

**ALGORITMO PARA EL APRENDIZAJE DE UNA  
RED BAYESIANA QUE REPRESENTA UNA RED  
DE REGULACIÓN DE LA EXPRESIÓN GENÉTICA  
BASADO EN OPTIMIZACIÓN POR COLONIA DE  
HORMIGAS**

ERWING FABIÁN CARDOZO OJEDA

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS  
BUCARAMANGA  
2012

# **ALGORITMO PARA EL APRENDIZAJE DE UNA RED BAYESIANA QUE REPRESENTA UNA RED DE REGULACIÓN DE LA EXPRESIÓN GENÉTICA BASADO EN OPTIMIZACIÓN POR COLONIA DE HORMIGAS**

ERWING FABIAN CARDOZO OJEDA

Monografía de grado para optar por el título de:  
Master en Ingeniería de Sistemas

Director:

HENRY ARGUELLO FUENTES, *PhD(c), Mpe.*

Codirector:

HERMÁN JOSÉ ARTEAGA, *MD., Phd.*

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS  
BUCARAMANGA  
2012

# Índice general

<b>Introducción</b>	<b>11</b>
<b>1. Aprendizaje de Redes Bayesianas</b>	<b>15</b>
1.1. Modelado Bayesiano para el Aprendizaje de Redes Bayesianas	15
1.2. Clases de Equivalencia de Redes Bayesianas	23
1.2.1. Representación de Clases de Equivalencia	24
1.2.2. Movimientos entre Clases de Equivalencia	25
1.3. Conclusiones	27
<b>2. Colonia de Hormigas en el Aprendizaje</b>	<b>28</b>
2.1. Colonia de Hormigas en el Aprendizaje	28
2.1.1. Representación del problema	29
2.1.2. Modelo de la Feromona	30
2.1.3. Información Heurística	30
2.1.4. Regla de Probabilidad	31
2.1.5. Descripción del Algoritmo	31
2.2. Evaluación y Resultados	35
2.3. Conclusiones	41
<b>3. Aproximación de una Red de Regulación de la Expresión Genética</b>	<b>44</b>
3.1. Generalidades de Biología molecular	45
3.2. Microarreglos de ADN	46
3.2.1. Selección de Genes Relevantes	48
3.2.2. Características del microarreglo utilizado	49
3.3. Obtención de una Red Semilla	50

---

3.4. Aproximación Red Biológica . . . . .	51
3.5. Conclusiones . . . . .	55
<b>4. Conclusiones</b>	<b>57</b>

# Índice de figuras

1.1. Ejemplo de la estructura y parámetros de una Red Bayesiana de tres nodos . . . . .	16
1.2. El problema de aprendizaje estructural, para el caso de 3 variables se reduce a encontrar la probabilidad $p(\mathbf{D} G)$ de que la matriz $\mathbf{D}$ de tres filas sea generado por $G$ . . . . .	18
1.3. Valores para $X_3$ y $\Xi_3$ en el caso 3 en $\mathbf{D}$ según $G$ , denotados como $\mathbf{D}_{33}$ y $\mathbf{D}_{\sigma_3 3}$ respectivamente, donde $\sigma_3 = (1, 2)$ . . . . .	19
1.4. Cómputo de $N_{ijk}$ para la configuración específica $X_3 = v_1$ y $\Xi_i = w_2$ . . . . .	21
2.1. El proceso de evaluación se puede resumir en cuatro pasos: (a), escoger una red bayesiana estándar; (b), generar un conjunto de datos a partir de dicha red; (c), por medio de un algoritmo aproximar una red bayesiana a partir de los datos; y (d) medir la cercanía de esta red con respecto a la inicial. . . . .	37
2.2. Ejemplo predicción de la estructura de la red bayesiana ASIA. (a), la red original, (b) la red aproximada por el algoritmo GES, (c) ACO-E y (d) el algoritmo propuesto en este artículo con 500 muestras. Los arcos rojos en el grafo aproximado por ACO-E son arcos que no corresponden al arco original, y los arcos en verde del grafo aproximado por el algoritmo propuesto son arcos los cuales uno de ellos no se encuentra en los grafos aproximados por los otros dos algoritmos. . . . .	38

---

2.3. Resultados significativos obtenidos en las métricas $S_f$ , $SHD$ y $NSC$ en el protocolo descrito para los tres algoritmos probados utilizando datos con 20 muestras. Los valores están normalizados.	40
2.4. Resultados significativos obtenidos en las métricas $S_f$ , $SHD$ y $NSC$ en el protocolo descrito para los tres algoritmos probados utilizando datos con 500 muestras. Los valores están normalizados. . . . .	42
3.1. Diferencia entre el nivel de expresión genética de la célula A con respecto a la B mayores a 2. . . . .	49
3.2. Resultados Interacciones obtenidas de la base de datos BioGrid.	51
3.3. Estructura inicial y final del proceso de aprendizaje con el algoritmo propuesto. . . . .	53
3.4. Comparación del valor de puntaje de la red semilla inicial y el valor de confianza estimado por la estructura inferida con el algoritmo propuesto. . . . .	54
3.5. Relación directa entre el gen KAT2B y Ccr3 . . . . .	54
3.6. Relación indirecta entre el gen SMAD3 y Pdc1 mediada por dos genes . . . . .	55
3.7. (a) Interacción aproximada entre el gen Ptpcap y KAT2B. (b) Relación inversa que posiblemente se encontraría experimentalmente. . . . .	55

# Resumen

TITULO: ALGORITMO PARA EL APRENDIZAJE DE UNA RED BAYESIANA QUE REPRESENTA UNA RED DE REGULACIÓN DE LA EXPRESIÓN GENÉTICA BASADO EN OPTIMIZACIÓN POR COLONIA DE HORMIGAS<sup>1</sup>

AUTOR: ERWING FABIÁN CARDOZO OJEDA <sup>2</sup>

PALABRAS CLAVE: Redes Bayesianas, Aprendizaje Estructural, Clases de Equivalencia, Colonia de Hormigas, Búsqueda Heurística.

La información biológica obtenida a partir de muestras del nivel de expresión genética, contiene gran cantidad de incertidumbre, pues implica información relativa a interacciones moleculares. Dicha información puede ser representada por medio de redes bayesianas y reconstruida a través de algoritmos heurísticos los cuales operan entre clases de equivalencia. Este documento propone un algoritmo de aprendizaje basado en búsqueda Greedy y modelos de búsqueda inspirados en hormigas competitivas. Los resultados mostraron que para redes pequeñas, el algoritmo puede obtener estructuras con menos interacciones erróneas que los algoritmos en la literatura, con mayor o igual probabilidad y con un orden de complejidad menor. Al aproximar una red biológica, los resultados para un microarreglo obtenido por muestras de pacientes con Leucemia Mieloide Aguda (LMA) aproximaron interacciones como potenciales interacciones biológicas directas, indirectas o inversas. Este último resultado es de gran importancia en la contribución de la generación de conocimiento en la investigación de genes que interactúan en relación con enfermedades tales como la AML

---

<sup>1</sup>Trabajo de Grado

<sup>2</sup>Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas. Director, Henry Arguello. Codirector, Hermán José Arteaga.

## Abstract

TITLE: LEARNING ALGORITHM FOR A BAYESIAN NETWORK THAT REPRESENTS A NETWORK OF GENETIC EXPRESSION REGULATION BASED ON ANT COLONY OPTIMIZATION <sup>1</sup>

AUTHOR:ERWING FABIÁN CARDOZO OJEDA <sup>2</sup>

KEYWORDS: Bayesian Networks, Structural Learning, Equivalence Classes, Ant Colony Techniques, Heuristic search.

The biological information obtained from samples of the level of expression of thousands of genes, contains a large amount of uncertainty because it involves information on molecular interactions. Such information can be represented using Bayesian networks and reconstructed through heuristic algorithms which operate between equivalence classes. This paper proposes a learning algorithm based on a Greedy search algorithm and models of competitive ant-inspired search. The results showed that the proposed algorithm, for small networks, it can get structures with fewer erroneous interactions than the algorithms in the literature, and in turn more likely or very similar and a lower order of complexity. To approximate a biological network, the results of the algorithm to a microarray samples obtained from patients with acute myeloid leukemia (AML) showed new interactions as potential biological interactions, direct, indirect or inverse. The latter result is of great importance in the contribution of the generation of knowledge in research and interacting genes related to diseases such as AML

---

<sup>1</sup>Research Work

<sup>2</sup>Faculty of Physical-Mechanical Engineering. Advisor, Henry Arguello. Co-advisor Hermán José Arteaga.

# Introducción

El surgimiento de tecnologías para obtener muestras de elementos de una célula a nivel molecular como los microarreglos de ADN -que miden la actividad genética de una célula-, ha traído consigo interés en inferir conocimiento estructural acerca de las interacciones de dichos elementos [1; 2; 3; 4]. Diversos tipos de análisis han sido realizados para obtener información de dichos datos y lograr así una interpretación [5], sin embargo, la mayoría no tienen en cuenta la estructura en las interacciones. Por otro lado, una interpretación consiste en suponer que los datos fueron generados por una estructura hipotética, asumiendo que los elementos muestreados hacen referencia a componentes que no están aislados sino que interactúan para formar una red con propiedades y funciones propias, y que se desea hallar. A este enfoque para analizar e interpretar datos biológicos de manera estructural se le denomina Biología de Sistemas [2; 6].

Dentro de diferentes modelos matemáticos que interpretan de manera estructural un fenómeno [5; 7; 8; 9; 10], los modelos estocásticos han sido ampliamente utilizados para representar el sistema de interacciones moleculares [6; 10; 11]. En estas aproximaciones se supone que el nivel de expresión de los genes que conforman la red tienen un grado de incertidumbre, por lo tanto pueden ser representados como variables aleatorias. De la misma forma, las relaciones entre los elementos, denominadas dependencias condicionales, tienen también un grado de incertidumbre, luego pueden ser descritas como probabilidades condicionales entre las variables aleatorias [6]. Así dada la incertidumbre presentada en los elementos y sus relaciones, un estado propio del sistema (red) se puede describir por una distribución de probabilidad con-

---

junta de las variables [6; 12; 13]. Una red bayesiana es un modelo estocástico que representa un sistema que cumple con las características mencionadas anteriormente donde las interacciones de los elementos conforman un digrafo acíclico y las probabilidades condicionales de sus relaciones sus parámetros. Luego el análisis e interpretación de los datos se define como el problema de encontrar una red bayesiana que mejor se ajuste a un conjunto de datos como es el de los microarreglos. Al conocer una red aproximada es posible inferir la probabilidad de que las variables se encuentren en algún estado, más específicamente se puede calcular la probabilidad de que un conjunto de genes tenga un cierto nivel de expresión, permitiendo hacer inferencias del comportamiento de la célula (en el caso de los microarreglos) a nivel genético [14; 15; 16; 17].

El proceso de encontrar una Red Bayesiana que se ajuste a un conjunto de datos es llamado *aprendizaje* y consiste en encontrar los parámetros y la estructura de la red [12; 13; 18; 19]. Obtener los parámetros, si se tiene conocimiento de la estructura, es una tarea sencilla [12], sin embargo, obtener la estructura es un problema NP-Duro [20] y se formaliza de la siguiente manera: dados los tres elementos  $(\mathbf{E}_B, S_f, M)$  donde  $\mathbf{E}_B$ , representa el conjunto de todas las estructuras posibles de redes bayesianas con  $n$  nodos;  $S_f$ , la *Función de Puntaje*, tal que  $S_f(G) : \mathbf{E}_B \rightarrow \mathbb{R}$  representa una medida para cada estructura  $\{G\}$  en  $\mathbf{E}_B$ ; y  $M$  es un *algoritmo* de búsqueda. Entonces el objetivo del aprendizaje consiste en encontrar la mejor estructura  $\hat{G}_n$  de una red bayesiana tal que:

$$\hat{G} = \max_{G \in \mathbf{E}_B} S_f(G). \quad (1)$$

Este enfoque es llamado *puntaje búsqueda*, y los algoritmos que han sido diseñados tienen en común que utilizan tres operadores para moverse dentro del espacio de búsqueda, esto es, *añadir*, *eliminar* o *invertir* un arco, y su diferencia radica en como se utilizan dichos operadores. Entre ellos están, el llamado *K2* [18] y sus extensiones [21; 22], algoritmos evolutivos [23; 24], colonias de hormigas [25], enjambre de partículas [26], sistemas inmunes artificiales [27] y uno de los más utilizados el llamado Hill Climbing [6; 19].

---

Además de lo anterior, el concepto de clases de equivalencia en redes bayesianas añade un componente para el problema de aprendizaje, ya que dos estructuras diferentes pueden describir la misma distribución de probabilidad [28; 29; 30; 31], y de esta forma se puede dividir  $\mathbf{E}_B$  en subconjuntos de redes equivalentes. La principal ventaja de buscar entre clases reside en que, dado un conjunto de operadores que permiten el paso de una clase a otra, se evita el movimiento dentro de una misma clase que desfavorece la búsqueda en el espacio  $\mathbf{E}_B$ , puesto que la función de puntaje  $S_f$  evalúa de la misma manera redes equivalentes [28; 29; 30; 31]. A pesar de la ventaja que tiene la búsqueda de clases de equivalencias (o más precisamente de grafos acíclicos parciales completos) la búsqueda no es trivial, puesto que el número de clases de equivalencia también crece en orden exponencial de acuerdo a como crece el número de nodos [32; 33; 34].

Entre los trabajos más recientes en el aprendizaje de redes bayesianas se destacan aquellos que utilizan Optimización basado en Colonia de Hormigas. Uno de ellos es el algoritmo llamado *ACO-E* propuesto en [35] que utiliza los operadores introducidos en [28] y otro el propuesto en [36] llamado *MMACO* que utiliza la búsqueda restringida utilizada en [37]. Estos algoritmos tienen un mayor desempeño en sus resultados que los demás algoritmos.

