



UNIVERSIDAD DEL PAPALOAPAN

Campus Loma Bonita

INGENIERÍA EN COMPUTACIÓN

APRENDIZAJE DE LA ESTRUCTURA EN MODELOS GRÁFICOS

TESIS

QUE PARA OBTENER EL TÍTULO DE:
INGENIERO EN COMPUTACIÓN

PRESENTA:

JESSICA MARÍA HERNÁNDEZ MARTÍNEZ

DIRECTOR DE TESIS:

M.C. EDUARDO ORTIZ HERNÁNDEZ

LOMA BONITA, OAXACA,

OCTUBRE 2013

Agradecimientos

Gracias a mi madre María del Rosario y a mi padre José Inés por todo el apoyo y comprensión que hasta ahora me siguen brindando. A mis tres hermanos: Liliana (la cual ha sido como otra madre para mi), Perla Janeth y José Arnulfo les doy las gracias, ya que sacrificaron algunas oportunidades que tuvieron para brindarmelas sin pedirme nada a cambio.

A mis dos grandes amigos Juan Erick y Freddy Arturo los cuales me apoyaron incondicionalmente durante la mayor parte de mi carrera profesional y me enseñaron tantas cosas valiosas que nunca olvidaré.

A mi director de tesis el M.C Eduardo Ortiz Hernández ya que el fue quien me propuso y motivo a realizar este trabajo, obsequiándome gran parte de su tiempo y paciencia. Al Dr. Eduardo Sánchez Soto por haberme asesorado, brindado su tiempo y conocimientos los cuales fueron de mucha ayuda durante el desarrollo de este trabajo. A la Dra. Beatriz Carely Luna por darme un panorama general de lo que se necesitaba para este trabajo.

A mis revisores la Dra. Beatriz Carely Luna, el Dr. Eduardo Sánchez Soto, el M.R.T. Carlos M. Copto del Puerto, el M.C. Fabián Ruiz Paz y el M.C. Ariel López por sus correcciones y críticas constructivas las cuales me sirvieron de mucho.

Para finalizar este apartado agradecemos a PROMEP por el apoyo otorgado para la realización del presente proyecto. Miles de gracias a todos y que Dios los bendiga siempre.

Resumen

En este trabajo se aborda el problema de la ingeniería inversa de redes regulatorias genéticas dentro de algún sistema biológico. En las últimas décadas han sido estudiadas las interacciones complejas entre los genes y las proteínas generando grandes cantidades de información, por lo que se requiere encontrar de forma automática un modelo gráfico que represente a la red regulatoria genética. Existen modelos utilizados para representar una red regulatoria tal como las redes booleanas o las redes bayesianas, por citar algunos. Las redes bayesianas son de dos tipos, estáticas o dinámicas, las dinámicas a comparación de las estáticas han sido muy poco estudiadas dentro del aprendizaje de los datos de expresión de los genes. Actualmente existe una función de puntuación teórica conocida como prueba de información mutua, la cual introduce el aprendizaje de una red bayesiana dinámica. Esta medida tiene la posibilidad de construir una prueba estadística de independencia basada en la distribución X^2 , la cual sirve para penalizar dentro de la red bayesiana dinámica los grados de interacción entre cada variable y sus variables padres. Por lo tanto seleccionamos la prueba de información mutua para obtener la red bayesiana dinámica mediante un algoritmo el cual aplicamos a la formación del flagelo de la bacteria *Escherichia coli* K-12 MG1655 y realizamos el análisis de la información obtenida.

Palabras clave: *redes regulatorias genéticas, aprendizaje, redes bayesianas dinámicas, prueba de información mutua.*

Abstract

This work approaches the problem of genetic regulatory networks reverse engineering inside any biological system. In the last decades have been studied the complex interactions between genes and proteins generating large amounts of information, therefore requires finding automatically a graphical model that represents a genetic regulatory network. There are models used to represent a regulatory network as a boolean or as bayesian networks, to cite a few. Bayesian networks are of two types, statics or dynamics, the dynamics compared to the statics have not been studied in depth to learn the gene expression data. At present there is a theoretical scoring function known by mutual information test, which introduces learning from a dynamic bayesian network. This measure have the possibility to build an independence statistical test based in the χ^2 distribution, which serves to penalize the interaction degrees between each variable and their parents inside a dynamic bayesian network. Therefore we select the mutual information test in order to obtain the dynamic bayesian network through an algorithm which we apply in the flagella formation of the bacterium *Escherichia coli* K-12 MG1655 and we realize the analysis of the obtained information.

