


Figura 3.28: Gráfica $H_b(\pi')$.

Gráfica de intersección asociada a la permutación no signada π' del ejemplo 18.

3.4.1. Complementación local e inversiones

La inversión correspondiente a un vértice orientado v de $H_b(\pi')$ se define como la complementación local no simple sobre la vecindad cerrada de v [5, 26]. Esta operación, denotada como $H_b^{inv}(\pi') * v$ en la subsección 2.5.5, convierte a v en un vértice aislado, ya que los pares orientados indican inversiones que producen elementos consecutivos que a su vez son enteros consecutivos (ejemplo 27).

Ejemplo 27

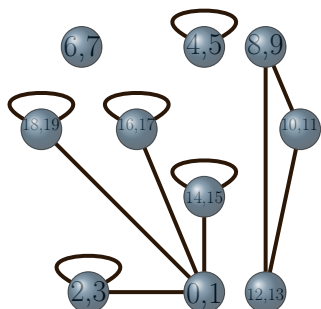


Figura 3.29: Inversión en $H_b(\pi')$.

Inversión sobre el vértice (4,5) en la gráfica $H_b(\pi')$ en el ejemplo (26).

Una componente conexa de $H_b(\pi')$ es orientada si al menos uno de sus vértices es un vértice con lazo y no orientada en otro caso. Una inversión segura es una inversión que no crea nuevas componentes no orientadas, con excepción de los vértices aislados.

3.4.2. Diagrama cordal

La gráfica $BG(\pi')$ puede representarse de manera lineal o circular [1] (ejemplo 28).

Ejemplo 28

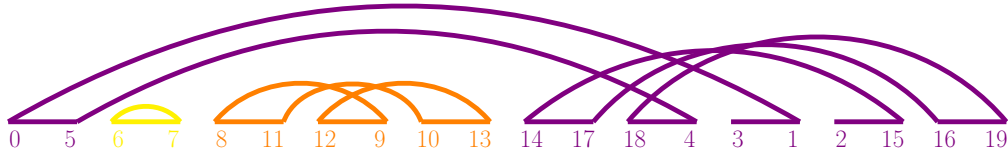


Figura 3.30: Representación lineal de la gráfica en el ejemplo (19).

La representación circular de $BG(\pi')$ es un diagrama cordal, donde las cuerdas son las aristas azules y las aristas rojas se encuentran sobre la circunferencia del diagrama. $BG(\pi')$ tiene asociada su gráfica de intersección (ver subsección 2.5.2), que en este contexto se denotará como $H_b(\pi')$.

Sea π' la imagen de una permutación signada y enmarcada π . Se considera la inversa de la transformación T_1 definida en la sección 3.3, sea esta T_1^{-1} . Las cuerdas unen enteros consecutivos empezando por el $(0,1)$ y T_1^{-1} se define asignando la etiqueta $(\frac{d}{2})$ a los extremos $(d, d+1)$ de cada cuerda.

$$(d, d+1) \mapsto \left(\frac{d}{2}, \frac{d}{2}\right) \quad (25)$$

La aplicación de la transformación en (25) permite recobrar π de π' y así etiquetar las cuerdas de la gráfica $BG(\pi')$ con las etiquetas de π . De la misma manera se recobra la palabra de doble ocurrencia relativa al diagrama de cuerdas, como se ilustra en el ejemplo 29.

Ejemplo 29

Consideremos la permutación no signada del ejemplo 18.

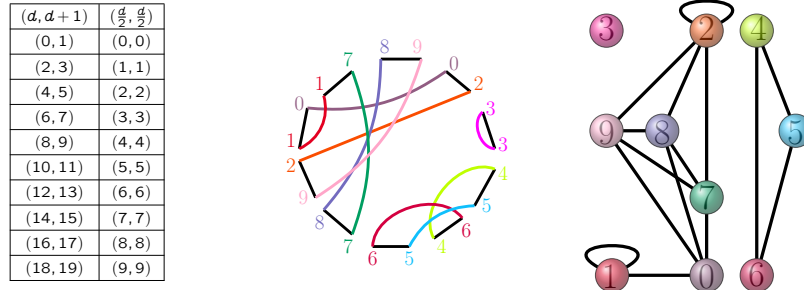


Figura 3.31: Ejemplo de aplicación de la transformación en (25).

Sea (v_1, \dots, v_k) una secuencia de vértices distintos de $H_b(\pi')$ y sea $\varphi = (*^{inv}v_1, \dots, *^{inv}v_k)$ una secuencia de inversiones aplicables sobre (v_1, \dots, v_k) (La asociatividad de $*^{inv}v$ es de izquierda a derecha). φ es una secuencia *óptima* si todos los vértices en la gráfica $H_b(\pi')_\varphi$ son aislados y no orientados.

Ejemplo 30

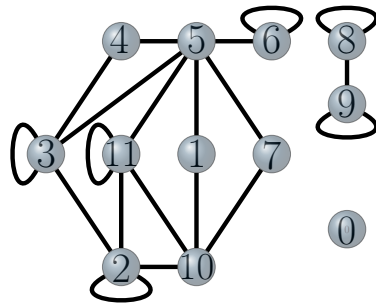


Figura 3.32: $H(\pi'_{ratón})$.

La figura en 3.32 muestra la gráfica de intersección asociada a la permutación del ejemplo 20. Las gráficas en la figura 3.33 presentan la aplicación de la secuencia en (26) sobre la gráfica $H(\pi'_{ratón})$.

$$\varphi = (*^{inv}v_6, *^{inv}v_5, *^{inv}v_4, *^{inv}v_2, *^{inv}v_3, *^{inv}v_8, *^{inv}v_1) \tag{26}$$

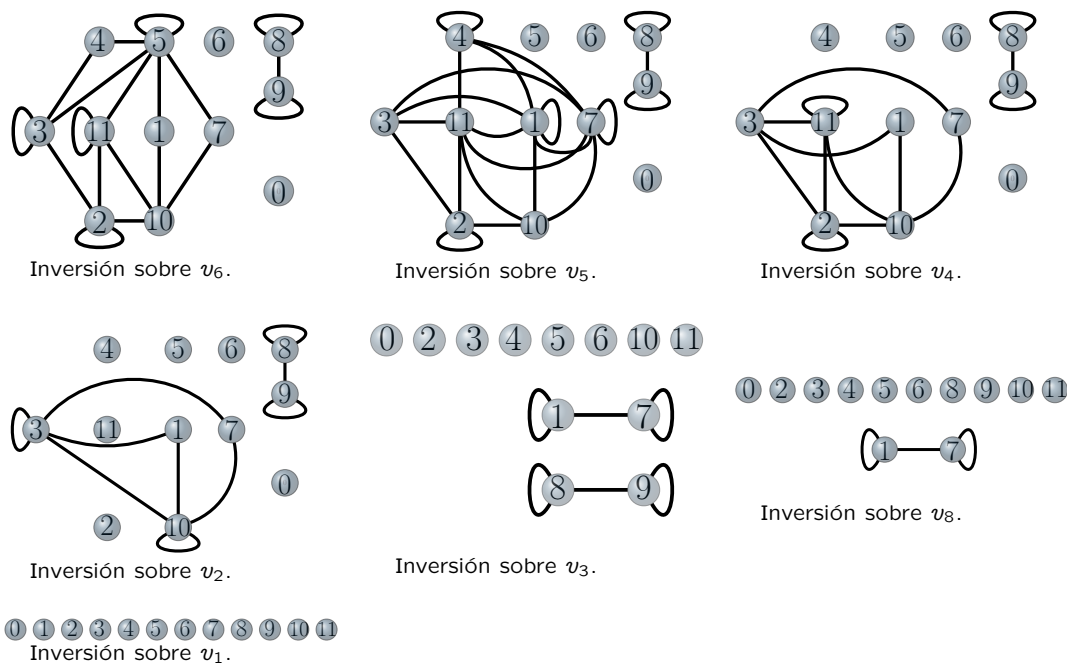


Figura 3.33: Ejemplo de secuencia óptima.

3.4.3. Matriz de intersección

Sea $A(H_b(\pi'))$ la matriz de adyacencia de $H_b(\pi')$ y sea $r(A(H_b(\pi')))$ el rango de $A(H_b(\pi'))$. El resultado en 3.1 aparece en [43], (Corolario 4) y en [26], (Corolario 12).

Corolario 3.1. *El número de vértices en cualquier secuencia orientada óptima para una gráfica H_b depende únicamente de H_b y es igual a $r(A(H_b))$.*

Tal que en términos de matrices simétricas (antisimétricas), la cota inferior del teorema 3.1 se define en el teorema 3.3.

Teorema 3.3 (Brijder 2017, [26]). *Para toda permutación signada π :*

$$d_r(\pi, \iota) \geq r(A(H_b))$$

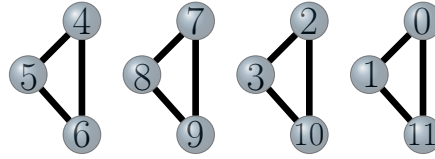
Ejemplo 31

Figura 3.34: Gráfica de intersección asociada a la permutación π_1 del ejemplo 17.

Sea $A(H_b(\pi'))$ en (27) la matriz de adyacencia de la gráfica en el extremo derecho de la figura 3.31 en el ejemplo 29 y sea $A(H_b(\pi'_1))$ en (28) la matriz de adyacencia de la gráfica $H_b(\pi'_1)$ en 3.34. Se tiene que la matriz $A(H_b(\pi'))$ es simétrica y $r(A(H_b(\pi'))) = 7$ mientras que $A(H_b(\pi'_1))$ es antisimétrica y $r(A(H_b(\pi'_1))) = 8$. Tal que $d_r(\pi, \iota) \geq 7$ y $d_r(\pi_1, \iota) \geq 8$.

$$A(H_b(\pi')) = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{matrix} \quad (27)$$

$$A(H_b(\pi'_1)) = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix} \quad (28)$$

Las submatrices principales correspondientes a obstáculos en la gráfica de intersección son matrices antisimétricas de tamaño impar y por consecuencia no son invertibles.