En este documento se propone un nuevo algoritmo para el aprendizaje de Redes Bayesianas. Este nuevo algoritmo combina el algoritmo de Colonia de Hormigas, según el enfoque competitivo propuesto en [36] y el uso de operadores para moverse entre clases de equivalencia de acuerdo al algoritmo descrito en [28]. El enfoque competitivo de hormigas difiere del modo clásico en el sentido que un conjunto de hormigas construye un solo camino o solución en lugar de que cada hormiga construya su propio camino como en [35], de esta forma se reduce la complejidad del problema y por tanto se obtienen mejores resultados que el algoritmo de búsqueda Greedy clásico [28] en términos del número de evaluaciones de la función de puntaje  $S_f$  y del número de interacciones erróneas en la estructura aproximada.

---

El orden del documento es: primero, se describe el modelo bayesiano para evaluar estructuras de redes bayesianas, siguiendo con una descripción breve de la forma de restringir la búsqueda usando clases de equivalencia, después se describe la propuesta de la aplicación de hormigas en la búsqueda, junto con un conjunto de resultados y su respectivo análisis. Por último se describe la aplicación del algoritmo propuesto para la inferencia de interacciones de naturaleza estocástica entre moléculas que responden en relación a cambios de la molécula EVI-1 y su aporte al descubrimiento de interacciones reales entre ellas.

# Capítulo 1

## Aprendizaje de Redes Bayesianas

### 1.1. Modelado Bayesiano para el Aprendizaje de Redes Bayesianas

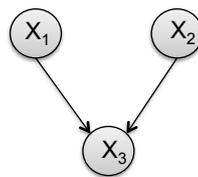
Una Red Bayesiana es una forma de representar el conocimiento de relaciones entre elementos a través de un Digrafo Acíclico<sup>1</sup> y un conjunto de probabilidades condicionales que cuantifican dichas relaciones, de tal forma que codifican una distribución de probabilidad conjunta de los elementos.

Como un ejemplo, una red bayesiana puede representar la regulación en la expresión de un gen por otros dos genes. Como se muestra en la figura 1.1(a), se puede asumir que los genes  $X_1$  y  $X_2$  regulan la expresión<sup>2</sup> del gen  $X_3$ . En este caso, cada gen  $X_i$  tiene dos posibles estados: estado 1 si el gen está activo o estado 2 si no lo está. La regulación en la red genética mencionada es cuantificada en la red bayesiana por sus parámetros. En la figura (1.1(b)) se puede observar que la probabilidad de que  $X_3$  esté activo cuando  $X_1$  y  $X_2$  están activos, denotado como  $\theta_{311}$ , es 0.91. Este resultado tiene la interpretación que cuando ambos genes,  $X_1$  y  $X_2$  están expresados pueden actuar sobre el gen  $X_3$  activando su expresión. Las demás relaciones siguen el mismo razonamiento, permitiendo representar y cuantificar las relaciones

---

<sup>1</sup>Un Digrafo Acíclico, es un tipo de grafo el cual sus arcos son dirigidos y no forman ciclos.

<sup>2</sup>Cuando un gen se 'expresa' se refiere a que fue utilizado para sintetizar una proteína, unidad funcional mas importante de una célula. Se dice que el gen está 'activo' si se expresa.



(a) Ejemplo de la estructura de una Red Bayesiana  $G$  de tres variables donde la variable  $X_3$  tiene dos padres. Una relación o arco es el par  $(X_1, X_3)$ .

	$X_1$	$X_2$	$X_3=1$	$X_3=2$	
j=1	1	1	$\theta_{311}=0.91$	$\theta_{312}=0.09$	$\Theta_{31}$
j=2	1	2	$\theta_{321}=0.82$	$\theta_{322}=0.18$	$\Theta_{32}$
j=3	2	1	$\theta_{331}=0.32$	$\theta_{332}=0.68$	$\Theta_{33}$
j=4	2	2	$\theta_{341}=0.11$	$\theta_{342}=0.89$	$\Theta_{34}$

(b) Ejemplo de los parámetros para  $X_3$  denotados como  $\Theta_3$ . Cada elemento de la tabla representa una probabilidad condicional, por ejemplo, la probabilidad de que  $X_3$  esté en el estado 1 dado que  $X_1$  y  $X_2$  también lo estén es 0.91, y se denota como  $\theta_{311}$  y es igual a  $p(X_3 = 1 | X_1 = 1, X_2 = 1) = p(X_3 = 1 | \Xi_3 = w_1)$ .

**Figura 1.1:** Ejemplo de la estructura y parámetros de una Red Bayesiana de tres nodos

entre los elementos.

Formalmente una Red Bayesiana es una *dupla*  $(G, \Theta)$  que representa la distribución de probabilidad conjunta:

$$p(X_1, X_2, \dots, X_n) = \prod_{i=1}^n p(X_i | \Xi_i, G, \Theta), \quad (1.1)$$

donde  $G = (\mathbf{X}, A)$  es la estructura de la red representada por un digrafo acíclico tal que sus nodos representan el conjunto de variables aleatorias  $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$  donde cada  $X_i \in \{1, \dots, r_i\}$  y  $A \subset \mathbf{X}^2$  define el conjunto de arcos dirigidos entre los nodos;  $\Xi_i = \{X_k | (X_k, X_i) \in A\}$  representa el conjunto de padres de  $X_i$ , y  $\Theta$  es el conjunto de relaciones entre las variables en  $\mathbf{X}$  definido como

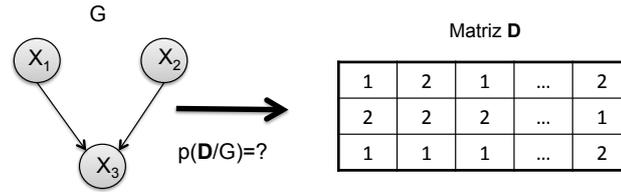
$$\begin{aligned} \Theta &= \bigcup_{i=1}^n \Theta_i, \\ \Theta_i &= \bigcup_{j=1}^{q_i} \{\Theta_{ij}\} \\ \Theta_{ij} &= \bigcup_{k=1}^{r_i} \{\theta_{ijk}\} \end{aligned} \quad (1.2)$$

donde  $\theta_{ijk}$  representa la probabilidad  $p(X_i = v_k | \Xi_i = w_j)$ , donde  $w_j$  es la  $j$ ésima combinación del conjunto  $\Omega = \{(v_1, \dots, v_\beta, \dots, v_{|\Xi_i|}) | \forall \beta, X_\beta = v_\beta \text{ y } X_\beta \in \Xi_i\}$  y  $q_i = |\Omega|$ . Por ejemplo, en la figura 1.1(b),  $\Theta_3$  es el conjunto de todos los valores dentro de la tabla,  $\Theta_{3j}$  es el conjunto de valores de la fila  $j$  en la tabla mostrada y el valor  $\theta_{311} = p(X_3 = 1 | \Xi_3 = w_1)$ . El modelado bayesiano para aproximar una estructura  $G$  de una red bayesiana se describe de la siguiente manera: Dado un conjunto de datos  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m]$ , donde  $\mathbf{d}_h$  describe el *estado* de las variables en  $\mathbf{X}$  para un caso  $h$ , y dada una hipótesis que dichos datos fueron generados por una red con estructura  $G$ , la aproximación bayesiana describe cómo se actualiza la *creencia*<sup>1</sup> de que ese grafo genera dichos datos, y así definir una función de puntaje  $S_f$ . Utilizando el Teorema de Bayes, se desea obtener la probabilidad a posteriori del grafo  $G$  dada la existencia de  $\mathbf{D}$ . Específicamente se desea calcular  $S_f(G) = p(G | \mathbf{D})$  dado por

<sup>1</sup>creencia denota una forma alternativa a la de frecuencia relativa para interpretar una probabilidad [18].

$$p(G|\mathbf{D}) = \frac{p(\mathbf{D}|G)}{p(\mathbf{D})}p(G) \quad (1.3)$$

donde  $p(G)$  es la probabilidad a priori de la estructura hipotética,  $p(\mathbf{D})$  es una constante de normalización y  $p(\mathbf{D}|G)$  es la probabilidad marginal de  $\mathbf{D}$  dada la estructura  $G$ . Asumiendo que  $p(G)$  es uniforme para todas las posibles estructuras, el problema de encontrar  $S_f$  se reduce a calcular la probabilidad que una estructura  $G$  pueda generar un conjunto de datos  $\mathbf{D}$ ,  $p(\mathbf{D}|G)$ . La figura 1.2) muestra la matriz  $\mathbf{D}$  asociada a la estructura de la figura (1.1(a)). Cada fila  $i$  de  $\mathbf{D}$  está asociada a la variable  $X_i$  y cada columna  $h$  de  $\mathbf{D}$  es un caso  $h$  en que todas las variables han sido generadas. La pregunta a resolver es, ¿cuál es la probabilidad que los valores en la matriz  $\mathbf{D}$  sean generados por la estructura  $G$ ?



**Figura 1.2:** El problema de aprendizaje estructural, para el caso de 3 variables se reduce a encontrar la probabilidad  $p(\mathbf{D}|G)$  de que la matriz  $\mathbf{D}$  de tres filas sea generado por  $G$ .

Si siguiendo el razonamiento realizado en [18] y [19] para obtener la probabilidad  $P(\mathbf{D}|G)$  en la ecuación (1.3), se tiene en cuenta que:

1. Se supone que el muestreo de las variables de  $\mathbf{X}$  en  $\mathbf{D}$  tiene distribución multinomial para cualquier estructura  $G$ . Entonces, dado  $\mathbf{D}$  como el arreglo de los  $d_{ih}$  y a  $\mathbf{D}_{\sigma_i h}$  como los valores de  $\Xi_i$  en  $\mathbf{d}_h$  donde  $\sigma_i = \{\ell | X_\ell \in \Xi_i\}$ , para todo  $G$  en  $\mathbf{E}_B$ , existe un conjunto de valor positivo  $\Theta$  tal que,

$$p(d_{ih} = v_k | D_{\sigma_i h} = w_j) = \theta_{ijk}, \quad (1.4)$$

donde  $\sum_k \theta_{ijk} = 1$ . En la figura 1.3 se muestra para el ejemplo de tres nodos, los valores para  $X_3$  y  $\Xi_3$  en el caso 3 en la matriz  $\mathbf{D}$  de acuerdo

a las relaciones definidas en  $G$ , denotados como  $d_{33}$  y  $D_{\sigma_33}$  respectivamente, donde  $\sigma_3 = (1, 2)$ . Dada entonces la existencia de los parámetros  $\Theta = \{\theta_{ijk}\}$ , se define entonces la ecuación de puntaje (1.3) teniendo en cuenta la incertidumbre que se tiene de ellos, calculando el promedio de la probabilidad  $P(\mathbf{D}|G)$  sobre todos los posibles valores de los parámetros en  $G$ , así:

$$p(\mathbf{D}|G) = \int_{\Theta} p(\mathbf{D}|G, \Theta) f(\Theta|G) d\Theta. \quad (1.5)$$

Entonces se requiere definir de (1.5) la probabilidad  $p(\mathbf{D}|G, \Theta)$  con respecto a los datos y  $f(\Theta|G)$  con respecto a los parámetros<sup>1</sup>.

					$\mathbf{D}_{\sigma_33} \leftarrow \Xi_3$
					2
					1
					$\mathbf{D}_{333} \leftarrow X_3$
					1

$\mathbf{D}$	$h=3$				
1	2	2	...	2	
2	2	2	...	1	
1	1	1	...	2	

**Figura 1.3:** Valores para  $X_3$  y  $\Xi_3$  en el caso 3 en  $\mathbf{D}$  según  $G$ , denotados como  $\mathbf{D}_{333}$  y  $\mathbf{D}_{\sigma_33}$  respectivamente, donde  $\sigma_3 = (1, 2)$ .

2. Se asume que cada caso  $\mathbf{d}_h$  es *independiente* de los demás y para cada variable  $X_i$  solo se puede tener un conjunto finito de estados, de tal forma que

$$p(\mathbf{D}|G, \Theta) = \prod_{h=1}^m p(\mathbf{d}_h|G, \Theta). \quad (1.6)$$

3. Se supone que *existe un valor* para todas las variables en todos los casos en  $\mathbf{D}$ , entonces la expresión (1.6) se puede escribir como,

<sup>1</sup>Nótese que  $p(\mathbf{D}|G, \Theta)$  es una función de masa de probabilidad y  $f(\Theta|G)$  es una función de densidad, ya que se está integrando con respecto a  $\Theta$  y las propiedades siguientes se basan en ello.

$$p(\mathbf{D}|G, \Theta) = \prod_{i=1}^n \prod_{h=1}^m p(d_{ih}|G, \Theta). \quad (1.7)$$

Sin embargo, dado que  $G$  establece los elementos de  $\Xi_i$ , se puede descomponer la expresión anterior de acuerdo a las contribuciones locales<sup>1</sup> de cada variable,

$$p(\mathbf{D}|G, \Theta) = \prod_{i=1}^n \prod_{h=1}^m p(X_i = d_{ih} | \Xi_i = D_{\sigma_{ih}}, \Theta). \quad (1.8)$$

Además, si se define a  $N_{ijk}$  como el número de veces que  $d_{ih}$  es igual a un valor  $v_k$  y  $D_{\sigma_{ih}}$  a un elemento  $w_j \in \Omega$  para todo  $h$  en  $\mathbf{D}$ , la ecuación anterior puede reagruparse de la siguiente manera,

$$p(\mathbf{D}|G, \Theta) = \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} p(d_{ih} = v_k | D_{\sigma_{ih}} = w_j)^{N_{ijk}}, \quad (1.9)$$

entonces por (1.4) y (1.9) se tiene que

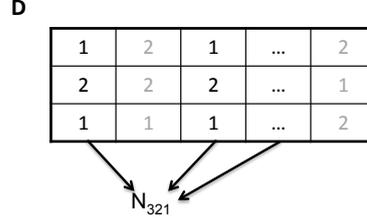
$$p(\mathbf{D}|G, \Theta) = \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{N_{ijk}}. \quad (1.10)$$

En la figura 1.4 se presenta la forma de calcular de  $N_{ijk}$  para la configuración específica  $X_3 = v_1$  y  $\Xi_i = w_2$  a partir de la matriz  $\mathbf{D}$ . Así, con la información mostrada en la figura 1.1(b), es posible, para este caso específico calcular  $N_{321}^{N_{321}}$ . Por lo tanto,  $N_{ijk}$  una estadística suficiente que da toda la información de la muestra multinomial  $\mathbf{D}$  para obtener un puntaje para una estructura  $G$ .