Keywords: *genetic regulatory networks, learn, dynamic bayesian networks, mutual information test.*

Índice general

Agradecimientos	II
Resumen	III
Abstract	IV
Lista de figuras	VIII
Lista de tablas	X
1. Introducción	1
1.1. Planteamiento del problema	3
1.2. Objetivos	5
1.2.1. Objetivo general	5
1.2.2. Objetivos específicos	5
1.3. Hipótesis	5
1.4. Justificación	6
1.5. Alcances y limitaciones	6
2. Marco de referencia	8
2.1. Antecedentes	8
2.2. Redes regulatorias genéticas	9
2.3. Bacteria Escherichia Coli	11
2.4. Modelos gráficos	13

<i>ÍNDICE GENERAL</i>	VI
2.4.1. Redes bayesianas	15
2.4.2. Redes bayesianas dinámicas	17
2.4.3. Aprendizaje en redes bayesianas	18
2.4.3.1. Prueba de información mutua (MIT)	22
2.4.3.2. Ji cuadrada o X^2	24
2.5. Tecnologías de software utilizadas	26
3. Protocolo experimental	28
3.1. Métodos y técnicas	28
3.1.1. Estructura del modelo	28
3.2. Prueba de información mutua (MIT)	40
4. Resultados	44
4.1. Prueba de información mutua	44
4.2. Aprendizaje de la DBN	49
4.3. Graficación del flagelo de la bacteria <i>Escherichia coli</i>	51
5. Conclusiones	56
Bibliografía	57

Índice de figuras

1.1. Representación de cuatro modelos gráficos correspondientes a dos variables.	4
2.1. Regulación de la expresión de un gen en diferentes etapas de la síntesis de una proteína. Imagen tomada de [Hidde2000].	11
2.2. Escherichia coli. Imagen tomada de la foto librería de ciencia de [ECDC].	11
2.3. Ejemplo de una red bayesiana estática de una alarma y los parámetros asociados a ella. Imagen tomada de [Valenzuela2008].	16
2.4. Representa el ejemplo de la BN de una alarma pero utilizando procesos temporales. Imagen tomada de [Valenzuela2008] para el inicio de los procesos temporales.	18
2.5. Alineamiento de datos para una DBN con un arco de $X_j \rightarrow X_i$ y el número efectivo de observaciones está conformado como N-1. Imagen tomada de [Nguyen2001].	24
3.1. Red regulatoria de la formación del flagelo. Imagen tomada de [Valdez].	29
3.2. Normalización de los valores del GFP/OD. Imagen tomada de [Valdez].	29
3.3. Red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_6 depende del nodo x_{10}	31
3.4. Red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_{10} depende del nodo x_6	32
3.5. Primera propuesta de la red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_6 depende del nodo x_{10}	32
3.6. Segunda propuesta de la red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_{10} depende del nodo x_6	33

3.7. Algoritmo de la prueba de independencia mutua. Imagen tomada de [CODEMIT]	42
3.8. Algoritmo de la independencia mutua condicional. Imagen tomada de [CODEMIT]	43
4.1. Decremento de información mutua de la red regulatoria.	47
4.2. Incremento de información mutua de la red regulatoria.	47
4.3. DBN obtenida a partir del MIT.	48
4.4. Red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_6 depende del nodo x_{10}	48
4.5. Diseño del modelo gráfico del flagelo de la bacteria Escherichia coli.	52
4.6. Datos pertenecientes a los resultados de laboratorio y los resultados de [Valdez].	53
4.7. Gráfica perteneciente a la variable aleatoria seleccionada.	53
4.8. Ecuación obtenida a partir de la variable aleatoria seleccionada de acuerdo al sistema de ecuaciones de [Valdez]	54
4.9. Subgrafo obtenido a partir de la ecuación de [Valdez].	54
4.10. Grafo completo del análisis del flagelo de la bacteria Escherichia coli. Imagen tomada de [Valdez] para respectiva graficación.	55