3.4.4. Complemento local modificado de M

Sean H_b una gráfica y M una $V \times V$ -matriz cuyas filas y columnas están etiquetadas por $V(H_b)$. Sea v un vértice de H_b , la matriz M queda expresada en (29).

$$M = \begin{bmatrix} d_{vv} & \rho_1 & \rho_2 \\ \kappa_1 & M_{11} & M_{12} \\ \kappa_2 & M_{21} & M_{22} \end{bmatrix} \quad (29)$$

La primer fila y la primer columna de M corresponden a v , las filas y columnas de M_{11} corresponden a los vértices en $n(v)$ mientras que las filas y

columnas de M_{22} se asignan a los vértices en $V(H_b) - n(v) - \{v\}$. El *complemento local modificado* de M con respecto a v , es la matriz M_{mod}^v en (30) obtenida de M al sumar la fila de v a toda fila correspondiente a un vecino de v [94].

$$M_{mod}^v = \begin{bmatrix} d_{vv} & \rho_1 & \rho_2 \\ \kappa_{1'} & M'_{11} & M'_{12} \\ \kappa_2 & M_{21} & M_{22} \end{bmatrix} \quad (30)$$

Sea $X \subseteq V$ el subconjunto $X = \{i, \dots, i+1\}$ que induce una matriz antisimétrica y singular $A(H_b^\pi)[X]$. La *inversión de corte* descrita en las subsecciones 3.2.5 y 3.3.3 corresponde al complemento local modificado de la subgráfica inducida por X , donde el vértice v_i corresponde al elemento $i \in X$.

Ejemplo 32

Consideremos la matriz $A(H_b(\pi'))$ en (27) del ejemplo 31 y el subconjunto $X = \{4, 5, 6\} \in V$. La submatriz $A(H_b(\pi'))[X]_{mod}^{v_4}$ se muestra en (31).

$$A(H_b(\pi'))[X]_{mod}^{v_4} = \begin{matrix} & 4 & 5 & 6 \\ 4 & \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \end{matrix} \quad (31)$$

$$A(H_b(\pi')) = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 7 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 8 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 9 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{matrix} \quad (32)$$

Observemos que las matrices en (27) y en (32) tienen el mismo rango. Esto se debe a que el complemento local modificado no altera el parámetro $c(\pi)$ lo que es congruente con la definición de inversión de corte dada en la subsección 3.3.4.

3.5. Paseos Eulerianos en multigráficas 4-regulares

3.5.1. Multigráfica 4-regular asociada a una permutación signada

Hasta ahora, la permutación signada y enmarcada π ha representado, indistintamente, el orden de los genes en un cromosoma lineal y el orden de los fragmentos MDS en un gen micronuclear. Se ha establecido una correspondencia directa entre los *punteros* (también llamados *puntos de recombinación*), descritos en la subsección 3.1.3, con los elementos (π'_{2i}, π'_{2i+1}) de la transformación τ_1 (sección 3.3). Para continuar con esta línea y como primer paso en la construcción de la multigráfica 4-regular G_π asociada a una permutación signada π , se hará la modificación mostrada en (33) y (34) a la notación de (16) y (17). Sea π una permutación signada y enmarcada,

$$i \rightarrow ((i-1)_b, i_a) \quad \text{si } i > 0 \quad (33)$$

$$i \rightarrow (i_a, (i-1)_b) \quad \text{si } i < 0 \quad (34)$$

Ejemplo 33

Sea π la permutación signada del ejemplo 16.

$$\pi' = (0_a, (2_b, 3_a), (3_b, 4_a), (5_b, 6_a), (4_b, 5_a), (6_b, 7_a), (8_b, 9_a), (2_a, 1_b), (0_b, 1_a), (7_b, 8_a), 9_b) \quad (35)$$

Aunado a lo anterior, seguiremos las siguientes convenciones:

- (1) Los genes en el cromosoma se representarán como fragmentos MDS.
- (2) La transformación definida en (33) y (34) se expandirá por medio de la inserción de un elemento intermedio I_i entre dos MDSs adyacentes en π .

Ejemplo 34

Consideremos la permutación π' en (35) del ejemplo 33 cuya permutación expandida se muestra en (35) (los punteros se marcan en rojo).

$$\pi' = M_0 0_a I_0 2_b M_3 3_a I_1 3_b M_4 4_a I_2 5_b M_6 6_a I_3 4_b M_5 5_a I_4 6_b M_7 7_a I_5 8_b M_9 9_a I_6 2_a \overline{M_2 1_b I_7 0_b M_1 1_a I_8 7_b M_8 8_a I_9 9_b M_{10}}$$
(36)

La gráfica D_π se define como una gráfica dirigida donde cada vértice se asocia con un único punto de recombinación y cada arco representa un segmento (de MDSs o IESs) entre dos punteros (ejemplo 35).

Ejemplo 35

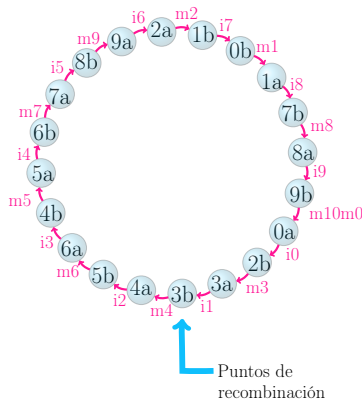


Figura 3.35: Gráfica dirigida correspondiente a la permutación expandida π' del ejemplo 34.

La multigráfica 4-regular G_π asociada a un cromosoma π se obtiene a partir de D_π modificando la transformación inversa τ_1^{-1} en (25) (ver subsección 3.4.2) como se expresa en (37):

$$(i_b, i_a) \mapsto i$$
(37)

Cada vértice v de G_π se relaciona con un elemento de la transformación en (37) 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 [26, 31].

Ejemplo 36

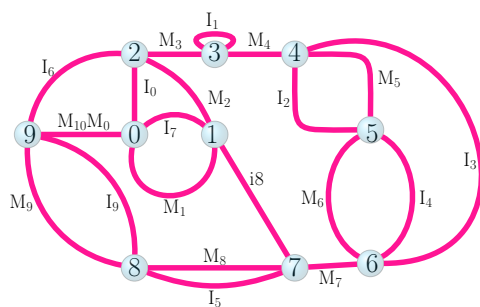


Figura 3.36: Multigráfica 4-regular obtenida de la gráfica dirigida del ejemplo 35.

3.5.2. Bitransiciones

Consideraremos que cada arista e de una multigráfica 4-regular G está compuesta por un par de *medias aristas* $\{h', h''\}$ con un vértice final cada una con solo un vértice de G tal que cada media arista incide en un extremo de e [8, 11, 16, 64]. Sea v un vértice 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 3.39). 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 . *Cortar* una bitransición incidente en v consiste en eliminar v de G y reemplazarlo con un vértice v' incidente a las medias aristas de t' y con un vértice v'' incidente a las medias aristas de t'' , obteniéndose una nueva gráfica denotada con $G|\omega$ [11, 64].

Sea TE un *paseo Euleriano* de G y sea $\Omega(TE)$ el conjunto de bitransiciones definido por TE . Si $\omega \in \Omega(TE)$, entonces TE sigue siendo un paseo Euleriano de $G|\omega$. El número de componentes conexas en G se denotará con $l(G)$.

$$l(G|\omega) = l(G) = 1 \quad \text{si } \omega \in \Omega(TE) \quad (38)$$

Cortar un conjunto $\Omega \in \hat{\Omega}$ consiste en la construcción de 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 [11, 20]).

Ejemplo 37

El orden de las secuencias de bases en el cromosoma representado por la permutación expandida π' en (36) del ejemplo 34 induce un paseo Euleriano en la multigráfica 4-regular conexas G_π de la figura 3.36. Este paseo, que denotaremos como $TE(\Phi)$, se muestra en (39)².

$$TE(\Phi) = M'_1 M''_1, I'_8 I''_8, M'_8 M''_8, I'_9 I''_9, M_{10} M'_0 M_{10} M''_0, I'_0 I''_0, M'_3 M''_3, I'_1 I''_1, M'_4 M''_4, \\ I'_2 I''_2, M'_6 M''_6, I'_3 I''_3, M'_5 M''_5, I'_4 I''_4, M'_7 M''_7, I'_5 I''_5, M'_9 M''_9, I'_6 I''_6, M'_2 M''_2, I'_7 I''_7 \quad (39)$$

²El uso de esta notación se justifica en la sección 3.8 (proposición 2).

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

- $\Omega(TE(\Phi))$
- v_0 $\{\{I_7'', M_1'\}, \{M_{10}M_0'', I_6'\}\}$
 - v_1 $\{\{M_1'', I_8'\}, \{M_2'', I_7'\}\}$
 - v_2 $\{\{I_6'', M_2'\}, \{I_0'', M_3'\}\}$
 - v_3 $\{\{M_3'', I_1'\}, \{I_1'', M_4'\}\}$
 - v_4 $\{\{M_4'', I_2'\}, \{I_3'', M_5'\}\}$
 - v_5 $\{\{I_2'', M_6'\}, \{M_5'', I_4'\}\}$
 - v_6 $\{\{M_6'', I_3'\}, \{I_4'', M_7'\}\}$
 - v_7 $\{\{I_8'', M_8'\}, \{M_7'', I_5'\}\}$
 - v_8 $\{\{M_8'', I_9'\}, \{I_5'', M_9'\}\}$
 - v_9 $\{\{I_9'', M_{10}M_0'\}, \{M_9'', I_6'\}\}$

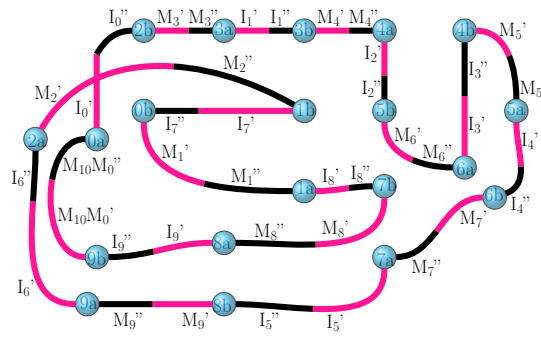


Figura 3.37: $G_\pi|\Omega(TE(\Phi))$

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 (40) ilustrada en la gráfica de la figura 3.38.