4. Se asume que los parámetros asociados con cada variable en la estructura son independientes, entonces se tiene que:

$$f(\Theta|G) = \prod_{i=1}^n f(\Theta_i|G). \quad (1.11)$$

<sup>1</sup>Por 'local' se hace referencia a un  $X_i$  y  $\Xi_i$  en  $G$ .



**Figura 1.4:** Cómputo de  $N_{ijk}$  para la configuración específica  $X_3 = v_1$  y  $\Xi_i = w_2$ .

A esta propiedad se le llama Independencia Paramétrica Global.

5. Se asume que los parámetros asociados con cada instancia de los valores de los padres de cada variable son independientes. Esto se puede expresar como

$$f(\Theta|G) = \prod_{i=1}^n \prod_{j=1}^{q_i} f(\Theta_{ij}|G). \quad (1.12)$$

Esta propiedad tiene el nombre de Independencia Paramétrica Local. Así a partir de (1.10) y (1.12), la ecuación (1.5) se puede reescribir de la siguiente forma

$$p(\mathbf{D}|G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \int_{\Theta_{ij}} \left[ \prod_{k=1}^{r_i} \theta_{ijk}^{N_{ijk}} \right] f(\Theta_{ij}|G) d\Theta_{ij}. \quad (1.13)$$

6. Por último, se supone que la función de densidad para los parámetros en  $\Theta_{ij}$ , sin conocer  $\mathbf{D}$  (o *a priori*), sigue una distribución de Dirichlet principalmente porque, al actualizarse cuando se hace un muestro multinomial  $\mathbf{D}$ , siguen siendo Dirichlet<sup>1</sup> [19]. Así, la función de densidad para los parámetros es como sigue,

$$f(\Theta_{ij}|G) = \frac{\Gamma(N'_{ij})}{\prod_{k=1}^r \Gamma(N'_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{N'_{ijk}-1} \quad (1.14)$$

<sup>1</sup>Por esta razón, en inferencia bayesiana, se dice que la Distribución de Dirichlet es un prior conjugado de la distribución Multinomial

donde  $N'_{ijk}$  es un sesgo para las probabilidades numéricas  $\theta_{ijk}$  sin tener **D**.

Basado en las suposiciones anteriores, y dado que la integral en (1.13) describe  $E \left[ \prod_{k=1}^{r_i} \theta_{ijk}^{N_{ijk}} \right]$  con respecto a  $f(\Theta_{ij}|G)$ , la ecuación resultante para (1.5) solucionando la integral según [18] y [19], es

$$p(\mathbf{D}|G) = p(G) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N_{ij} + N'_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N_{ijk} + N'_{ijk})}{\Gamma(N'_{ijk})}. \quad (1.15)$$

o de otra forma,  $S_f(G) = \log(p(\mathbf{D}|G))$  está dado por

$$S_f(G) = \sum_{i=1}^n s(X_i|\Xi_i) \quad (1.16)$$

donde

$$s(X_i|\Xi_i) = \log \left[ \prod_{j=1}^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N_{ij} + N'_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N_{ijk} + N'_{ijk})}{\Gamma(N'_{ijk})} \right]. \quad (1.17)$$

La ecuación (1.15) y (1.16) son llamadas Ecuaciones bayesianas de Dirichlet (BD), donde  $\Gamma(\cdot)$  es la función Gama definida como  $\Gamma(c) = \int_0^\infty e^{-u} u^{c-1} du$  que describe una extensión del concepto de factorial a los números reales.  $N_{ijk}$  es el número de veces que la variable  $X_i$  tiene el valor  $v_k$  y los padres  $\Xi_i$  tienen la combinación de estado  $w_j$ . el valor de  $N'_{ijk}$  tiene el mismo significado de  $N_{ijk}$  pero con la diferencia que es un valor asumido *antes* de tener el conjunto de muestras **D** utilizado como información inicial. Este parámetro ha sido ajustado a  $N'_{ijk} = \eta \cdot \frac{1}{r_i q_i}$  asumiendo que dos estructuras equivalentes (concepto que se explicará en la siguiente sección) pueden tener el mismo puntaje y que en  $\eta$  posibles muestras iniciales la probabilidad de que  $X_i = w_k$  es  $\frac{1}{r_i}$  y que  $\Xi_i = v_j$  es  $\frac{1}{q_i}$  en cada muestra [35; 38].

Teniendo la función que evalúa las estructuras, se define entonces como está restringida la búsqueda de éstas por medio del enfoque híbrido y se describe el espacio de las clases de equivalencia.

---

## 1.2. Clases de Equivalencia de Redes Bayesianas

El concepto de clases de equivalencia en redes bayesianas provee un componente para el problema de aprendizaje, ya que dos estructuras diferentes pueden describir la misma distribución de probabilidad, y una función de puntaje puede evaluar dos estructuras equivalentes con el mismo valor [28; 29; 30; 31], lo que repercutiría en la búsqueda de dicha estructura.

El concepto de equivalencia viene dado dentro de las redes bayesianas en la codificación del *flujo de información* de acuerdo a las relaciones que hay dentro de los elementos. Más formalmente se dice que dos variables  $X_\alpha, X_\beta \in X$  de una Red Bayesiana están *direccionalmente separadas* (o *d-separadas*) si existe una variable intermedia  $X_\gamma$  (diferente de  $X_\alpha$  y  $X_\beta$ ) y se cumple una de las siguientes condiciones:

- Existe una configuración de la forma  $X_\alpha \rightarrow X_\gamma \rightarrow X_\beta$  o de la forma  $X_\alpha \leftarrow X_\gamma \leftarrow X_\beta$  y se conoce el estado de la variable  $X_\gamma$ .
- Existe una configuración de la forma  $X_\alpha \leftarrow X_\gamma \rightarrow X_\beta$  y se conoce el estado de la variable  $X_\gamma$ .
- Existe una configuración de la forma  $X_\alpha \rightarrow X_\gamma \leftarrow X_\beta$  y *no* se conoce el estado de la variable  $X_\gamma$ .

Si  $X_\alpha$  está *d-separada* de  $X_\beta$  dado el conocimiento del estado de la variable  $X_\gamma$  se dice que  $X_\alpha$  y  $X_\beta$  son *condicionalmente independientes* dado  $X_\gamma$  y se denota  $(X_\alpha \perp X_\beta | X_\gamma)$ . Así  $X_\alpha, X_\beta$  y  $X_\gamma$  satisfacen la siguiente relación,

$$p(X_\alpha | X_\gamma, X_\beta) = p(X_\alpha | X_\gamma). \quad (1.18)$$

Dados estos conceptos, dos estructuras  $G$  y  $G'$  son *Equivalentes* si y sólo si para toda Red Bayesiana  $B = (G, \Theta)$  existe un red Bayesiana  $B' = (G', \Theta')$  tal que  $B$  y  $B'$  describen el mismo conjunto de independencias y tienen la misma distribución de Probabilidad. De forma similar, dos estructuras  $G$  y  $G'$  son equivalentes *si y sólo si* tienen el mismo *esqueleto* y las mismas *estructuras-*

---

$v^1$ .

### 1.2.1. Representación de Clases de Equivalencia

Dado el espacio no vacío  $\mathbf{E}_B$  de estructuras de redes bayesianas, la relación entre redes equivalentes en dicho espacio, denotada como  $\sim$ , cumple las siguientes propiedades:

*Reflexividad:*

$$\forall G \in \mathbf{E}_B (G \sim G). \quad (1.19)$$

*Simetría:*

$$\forall G, G' \in \mathbf{E}_B (G \sim G' \Rightarrow G' \sim G). \quad (1.20)$$

*Transitividad:*

$$\forall G, G', G'' \in \mathbf{E}_B (G \sim G' \wedge G' \sim G'' \Rightarrow G \sim G''). \quad (1.21)$$

Luego las redes con dicha relación forman una *clase de equivalencia*. Así, para representar de una sola forma, una clase de equivalencia, se define a un *digrafo acíclico parcial*  $P$  como un grafo acíclico que contiene aristas dirigidas y no dirigidas, de tal forma que, una *clase de equivalencia* se define como:

$$[P] = \{G \in \mathbf{E}_B \mid G \sim P\}. \quad (1.22)$$

Si la arista  $X_\alpha \rightarrow X_\beta$  está presente en todo  $G \in [P]$ , se dice que  $X_\alpha \rightarrow X_\beta$  es *forzada* y si una arista  $X_\alpha \rightarrow X_\beta$  no es forzada, se dice que es *reversible*. Un *digrafo acíclico parcial completo*  $P^c$  es un digrafo acíclico parcial tal que toda arista forzada es una arista dirigida, y toda arista reversible es una arista no dirigida. De esta forma, dado un espacio de estructuras de Redes Bayesianas  $\mathbf{E}_B$  existe un único digrafo acíclico parcial completo por cada clase de equivalencia en dicho espacio.

---

<sup>1</sup>El *esqueleto* de  $G$  es el grafo no dirigido que se obtiene al omitir la dirección de las aristas de  $G$  y una *estructura-v* es una tripleta es una estructura de la forma  $X_\alpha \rightarrow X_\beta \leftarrow X_\gamma$  tal que  $X_\alpha$  y  $X_\gamma$  no son adyacentes [28].

---

## 1.2.2. Movimientos entre Clases de Equivalencia

El *aprendizaje de Clases de Equivalencia* difiere del aprendizaje estructural en que el Espacio de Búsqueda de las clases de equivalencia, denominado  $\mathbf{E}_E$ , tiene las siguientes propiedades:

- Contiene un conjunto de Estados, que son las clases de equivalencia que conforman dicho espacio.
- Cada estado tiene una representación, que es un grafo aciclico parcial completo.
- Existen un conjunto de operadores para hacer posible el movimiento entre los estados en un solo paso.

Un operador define una transformación de un estado en  $\mathbf{E}_E$  a otro, o lo que es lo mismo de un *digrafo aciclico parcial completo* a otro. Este trabajo utiliza un conjunto de seis operadores con una prueba de validación para cada uno con la ventaja de que se puede calcular el puntaje del nuevo grafo generado solo con el puntaje local  $s(X_i|\Xi_i)$  de acuerdo al cuadro 1.1 y la ecuación (1.17) [28].

La validación tiene en cuenta que:

- $\Xi_\alpha$  es el conjunto de Padres de  $X_\alpha$ , es decir, el conjunto de nodos que tienen arcos dirigidos hacia  $X_\alpha$ .
- $\Psi_\alpha$  es el conjunto de vecindad de  $X_\alpha$ , es decir, el conjunto de nodos que tiene arcos no dirigidos hacia  $X_\alpha$ .
- $\Psi_{\alpha,\beta}$  es la intersección entre  $\Psi_\alpha$  y  $\Psi_\beta$ .
- $\Omega_{\alpha,\beta}$  es el conjunto de padres de  $X_\alpha$  que son vecinos de  $X_\beta$ .
- Un clique en un grafo no dirigido  $G$  es un conjunto de vértices  $V$  tal que para todo par de vértices de  $V$ , existe una arista que las conecta.

Operador	Efecto	Prueba de Validación	Cambio en el Puntaje
$X_\alpha - X_\beta$	Añadir un arco no dirigido entre $X_\alpha$ y $X_\beta$ .	1. Todo camino no dirigido de $X_\alpha$ a $X_\beta$ contiene un nodo en $\Psi_{\alpha,\beta}$ . 2. $\Xi_\alpha = \Xi_\beta$ .	$s(X_\beta \Psi_{\alpha,\beta}^{+\alpha} \cup \Xi_\beta) - s(X_\beta \Psi_{\alpha,\beta} \cup \Xi_\beta)$
$X_\alpha \approx X_\beta$	Eliminar un arco no dirigido entre $X_\alpha$ y $X_\beta$ .	$\Psi_{\alpha,\beta}$ es un clique.	$s(X_\beta \Psi_{\alpha,\beta} \cup \Xi_\beta) - s(X_\beta \Psi_{\alpha,\beta}^{+\alpha} \cup \Xi_\beta)$
$X_\alpha \rightarrow X_\beta$	Añadir un arco dirigido entre $X_\alpha$ y $X_\beta$ .	1. Todo camino semi-dirigido de $X_\beta$ a $X_\alpha$ contiene un nodo en $\Omega_{\alpha,\beta}$ . 2. $\Omega_{\alpha,\beta}$ es un clique. 3. $\Xi_\alpha \neq \Xi_\beta$ .	$s(X_\beta \Omega_{\alpha,\beta} \cup \Xi_\beta^{+\alpha}) - s(X_\beta \Omega_{\alpha,\beta} \cup \Xi_\beta)$
$X_\alpha \rightarrow X_\beta$	Eliminar un arco dirigido entre $X_\alpha$ y $X_\beta$ .	$\Psi_\beta$ es un clique.	$s(X_\beta \Psi_\beta \cup \Xi_\beta^{-i}) - s(X_\beta \Psi_\beta \cup \Xi_\beta)$
$X_\alpha \leftarrow X_\beta$	Invertir el arco dirigido de $X_\alpha$ a $X_\beta$ .	1. Todo camino semi-dirigido de $X_\alpha$ a $X_\beta$ que no incluya el arco $X_\alpha \rightarrow X_\beta$ contiene un nodo en $\Omega_{\alpha,\beta} \cup \Psi_\beta$ . 2. $\Omega_{\beta,\alpha}$ es un clique.	$s(X_\beta \Xi_\beta^{-\alpha}) + s(X_\alpha \Xi_\alpha^{+\beta} \cup \Omega_{\beta,\alpha}) - s(X_\beta \Xi_\beta) - s(X_\alpha \Xi_\alpha \cup \Omega_{\beta,\alpha})$
$X_\alpha \rightarrow X_\gamma \leftarrow X_\beta$	Convertir en un arco dirigido los arcos no dirigidos de $X_\alpha$ y $X_\beta$ a $X_\gamma$ .	Todo camino semi-dirigido entre $X_\alpha$ y $X_\beta$ contiene un nodo en $\Psi_{\alpha,\beta}$ .	$s(X_\gamma \Xi_k^{+\beta} \cup \Psi_{\alpha,\beta}^{-\gamma+\alpha}) + s(X_\beta \Xi_\beta \cup \Psi_{\alpha,\beta}^{-\gamma}) - s(X_\gamma \Xi_\gamma \cup \Psi_{\alpha,\beta}^{-\gamma+\alpha}) - s(X_\beta \Xi_\beta \cup \Psi_{\alpha,\beta})$

**Cuadro 1.1:** Conjunto de operadores que permiten pasar de una clase de equivalencia a otra [28]. Cada operador tiene como entrada dos o tres nodos de un grafo acíclico parcial completa y produce un grafo acíclico parcial de acuerdo a la operación.