Índice de cuadros

2.1. Escherichia coli K12. Descripción tomada de [IBIS]	13
2.2. Problemas de aprendizaje de BN. Tabla tomada del autor [Murphy1999]	20
3.1. Primera matriz de adyacencia de la red bayesiana dinámica elaborada. Tabla tomada a partir de la figura (ver Figura 3.3).	33
3.2. Segunda matriz de adyacencia de la red bayesiana dinámica elaborada. Tabla tomada a partir de la figura (ver Figura 3.4).	34
3.3. Resultados de la discretización obtenidos a partir de la figura (ver Figura 3.2)	34
3.4. Decremento del nodo x_1	37
3.5. Decremento del nodo x_2	37
3.6. Decremento del nodo x_4	37
3.7. Decremento del nodo x_6	37
3.8. Decremento del nodo x_7	38
3.9. Decremento del nodo x_9	38
3.10. Decremento del nodo x_{10}	38
3.11. Incremento de las aristas de x_1 no tiene.	38
3.12. Incremento de las aristas de x_2 no tiene.	39
3.13. Incremento de las aristas de x_4 no tiene.	39
3.14. Incremento de las aristas de x_6 dado x_2 y x_{10} . Incremento de la arista de x_6 dado x_2	39
3.15. Incremento de las aristas de x_7 dado x_2 y x_6 . Incremento de las aristas de x_7 dado x_2 y x_6	39
3.16. Incremento de las aristas de x_9 dado $x_1, x_2, x_3, x_4, x_5, x_7$ y x_8 . Incremento de las aristas de x_9 dado $x_1, x_2, x_3, x_4, x_5, x_7$ y x_8	40

3.17. Incremento de las aristas de x_{10} no tiene. Incremento de las aristas de x_{10} dado x_2 y x_6	40
4.1. Resultados del decremento de cada nodo	46
4.2. Resultados del decremento de cada nodo	46
4.3. Resultados del incremento de cada nodo	46
4.4. Probabilidad del nodo A.	49
4.5. Probabilidad del nodo D.	49
4.6. Probabilidad del nodo G dado el nodo F.	49
4.7. Probabilidad del nodo H dado el nodo G.	49
4.8. Probabilidad del nodo G dado los nodos C y nodo G.	50
4.9. Probabilidad del nodo C dado el nodo B y el nodo J.	50
4.10. Probabilidad del nodo C dado los nodos B y nodo J.	50
4.11. Probabilidad del nodo F dado el nodo A, el nodo B, el nodo D, el nodo E, el nodo G y el nodo H.	50

Capítulo 1

Introducción

La biología de sistemas, también conocida como biosistemas o biología integral, estudia y trata de entender las funciones moleculares de un sistema biológico mediante el aprendizaje dinámico de su estructura y las interacciones complejas entre genes y proteínas. Para obtener información de un sistema biológico, en ésta área se analizan conjuntos de datos complejos, estos provienen de la genómica, transcriptómica, proteómica y metabolómica [Tiessen2012].

La biología de sistemas utiliza las áreas de biología clásica, matemáticas y computación para crear y entender el modelo de un sistema biológico mediante las técnicas y conceptos de esas áreas. La complejidad de los procesos celulares y la gran cantidad de datos de un sistema biológico obtenidos de los experimentos realizados en el laboratorio, hacen difíciles las tareas de modelar, simular y analizar matemáticamente o computacionalmente el sistema [Ideker2001]. El desarrollo de un modelo para un sistema biológico sirve para facilitar el estudio de la dinámica y las propiedades de procesos celulares que posee, mediante algún método o función de aprendizaje. Por tal motivo el desarrollo de este trabajo forma parte de la biología de sistemas.