$\Omega(\Psi)$	
v_0 $\{\{M_{10}M_0'', M_1'\}, \{I_0'', I_7'\}\}$	v_5 $\{\{M_5'', M_6'\}, \{I_2'', I_4'\}\}$
v_1 $\{\{M_1'', M_2'\}, \{I_7'', I_8'\}\}$	v_6 $\{\{M_6'', M_7'\}, \{I_4'', I_3'\}\}$
v_2 $\{\{M_2'', M_3'\}, \{I_6'', I_0'\}\}$	v_7 $\{\{M_7'', M_8'\}, \{I_8'', I_5'\}\}$
v_3 $\{\{M_3'', M_4'\}, \{I_1'', I_1'\}\}$	v_8 $\{\{M_8'', M_9'\}, \{I_5'', I_9'\}\}$
v_4 $\{\{M_4'', M_5'\}, \{I_3'', I_2'\}\}$	v_9 $\{\{M_9'', M_{10}M_0'\}, \{I_9'', I_6'\}\}$

$$P(\Omega(\Psi)) = \{\{M_1' M_1'', M_2' M_2'', M_3' M_3'', M_4' M_4'', M_5' M_5'', M_6' M_6'', M_7' M_7'', M_8' M_8'', M_9' M_9'', M_{10} M_{10}' M_{10} M_{10}'\}, \{I_3' I_3'', I_2' I_2'', I_4' I_4'', \{I_1' I_1''\}, \{I_7' I_7'', I_8' I_8'', I_5' I_5'', I_9' I_9'', I_6' I_6'', I_0' I_0''\}\} \quad (40)$$

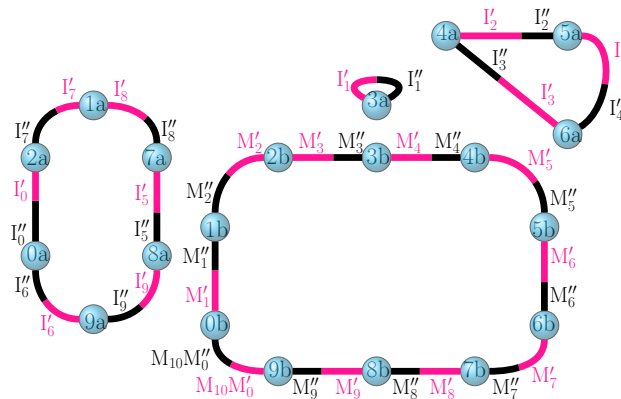


Figura 3.38: $G_\pi|\Omega(\Psi)$

$P(\Omega(\Psi))$ contiene el circuito con la permutación identidad ι . Los conjuntos $\Omega(TE(\Phi))$ y $\Omega(\Psi)$ no tienen bitransiciones en común, en este caso se dice que los conjuntos son *suplementarios*.

En la subsección 3.5.3 se describe la asociación de un sistema de isotropía (ver apéndice A) con una multigráfica 4-regular. Los sistemas de isotropía asociados a multigráficas 4-regulares se conocen como sistemas gráficos. Es importante mencionar que no todos los sistemas de isotropía son gráficos [16].

3.5.3. Multigráfica codificada

Sea G una multigráfica conexa 4-regular y sea $TE = h'_0 h''_0, h'_1 h''_1, \dots, h'_{n-1} h''_{n-1}$ 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 $\{\{h''_{j-1}, h'_i\}, \{h'_i, h'_j\}\}$ los cuáles se ilustran en la figura 3.39.

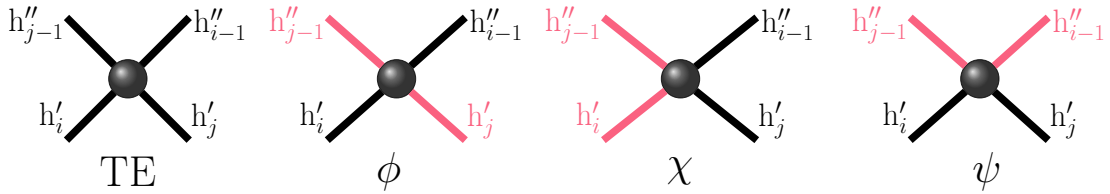


Figura 3.39: Bitransiciones.

La bitransición $\{\{h''_{j-1}, h'_j\}, \{h''_{i-1}, h'_i\}\}$ está contenida en $\Omega(TE)$. 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 . Sea $K^V = \{\phi, \chi, \psi\}$ [8, 11, 15].

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

Cuadro 3.1: Codificación de bitransiciones.

Una *codificación de bitransiciones* es una familia $\tau = (\tau_v : v \in V)$ [8, 16].

Ejemplo 38

Asignación de etiquetas para los conjuntos disjuntos de bitransiciones $\Omega(TE(\Phi))$ y $\Omega(\Psi)$ del ejemplo 37.

	v_0	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
$\tau(\Omega(TE(\Phi)))$	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 bitransiciones [11, 15]. 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 , $\forall v \in V$ [8]. 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 $A \in K^V$ en un vértice v de G es una bitransición tal que $\tau_v(\omega) = A_v$. Denotamos con $\Omega(A)$ al conjunto de todas las bitransiciones inducidas por A . Todo conjunto de bitransiciones (partición Euleriana) es determinado por un único vector de K^V [16]. Un vector completo A puede dar origen a un paseo Euleriano o a una partición de G en circuitos. La ecuación (41) determina la cardinalidad de cualquier partición Euleriana P de la multigráfica 4-regular conexa G [14]. Sea $|P|$ el número de circuitos en P .

$$|P| - l(G) = r(A) \quad \text{si} \quad A = T(P) \quad (41)$$

Donde $r(A)$ es el rango de A con respecto a S (ver apéndice A).

Ejemplo 39

Sea $\Omega(\Psi)$ el conjunto de bitransiciones en el ejemplo 37 que induce la partición Euleriana $P(\Omega(\Psi))$ en (40), de la multigráfica 4-regular G_π en la figura 3.36. La codificación de $\Omega(\Psi)$ se muestra en el ejemplo 38. Sea Ψ un vector en K^V tal que para cada $v \in V$, $\tau(\Omega(\Psi))_v = \Psi_v$. Además de contener el circuito con el orden en la permutación identidad, $P(\Omega(\Psi))$ contiene (en el ejemplo) otros tres circuitos compuestos por ADN no genético (IESs) denotados por C_{IES} . Sea $l(G_\pi)$ el número de componentes conexas de G_π (ecuación (38)).

$$|P(\Omega(\Psi))| - l(G_\pi) = |C_{IES}| = 3 \quad (42)$$

De acuerdo con (42), el número de circuitos conformados por segmentos no codificadores en la partición inducida por el vector que representa al cromosoma identidad es igual al rango de dicho vector con respecto a S . Sea $c(\pi)$ el número de ciclos en una decomposición de $BG(\pi')$ (ver sección 3.3). $c(\pi) = |C_{IES}|$, por lo que el número de circuitos en $BG(\pi')$ es igual a $r(\Psi)$ (ejemplo 19).

Si S es un sistema de isotropía, entonces un vector $A \in K^V$ es un *vector*

Euleriano de S , si $r(A) = 0$. Sea TE un paseo Euleriano de G , Bouchet [16] demostró que para toda multigráfica codificada (G, τ) , existe exactamente un sistema de isotropía S definido de manera que:

$$|\Omega(TE)| - l(G) = r(A) = 0 \quad \text{si} \quad A = T(\Omega(TE)) \quad (43)$$

Ejemplo 40

Ahora consideremos el conjunto de bitrasciones $\Omega(TE(\Phi))$ que induce el paseo Euleriano $TE(\Phi)$ en (39) del ejemplo 37 cuya codificación se muestra en el ejemplo 38. Sea Φ un vector en K^V tal que para cada $v \in V$, $\tau(\Omega(TE(\Phi)))_v = \Phi_v$ y sea $P(\Omega(TE(\Phi)))$ la partición Euleriana de G_π inducida por Φ (en la figura 3.37). $P(\Omega(TE(\Phi)))$ satisface la igualdad en (44):

$$|P(\Omega(TE(\Phi)))| - l(G_\pi) = r(\Phi) = 0 \quad (44)$$

Un sistema de isotropía S que se puede asociar de la manera descrita con una multigráfica codificada es un *sistema gráfico* [15, 19].

3.6. Transformaciones en paseos Eulerianos

En esta sección se describen dos transformaciones aplicables sobre un paseo Euleriano en un gráfica 4-regular. Estas transformaciones fueron descritas por primera vez por Kotzig [71] y extendidas posteriormente a sistemas de isotropía por Bouchet [19].

Consideremos una multigráfica codificada (G, τ) , su sistema gráfico asociado S , y un vector Euleriano A de S . Sea $TE(A) = h'_0 h''_0, h'_1 h''_1, \dots, h'_{n-1} h''_{n-1}$ una sucesión de medias aristas que forman un paseo Euleriano en G inducido por A . La secuencia de vértices de $TE(A)$ es $V(A) = v_0, v_1, \dots, v_{n-1}$ donde v_i es el vértice incidente a $TE(A) = h'_{i-1} h''_{i-1}, h'_i$ para $i \in \{0, \dots, n-1\}$. $V(A)$ es una *palabra de doble ocurrencia* sobre $V(G)$ [7]. Sea $v \in V(G)$ y supongamos $v = v_i = v_j$, ($i \neq 0$). Se definen dos tipos de transformaciones, la *separación* y la *conmutación*.

3.6.1. Separación

Si sustituimos la bitransición codificada por el vector Euleriano A en v con la bitransición etiquetada con χ en v se obtiene el conjunto de bitrasciones de una nueva partición Euleriana $\{U', U''\}$.

$$U' = \{h'_i h''_i, h'_{i+1} h''_{i+1}, \dots, h'_{j-1} h''_{j-1}\}$$

$$U'' = \{h'_j h''_j, h'_{j+1} h''_{j+1}, \dots, h'_{n-1} h''_{n-1}, h'_0 h''_0, \dots, h'_{i-1} h''_{i-1}\}$$

Se dice que la bitransición con la etiqueta χ , separa al paseo Euleriano $TE(A)$ en v .

3.6.2. Conmutación

El resultado de reemplazar la bitransición codificada por A en v por la bitransición con la etiqueta ψ en v , es el conjunto de bitransiciones de un nuevo paseo Euleriano denotado con $TE(A)^*v$. Se dice que $TE(A)^*v$ es obtenido por la *conmutación* de $TE(A)$ en v [15, 19].

Ejemplo 41

La bitransiciones inducidas por el vector completo Ψ del ejemplo 39 y el vector Euleriano Φ del ejemplo 40 en el vértice v_4 de la multigráfica 3.36 son respectivamente:

$\{M''_4, I''_2\}\{I''_3, M'_5\}$ y $\{M''_4, I''_3\}\{I''_2, M'_5\}$. La conmutación de $TE(\Phi)$ en v_4 se muestra en (45).

$$TE(\Phi)^*v_4 = \begin{matrix} M'_1 M''_1, I'_8 I''_8, M'_8 M''_8, I'_9 I''_9, M_{10} M'_0 M_{10} M''_0, I'_0 I''_0, M'_3 M''_3, I'_1 I''_1, M'_4 M''_4, \\ I''_3 I'_3, M''_6 M'_6, I''_2 I'_2, M'_5 M''_5, I'_4 I''_4, M'_7 M''_7, I'_5 I''_5, M'_9 M''_9, I'_6 I''_6, M'_2 M''_2, I'_7 I''_7 \end{matrix} \quad (45)$$

Proposición 1 (Bouchet 1988, [19]). *Si $TE(A')$ y $TE(A'')$ son paseos Eulerianos de una multigráfica codificada (G, τ) entonces existe una palabra m sobre V tal que $TE(A'') = TE(A')^*m$.*

3.7. Gráficas fundamentales

Sea $S = (L, V)$ el sistema de isotropía asociado a una multigráfica codificada (G, τ) , y sea A la codificación de un paseo Euleriano TE de G . Si H es la gráfica de alternancia de TE , vista como una palabra de doble ocurrencia, se dice que H es la *gráfica fundamental* de S . Existe exactamente un vector Euleriano A' de S que satisface $A'(v) \neq A(v)$ y $A'(w) = A(w)$, $\forall w \in V \setminus \{v\}$ [11]. La gráfica fundamental de S derivada de $TE(A')$ se obtiene por la complementación local de H en un vértice v .