---

Además, dentro de la función local de puntaje en la última columna, se utiliza la notación  $\Psi_{\alpha}^{+\beta}$  y  $\Psi_{\alpha}^{-\beta}$ , que define el conjunto  $\Psi_{\alpha} \cup \{X_{\beta}\}$  y  $\Psi_{\alpha} \setminus \{X_{\beta}\}$  correspondientemente.

Finalmente, teniendo un operador y dado un *digrafo aciclico parcial*  $P$ , un *movimiento*  $l$  es una instancia de un operador aplicado a un conjunto de nodos de  $P$ , y  $M$  como al conjunto de todos los movimientos  $l$  válidos a partir de  $P$ . Así utilizando dichos operadores es posible hacer más eficientes los algoritmos moviéndose entre subconjuntos dentro de  $\mathbf{E}_B$ . Cada operador tiene como entrada dos o tres nodos de un grafo acíclico parcial completo y produce un grafo acíclico parcial de acuerdo a la operación [28]<sup>1</sup>.

### 1.3. Conclusiones

Se realizó la descripción formal del modelado bayesiano para el aprendizaje de redes bayesiana, produciendo como resultado la conocida función de puntaje bayesiana de dirichlet. Además se describieron las ventajas del enfoque de puntaje y búsqueda entre clases de equivalencia de redes bayesianas. Principalmente, se evita por medio de operadores el movimiento redundante de búsqueda entre la misma clase de equivalencia.

---

<sup>1</sup>Es necesario transformar el grafo acíclico parcial a uno completo [28]. Los algoritmos para ellos se encuentran descritos en [28].

## Capítulo 2

# Colonia de Hormigas en el Aprendizaje

En este capítulo se presenta uno de los aportes principales de este documento, el algoritmo propuesto para el aprendizaje de clases de equivalencia de redes bayesianas utilizando un modelo basado en colonias de hormigas con la singularidad de que las hormigas compiten en la construcción de un solo camino.

### 2.1. Colonia de Hormigas en el Aprendizaje

El método de Optimización Basado en Colonia de hormigas es un algoritmo heurístico multiagente que fue propuesto por Marco Dorigo [39; 40; 41]. Estos algoritmos son inspirados en el comportamiento colaborativo de las hormigas para encontrar la ruta más corta desde su nicho hasta la comida, y son utilizados con el fin de resolver problemas combinatorios de búsqueda y optimización con alta complejidad (o llamados NP-Duros). El componente principal de interacción entre las hormigas es la comunicación *indirecta*, que emplean gracias a un medio específico llamado *feromona*.

El principal factor para llegar a una aproximación de la solución con Colonia de hormigas es la representación del problema. El algoritmo propuesto

---

para aplicar Colonia de Hormigas al aprendizaje de clases de equivalencia es esencialmente basado en dos trabajos de aplicación de colonia de Hormigas presentados en [36] y [35]. Del primero es tomado el método por el cual las hormigas construyen el camino, y del segundo se toma la forma de representar el problema.

### 2.1.1. Representación del problema

La Representación del problema de aprendizaje de clases de equivalencia está descrito de la siguiente manera:

- El primer elemento del problema es el conjunto de componentes  $C = \{c_1, c_2, \dots, c_n\}$  que se pueden combinar para obtener una solución. En este problema corresponde al conjunto de movimientos  $l$  que pueden ser hechos en un estado representado por el digrafo acíclico parcial  $P$ .
- Un camino (o solución)  $\mathbf{s}$  es una secuencia de movimientos  $\langle l_i, \dots, l_j \rangle$  de tal forma que si se aplica cada uno de ellos (en el exacto orden) en un *digrafo aciclico parcial*  $P$  dará como resultado un *digrafo aciclico parcial completo* que representa el estado actual.
- Un conjunto de  $m$  hormigas que son los agentes que contruyen una solución. Así cada hormiga compite para incluir un nuevo movimiento al camino, hasta que todas construyen una solución  $\mathbf{s}$ .
- Existe una feromona  $\tau_i$  asociada a cada movimiento  $l_i \in C$  en una solución  $\mathbf{s}$ . El comportamiento de cada  $\tau_i$  es representado por un modelo que define como incrementa de acuerdo a si el mejor camino  $\mathbf{s}$  contiene el movimiento  $l_i$  y disminuye exponencialmente de acuerdo a un factor  $\rho$  de evaporación. Esta actualización es tanto para la mejor solución de una iteración como al mejor camino de todas las iteraciones.

En el algoritmo todas las hormigas construyen una única solución o camino escogiendo cada una, un sólo movimiento, y aplicándolo solo aquel que produzca un mejor puntaje  $S_f$  hasta que no haya posibilidad de un mejor movimiento para cualquier hormiga.

---

### 2.1.2. Modelo de la Feromona

El modelo dinámico de la actualización de la feromóna se describe de la siguiente manera:

1. Se inicializa la feromona  $\tau_l$  asociada a todos los movimientos posibles desde un digrafo  $\hat{P}$  vacío de la siguiente manera:

$$\tau_l = \frac{1}{|S_f(\hat{P})|}. \quad (2.1)$$

donde  $S_f$  está dada por la ecuación (1.16).

2. Se actualiza la feromona asociada a todos los movimientos que se han realizado utilizando una función llamada evaporación definida como:

$$\tau_l = (1 - \rho)\tau_l. \quad (2.2)$$

3. Después de que todas las hormigas han construido el mejor camino  $\mathbf{S}$ , cada feromona  $\tau_l$  es incrementada si el movimiento se encuentra allí. Así ,

$$\tau_l = (1 - \rho)\tau_l + \frac{1}{|S_f(P^k)|}, l \in \mathbf{s}. \quad (2.3)$$

4. También se actualiza la feromona de aquellos elementos que se relacionan con los movimientos del mejor camino de todas las iteraciones  $\hat{\mathbf{s}}$  así:

$$\tau_l = (1 - \rho)\tau_l + \frac{\rho}{|S_f(\hat{P})|}, l \in \hat{\mathbf{s}}. \quad (2.4)$$

### 2.1.3. Información Heurística

La información heurística  $\eta_l$  define un puntaje para un movimiento  $l$ . Dado que un movimiento consiste en la transformación de un digrafo acíclico parcial a otro, ese cambio trae consigo un cambio en el puntaje de todo el digrafo. Sin embargo, como se vio en secciones anteriores, puede ser simplificado como

---

el cambio en la función local de puntaje  $s(X_i|\Xi_i)$  definida en la ecuación (1.17) dependiendo del operador aplicado. Este cómputo está definido en la última columna del cuadro 1.1 en la sección anterior.

#### 2.1.4. Regla de Probabilidad

Cuando cualquier hormiga compite por escoger el mejor movimiento, cada una utiliza una regla de transición pseudoaleatoria [40] que permite hacer un balance entre exploración y explotación para obtener el posible siguiente movimiento.

Formalmente el siguiente movimiento  $l$  para una hormiga se escoge así,

$$l = \begin{cases} \arg \max_{l \in M} \tau_l^\alpha \eta_l^\beta & \text{si } q \leq q_0 \\ \text{Aleatoriamente de acuerdo a } p(l) & \text{si } q > q_0 \end{cases} \quad (2.5)$$

donde

$$p(l) = \frac{\tau_l^\alpha \eta_l^\beta}{\sum_{\mu \in M} \tau_\mu (\eta_\mu^\beta)}, \quad (2.6)$$

los parámetros  $\alpha$  y  $\beta$  son la potencia asignada a la feromona y a la información heurística correspondientemente y  $q_0$  es un número aleatorio con distribución uniforme en el intervalo  $[0, 1]$ .

#### 2.1.5. Descripción del Algoritmo

El algoritmo propuesto, mostrado en la tabla (1), está principalmente basado en el procedimiento para el aprendizaje de estructuras de Redes bayesianas donde todas las hormigas colaborativamente construyen un camino por cada iteración a partir de una estructura vacía [36]. Además, el aprendizaje se hace en el espacio de clases de equivalencia utilizando una representación en la cual cada solución conduce a un dígrafo acíclico parcial [35]. La notación del algoritmo es presentada en el cuadro 2.1.

---

<b>O</b>	Conjunto de todos los operadores descritos en el cuadro 1.1.
$\hat{t}$	número máximo de iteraciones.
$N$	número de hormigas.
$\rho$	tasa de evaporación de la ferómona.
$q_0$	valor entre $[0,1]$ que hace un balance entre la explotación y la exploración al escoger un nuevo movimiento por una hormiga.
$\beta$	influencia que tiene la información heurística en la regla de probabilidad número de nodos de la red que se desea inferir.
$M$	conjunto de todos los posibles movimientos a partir de un digrafo acíclico parcial.
$\hat{P}$	el mejor digrafo acíclico parcial de todas las iteraciones
$\hat{s}$	mejor solución (conjunto de movimientos) de las hormigas de todas las iteraciones.
$P$	digrafo acíclico parcial construido por todas las hormigas en una iteración.
<b>s</b>	camino construido por todas las hormigas en una iteración.

**Cuadro 2.1:** Notación utilizada para cada uno de las variables, grafos y parámetros dentro del algoritmo.

---

**Algoritmo 1** Algoritmo ACO-CE: Algoritmo principal de hormigas competitivas para el aprendizaje de clases de equivalencia de redes bayesianas.

---

**Entrada:**  $\mathbf{D}, O, t_{max}, m, \rho, q_0, \alpha, \beta, n$

**Salida:**  $\hat{P}, \hat{h}$

$M = o(P); \forall o \in \mathbf{O}$

$\tau_l = \frac{1}{|S_f(P)|}; \forall l \in M$

**para**  $t = 0$  hasta  $\hat{t}$  **hacer**

$P =$  Grafo vacio

$\mathbf{s} = \langle \rangle$

$(P, \mathbf{s}) = ANTS(P, \mathbf{s}, \mathbf{O}, q_0, \alpha, \beta)$

**si**  $S_f(P) > S_f(\hat{P})$  **entonces**

$\hat{P} = P$

$\hat{\mathbf{s}} = \mathbf{s}$

**fin si**

$\tau_l = \tau_l(1 - \rho) + \frac{\rho}{|S_f(P)|}; \forall l \in \mathbf{s}$

$\tau_l = \tau_l(1 - \rho) + \frac{\rho}{|S_f(\hat{P})|}; \forall l \in \hat{\mathbf{s}}$

**fin para**

---

A partir de un digrafo acíclico parcial  $P$  vacío, las hormigas construyen el camino de tal forma que cada una realiza, como se muestra en la tabla 2, lo siguiente:

- Halla todos los posibles movimientos a partir del actual  $P$ .
- Escoge uno de los movimientos de manera pseudoaleatoria de acuerdo a la regla de probabilidad de acuerdo a la ecuación (2.5).
- Se aplica dicho movimiento a  $P$  obteniendo un nuevo digrafo  $P^k$  para la hormiga  $k$ .

Después que todos las hormigas han escogido un posible siguiente movimiento, se escoge la solución  $s_r$  donde  $r \sim U(1, m)$ . Si hubo al menos una hormiga que realizó un mejor movimiento, es decir, si  $S_f(P_r) > S_f(P)$ , las hormigas vuelven a competir por el siguiente movimiento. De otro modo, se retorna a la solución  $P$ , se actualiza el mejor camino de todas las iteraciones  $\hat{P}$  y se actualiza el valor de la feromona de acuerdo a las ecuaciones (2.3) y (2.4).

---

**Algoritmo 2** Algoritmo ANTS: Algoritmo que describe la competencia de las hormigas para generar un digrafo acíclico parcial en una iteración del algoritmo principal.