La ingeniería inversa de redes regulatorias genéticas (en inglés conocidas como GRNs, Genetic Regulatory Networks) estudia las interacciones complejas entre los genes y proteínas de un sistema biológico. Por medio de ésta se puede obtener de forma automática un modelo gráfico a partir de los datos de ese sistema.

Una GRN está compuesta de genes que interactúan entre sí. Cada gen se transcribe y traslada en una proteína el factor de transcripción (en inglés conocido como TF, Transcription Factor) y a la vez regula los niveles de expresión de otros genes de la red por medio de la unión de regiones promotoras que cada gen posee [Rau2011]. La proteína que utiliza cada gen activa o inhibe la expresión de otro gen, obteniendo con ello interacciones complejas. Las interacciones pueden ser representadas en una GRN y modelarse por medio de un sistema dinámico. Existen diversos modelos utilizados para comprender, predecir y representar el comportamiento de fenómenos dinámicos en las GRNs, algunos de ellos son ecuaciones diferenciales, redes booleanas o redes bayesianas.

Las redes bayesianas (en inglés conocidas como BNs, Bayesian Networks) modelan estructuras estáticas y dinámicas, las estáticas sirven para modelar la causa-efecto de redes biomoleculares simples, las dinámicas son llamadas redes bayesianas dinámicas (en inglés conocidas como DBNs, Dynamic Bayesian Networks), a través de ellas también se obtiene la estructura y el comportamiento de las interacciones dinámicas entre los genes de una GRN.

El modelo de un sistema biológico por medio de BNs emplea un análisis causal de los niveles de expresión de los genes que influyen sobre él. Para poder analizar e interpretar los datos obtenidos es posible utilizar inferencia bayesiana. La inferencia bayesiana sirve para tener una idea de cómo se está comportando el sistema biológico de acuerdo a la modificación de los datos conocidos de sus genes, utilizando parte de la dinámica del sistema y generando con ello el aprendizaje del modelo. Para el aprendizaje dinámico de una BN, es necesario verificar la dependencia o independencia de los datos obtenidos en el análisis, respecto al conjunto de interacciones obtenidas por medio del comportamiento dinámico.

Hoy en día para algunos investigadores ya es posible medir de forma experimental los niveles de expresión de un gran número de genes en forma simultánea para obtener la estructura y comportamiento de una GRN [Murphy1999]. Otras personas también se han dedicado a inferir y verificar la calidad de una GRN de forma computacional, creando o usando herramientas que le permitieran realizar estas actividades. Una forma alternativa de medir su calidad es por medio de una función de

puntuación (bayesiana o de información teórica) [Carvalho2009], la cual es importante para esta tesis.

Estudios recientes han encontrado una función de puntuación de información teórica conocida como la Prueba de Información Mutua (en inglés conocida como MIT, Mutual Information Test) la cual es utilizada para encontrar la dinámica de una BN por medio de la información mutua entre los datos que la conforman. La parte esencial del MIT es comprobar que la dinámica de la BN sea correcta por medio de la prueba de independencia de X^2 entre los datos [Campos2006].

En este trabajo se pretende inferir sobre el modelo una GRN, utilizando métodos que permitan describir su funcionamiento de forma computacional, a través de la obtención de la estructura, inferencia y verificación de la calidad, tomando para este propósito la medida de la prueba de información mutua.

Para ello se hará un modelo que represente la estructura del flagelo de la bacteria *Escherichia coli* (también conocida como *E.coli*), debido a que es la bacteria más estudiada y citada en diversos trabajos de investigación.

1.1. Planteamiento del problema

El proceso para crear la estructura del modelo de una GRN, depende de la cantidad de información que se tenga disponible, la forma en que se acomode la información dentro del modelo y sobre todo si este cuenta con fenómenos dinámicos.