Sea $A(H)$ la matriz de adyacencia de la gráfica fundamental H de un sistema de isotropía S asociado a (G, τ) . Se dice que $A(H)$ es la *matriz de alternancia* de G con respecto a A .

Ejemplo 42

La figura 3.40 muestra la gráfica de alternancia del paseo Euleriano $TE(\Phi)$ del ejemplo 37 inducido por el vector Euleriano Φ del ejemplo 40. La matriz en (46) es la matriz

de alternancia de G_π con respecto a Φ .

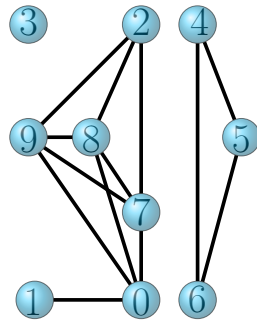


Figura 3.40: Gráfica de alternancia del paseo Euleriano $TE(\Phi)$ del ejemplo 37.

$$A(H) = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 7 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 8 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 9 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{matrix} \quad (46)$$

3.7.1. Gráfica de alternancia relativa

En la teoría de alternancia de multigráficas 4-regulares, la gráfica fundamental H de un sistema de isotropía S asociada a una multigráfica codificada (G, τ) no se maneja como una gráfica con lazos permitidos. Sin embargo el modelado de la estructura de doble hélice de la molécula de ADN requiere la codificación de la orientación de los vértices por lo cuál a partir de ahora nos referiremos a la gráfica de alternancia de un paseo Euleriano en G como una gráfica simple *con lazos permitidos* H_b .

Sea $\Pi = (H_b, A, B)$ la presentación gráfica de un sistema de isotropía S asociado a una multigráfica codificada (G, τ) y sea $P(\Omega(B))$ la partición Euleriana inducida por el vector completo B . Para simplificar la notación desde este momento $P(\Omega(B))$ se denotará como P_B . La *gráfica de alternancia relativa* [32, 90, 91, 93] de P_B con respecto a $TE(A)$ es la gráfica $H_b(A, B)$ que se obtiene de H_b modificando cada vértice $v \in V(G)$ tal que $B_v \neq \chi$, como sigue:

- (i) Eliminando v si $B_v = \phi$.
- (ii) Añadiendo un lazo a v si $B_v = \psi$.

La operación (ii) corresponde a la complementación de lazo descrita en la subsección 2.5.4.

Ejemplo 43

Sea P_Ψ la partición Euleriana inducida por el vector completo Ψ del ejemplo 39. La gráfica de alternancia relativa de P_Ψ con respecto al paseo Euleriano $TE(\Phi)$ del ejemplo 37 corresponde a la gráfica en el extremo derecho de la figura 3.31 en el ejemplo

29.

En adelante $H_b(\pi')$ será referida como la gráfica de alternancia relativa $H_b(\Phi, \Psi)$. Sea (G, τ) una multigráfica codificada y sea P una partición Euleriana en G . Se dice que un circuito $\xi \in P$ es *incidente de forma única* en un vértice v si contiene exactamente dos de las cuatro medias aristas incidentes en v [93].

Si $\Pi = (H_b, \Phi, \Psi)$ es la presentación gráfica de un sistema de isotropía S asociado a (G, τ) y $A(\Phi, \Psi)$ es la matriz de adyacencia de $H_b(\Phi, \Psi)$. Entonces el *vector relativo principal* (relative core vector) de un circuito $\xi \in P_\Psi$ con respecto a Φ se define como el vector $\Xi(\xi, \Phi) \in GF(2)^{V(G)}$ cuyas entradas distintas de cero corresponden a los vértices de G en los cuáles ξ es incidente de forma única y se cumple que $\Psi_v = \chi$ o $\Psi_v = \psi$, $\forall \xi \in P_\Psi$.

Teorema 3.4 (Traldi 2012, [93]). *Sea $A(\Phi, \Psi)$ la matriz de adyacencia de la gráfica $H_b(\Phi, \Psi)$.*

- (i) *El espacio nulo de $A(\Phi, \Psi)$ es generado por los vectores relativos principales de los circuitos de P_Ψ .*
- (ii) *Para cada componente conexa de G , los vectores relativos principales de los circuitos incidentes de P_Ψ suman 0.*
- (iii) *Si $Q \subset P_\Psi$ y no existe una componente conexa de G para la cuál Q contenga a cada circuito incidente de P_Ψ , entonces los vectores relativos de los circuitos de Q son linealmente independientes.*

Sea $\nu(A(\Phi, \Psi))$ la nulidad de $A(\Phi, \Psi)$. De acuerdo con el teorema 3.4 y los resultados de la subsección 3.5.3 se tiene,

$$|C_{IES}| = \nu(A(\Phi, \Psi)) \tag{47}$$

Ejemplo 44

Consideremos la matriz de adyacencia de la gráfica $H_b(\Phi, \Psi)$ referenciada en el ejemplo 43. $A((\Phi, \Psi))$ se muestra en (27) del ejemplo 31. Tenemos que $\nu(A(\Phi, \Psi)) = 3$ lo que coincide con los resultados obtenidos anteriormente.

3.8. Δ -matroides y sistemas de isotropía

Definición 3.8.1 ([23]). *Si H_b es una gráfica simple con lazos permitidos, entonces el Δ -matroide asociado a H_b es,*

$\mathcal{D}_{H_b} = \{X \subseteq V : A[X] \text{ es no singular sobre } GF(2)\}$

Definición 3.8.2. Sea H_b una gráfica simple con lazos permitidos, y sea $W(H_b)$ el conjunto finito del matroide binario asociado a H_b . Los transversales (resp. subtransversales) de $W(H_b)$ son los elementos de:

$$\mathcal{T}(W(H_b)) = \{T \subseteq W(H_b) : |T \cap \{v_\phi, v_\chi, v_\psi\}| = 1 \quad \forall v \in V(H_b)\}$$

$$\mathcal{S}(W(H_b)) = \{S \subseteq W(H_b) : |S \cap \{v_\phi, v_\chi, v_\psi\}| \leq 1 \quad \forall v \in V(H_b)\}$$

El conjunto potencia $\mathcal{P}(W(H_b))$ se considera un espacio vectorial sobre $GF(2)$ con la suma, la multiplicación usual por los escalares del campo $GF(2)$ y una forma bilineal definidas en el apéndice A. Sea Q el subespacio de $\mathcal{P}(W(H_b))$ generado por tripletas de vértices. Entonces cada vértice del espacio cociente $\mathcal{P}(W(H_b))/Q$ incluye un elemento de $\mathcal{S}(W(H_b))$. De manera que podemos identificar a $\mathcal{S}(W(H_b))$ con $\mathcal{P}(W(H_b))/Q$. La suma resultante en $\mathcal{S}(W(H_b))$ se denota con \boxplus y se define en el cuadro 3.2, para cada componente correspondiente al par de transversales que se consideran.

\boxplus	0	ϕ	χ	ψ
0	0	ϕ	ψ	χ
ϕ	ϕ	0	χ	ψ
χ	χ	ψ	0	ϕ
ψ	ψ	χ	ϕ	0

Cuadro 3.2: Suma \boxplus .

El espacio ciclo $Z(M(IAS(H_b)))$ es el subespacio de $\mathcal{P}(W(H_b))$ en $GF(2)$ compuesto por los subconjuntos de $W(H_b)$ correspondientes a los conjuntos de columnas de $IAS(H_b)$ que suman 0.

Definición 3.8.3 ([91]). Un ciclo transverso de H_b es un elemento de:

$$\mathcal{L}(H_b) = \mathcal{S}(W(H_b)) \cap Z(M(IAS(H_b)))$$

Si $X \subseteq H_b$ y $S \in \mathcal{S}(W(H_b))$, entonces $X \cdot S$ denota a $S \cap \{v_\phi, v_\chi, v_\psi : v \in X\}$.

Proposición 2 (Traldi 2015, [91]). Sea $\Phi(H_b) = \{v_\phi : v \in V(H_b)\}$, y sea $\Psi(H_b) = \{v_\psi : v \in V(H_b) \mid v \text{ es un vértice con lazo}\} \cup \{v_\chi : v \in V(H_b) \mid v \text{ es un vértice sin lazo}\}$. Entonces

$$\mathcal{L}(H_b) = \{(X \cdot \Psi(H_b)) \boxplus (n(X) \cdot \Phi(H_b)) : X \subseteq V(H_b)\}$$

$\Phi(H_b)$ y $\Psi(H_b)$ son elementos disjuntos de $\mathcal{S}(W(H_b))$ de tamaño $|V(H_b)|$, de manera que satisfacen la definición de vectores suplementarios dada en el apéndice A. De acuerdo con la proposición 2, $\mathcal{L}(H_b(\Phi, \Psi))$ es un sistema de

isotropía con gráfica fundamental $H_b(\Phi, \Psi)$. Por sencillez en la notación los vectores $\Phi(H_b(\Phi, \Psi))$ y $\Psi(H_b(\Phi, \Psi))$ se denotarán con Φ y Ψ . El Δ -matroide \mathcal{D}_{H_b} también se puede definir como (48):

$$\mathcal{D}_{H_b} = \{S \subseteq V(H_b) : \{s_\chi : s \in S\} \cup \{v_\phi : v \notin S\} \text{ es una base de } M(IA_S(H_b))\} \quad (48)$$

El índice ψ no aparece en la descripción (48), ya que el matroide binario restringido $M(IA(H_b))$ contiene información suficiente para determinar a \mathcal{D}_{H_b} [27, 91]. Si H_b es una gráfica simple con lazos permitidos, entonces H_b está completamente determinada por el Δ -matroide \mathcal{D}_{H_b} : un vértice v es un vértice con lazo si y sólo si $\{v\} \in \mathcal{D}_{H_b}$. Dos vértices con lazo v y $w \in \mathcal{D}_{H_b}$, son adyacentes si y sólo si $\{v, w\} \notin \mathcal{D}_{H_b}$. Dos vértices v y w sin lazo son adyacentes si y sólo si $\{v, w\} \in \mathcal{D}_{H_b}$ [91]. Como consecuencia \mathcal{D}_{H_b} determina a los sistemas de isotropía con gráfica fundamental H_b con respecto a isomorfismo fuerte.