---

**Entrada:**  $P, \mathbf{s}, \mathbf{O}, q_0, \alpha, \beta$ .

**Salida:**  $P, \mathbf{s}$

Fin=0

**mientras** Fin=0 **hacer**

Fin=1

**para**  $k = 1$  hasta  $m$  **hacer**

$M = o(P); \forall o \in \mathbf{O}$

**si**  $|M| = 0 \vee \max_{l \in M} \tau_l^\alpha \eta_l^\beta \leq 0$  **entonces**

Devolver  $(P, \mathbf{h})$

**fin si**

$q \sim U[0, 1]$

$p(l) = \frac{\tau_l^\alpha \eta_l^\beta}{\sum_{\mu \in M} \tau_\mu^\alpha \eta_\mu^\beta}; \forall l \in M$

$l = \begin{cases} \arg \max_{l \in M} \tau_l^\alpha \eta_l^\beta & \text{si } q \leq q_0 \\ \text{Aleatoriamente de acuerdo a } p(l) & \text{si } q > q_0 \end{cases}$

$P_k = \text{Aplicar } l \text{ a } P$

$s_k = \langle \mathbf{s}, l \rangle$

$S_f(P_k) = S_f(P) + n_l$

**fin para**

$r \sim U[1, \dots, m]$

$S_b = S_f(P_r)$

**si**  $S_b > S_f(P)$  **entonces**

$P = P_r$

$\mathbf{s} = s_r$

**si no**

Fin=0

**fin si**

**fin mientras**

---

---

## 2.2. Evaluación y Resultados

La metodología de Evaluación del aprendizaje que se utilizó se basó en gran medida en la descrita en la literatura [35; 36; 37] sigue, en general, el siguiente protocolo:

1. Se selecciona la red ASIA como red teórica o estándar <sup>1</sup> para comparar los algoritmos Greedy [28], ACO-E [35] y el algoritmo propuesto. La red ASIA como muestra la figura 2.2(a) es una red de 8 nodos y 8 arcos.
2. Se selecciona un conjunto de valores para cada uno de los parámetros del algoritmo propuesto, manteniendo el número de hormigas  $m$  igual a 15. Los conjuntos de valores son:

$$\begin{aligned}\rho &= [0,05 \quad 0,1 \quad 0,3 \quad 0,5] \\ q_0 &= [0,3 \quad 0,5 \quad 0,7 \quad ] \\ \alpha &= [2 \quad 3 \quad 4 \quad 5] \\ \beta &= [1 \quad 1,5 \quad 2 \quad ]\end{aligned}\tag{2.7}$$

3. Para el algoritmo ACO-E se mantuvieron los parámetros obtenidos en [35], es decir,  $\rho = 0,3$ ,  $q_0 = 0,7$ ,  $\alpha = 1$  y  $\beta = 2,5$ .
4. Para cada una de las combinaciones de los valores de los parámetros se hicieron las simulaciones respectivas para los tres algoritmos para 20 diferentes conjuntos de datos con 20, y 500 muestras. El rendimiento así como la red o estructura obtenida por cada uno de los algoritmos se evaluaron utilizando las siguientes métricas:
  - **Función de Puntaje:** Es la métrica utilizada en el algoritmo para evaluar la cercanía de la red construida y los datos según la ecuación (1.15) o (1.16), y como tal, es la que primero evalúa los resultados del algoritmo. Un valor mayor indica que el grafo es más probable de generar los datos.

---

<sup>1</sup>Las redes estándar mas usadas se pueden encontrar en el repositorio <http://compbio.cs.huji.ac.il/Repository/>.

- 
- Distancia Estructural de Hamming (SHD): Es el promedio entre los arcos añadidos, omitidos e invertidos (en dirección) entre la estructura generada por el algoritmo y la estructura de la cual se generaron los datos iniciales. La obtención de esta métrica se describe en el Algoritmo 3, donde  $H$  es el digrafo aproximado por el algoritmo y  $G$  es el digrafo teórico el inicial del cual se generaron los datos. Una distancia menor significa entonces, que el digrafo aproximado es mas cercano al teórico.
  - Número de evaluaciones o llamadas estadísticas (NSC): Devuelve el número de llamadas que hace el algoritmo a la función objetivo. Esta es una métrica que mide la complejidad del algoritmo.

---

**Algoritmo 3** Algoritmo SHD tomado de [37]

---

**Entrada:**  $PDAG H$  construida,  $PDAG G$  Teórica

**Salida:**  $shd$

$shd = 0$

**para** cada arco  $E$  en  $H$  diferente en  $G$  **hacer**

**si**  $E$  no esta en  $H$  **entonces**

$shd = shd + 1$

**fin si**

**si**  $E$  esta en  $H$  y no en  $G$  **entonces**

$shd = shd + 1$

**fin si**

**si**  $E$  esta en  $H$  en diferente dirección que  $G$  **entonces**

$shd = shd + 1$

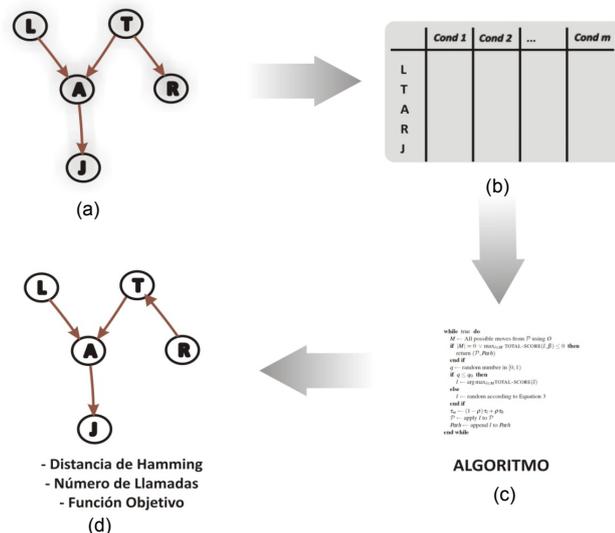
**fin si**

**fin para**

---

En la figura 3.2.1 se muestra el protocolo descrito anteriormente para la red ASIA, para una simulación utilizando 500 muestras. La primera red, es la red teórica de la cual se obtienen los datos, las demás son las redes aproximadas por los algoritmos GES, ACO-E y el algoritmo propuesto respectivamente.

5. Para la función de puntaje se escogió un tamaño de muestreo equivalente  $\eta = 4$  como fue realizado en [35].

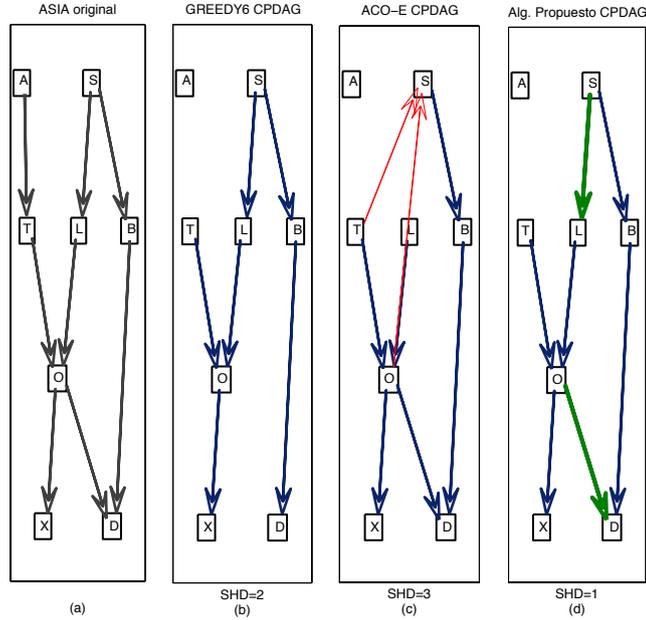


**Figura 2.1:** El proceso de evaluación se puede resumir en cuatro pasos: (a), escoger una red bayesiana estándar; (b), generar un conjunto de datos a partir de dicha red; (c), por medio de un algoritmo aproximar una red bayesiana a partir de los datos; y (d) medir la cercanía de esta red con respecto a la inicial.

El número total de combinaciones para los valores de los parámetros del algoritmo según el ítem 4 fueron 144. Como se muestra en el cuadro 2.2 estas combinaciones se clasificaron de acuerdo a un criterio basado en los valores obtenidos en la función de puntaje  $S_f$ , la distancia estructural de hamming  $SHD$  y el número de llamadas estadísticas  $NSC$  para los tres algoritmos probados.

Es notorio, de acuerdo a la segunda fila del cuadro 2.2, que sin importar el número de muestras y la combinación de parámetros, el algoritmo propuesto tiene más probabilidad de generar los datos que el algoritmo Greedy.

En la columna del cuadro para 20 muestras, se describe que para sólo una combinación de los parámetros  $\rho = 0,1$ ,  $q_0 = 0,7$ ,  $\alpha = 3$  y  $\beta = 1$ , el resultado del algoritmo propuesto obtuvo un mayor valor en la métrica  $S_f$  con respecto al algoritmo ACO-E. Como muestra la figura 2.3(a), la diferencia entre el valor de  $S_f$  del algoritmo propuesto con respecto a ACO-E es muy pequeña en compa-



**Figura 2.2:** Ejemplo predicción de la estructura de la red bayesiana ASIA. (a), la red original, (b) la red aproximada por el algoritmo GES, (c) ACO-E y (d) el algoritmo propuesto en este artículo con 500 muestras. Los arcos rojos en el grafo aproximado por ACO-E son arcos que no corresponden al arco original, y los arcos en verde del grafo aproximado por el algoritmo propuesto son arcos los cuales uno de ellos no se encuentra en los grafos aproximados por los otros dos algoritmos.

**Cuadro 2.2:** Clasificación de los resultados al protocolo descrito dependiendo de las métricas mencionadas para 20 y 500 muestras.

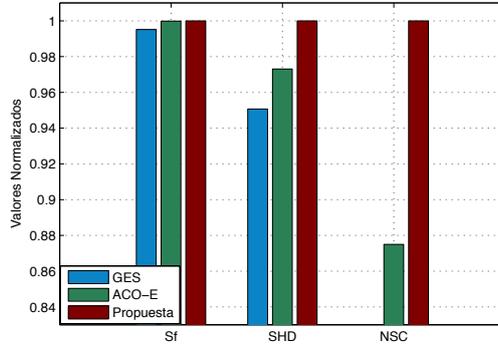
	Criterio Clasificación	20 Muestras	500 Muestras
1.	$Sf_{Propuesta} > Sf_{ACO-E}$	1	4
2.	$Sf_{Propuesta} > Sf_{GES}$	144	144
3.	$SHD_{Propuesta} < SHD_{ACO-E}$ y $SHD_{Propuesta} < SHD_{GES}$	21	41
4.	$SHD_{Propuesta} < SHD_{GES}$	24	66
5.	$Sf_{Propuesta} > Sf_{ACO-E}$ y $SHD_{Propuesta} < SHD_{ACO-E}$	0	2
6.	$NSC_{ACO-E} > NSC_{Propuesta}$	0	129

---

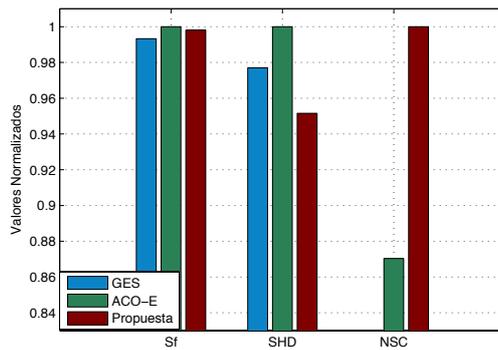
ración a la diferencia entre las otras dos métricas. Para el criterio utilizado con respecto a la métrica *SHD*, hubo 21 combinaciones de los parámetros para los cuales el algoritmo propuesto tuvo menor valor con respecto a los otros dos algoritmos. En promedio, como puede verse en la figura 2.3(b), para las 21 combinaciones se obtuvo un 2,6% y 4,8% menor valor que para los algoritmos Greedy y ACO-E respectivamente. De acuerdo a la suma cuadrática de las diferencias en los valores del SHD de estas 21 combinaciones, la mayor diferencia fue para la combinación  $\rho = 0,1$ ,  $q_0 = 0,3$ ,  $\alpha = 5$  y  $\beta = 1$ . Como es descrito en la figura 2.3(c), la diferencia con respecto al algoritmo ACO-E fue del 10,7% y con el algoritmo Greedy del 5,3%.

Los resultados anteriores muestran que el algoritmo propuesto para redes pequeñas, usando 20 muestras, y en específicas combinaciones de sus parámetros, tiene la posibilidad de obtener estructuras con menos interacciones erróneas (falsos positivos) que los algoritmos ACO-E y Greedy.

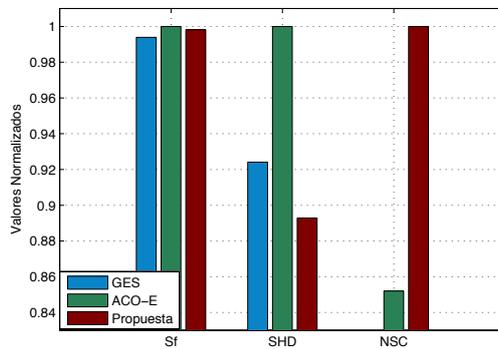
Para el caso de un conjunto de datos de 500 muestras, como lo muestra el cuadro 2.2, de las 144 combinaciones de los parámetros, en cuatro de ellos, según la primera fila, el algoritmo propuesto obtuvo mayor valor en la función de puntaje con respecto al algoritmo ACO-E. De estas cuatro, aquella combinación con mayor diferencia en la métrica ACO-E,  $\rho = 0,5$ ,  $q_0 = 0,7$ ,  $\alpha = 5$  y  $\beta = 1$ , obtuvo un valor del 10,4% y 4% menor con respecto al algoritmo GES y ACO-E respectivamente, como es mostrado en la figura 2.4(a). En la tercera fila del cuadro 2.2 para 500 muestras, de las 144 combinaciones se obtuvieron 41 para las cuales el valor de la métrica *SHD* del algoritmo propuesto fue menor que de los otros dos. Específicamente en promedio el valor fue de un 5% y 9% con respecto al algoritmo ACO-E y Greedy respectivamente como se muestra en la figura 2.4(b). De estos 41 combinaciones, como lo muestra la quinta fila del cuadro, en dos de ellas se obtuvo un mayor valor para la función de puntaje. Una de ellas es la mostrado en la figura 2.4(a), mencionado arriba y la otra es la combinación  $\rho = 0,5$ ,  $q_0 = 0,3$ ,  $\alpha = 5$  y  $\beta = 1$ , la cual obtuvo un valor del *SHD* 3% menor para el algoritmo propuesto con respecto a los otros dos algoritmos.