Aplicación en el ordenamiento por inversiones

4.1. Inversiones orientadas y Δ -matroides

Sea $\Pi = (H_b, \Phi, \Psi)$ la presentación gráfica del sistema de isotropía $\mathcal{L}(H_b)$ asociado a una multigráfica codificada (G_π, τ) y determinado por el Δ -matroide $\mathcal{D}_{H_b(\Phi, \Psi)}$ donde el vector Euleriano Φ induce el paseo Euleriano $TE(\Phi)$ con el orden de los pares de bases en el cromosoma representado por la permutación expandida π' (ver subsección 3.5.2) y Ψ induce la partición Euleriana P_Ψ que contiene el circuito correspondiente a la permutación identidad y los circuitos formados por ADN no genético. Sea $X \subseteq V$, en la sección 2.4 el torcimiento \mathcal{D}^*X se definió como el Δ -matroide $(V, \{F \Delta X : F \in \mathcal{F}\})$.

Teorema 4.1 (Brijder 2017, [26]). *Sea una gráfica H_b y sea $v \in V(H_b)$ un vértice con lazo. Entonces $\mathcal{D}_{H_b *^{ns} v} = \mathcal{D}_{H_b} *^{ns} v$.*

El teorema 4.1 y la propiedad 1 permiten definir la inversión orientada como la operación de contracción descrita en la subsección 2.4.4.

$$\mathcal{D}_{H_b *^{inv} v} = (\mathcal{D}_{H_b} \Delta F) \setminus v \quad \text{si } v \in F \quad (49)$$

Ejemplo 45

Sea $\mathcal{F}_{H_b(\Phi, \Psi)}$ en (50) la familia de conjuntos factibles del Δ -matroide $\mathcal{D}_{H_b(\Phi, \Psi)}$ que determina a la gráfica simple con lazos permitidos de la figura 3.31.

$$\begin{aligned}
\mathcal{F}_{H_b(\Phi, \Psi)} = & \{\emptyset, \{1\}, \{2\}, \{0, 1\}, \{0, 7\}, \{0, 8\}, \{0, 9\}, \{1, 2\}, \{2, 7\}, \{2, 8\}, \{2, 9\} \\
& \{4, 5\}, \{4, 6\}, \{5, 6\}, \{7, 8\}, \{7, 9\}, \{8, 9\}, \{0, 1, 2\}, \{0, 1, 7\}, \{0, 1, 8\}, \\
& \{0, 1, 9\}, \{0, 2, 7\}, \{0, 2, 8\}, \dots, \{0, 1, 2, 4, 5, 7, 8\}, \{0, 1, 2, 5, 6, 7, 8\}, \\
& \{0, 1, 2, 4, 5, 7, 9\}, \{0, 1, 2, 4, 5, 8, 9\}, \{0, 1, 4, 5, 7, 8, 9\}, \{0, 1, 4, 6, 7, 8, 9\} \\
& \{0, 1, 5, 6, 7, 8, 9\}, \{0, 2, 4, 5, 7, 8, 9\}, \{0, 2, 4, 6, 7, 8, 9\}, \{0, 2, 5, 6, 7, 8, 9\}, \\
& \{1, 2, 4, 5, 7, 8, 9\}, \{1, 2, 4, 6, 7, 8, 9\}, \{1, 2, 4, 5, 6, 8, 9\}\}
\end{aligned} \tag{50}$$

La operación $(\mathcal{D}_{H_b(\Phi, \Psi)} \triangle \{2\}) \setminus 2$ en (51) corresponde a la inversión sobre el vértice v_2 de $H_b(\Phi, \Psi)$ (vértice (4, 5) en el ejemplo 27).

$$\begin{aligned}
(\mathcal{F}_{H_b(\Phi, \Psi)} \triangle \{2\}) \setminus 2 = & \{\emptyset, \{1\}, \{7\}, \{8\}, \{9\}, \{0, 1\}, \{0, 7\}, \{0, 8\}, \dots, \{0, 1, 4, 5, 7, 8\}, \\
& \{0, 1, 5, 6, 7, 8\}, \{0, 1, 4, 5, 7, 9\}, \{0, 1, 4, 5, 8, 9\}, \\
& \{0, 4, 5, 7, 8, 9\}, \{0, 4, 6, 7, 8, 9\}, \{0, 5, 6, 7, 8, 9\}, \\
& \{1, 4, 5, 7, 8, 9\}, \{1, 4, 6, 7, 8, 9\}, \{1, 4, 5, 6, 8, 9\}\}
\end{aligned} \tag{51}$$

Una secuencia (v_1, v_2, \dots, v_k) de elementos distintos de V es una *secuencia óptima* para $\mathcal{D}_{H_b(\Phi, \Psi)}$ si para todo $i \in \{0, \dots, k\}$, $\{v_1, \dots, v_i\} \in \mathcal{F}_{H_b(\Phi, \Psi)}$. En particular $\emptyset, \{v_1, \dots, v_k\} \in \mathcal{F}_{H_b(\Phi, \Psi)}$ [26]. Sea $\max(\mathcal{F}_{H_b(\Phi, \Psi)})$ la familia de conjuntos maximales de $\mathcal{F}_{H_b(\Phi, \Psi)}$ con respecto a inclusión. Una secuencia (v_1, v_2, \dots, v_k) es *óptima* si $\{v_1, \dots, v_k\} \in \max(\mathcal{F}_{H_b(\Phi, \Psi)})$ [26]. El teorema 4.2 aplica el concepto de conexidad de Δ -matroides definido en la subsección 2.4.5, en la determinación de secuencias óptimas.

Teorema 4.2 (Brijder 2017, [26]). *Sea \mathcal{D} un Δ -matroide binario normal. Entonces existe una secuencia óptima para \mathcal{D} si y sólo si cada sumando conexo par de \mathcal{D} con conjunto base no vacío es de la forma $(\{v\}, \emptyset)$ para algún elemento v en el conjunto base de \mathcal{D} .*

Sea π una permutación signada sin obstáculos y sea $\mathcal{D}_{H_b(\Phi, \Psi)}$ el Δ -matroide que determina la gráfica de intersección de π . Entonces el $\mathcal{D}_{H_b(\Phi, \Psi)}$ cumple con lo postulado en el teorema 4.2 y la distancia de inversión para π en términos de Δ -matroides queda expresada en (52).

$$d_r(\pi, \iota) = \{|F| : F \in \max(\mathcal{F}_{H_b(\Phi, \Psi)})\} \tag{52}$$

4.2. Inversión de corte

El teorema 4.3 relaciona el complemento modificado (ver subsección 3.4.4) con la conmutación definida en la sección 3.6).

Teorema 4.3 (Traldi 2012, [94]). (Modificado) Sea $A(\Phi, \Psi)$ la matriz de adyacencia de la gráfica $H_b(\Phi, \Psi)$ y sea v un vértice de la multigráfica 4-regular G .

$$M_{mod}^v(\Phi, \Psi) = A(\Phi^*v, \Psi)$$

Si $H_b(\Phi, \Psi)$ es la matriz de alternancia relativa de P_Ψ con respecto a $TE(\Phi)$ y $X \subseteq V$ es un subconjunto X tal que $A(H_b(\Phi, \Psi))[X]$ es un obstáculo, entonces de acuerdo al teorema 4.3 es posible definir la inversión de corte descrita en las subsecciones 3.2.5 y 3.3.4 como la conmutación del paseo Euleriano $TE(\Phi)$ en el vértice $v_i \in X$.

Ejemplo 46

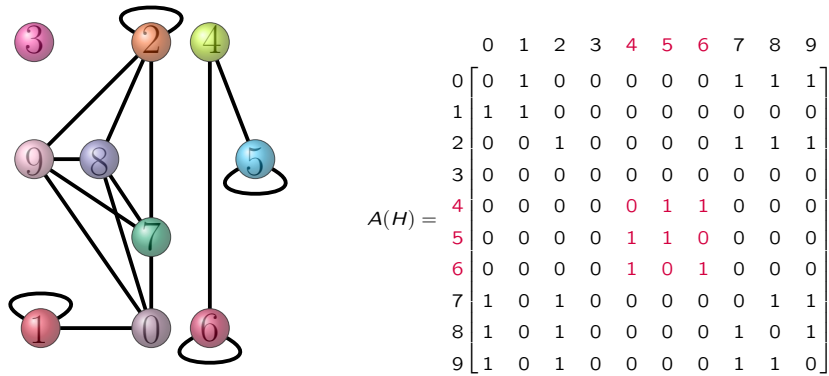


Figura 4.1: Gráfica de alternancia del paseo Euleriano $TE(\Phi)^*v_4$ del ejemplo 41.

La matriz de adyacencia de gráfica ilustrada en la figura 4.1 corresponde a la matriz (32) del ejemplo 32.

4.3. Ecuación exacta para la distancia de inversión

Sea $\mathcal{D}_{H_b(\Phi, \Psi)}$ un Δ -matroide que no cumple con la condición del teorema 4.2. Entonces la fórmula exacta para la distancia de inversión del teorema 3.2 queda expresada en (53).

$$d_r(\pi, \iota) = |F| + h(\pi) + f(\pi) \quad \forall F \in \max(\mathcal{F}_{H_b(\Phi, \Psi)}) \tag{53}$$

Ejemplo 47

Sea $\mathcal{D}_{H_b(\Phi, \Psi)}$ el Δ -matroide del ejemplo 45. La cardinalidad de los miembros de $\max(\mathcal{F}_{H_b(\Phi, \Psi)})$ es 7. De acuerdo con la ecuación (53) la distancia de inversión para la permutación signada del ejemplo 16 es $d_r(\pi, \iota) = 7 + 1 + 0$.

4.4. Ciliados

En la subsección 3.1.3 se mencionó que las operaciones de ensamblamiento genético en ciliados ordenan los fragmentos del gen micronuclear para obtener la forma macronuclear del gen. Los genes no presentan las cubiertas protectoras del cromosoma lineal por lo cuál no es necesario añadir los extremos 0 y $n+1$. Está representación sin telómeros corresponde también a los cromosomas circulares presentes en las bacterias [54], de modo que el modelo que exponemos en este trabajo se puede emplear en cromosomas lineales, en cromosomas circulares y en genes.

La multigráfica 4-regular asociada a la permutación que representa el orden de los segmentos MDS en el gen micronuclear, es muy similar a la descrita en la subsección 3.5.1 en este caso se eliminan los punteros en los extremos de los MDSs M_0 y M_{10} y se fusionan sus segmentos adyacentes. Dicha fusión produce aristas mixtas (ejemplo 48).

Ejemplo 48

Sea ϖ' en (54) la permutación extendida correspondiente al gen Actina I en el *Sterkiella Nova* del ejemplo 14. La multigráfica 4-regular asociada se ilustra en la figura 4.2.

$$\varpi' = I_0 2_b M_3 3_a I_1 3_b M_4 4_a I_2 5_b M_6 6_a I_3 4_b M_5 5_a I_4 6_b M_7 7_a I_5 8_b M_9 M_9 I_6 2_a \overline{M_2} 1_b I_7 M_1 1_a I_8 7_b M_8 8_a \quad (54)$$

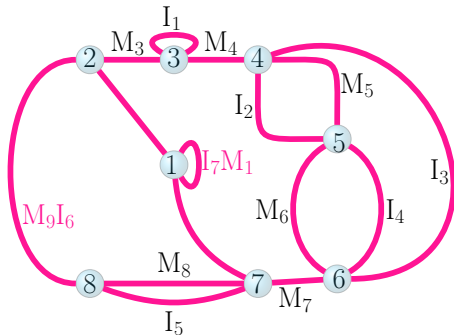


Figura 4.2: Multigráfica 4-regular G_{ϖ} de la permutación extendida ϖ' en (54).

El resto del modelo funciona de forma análoga a lo ya descrito.

Conclusiones y trabajo futuro

5.1. Conclusiones

Este trabajo presenta un enfoque combinatorio al problema biológico conocido como problema de la distancia de inversión el cuál estudia las relaciones evolutivas entre genomas mediante la comparación del orden cromosómico. Se demostró, que dos cromosomas sobre el mismo número de genes se pueden representar como los vectores suplementarios Φ y Ψ del sistema de isotropía $\mathcal{L}(H_b(\Phi, \Psi))$ con gráfica fundamental $H_b(\Phi, \Psi)$.

Sea π una permutación con obstáculos y sea v_i el vértice asociado al elemento i dentro de un obstáculo $i, \pi_{j+1}, \pi_{j+2}, \dots, i+1, \dots, \pi_{j+k-1}, i+k$ en π . Sea Π la presentación gráfica de $\mathcal{L}(H_b(\Phi, \Psi))$ y sea Π' la presentación gráfica asociada al vector Euleriano Φ^*v_i . En esta tesis se concluye que la matriz de adyacencia de $H_b(\Phi, \Psi)^*v_i$ corresponde a la inversión de corte en la permutación signada π .

El sistema de isotropía $\mathcal{L}(H_b(\Phi, \Psi))$ queda determinado por el Δ -matroide $\mathcal{D}_{H_b(\Phi, \Psi)}$ cuyo rango equivale a la distancia de inversión del cromosoma abstraído por el vector Euleriano Φ .

5.2. Trabajo futuro

Enriquecer el modelo que se maneja en este trabajo, introduciendo el reordenamiento de transposición que es equivalente a la complementación de aristas en la gráfica $H_b(\Phi, \Psi)$.

Dado que el problema de programación entera planteado en la tesis es polinomial, éste puede relajarse a un problema de programación lineal, mostrando que los vértices son enteros y que corresponden a las soluciones de programación entera. Será deseable, en un trabajo posterior, hacer la relajación mencionada y desarrollar un algoritmo para implementar la

solución al problema de la distancia de inversión.

Apéndice A

Sistemas de isotropía

Sea K un espacio vectorial de dimensión 2 sobre $GF(2)$ con la forma bilineal, antisimétrica, no degenerada $(a \wedge b)$, tal que $(a \wedge b = 1 \iff 0 \neq a \neq b \neq 0)$. 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 $(A, B) \rightarrow A \wedge B = \sum (A_v \wedge B_v : v \in V)$. Si L es un subespacio totalmente isotrópico (esto es, $A, B \in L \implies A \wedge B = 0$) de K^V 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 $A \in K^V$ y un subconjunto $X \subseteq V$, el vector $Y \in K^V$ definido por $Y_v = A_v$ si $v \in X$ y $Y_v = 0$ si $v \notin X$, se denotará con AX . Sea $\hat{A} = \{AX : X \subseteq V\}$, \hat{A} es un subespacio de K^V y la dimensión de \hat{A} es igual al número de elementos v tales que $A_v \neq 0$.

Sea $K' = K \setminus \{0\}$, se dice que A es completo si $A \in K'^V$, tal que $\dim(\hat{A}) = |V|$. Dos vectores completos A y B son *suplementarios* si $A_v \neq B_v \quad \forall v \in V$, esto es si \hat{A} y \hat{B} son subespacios suplementarios de K^V . El *rango* de un vector $A \in K^V$ con respecto a un sistema de isotropía $S = (L, V)$ se define como $r(A) = \dim(\hat{A} \cap L)$. Un *vector Euleriano* de S es un vector $A \in K'^V$ tal que $r(A) = 0$ [11, 19].

.0.1. Presentaciones gráficas de sistemas de isotropía

Sea H una gráfica simple sobre el conjunto de vértices $V(G)$ y el conjunto de aristas E . El conjunto potencia $\mathcal{P}(V)$ se considera un espacio vectorial sobre $GF(2)$ con la suma $(P, Q) \rightarrow P + Q = P \Delta Q$. Además se considera la forma bilineal $(P, Q) \rightarrow P \wedge Q = P \cap Q$ sobre $\mathcal{P}(V)$ [19].

Teorema .1 (Bouchet 1988, [19]). Sean A y B vectores suplementarios de K'^V y sea $L = \{AX + Bn(X) : X \subseteq V\}$. Entonces $S = (L, V)$ es un sistema de isotropía. La tripleta $\Pi = (H, A, B)$ se conoce como *presentación gráfica* de S .

Se dice que la gráfica H es una *gráfica fundamental* de S .

Propiedad 2 (Bouchet 1990, [15]). *Para todo vector Euleriano A de un sistema de isotropía $S = (L, V)$ existe exactamente una presentación gráfica $\Pi = (H, A, B)$ de S . Si $v \in V$, $A' = A * v$ y $\Pi' = (H', A', B')$ es la presentación gráfica asociada al vector Euleriano A' , entonces $H' = H * v$.*

La propiedad 2 implica que cualquier clase de equivalencia local es el conjunto de gráficas fundamentales de algún sistema de isotropía. La propiedad 3, enuncia la relación entre las gráficas de alternancia y las gráficas fundamentales de un sistema de isotropía

Propiedad 3 (Bouchet 1990, [15]). *Sea $S = (L, V)$ el sistema gráfico asociado a una multigráfica codificada (G, τ) , y sea A la codificación de un paseo Euleriano TE de G . Si $\Pi = (H, A, B)$ es la presentación gráfica de S asociada a A , entonces H es la gráfica de alternancia de TE .*

Bibliografía

- [1] Adamyk, K. L. M., Holmes, E., Mayfield, G. R., Moritz, D. J., Scheepers, M., Tenner, B. E., and Wauck, H. C. *Sorting Permutations: Games, Genomes, and Cycles*. arXiv preprint arXiv:1410.2353. (2014).
- [2] Arratia, R., Bollobás, B., and Sorkin, G. B. *The interlace polynomial of a graph*. *Journal of Combinatorial Theory, Series B*, 92(2), 199-233.(2004).
- [3] V. Bafna and P. A. Pevzner. *Genome rearrangements and sorting by reversals*. *SIAM Journal on Computing*, 25(2):272–289, (1996).
- [4] Bergeron, A., Mixtacki, J., and Stoye, J. (2006, September). *A unifying view of genome rearrangements*. In *International Workshop on Algorithms in Bioinformatics* (pp. 163-173). Springer Berlin Heidelberg.
- [5] Bergeron, A. *A very elementary presentation of the Hannenhalli–Pevzner theory*. *Discrete Applied Mathematics*, 146(2), 134-145. (2005).
- [6] Bouchet, A. *Graph polynomials derived from Tutte–Martin polynomials*. *Discrete Mathematics*, 302(1), 32-38. (2005).
- [7] Bouchet, A. *Multimatroids III. Tightness and fundamental graphs*, *European Journal of Combinatorics*, 22(5), 657-677. (2001).
- [8] Bouchet, A., and Ghier, L. *Connectivity and β -invariants of isotropic systems and 4-regular graphs*. *Discrete Mathematics*, 161(1-3), 25-44. (1996).
- [9] Bouchet, A., and Cunningham, W. H. *Delta-matroids, jump systems, and bisubmodular polyhedra*, *SIAM Journal on Discrete Mathematics*, 8(1), 17-32. (1995).
- [10] Bouchet, A. *Circle graph obstructions*. *Journal of Combinatorial Theory, Series B*, 60(1), 107-144. (1994).

-
- [11] Bouchet, A. *Compatible Euler tours and supplementary Eulerian vectors*. European journal of combinatorics, vol. 14, no 6, p. 513-520, 1993.
- [12] Bouchet, A. *A characterization of unimodular orientations of simple graphs*. Journal of Combinatorial Theory, Series B, 56(1), 45-54. (1992).
- [13] Bouchet, A., and Duchamp, A. *Representability of Δ -matroids over GF (2)*. Linear Algebra and its Applications, 146, 67-78. (1991).
- [14] Bouchet, A. *Tutte-Martin polynomials and orienting vectors of isotropic systems*. Graphs and Combinatorics, 7(3), 235-252. (1991).
- [15] Bouchet, A. *k-transformations, local complementations and switching*. In Cycles and Rays (pp. 41-50). Springer Netherlands. (1990).
- [16] Bouchet, A. *Isotropic systems*. European J. Combin. 8 (1987) 231-244.
- [17] Bouchet, A. *Matchings and Δ -matroids*. Discrete Applied Mathematics, 24(1-3), 55-62. (1989).
- [18] Bouchet, A. *Maps and Δ -matroids*. Discrete Mathematics, 78(1), 59-71. (1989).
- [19] Bouchet, A. *Graphic presentations of isotropic systems*. Journal of Combinatorial Theory, Series B, 45(1), 58-76. (1988).
- [20] Bouchet, A. *Greedy Algorithm and Symmetric Matroids*, Math. Program. 38, 2 (September 1987), 147-159.
- [21] Bouchet, A. *Unimodularity and circle graphs*. Discrete mathematics, 66(1-2), 203-208. (1987).
- [22] Bouchet, A. *Digraph decompositions and Eulerian systems*. SIAM Journal on Algebraic Discrete Methods, 8(3), 323-337. (1987).
- [23] Bouchet, A. *Representability of Δ -matroids*, Colloquia Mathematica Societatis János Bolyai 52. Combinatorics, Eger (Hungary), 167-182., 1987.
- [24] A. Bouchet, *Caractérisation des symboles croisés de genre nul*, C. R. Acad. Sci. Paris Sér. A-B 274 (1972) A724-A727.
- [25] Braga, M. D., Sagot, M. F., Scornavacca, C., and Tannier, E. *Exploring the solution space of sorting by reversals, with experiments and an application to evolution*. IEEE/ACM Transactions on Compu-