(a) Resultados para la combinación  $\rho = 0,1$ ,  $q_0 = 0,7$ ,  $\alpha = 3$  y  $\beta = 1$  referido en la fila 1 del cuadro 2.2 para 20 muestras.



(b) Resultados promedio para las combinaciones referidas en la fila 3 del cuadro 2.2 para 20 muestras.



(c) Resultados para la mejor combinación  $\rho = 0,1$ ,  $q_0 = 0,3$ ,  $\alpha = 5$  y  $\beta = 1$  de las referidas en la fila 3 del cuadro 2.2 para 20 muestras.

**Figura 2.3:** Resultados significativos obtenidos en las métricas  $S_f$ ,  $SHD$  y  $NSC$  en el protocolo descrito para los tres algoritmos probados utilizando datos con 20 muestras. Los valores están normalizados.

---

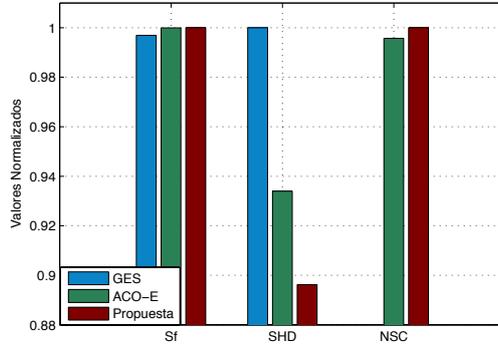
La última fila del cuadro 2.2 hace referencia a la medida de complejidad del algoritmo NSC. El número de llamadas a la función de puntaje por el algoritmo propuesto fue menor que en el algoritmo ACO-E en el 90% de las combinaciones utilizando 500 muestras. En promedio esta diferencia fue del 6,8% menor. Anãdido a esto, para las combinaciones referidas en la fila 3 del cuadro, mostradas en la figura 2.4(b), en promedio el algoritmo propuesto obtuvo una disminuci3n del 4,6% con respecto al algoritmo ACO-E. Y por ultiimo en promedio para la combinaci3n mostrada en la figura 2.4(c) referida en la fila cinco del cuadro obtuvo un valor del NSC del 5,8% menor que el algoritmo ACO-E.

Todo lo anterior indica que el algoritmo propuesto, para la mayoría de combinaciones, para 500 muestras de redes pequeñãs tiene un orden de complejidad menor que el algoritmo ACO-E. Ademãs para las combinaciones referidas en la fila 5 del cuadro 2.2 y mostradas en la figura 3.4 tienen la posibilidad de obtener estructuras con menos interacciones err3neas (falsos positivos) que los algoritmos ACO-E y Greedy, y a su vez con mayor o muy semejante probabilidad como indica la funci3n de puntaje.

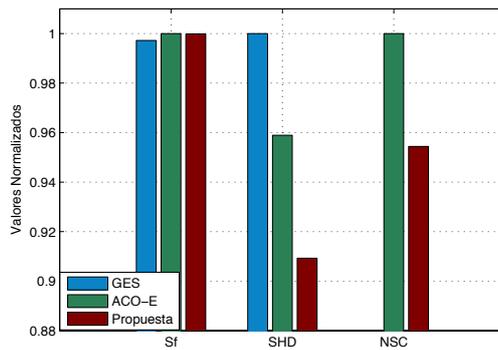
## 2.3. Conclusiones

Uno de los principales resultados de este trabajo fue el desarrollo de un algoritmo de aprendizaje de clases de equivalencia basado en modelos de b3squeda inspirado en colonia de hormigas. El algoritmo incluy3 operadores para el movimiento entre clases de equivalencias, una representaci3n de cada movimiento basada en colonia de hormigas, el concepto de que las hormigas construyen un solo camino y compiten por cada elemento de dicho camino.

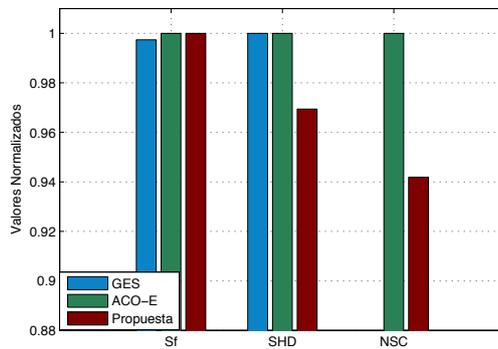
Los resultados mostraron que el algoritmo propuesto para redes pequeñãs, usando 20 muestras, y en específcas combinaciones de sus parámetros, tiene la posibilidad de obtener estructuras con menos interacciones err3neas (falsos positivos) que los algoritmos ACO-E y Greedy. Ademãs, los resultados tambi3n mostraron que, para 500 muestras, y redes pequeñãs tiene la



(a) Resultados para la mejor combinación  $\rho = 0,5$ ,  $q_0 = 0,7$ ,  $\alpha = 3$  y  $\beta = 1$  referido en la fila 1 del cuadro 2.2 para 500 muestras.



(b) Resultados promedio para las combinaciones referidas en la fila 3 del cuadro 2.2 para 500 muestras.



(c) Resultados para la combinación  $\rho = 0,5$ ,  $q_0 = 0,3$ ,  $\alpha = 3$  y  $\beta = 1$ . Es una de las 2 referidas en la fila 5 del cuadro 2.2 para 500 muestras, la otra es mostrada en la figure 2.4(a).

**Figura 2.4:** Resultados significativos obtenidos en las métricas  $S_f$ ,  $SHD$  y  $NSC$  en el protocolo descrito para los tres algoritmos probados utilizando datos con 500 muestras. Los valores están normalizados.

---

posibilidad de obtener estructuras con menos interacciones erróneas (falsos positivos) que los algoritmos ACO-E y Greedy, y a su vez con mayor o muy semejante probabilidad como indica la función de puntaje y con un orden de complejidad menor que el algoritmo ACO-E.

## Capítulo 3

# Aproximación de una Red de Regulación de la Expresión Genética

El entendimiento apropiado de como una célula normal o anormal responden a diferentes estímulos requiere de una comprensión de como sus componentes interactúan entre si. Una forma de medir en que nivel cada componente molecular responde ante estos estímulos dentro de la célula es a través de los microarreglos de ADN. Por medio de una matriz de datos, estos microarreglos cuantifican el nivel de respuesta de un gen dentro del ADN bajo ciertas condiciones. A partir de esta matriz es posible es construir una estructura de una red bayesiana que aproxime las interacciones involucradas en la regulación de las respuestas genéticas de manera estadística. Esta propiedad de las redes bayesianas, de representar relaciones no lineales y estocásticas entre muchas moléculas es tal vez la propiedad mas atrayente. La razón de lo anterior es porque esta representación hace altamente probable el descubrimiento de interacciones reales de moléculas para células bajo cierta condición, y así realizar un aporte significativo dentro del campo de la biología molecular. En este capítulo se describirá la aplicación de los capítulos anteriores en cuanto a la inferencia de interacciones de naturaleza estocástica entre moléculas que responden en relación a cambios de la molécula EVI-1 y su aporte al descubri-

---

miento de interacciones reales entre ellas. La razón por la cual se escogió este gen, es por los recientes estudios de como la producción anormal de éste se ve asociada a una de las formas más severas de un tipo de patología tumoral denominado Leucemia Mieloide Aguda (LMA)[42].

### 3.1. Generalidades de Biología molecular

La Célula es la unidad fundamental de cualquier sistema viviente. Esta se compone de un conjunto de redes de interacciones biomoleculares que definen su estructura y funciones. Dichas redes de relaciones, son básicamente compuestas por biomoléculas llamadas proteínas, que de acuerdo a la red en que pertenezca desarrollan una función, ya sea, estructural, metabólica, reproductiva, entre otras.

La forma como las proteínas son formadas o “expresadas” dependen de la información que se encuentre en el código genético en la célula y de como otras proteínas regulan los “factores de transcripción” del código. Dicho de otra manera, las proteínas son “los efectores de la información genética”[43]. Formalmente la expresión de una proteína se define así:

- Transducción: Cuando una parte del código genético (ADN - ácido desoxirribonucleico) es copiado (la copia es llamada ARN - ácido ribonucleico)
- Traslación: La copia del código genético es procesado en el ribosoma para obtener a partir de ella una proteína

Cabe resaltar que se conocen hasta 25000 genes dentro del código del ADN, de los cuales se pueden obtener hasta 450000 y de ellas sintetizar en una célula desde 1 hasta 10 *millones* de proteínas [43]. Y mas aun, las funciones celulares efectuadas por la maquinaria de las redes de proteínas define son las que definen de acuerdo al contexto celular los procesos fisiológicos de un sistema viviente. De esta forma, es posible decir que, de acuerdo a como

---

se expresen los genes, esto es, de acuerdo a como el conjunto de interacciones dentro de una celular se establezca, y el ambiente en que se encuentra, define la emergencia de la actividad de un sistema viviente.

Es por ellos que el propósito de este estudio es, basado en información de la expresión genética, desarrollar aproximaciones de las interacciones llevadas a cabo para su regulación. Para ello entonces se definirá como se llevan a cabo esas mediciones, el enfoque escogido para interpretar dichos datos, y el método para obtener la aproximación de las interacciones mencionadas.

### **3.2. Microarreglos de ADN**

Un microarreglo de ADN es un chip que contiene información relativa *al nivel de expresión* de un conjunto de genes en una célula. Fundamentalmente es una técnica en la cual se compara el nivel de expresión genética de una célula de *referencia* con la información genética de una célula *blanco* o que la que se desea estudiar. Por ejemplo, una célula de referencia puede ser una célula sana, y la célula blanco una anormal [43; 44]. El proceso que se realiza para imprimir la información genética en el chip y para obtener la información de dicho chip, aun sigue siendo objeto de estudio con el fin de minimizar los errores en las mediciones [45]. Surgen entonces dos preguntas con respecto a un microarreglo: ¿Cómo se imprimió la información en dicho Chip? y ¿Cómo se obtiene la información incluida allí?

La respuesta a las preguntas anteriores se resume en un proceso de cuatro pasos:

1. Se obtienen copias del código genético, esto es cadenas de ARNm (la 'm' significa 'mensajero') por separado de la célula de referencia y la célula blanco
2. Se obtiene el ADN relativo a cada una de las células a partir del ARNm, y se etiquetan con fluorescencia de diferente color (este proceso es llevado

---

a cabo por enzimas llamadas 'transcriptasas inversas' <sup>1</sup>). Normalmente las cadenas de ADN de la célula de referencia se etiquetan de color verde y las de la célula blanco de rojo.

3. Se mezclan las cadenas de ADN etiquetadas y se imprimen en un chip con una matriz donde cada elemento de la matriz contiene ácidos nucleicos de tal forma que, si son cadenas complementarias de algunas de las cadenas de ADN de la célula de referencia o blanco, estas últimas se unen a ellas, imprimiendo en el elemento de la matriz la información genética, y de la misma forma en todo el chip.
4. El chip es puesto en una máquina que escanea la imagen y transforma el color del fluorescente en información digital de colores entre verde y rojo. Si un elemento obtenido es de color verde quiere decir que ese gen en la célula blanco no se expresó. De la misma manera si un elemento es rojo, quiere decir que ese gen está expresado solo en la célula blanco y no la de referencia. La gama de colores entre verde y rojo refleja la proporción en que se expresó un gen, si es más tendiente a verde, quiere decir que la expresión del gen está *inhibido* en la célula blanco, y cuando es más tendiente a rojo se dice que el gen está *activado* en la célula blanco.

Matemáticamente, se puede expresar la relación entre la expresión del gen  $i$  en la célula blanco  $R_i$  con respecto a su expresión en la célula de referencia  $G_i$  así:

$$T_i = \frac{R_i}{G_i} \quad (3.1)$$

Que se convierte, por conveniencia a escala  $\log_2$ ,

$$\log_2(T_i) = \log_2\left(\frac{R_i}{G_i}\right) \quad (3.2)$$

puesto que cuando  $\log_2(T_i) > 1$  se dice que el gen  $i$  está activado en la célula blanco, y si  $\log_2(T_i) < 1$  se dice que el gen  $i$  en la célula blanco está inhibida. De esa forma la discretización más simple, esto es de tres estados

---

<sup>1</sup>llamado así porque la transcriptasa realiza el proceso de transducción

---

(inhibido, neutro, activado) es sencilla de hacer[46].

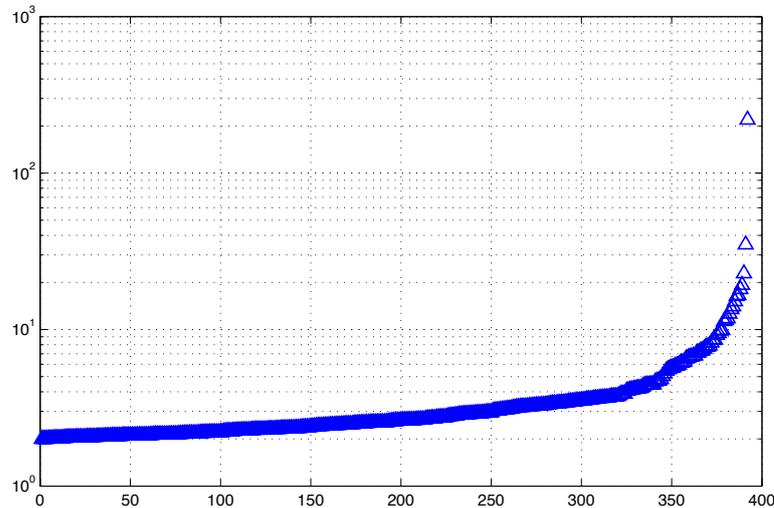
Generalizando, si el número de células blanco es igual a  $m$ , para las cuales se desea obtener la información del nivel de expresión de  $n$  genes, el conjunto de datos final es una matriz  $\mathbf{D}$  de  $n \times m$  donde  $d_{ij}$  representa el radio  $\log_2(T_i)$  para la célula blanco  $j$ , donde  $i = 1, \dots, n$  y  $j = 1, \dots, m$ . A partir de esta matriz es posible obtener una aproximación de una red bayesiana que represente la red de regulación genética deseada, dependiendo de que genes se escojan.