- tational Biology and Bioinformatics (TCBB), 5(3), 348-356. (2008).
- [26] Brijder, R. *Sorting by Reversals and the Theory of 4-Regular Graphs*. arXiv preprint arXiv:1701.07463. (2017).
- [27] Brijder, R., Traldi, L. *A characterization of circle graphs in terms of multimatroid representations*. arXiv eprint arXiv:1703.05960. (2017).
- [28] Brijder, R. *Orienting transversals and transition polynomials of multimatroids*. arXiv preprint arXiv:1605.04244.(2016).
- [29] Traldi, L., and Brijder, R. *Isotropic matroids. III. Connectivity*. arXiv preprint arXiv:1602.03899. (2016).
- [30] Brijder, R. and Traldi, L. *Isotropic matroids. II. Circle graphs*, ArXiv e-prints, 1504.04299, (2015).
- [31] Brijder, R. *Recombination faults in gene assembly in ciliates modeled using multimatroids*. Theoretical Computer Science, 608, 27-35. (2015).
- [32] Brijder, R., and Hoogeboom, H. J. *Interlace polynomials for multimatroids and delta-matroids*. European Journal of Combinatorics, 40, 142-167. (2014).
- [33] Brijder, R., and Hoogeboom, H. J. *The algebra of gene assembly in ciliates*. In Discrete and Topological Models in Molecular Biology (pp. 289-307). Springer Berlin Heidelberg. (2014).
- [34] Brijder, R., and Hoogeboom, H. J. *Binary symmetric matrix inversion through local complementation*. Fundamenta Informaticae, 116(1-4), 15-23. (2012).
- [35] Brijder, R., and Hoogeboom, H. J. *Nullity invariance for pivot and the interlace polynomial*. Linear Algebra and its Applications, 435(2), 277-288. (2011).
- [36] Brijder, R., Hoogeboom, H. J., and Traldi, L. *The adjacency matroid of a graph*. arXiv preprint arXiv:1107.5493. (2011).
- [37] Brijder, R., and Hoogeboom, H. J. *The group structure of pivot and loop complementation on graphs and set systems*. European Journal of Combinatorics, 32(8), 1353-1367. (2011).
- [38] Brijder, R., and Hoogeboom, H. J. *Maximal pivots on graphs with an application to gene assembly*. Discrete Applied Mathematics, 158(18), 1977-1985. (2010).
- [39] Brijder, R., Hoogeboom, H. J., and Muskulus, M. *Strategies of Loop*

- recombination in ciliates*. Discrete Applied Mathematics, 156(10), 1736-1753. (2008).
- [40] Caprara, A. *Sorting permutations by reversals and Eulerian cycle decompositions*. SIAM journal on discrete mathematics, 12(1), 91-110. (1999).
- [41] Chen, X., Bracht, J. R., Goldman, A. D., Dolzhenko, E., Clay, D. M., Swart, E. C., ... and Sebra, R. P. *The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development*. Cell, 158(5), 1187-1198. (2014).
- [42] Chun, C., Chun, D., and Noble, S. D. *Inductive tools for connected delta-matroids and multimatroids*. European Journal of Combinatorics, 63, 59-69. (2017).
- [43] Cooper, J., and Davis, J. *Successful pressing sequences for a bicolored graph and binary matrices*. Linear Algebra and its Applications, 490, 162-173. (2016).
- [44] Drillon, Guénola, Alessandra Carbone, and Gilles Fischer. *Combinatorics of chromosomal rearrangements based on syntenic blocks and syntenic packs*. Journal of Logic and Computation 23.4 (2011): 815-838.
- [45] Ehrenfeucht, A., Harju, T., Petre, I., Prescott, D. M., and Rozenberg, G. *Gene assembly through cyclic graph decomposition*. In Computation in Living Cells (pp. 151-175). Springer Berlin Heidelberg. (2004).
- [46] Ehrenfeucht, A., Harju, T., Petre, I., Prescott, D. M., and Rozenberg, G. *Computation in living cells: gene assembly in ciliates*. Springer Science and Business Media. (2003).
- [47] Ellis-Monaghan, J. A. *Transition polynomials, double covers, and biomolecular computing*. Congressus Numerantium, 166, 181. (2004).
- [48] Ellis-Monaghan, J. A., and Sarmiento, I. *Generalized transition polynomials*. Congressus Numerantium, 57-70. (2002).
- [49] Fleming, T., and Mellor, B. *Chord Diagrams and Gauss Codes for Graphs*. arXiv preprint math/0508269. (2005).
- [50] Frausto D., Saucedo O., and Valenzuela E., *Paquete computacional para la investigación en Δ -matroides*. UAM 2017.
- [51] Geelen, J. F., Iwata, S., and Murota, K. *The linear delta-matroid parity problem*. Journal of Combinatorial Theory, Series B, 88(2),

- 377-398. (2003).
- [52] Genest, F. *On the complementation orbits of graphs*.
- [53] Gordon, G., and McNulty, J. *Matroids: a geometric introduction*. Cambridge University Press. (2012)
- [54] Griffiths AJF, Gelbart WM, Miller JH, et al. *Modern Genetic Analysis*. New York: W. H. Freeman; 1999. Chromosomal Rearrangements. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21367/>
- [55] Grusea, S., and Labarre, A. *The distribution of cycles in breakpoint graphs of signed permutations*. Discrete Applied Mathematics, 161(10), 1448-1466. (2013).
- [56] Hannenhalli, S., and Pevzner, P. A. *Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals*. Journal of the ACM (JACM), 46(1), 1-27. (1999).
- [57] Hannenhalli, S., and Pevzner, P. A. (1996, January). *To Cut... or Not to Cut (Applications of Comparative Physical Maps in Molecular Evolution)*. In SODA (Vol. 7, pp. 304-313).
- [58] Hannenhalli, S. *Polynomial-time algorithm for computing translocation distance between genomes*. Discrete applied mathematics, 71(1-3), 137-151. (1996).
- [59] Hannenhalli, S., and Pevzner, P. A. (1995, October). *Transforming men into mice (polynomial algorithm for genomic distance problem)*. In Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on (pp. 581-592). IEEE.
- [60] Hartman, T., and Verbin, E. *Matrix tightness: A linear-algebraic framework for sorting by transpositions*. In String Processing and Information Retrieval (pp. 279-290). Springer Berlin/Heidelberg.(2006).
- [61] Herlin, J. L., Nelson, A., and Scheepers, M. *Using ciliate operations to construct chromosome phylogenies*. Involve, a Journal of Mathematics, 9(1), 1-26. (2015).
- [62] Jaeger, F. *Symmetric representations of binary matroids*. North-Holland Mathematics Studies, 75, 371-376. (1983).
- [63] Jansen, C. L., Scheepers, M., Simon, S. L., and Tatum, E. *Context Directed Reversals and the Ciliate Decryptome*. arXiv preprint arXiv:1603.06149. (2016).
- [64] Jackson, B. *A characterisation of graphs having three pairwise compa-*

- tible Euler tours*. Journal of Combinatorial Theory, Series B, 53(1), 80-92. (1991).
- [65] Jiménez Garcia, Tapia-Vieyra, and Mas-Oliva. *El espliceosoma: Corte y empalme del pre-ARNm.* Revista de Educación Bioquímica 23 (2004): 59-63.
- [66] Kari, L., and Mahalingam, K. *Watson-Crick palindromes in DNA computing*. Natural Computing, 9(2), 297-316. (2010).
- [67] Kari, L., and Mahalingam, K. *Watson-Crick conjugate and commutative words*. In International Workshop on DNA-Based Computers (pp. 273-283). Springer Berlin Heidelberg. (2007, June).
- [68] Kaplan, H., Shamir, R., and Tarjan, R. E. *A faster and simpler algorithm for sorting signed permutations by reversals*. SIAM Journal on Computing, 29(3), 880-892. (2000).
- [69] Kececioglu, J., and Sankoff, D. *Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement*. Algorithmica, 13(1), 180-210. (1995).
- [70] Kececioglu, J., and Sankoff, D. *Efficient bounds for oriented chromosome inversion distance*. In Combinatorial Pattern Matching (pp. 307-325). Springer Berlin/Heidelberg. (1994).
- [71] Kotzig, A. *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)
- [72] Lancia, G., Rinaldi, F., and Serafini, P. (2015, April). *A Unified Integer Programming Model for Genome Rearrangement Problems*. In IWBBIO (1) (pp. 491-502).
- [73] Lancia, A.C.G., and Ng, S.K. *A column-generation based branch-and-bound algorithm for sorting by reversals*. Mathematical Support for Molecular Biology: Papers Related to the Special Year in Mathematical Support for Molecular Biology, 1994-1998, 47, 213. (1999).
- [74] Lehninger, A.L. and Cox, M.M. *Principios de bioquímica*, isbn:9788428214100, Omega. (2006).
- [75] Liu, C.L. *Elementos de Matemáticas Discretas*. Mc Graw Hill. (1985)
- [76] Mack, Katya L., and Michael W. Nachman. *"Gene regulation and speciation."* Trends in Genetics (2016).

- [77] Michel, B., and Leach, D. *Homologous Recombination-Enzymes and Pathways*. EcoSal Plus, 5(1). (2012).
- [78] Miklós, I., and Smith, H. *Sampling and counting genome rearrangement scenarios*. BMC bioinformatics, 16(14), S6. (2015).
- [79] Miranda-Rodriguez, J.R. Salas-Vidal, E. Lomeli, H. Zurita, M. Schnabel, D. *RhoA/ROCK pathway activity is essential for the correct localization of the germ plasm mRNAs in zebrafish embryos* Developmental Biology, 421, 27-42. (2017)
- [80] Pevzner, P. A. *Computational Molecular Biology*.
- [81] Pevzner, P., and Tesler, G. *Genome rearrangements in mammalian evolution: lessons from human and mouse genomes*. Genome research, 13(1), 37-45. (2003).
- [82] Prescott, David M. "The DNA of ciliated protozoa." Microbiological reviews 58.2 (1994): 233-267.
- [83] Prescott, David M., Andrzej Ehrenfeucht, and Grzegorz Rozenberg. "Molecular operations for DNA processing in hypotrichous ciliates." European Journal of Protistology 37.3 (2001): 241-260.
- [84] Rodríguez Sánchez, M. G. *Delta-matroides rueda ternarios*. Universidad Autónoma Metropolitana, Departamento de Ciencias Básicas, México, Distrito Federal. México. [Morfismos, México, 2002 Vól. 6 Núm. 1 Jun, Pág. 31-55]
- [85] R. C. Read, P. Rosenstiehl, *On the Gauss crossing problem*, in: Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai, 18, North-Holland, Amsterdam-New York, 1978, pp. 843-876.
- [86] Whitney, H. *On the abstract properties of linear dependence*. American Journal of Mathematics, 57(3), 509-533. (1935).
- [87] Oxley, J. *Matroid Theory*, Oxford Univ. Press, Oxford, (2011).
- [88] Sankoff, D., Cedergren, R., and Abel, Y. *Genomic divergence through gene rearrangement*. Methods in enzymology, 183, 428-438. (1990).
- [89] Sankoff, D. *Edit distance for genome comparison based on non-local operations*. In Combinatorial Pattern Matching (pp. 121-135). Springer Berlin/Heidelberg. (1992).
- [90] Traldi, L. *Circuit partitions and signed interlacement in 4-regular graphs*. arXiv preprint arXiv:1607.04233. (2016).

- [91] Traldi, L. *Binary matroids and local complementation*. European Journal of Combinatorics, 45, 21-40. (2015).
- [92] Traldi, L. *The transition matroid of a 4-regular graph: an introduction*. European Journal of Combinatorics, 50, 180-207. (2015).
- [93] Traldi, L. *On the linear algebra of local complementation*. Linear Algebra and Its Applications, 436(5), 1072-1089. (2012).
- [94] Traldi, L. *Interlacement in 4-regular graphs: a new approach using nonsymmetric matrices*. arXiv preprint arXiv:1204.0482. (2012).
- [95] Traldi, L. *Binary nullity, Euler circuits and interlace polynomials*. European Journal of Combinatorics, 32(6), 944-950. (2011).
- [96] Tsang, O. W. *Sorting by Reversals in Genome Rearrangements*. (1996)
- [97] Truemper, K. *Matroid decomposition* (Vol. 6). Boston: Academic Press. (1992).
- [98] Tucker, A.W. *A combinatorial equivalence of matrices*, in: Combinatorial Analysis, Proceedings of Symposia in Applied Mathematics, vol. X, American Mathematical Society, 1960, pp. 129-140.
- [99] Watterson, G.A. et al. *The chromosome inversion problem*. J. Theor. Biol., 99,1-7. (1982)
- [100] Yin, Z., Tang, J., Schaeffer, S.W., and Bader, D.A. *Streaming breakpoint graph analytics for accelerating and parallelizing the computation of dcj median of three genomes*. Procedia Computer Science, 18, 561-570. (2013).
- [101] Zerbino, D.R., Ballinger, T., Paten, B., Hickey, G., and Hausler, D. *Representing and decomposing genomic structural variants as balanced integer flows on sequence graphs*. BMC bioinformatics, 17(1), 400. (2016).
- [102] <https://s-media-cache-ak0.pinning.com/originals/05/26/d9/0526d9ec3520afd8eb4a0d4ab63b2db3.jpg>
- [103] https://www.researchgate.net/figure/255736580_fig1_A-From-left-to-right-Oxytricha-trifallax-Paramecium-tetraurelia-Tetrahymena
- [104] <https://es.khanacademy.org/science/biology/dna-as-the-genetic-material/dna-discovery-and-structure/a/discovery-of-the-structure-of-dna>

Índice alfabético

- Δ -matroides, 3, 55, 57, 59, 60
 - Δ -matroide binario, 12
 - Δ -matroide normal, 12
 - Δ -matroide par, 12
 - Axioma de Intercambio
 - Simétrico, 11
 - Conexidad, 14
 - Familia de conjuntos
 - maximales, 60
 - Gráfica fundamental, 12, 13
 - Menores, 13
 - Operador diferencia simétrica, 11
 - Representación lineal, 12
 - Sistema de conjuntos, 11
 - Sistema de conjuntos propio, 11
 - Torcimiento, 12
 - Transformación de pivote principal, 4, 13, 18
- ADN, 1, 21
 - Transcripción, 22
 - Alfabeto, 14
 - Alfabeto complementario, 14, 22
 - Alfabeto de MDSs, 30
 - Alfabeto de nucleótidos, 1, 22
 - Antecesor común, 2, 24
 - ARN, 1, 21, 23
 - Corte y empalme, 23
 - Bafna V., 3
 - Bases nitrogenadas, 1, 21, 31
 - Úracilo, 21
 - Adenina, 1, 21, 22
 - Citosina, 1, 21, 22
 - Enlaces de hidrógeno, 22
 - Guanina, 1, 21, 22
 - Tiamina, 1, 21, 22
 - Bitransiciones, 48, 53
 - Codificación, 50, 51
 - Conjunto de, 48, 49
 - Corte, 48
 - Pares incidentes con un vértice, 50
 - Bloques de sintenia, 24
 - Bouchet A., 4, 52
 - Brijder R., 4
 - Campo binario $GF(2)$, 3, 8, 39, 55, 56
 - Suma y producto, 8
 - Ciclo transverso, 56
 - Ciliados, 26
 - Ensamblamiento genético, 26, 62
 - Macronúcleo, 26
 - Micronúcleo, 26
 - Puntero de entrada, 27
 - Puntero de salida, 27
 - Punteros, 27, 46, 62
 - Secuencias destinadas al

- macronúcleo MDSs, 27, 30, 46, 47, 62
- Secuencias internamente eliminadas IESs, 27, 29, 47
- Complementación Local
 - Complementación de lazo, 17, 54
 - Complementación local no simple, 18, 40
 - Complementación local simple, 3, 18
 - Complemento local modificado, 5, 44, 60
 - Inversión, 40
- Conjunto potencia, 56
- Crick F., 22
- Cromosoma, 1, 2, 59
 - Cromosoma circular, 62
 - Cromosoma lineal, 23, 32, 33, 46, 62
 - Representación, 31
 - Telómeros, 24
- Distancia de inversión, 2, 31
 - Cota inferior, 3, 35, 43
 - Ecuación exacta, 3, 39, 61
- Escenario evolutivo, 25
- Espacio ciclo, 56
- Espacio cociente, 56
- Espacio vectorial, 9, 56
- Forma bilineal, 56
- Gen, 1, 23, 33, 62
 - Codones, 22
 - Orientación, 23
- Genoma, 1, 2
 - Eucariota, 23
- Gráfica de alternancia, 15, 53
 - Gráfica de alternancia signada, 17
 - Relativa, 54, 55
- Gráfica de cubierta, 4
- Gráfica de intersección, 3, 15, 40, 41, 44
 - Diagrama cordal, 15, 41
 - Transposición, 17
- Gráfica de punto de rompimiento, 3, 33
 - Cardinalidad de una partición, 3, 34, 36, 45, 51
 - Ciclos alternantes, 33
 - Partición en ciclos, 3
- Gráfica de superposición, 3
- Gráfica dirigida, 47
- Gráfica simple con lazos permitidos, 3, 10, 11, 17, 18, 55, 57, 59
- Grupo de permutaciones, 11
- Hannenhalli S., 3
- Inversión, 2, 25, 30, 59
 - Inversión segura, 40
 - Pares orientados, 31
 - Secuencia óptima, 42, 60
- Involución, 14
- Isomorfismo compatible, 11, 17
- Kaplan H., 3
- Kececioglu, J., 3
- Kotzig A., 4, 52
- Matrices
 - Matrices de tipo simétrico, 3, 5, 12, 13, 45
 - Matriz de adyacencia, 3, 10, 12, 53, 55
 - Matriz de alternancia, 53
 - Matriz de alternancia relativa, 61
 - Matriz de intersección, 43
 - Matriz identidad, 10

- Matriz suma, 10
- Submatrices principales, 4, 12
- Matroide binario, 9, 11, 56
 - Conjunto base, 11
 - Subtransversales, 56
 - Transversales, 56
- Isomorfo, 10
- Representación estándar, 10
- Restringido, 11, 57
- Matroides, 8
 - Axioma de incremento de independencia, 9
 - Circuito fundamental, 9
 - Conjunto base, 9
 - Conjuntos dependientes, 9
 - Conjuntos independientes, 9
 - Familia de Bases, 9
 - Matroide binario, 4
 - Matroide lineal, 9
- Modelo de programación lineal entera, 36
- Modelo intramolecular de ensamblamiento genético, 28
- Multigráfica 4-regular, 4, 47, 48, 50
 - Asociada a una permutación signada, 46–48, 51, 52, 62
- Multigráfica codificada, 50–52, 54, 59
- Número de componentes, 48
- Partición Euleriana, 8, 48, 51, 54, 59
 - Cardinalidad de, 51
- Partición Euleriana cardinalidad de, 51
- Partición Euleriana, 55
- Paseo Euleriano, 8, 48, 51–53, 59
- Transformaciones en paseos Eulerianos, 4, 52
- Conmutación, 52, 53, 60
- Separación, 52
- Nucleótidos, 1, 21
- Obstáculos, 3, 32, 37
 - Fortaleza, 39
 - Inversión de corte, 4, 32, 38, 45, 60
 - Inversión de fusión, 4, 32, 38
 - Súper obstáculo, 39
 - Submatrices principales singulares, 44
- Operaciones de ADN en ciliados, 28
 - Escisión de lazo, 28
 - Recombinación de doble lazo, 28
 - Recombinación de horquilla, 28
- Ordenamiento por inversiones, 2
- Palabra de doble ocurrencia, 14, 53
 - Palabra de doble ocurrencia signada, 14
- Parsimonia máxima, 26
- Permutaciones, 3, 14
 - Permutación expandida, 46, 59
 - Permutación identidad, 3, 26, 59
 - Permutación imagen de una permutación signada, 33
 - Permutación signada, 3, 15, 26, 30
- Pevzner A., 3
- Problema del reordenamiento cromosómico, 2, 24, 26
- Puntos de rompimiento, 24, 25, 33
- Recombinación homóloga, 24

- Reordenamientos cromosómicos,
2, 24
- Sankoff, D., 3
- Shamir R., 3
- Sistemas de isotropía, 3, 50, 52,
53, 57
- Gráfica fundamental, 4, 53,
57
- Presentaciones gráficas, 54,
55, 59
- Sistemas gráficos, 50, 52
- Vector completo, 51, 54
- Vector completo rango de, 51
- Vector Euleriano, 52, 53, 59
- Vectores suplementarios, 4,
56
- Suma \boxplus , 56
- Tarjan R.E., 3
- Traldi L., 5
- Transformación T_1^{-1} , 41, 47
- Transformación T_1 , 33
- Transición, 48
- Transposición (Reordenamiento
cromosómico), 2, 25, 26
- Transposición invertida, 2, 25
- Tripleta de vértices $\{v_\phi, v_\chi, v_\psi\}$,
11, 56
- Tucker A.W, 4
- Vector relativo principal, 55
- Watson J., 22