### 3.2.1. Selección de Genes Relevantes

Con el fin de filtrar cuáles genes se tendrán en cuenta para la inferencia de la red, fue obtenido un microarreglo a partir de un experimento con condiciones específicas para dos tipos de células. Este microarreglo contiene la información relativa al nivel de expresión genética en dos células diferentes  $A$  y  $B$ . Estas células fueron condicionadas en un cultivo de tal forma que la célula  $A$  tiene características de una célula "mortal", es decir, estable, con una pequeña alteración en el gen EVI-1; la célula  $B$  fue condicionada como célula "inmortal", es decir, es tumoral. Para cada célula se midieron los niveles de expresión genética. Este procedimiento se hizo tres veces para cada tipo de célula.

Con la información fue calculado el número de veces que el gen en cada sonda se ha expresado en la célula  $A$  con respecto a la célula  $B$ . Este cambio puede ser negativo o positivo. La figura 3.2.1 muestra el cambio que hubo en todos los genes con cambios mayores a 2.

Al evaluar la confianza de este cambio por medio del test t-student, se puede confirmar, para los datos en la figura 3.2.1 todas estas diferencias se mantendrán con un valor de confianza mayor o igual al 95%. Entre mas grande la distancia, mayor es el nivel de expresión del gen en las condiciones dadas. De estos genes son entonces seleccionados aquellos para realizar la inferencia e inferir las interacciones.



**Figura 3.1:** Diferencia entre el nivel de expresión genética de la célula A con respecto a la B mayores a 2.

### 3.2.2. Características del microarreglo utilizado

Después de haber sido seleccionados los genes relevantes de acuerdo a los criterios en la sección anterior se seleccionó un microarreglo en particular de los encontrados en la base de datos *Gene Expression Omnibus* ([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) con número de acceso GSE6891). Este microarreglo contiene datos de expresión genética de grupos de pacientes con Leucemia Mieloide Aguda (LMA) con el objetivo de estudiar su aplicabilidad en la predicción de subclases asociadas a una anomalía específica en la LMA [47]. Dentro de sus resultados se encontró que dentro de una de aquellas subclases, uno de los genes más discriminantes para detectar anomalías de dicha clase fue el gen EVI-1. Estas características hacen de este conjunto de datos propicio y viable para ser utilizado en el aprendizaje de la red. Específicamente el conjunto de datos cuenta con información de 461 muestras tomadas de la sangre o de la médula ósea. Las muestras fueron procesadas de la forma mencionada en la sección 3.2 con la diferencia que los valores cuantificados menores a 30 se mantuvieron en ese valor como valor mínimo. Después de ellos fueron transformados a la escala  $\log_2$ , que es en el formato que se pre-

---

sentan finalmente.

De este microarreglo se obtuvieron solamente las filas (genes) que hacen parte a la lista de genes relevantes seleccionados seleccionados a partir de allí. De esta forma este es el conjunto de datos **D** es el utilizado por el algoritmo para aproximar la red de interacciones.

### 3.3. Obtención de una Red Semilla

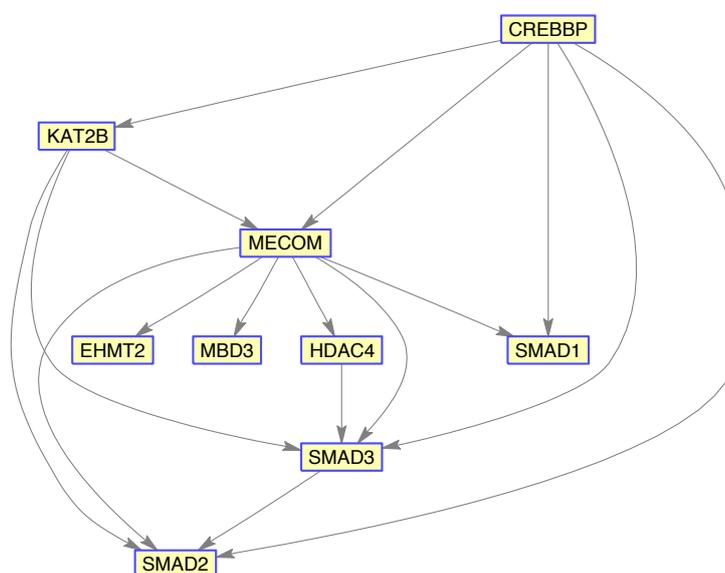
Antes de realizar el proceso de aprendizaje es posible y necesario extraer conocimiento de interacciones moleculares a partir de la literatura científica. Esta información es almacenada en bases de datos. Se escogió que la base de datos Biogrid (<http://thebiogrid.org/>) por ser la base de datos libre más completa, en la que se encontraban mas interacciones, y la que mantenía el formato de los nombres de las moléculas como en los microarreglos.

Para extraer una red inicial a partir de dicha información método eficiente para obtener una estructura de una red bayesiana inicial donde no haya redundancia en los datos basado en el método descrito en [48] descrito de la siguiente manera:

1. Obtener Red inicial de la Biomolecula en studio (EVI-1) (<http://thebiogrid.org/>)
2. De cada uno de los elementos de la red anterior obtener su propia red (ya que las interacciones de EVI-1 son muy pocas).
3. Unir las redes filtrando las interacciones repetidas.
4. Filtrar las interacciones de acuerdo a los genes listados en el microarreglo.
5. Se forma una red sin unidireccional.
6. Se aplica el algoritmo de [48] basado en depth-first para convertir la red anterior en una grafo aciclico no dirigido.

---

Teniendo entonces de la sección anterior el conjunto de datos con los genes relevantes, entonces de la red semilla son utilizados aquellas interacciones cuyos genes se encuentran en el microarreglo. La figure 3.2 muestra las interacciones resultantes que son utilizados como información inicial para el aprendizaje, para genes relevantes con diferencia mayor o igual a 6.



**Figura 3.2:** Resultados Interacciones obtenidas de la base de datos BioGrid.

### 3.4. Aproximación Red Biológica

El procedimiento para obtener nuevas interacciones utilizando el algoritmo de aprendizaje propuesto en el capítulo anterior es realizado de la siguiente manera. Dado que es posible que haya información espuria en varios casos en la tabla de datos, que produzcan posibles falsos positivos, se obtienen un conjunto de  $m$  estructuras que que tengan la mayor probabilidad de aproximar la red de regulación genética y se seleccionan las interacciones en común de todas ellas. Para cada estructura  $G_i$  obtenida se obtiene su función de puntaje

---

o probabilidad de aproximar la información en los datos  $S_f(G_i)$  y a partir de ellas se obtiene un medidor de confianza por el promedio de las probabilidades, así:

$$conf(G) = \frac{1}{m} \sum_{i=1}^m S_f(G_i) \quad (3.3)$$

El resultado es entonces un grafo  $G$  con las interacciones en común mencionadas, con un valor de confianza  $conf(G)$ .

Para los datos del microarreglo con los genes relevantes filtrados, el resultante modelo  $G$  inferido por el algoritmo propuesto, con un número de 20 arcos entre 16 componentes para los genes con relevantes filtrados con diferencia igual a 6.

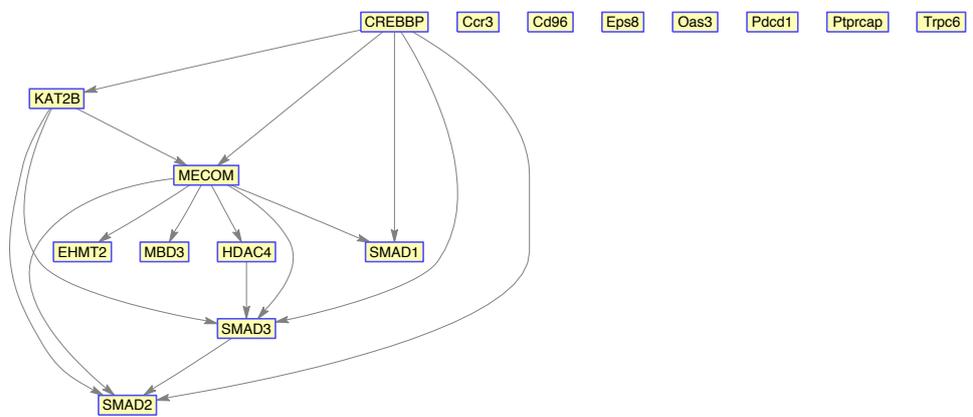
La figura 3.4 describe la comparación del valor de confianza obtenido para las estructuras inferidas con el algoritmo propuesto mostrado en la figura 3.3(b) y el valor de la función de puntaje para la red inicial mostrada en la figura 3.3(a). Como muestran los valores, el valor de confianza obtenido por el algoritmo propuesto tiene un puntaje del 10 % mayor que el inicial. Dados los resultados del capítulo anterior, el algoritmo infiere una red de interacciones de naturaleza estadística, y a pesar de que tiene la posibilidad de obtener estructuras con menos interacciones erróneas, sólo es comprobable a través de métodos experimentales.

Sin embargo, estas nuevas interacciones obtenidas con el algoritmo propuesto son de potencial importancia al elucidar la exacta interacción que hay entre los elementos. Es decir, las interacciones encontradas corresponder a 3 clases al ser probadas en laboratorio <sup>1</sup>.

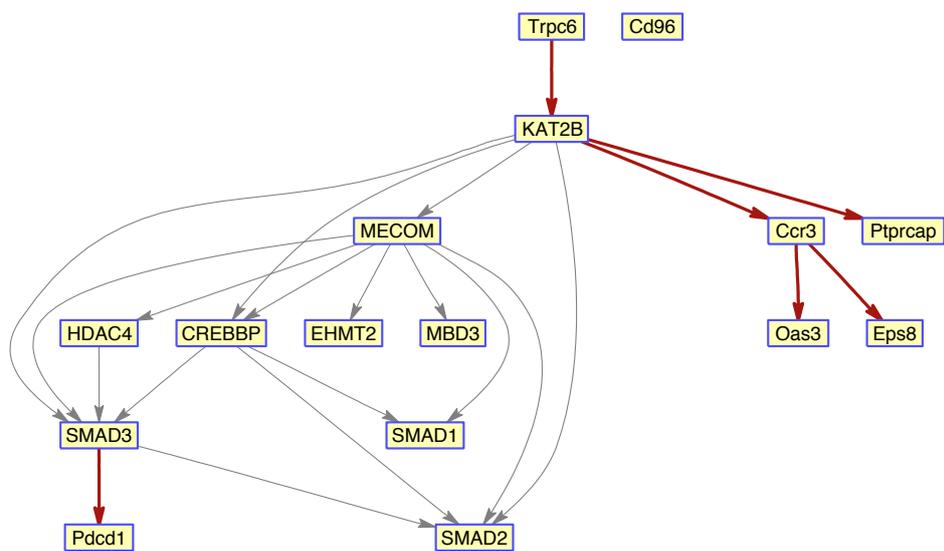
La primera clase es aquella en la que se encuentran las interacciones *directas*, es decir, aquellos genes que al expresarse uno de ellos directamente regula al otro en la dirección aproximada por el algoritmo. La figura 3.5 muestra un ejemplo de relación directa, es sencillamente aquella que es encontrada experimentalmente tal y como se aproximó.

---

<sup>1</sup>Procedimiento que va más allá del alcance este trabajo.

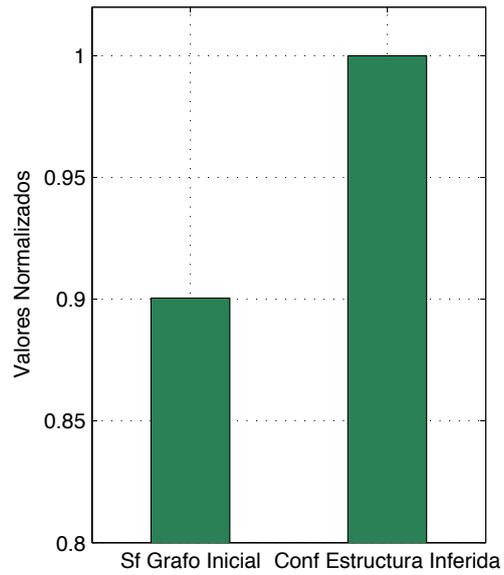


(a) Estructura inicial antes del proceso de aprendizaje.



(b) Estructura inferida por el algoritmo a partir de los datos del microarreglo con los genes relevantes. Se inferieron 6 interacciones nuevas, de las cuales tres fueron con el gen SMAD3 y 5 con KAT2B.

**Figura 3.3:** Estructura inicial y final del proceso de aprendizaje con el algoritmo propuesto.



**Figura 3.4:** Comparación del valor de puntaje de la red semilla inicial y el valor de confianza estimado por la estructura inferida con el algoritmo propuesto.



**Figura 3.5:** Relación directa entre el gen KAT2B y Ccr3

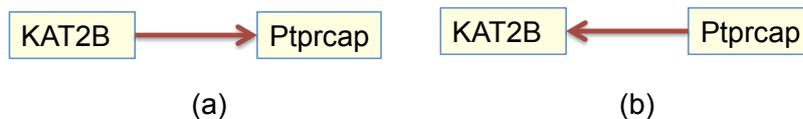
---

En el segundo grupo, están las interacciones *indirectas*, es decir, aquellas en las cuales la regulación entre los genes, aunque en la dirección aproximada, pueden ser mediada por otros genes que no fueron tenidos en cuenta como genes relevantes en la sección anterior. La figura 3.6 muestra un ejemplo en el cual los genes SMAD3 y Pcd1 aproximados por el algoritmo podrías estar mediados por dos genes realmente.



**Figura 3.6:** Relación indirecta entre el gen SMAD3 y Pcd1 mediada por dos genes .

Por último, las interacciones pueden llegar a ser *inversas*, es decir, en cualquiera de los casos anteriores es posible que la dirección de la regulación sea contraria a la que se paroximó con el algoritmo, como es mostrado en la figura 3.7.



**Figura 3.7:** (a) Interacción aproximada entre el gen Ptpcap y KAT2B. (b) Relación inversa que posiblemente se encontraría experimentalmente.

### 3.5. Conclusiones

En este capítulo se describió la aplicación de los capítulos anteriores en cuanto a la inferencia de interacciones de naturaleza estocástica entre moléculas que responden en relación a cambios de la molécula EVI-1 y su aporte al descubrimiento de interacciones reales entre ellas.

Se describieron los conceptos básicos sobre biología molecular y sobre microarreglos necesarios para los métodos utilizados. Además se mostró como es posible emplear datos de microarreglo para la selección de genes relevantes de acuerdo a cambios significativos cuando hay cambios en las condiciones celulares. También se mostró la utilidad de utilizar información a priori de la

---

literatura científica para direccionar mas precisamente el aprendizaje.

Por último se mostró los resultados de la aplicación del algoritmo a un microarreglo obtenido por muestras de pacientes con Leucemia Mieloide Aguda (LMA) con el fin de encontrar interacciones con dependencias estadísticas como potenciales interacciones biológicas con alta probabilidad. Este último resultado es de gran importancia puesto que aporta en la generación de conocimiento en investigaciones de como interactúan los genes relacionados con enfermedades como la LMA.

# Capítulo 4

## Conclusiones

De todo el trabajo realizado se tienen las siguientes conclusiones en cuanto su aporte:

Se realizó la descripción formal del modelado bayesiano para el aprendizaje de redes bayesiana, produciendo como resultado la conocida función de puntaje bayesiana de dirichlet. Además se describieron las ventajas del enfoque de puntaje y búsqueda entre clases de equivalencia de redes bayesianas. Principalmente, se evita por medio de operadores el movimiento redundante de búsqueda entre la misma clase de equivalencia.

Además, uno de los principales resultados de este trabajo fue el desarrollo de un algoritmo de aprendizaje de clases de equivalencia basado en modelos de búsqueda inspirado en colonia de hormigas. El algoritmo incluyó operadores para el movimiento entre clases de equivalencias, una representación de cada movimiento basada en colonia de hormigas, el concepto de que las hormigas construyen un solo camino y compiten por cada elemento de dicho camino.

Los resultados mostraron que el algoritmo propuesto para redes pequeñas, usando 20 muestras, y en específicas combinaciones de sus parámetros, tiene la posibilidad de obtener estructuras con menos interacciones erróneas (falsos positivos) que los algoritmos ACO-E y Greedy. Además, los resultados también mostraron que, para 500 muestras, y redes pequeñas tiene la

---

posibilidad de obtener estructuras con menos interacciones erróneas (falsos positivos) que los algoritmos ACO-E y Greedy, y a su vez con mayor o muy semejante probabilidad como indica la función de puntaje y con un orden de complejidad menor que el algoritmo ACO-E.

También se mostró la aplicación del modelo y el algoritmo propuesto en cuanto a la inferencia de interacciones de naturaleza estocástica entre moléculas que responden en relación a cambios de la molécula EVI-1 y su aporte al descubrimiento de interacciones reales entre ellas.

Se describieron los conceptos básicos sobre biología molecular y sobre microarreglos necesarios para los métodos utilizados. Además se mostró como es posible emplear datos de microarreglo para la selección de genes relevantes de acuerdo a cambios significativos cuando hay cambios en las condiciones celulares. También se mostró la utilidad de utilizar información a priori de la literatura científica para direccionar mas precisamente el aprendizaje.

Por último se mostró los resultados de la aplicación del algoritmo a un microarreglo obtenido por muestras de pacientes con Leucemia Mieloide Aguda (LMA) con el fin de encontrar interacciones con dependencias estadísticas como potenciales interacciones biológicas con alta probabilidad. Este último resultado es de gran importancia puesto que aporta en la generación de conocimiento en investigaciones de como interactúan los genes relacionados con enfermedades como la LMA.

# Referencias

- [1] W Tong. Analyzing the biology on the system level. *Genomics Proteomics Bioinformatics*, 2(1):6–14, 2004.
- [2] Hiroaki Kitano. Systems Biology: A Brief Overview. *Science*, 295(5560):1662–1664, 2002.
- [3] Anthony Finkelstein, James Hetherington, Linzhong Li, Ofer Margoninski, Peter Saffrey, Rob Seymour, and Anne Warner. Computational challenges of systems biology. *Computer*, 37(5):26–33, 2004.
- [4] P.K. Dhar, Hao Zhu, and S.K. Mishra. Computational approach to systems biology: from fraction to integration and beyond. *NanoBioscience, IEEE Transactions on*, 3(3):144–152, Sept. 2004.
- [5] K.-H. Cho, S.-M. Choo, S.H. Jung, J.-R. Kim, H.-S. Choi, and J. Kim. Reverse engineering of gene regulatory networks. *Systems Biology, IET*, 1(3):149–163, May 2007.
- [6] Dana Pe'er. Bayesian Network Analysis of Signaling Networks: A Primer. *Sci. STKE*, 2005(281):pl4–, 2005.
- [7] Hidde de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):67–103, January 2002.
- [8] Alberto Ambesi-Impiombato & Diego di Bernardo Mukesh Bansal, Vincenzo Belcastro. How to infer gene networks from expression profiles. *Molecular Systems Biology*, 3, 2007.

- 
- [9] Feng He, Rudi Balling, and An-Ping Zeng. Reverse engineering and verification of gene networks: Principles, assumptions, and limitations of present methods and future perspectives. *Journal of Biotechnology*, In Press, Corrected Proof:–, 2009.
- [10] Valdimir Filkov. Identifying gene regulatory networks from gene expression data. *Handbook of Computational Molecular Biology*, Handbook of Computational Molecular Biology, 2005.
- [11] Karen Sachs, Omar Perez, Dana Pe’er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. *Science*, 308(5721):523–529, 2005.
- [12] Finn V. Jensen and Thomas D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer Science + Business Media, LLC, 2007.
- [13] Richard E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall; illustrated edition, 2003.
- [14] Dana Pe’er. *FROM GENE EXPRESSION TO MOLECULAR PATHWAYS*. PhD thesis, The Hebrew University, 2003.
- [15] Alexander J. Hartemink, David K. Gifford, Tommi S. Jaakkola, and Richard A. Young. Bayesian methods for elucidating genetic regulatory networks. *IEEE Intelligent Systems*, 17(2):37–43, 2002.
- [16] Alexander John Hartemink. *Principled Computational Methods for the Validation and Discovery of Genetic Regulatory Networks*. PhD thesis, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, September 2001.
- [17] Chris Needham, Iain Manfield, Andrew Bulpitt, Philip Gilmartin, and David Westhead. From gene expression to gene regulatory networks in arabidopsis thaliana. *BMC Systems Biology*, 3(1):85, 2009.
- [18] Gregory F. Cooper and Edward Herskovits. A bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9, Number 4:309–347, 1992.

- 
- [19] David Heckerman. A tutorial on learning with bayesian networks. Technical report, Microsoft Research, 1995.
- [20] David Maxwell Chickering. Learning bayesian networks is np-complete. pages 121–130, 1996.
- [21] Xue wen Chen, Gopalakrishna Anantha, and Xinkun Wang. An effective structure learning method for constructing gene networks. *Bioinformatics*, 22(11):1367–1374, 2006.
- [22] K. Numata, S. Imoto, and S. Miyano. A structure learning algorithm for inference of gene networks from microarray gene expression data using bayesian networks. In *Bioinformatics and Bioengineering, 2007. BIBE 2007. Proceedings of the 7th IEEE International Conference on*, pages 1280–1284, Oct. 2007.
- [23] Pedro Larrañaga, Mikel Poza, Yosu Yurramendi, Roberto H. Murga, and Cindy M.H. Kuijpers. Structure learning of bayesian networks by genetic algorithms: A performance analysis of control parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):912–926, 1996.
- [24] Man Leung Wong and Kwong Sak Leung. An efficient data mining method for learning bayesian networks using an evolutionary algorithm-based hybrid approach. *Evolutionary Computation, IEEE Transactions on*, 8(4):378–404, Aug. 2004.
- [25] Luis M. de Campos, Juan M. Fernández-Luna, José A. Gámez, and José M. Puerta. Ant colony optimization for learning bayesian networks. *International Journal of Approximate Reasoning*, 31(3):291 – 311, 2002.
- [26] Zhihua Du, Yiwei Wang, and Zhen Ji. A new structure learning method for constructing gene networks. In *Bioinformatics and Biomedical Engineering , 2009. ICBBE 2009. 3rd International Conference on*, pages 1–4, June 2009.
- [27] Pablo A. D. Castro and Fernando J. Von Zuben. An immune-inspired approach to bayesian networks. In *HIS '05: Proceedings of the Fifth In-*

- 
- ternational Conference on Hybrid Intelligent Systems*, pages 23–28, Washington, DC, USA, 2005. IEEE Computer Society.
- [28] David Maxwell Chickering. Learning equivalence classes of bayesian-network structures. *The Journal of Machine Learning Research*, 2:445 – 498, March 2002.
- [29] Paul Munteanu and Denis Cau. Efficient score-based learning of equivalence classes of bayesian networks. In *PKDD '00: Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery*, pages 96–105, London, UK, 2000. Springer-Verlag.
- [30] P. Munteanu and M. Bendou. The eq framework for learning equivalence classes of bayesian networks. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 417–424, 2001.
- [31] Silva Acid and Luis M. de Campos. Searching for bayesian network structures in the space of restricted acyclic aprtially directed graphs. *Journal of Artificial Intelligence Research*, 18:445–490, 2003.
- [32] Steven Gillispie Department and Steven B. Gillispie. Enumerating markov equivalence classes of acyclic digraph models. In *Proc. of the Conf. on Uncertainty in Artificial Intelligence*, pages 171–177. Morgan Kaufmann, 2001.
- [33] Steven B. Gillispie and Michael D. Perlman. The size distribution for markov equivalence classes of acyclic digraph models. *Artificial Intelligence*, 141(1-2):137 – 155, 2002.
- [34] Steven B. Gillispie. Formulas for counting acyclic digraph markov equivalence classes. *Journal of Statistical Planning and Inference*, 136(4):1410 – 1432, 2006.
- [35] Rónán Daly and Qiang Shen. Learning bayesian network equivalence classes with ant colony optimization. *Artificial Intelligence Research*, 35:391–447, 2009.

- 
- [36] P.C. Pinto, A. Nagele, M. Dejori, T.A. Runkler, and J.M.C. Sousa. Using a local discovery ant algorithm for bayesian network structure learning. *Evolutionary Computation, IEEE Transactions on*, 13(4):767–779, Aug. 2009.
- [37] Laura E. Brown Ioannis Tsamardinos and Constantin F. Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, 2006.
- [38] Luis M. de Campos. A scoring function for learning bayesian networks based on mutual information and conditional independence tests. *The Journal of Machine Learning Research*, 7:2149 – 2187, December 2006.
- [39] Marco Dorigo and Thomas Stutzle. *Ant colony Optimization*. Massachusetts Institute of Technology, 2004.
- [40] Marco Dorigo and Christian Blum. Ant colony optimization theory: A survey. *Theoretical Computer Science*, 344(2-3):243 – 278, 2005.
- [41] Christian Blum. Ant colony optimization: Introduction and recent trends. *Physics of Life Reviews*, 2(4):353 – 373, 2005.
- [42] Rotraud Wieser. The oncogene and developmental regulator *evi1*: Expression, biochemical properties, and biological functions. *Gene*, 396(2):346 – 357, 2007.
- [43] J©rme and Lamartine. The benefits of dna microarrays in fundamental and applied bio-medicine. *Materials Science and Engineering: C*, 26(2,Äì3):354 – 359, 2006. |ce:title¿Selected Papers Presented at the MADICA 2004 Conference, Fourth Maghreb-Europe Meeting on Materials and their Applications for Devices and Physical, Chemical and Biological Sensors Tunis-Ghammart, Tunisia, 29 November- 1 December 2004. |/ce:title¿.
- [44] Rolando Meloni, Olfa Khalfallah, and Nicole Faucon Biguet. Dna microarrays and pharmacogenomics. *Pharmacological Research*, 49(4):303 – 308, 2004. Trends in pharmacogenomics.

- 
- [45] Lin Wang and Paul C.H. Li. Microfluidic dna microarray analysis: A review. *Analytica Chimica Acta*, 687(1):12 – 27, 2011.
- [46] John Quackenbush. Microarray data normalization and transformation. *Nat Genet VL* -.
- [47] Roel G.W. Verhaak, Bas J. Wouters, Claudia A.J. Erpelinck, Saman Abbas, H. Berna Beverloo, Sanne Lugthart, Bob Lwenberg, Ruud Delwel, and Peter J.M. Valk. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*, 94(1):131–134, 2009.
- [48] Amira Djebbari and John Quackenbush. Seeded bayesian networks: Constructing genetic networks from microarray data. *BMC Systems Biology*, 2(1):57, 2008.