Los datos que se obtienen del sistema biológico de una GRN pueden tener concentradas muchas variables, entre mayor sea el número de variables, se vuelve muy complejo proponer un modelo gráfico para esa GRN. Por ejemplo, si se tienen dos variables de una GRN x_1 y x_2 , existen cuatro modelos que se pueden proponer para su aprendizaje (ver Figura 1.1). Sin embargo el hecho no es sólo proponer cualquier modelo para el aprendizaje de la red regulatoria (ya sea por medio de una BN para el caso de fenómenos estáticos o por medio de una DBN para el caso de fenómenos dinámicos) y asimilar que ese es el más adecuado para resolver el problema sino, se trata de comprobar la calidad de esa red respecto a una medida de calidad.

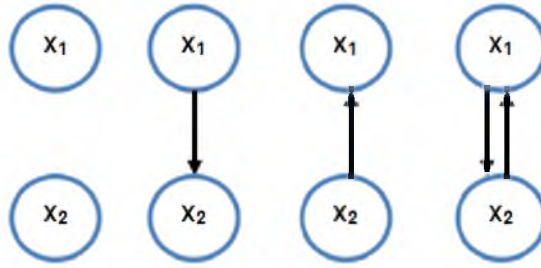


Figura 1.1: Representación de cuatro modelos gráficos correspondientes a dos variables.

Algunos autores mencionan que para encontrar un modelo que describa la estructura de un sistema biológico, se puede realizar una búsqueda exhaustiva, es decir este proceso radica en evaluar cada posible estructura que se proponga y escoger la mejor, de acuerdo a un criterio de evaluación. Cuando el número de variables es pequeño entonces si es posible aplicar la búsqueda, sin embargo, cuando no lo es entonces la búsqueda se convierte en intratable (desde el punto computacional). [Robinson1977] investigó y demostró cuántos grafos dirigidos acíclicos se pueden obtener para n variables mediante una función recursiva, como se muestra a continuación:

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i) \quad n > 2, \quad (1.1)$$

Por ejemplo, si tenemos dos variables sería $f(2)=3$, con tres variables $f(3)=25$, con cinco variables $f(5)=29000$, por lo tanto es muy claro que el número de grafos dirigidos acíclicos es elevado. Entonces para evitar toda la búsqueda exhaustiva, otros autores han propuesto criterios de evaluación basados en funciones de puntuación de información teórica tal como la prueba de información mutua.

Debido a las características de los datos de expresión de los genes involucrados en la formación del flagelo de la bacteria E.coli, su estructura puede ser modelada dinámicamente. En este caso serán utilizadas las DBNs para modelar la dinámica de este sistema biológico, se propondrá una red dinámica para los datos de expresión de los genes del flagelo de la bacteria E.coli mediante la prueba de información mutua, usando para ello un algoritmo que sirva para comprobar su calidad respecto a los datos obtenidos del laboratorio.

1.2. Objetivos

1.2.1. Objetivo general

El objetivo general de este trabajo es construir un modelo de una GRN del flagelo de la bacteria *E.coli* utilizando DBNs y una medida de calidad.

1.2.2. Objetivos específicos

Los objetivos específicos que nos permitirán lograr el objetivo general se describen a continuación:

- Estudiar las medidas de calidad contenidas en las funciones de puntuación de información teórica.
- Plantear una estructura con redes bayesianas dinámicas de la GRN del flagelo, a partir de la red propuesta en [Valdez] y [Martinez2012].
- Si hay más de dos redes aprendidas de la red bayesiana dinámica, entonces se debe de comprobar la calidad de las redes propuestas.
- Programar un algoritmo para verificar la medida de calidad de una DBN.
- Realizar un programa para visualizar el comportamiento de cada factor de transcripción del flagelo.

1.3. Hipótesis

Las DBNs pueden modelar la estructura de la GRN de algún organismo, gracias a su capacidad para representar relaciones causa-efecto a través de las independencias o dependencias condicionales obtenidas de los elementos del modelo. La prueba de información mutua es usada para verificar la independencia condicional entre sus elementos mediante la prueba estadística de X^2 .

1.4. Justificación

La tarea de modelar, inferir, simular el comportamiento de la estructura de un sistema biológico y utilizar alguna función de puntuación para comprobar la complejidad de esa estructura actualmente tiene un costo computacional alto. Representar la estructura de un sistema biológico es una tarea difícil, debido a que no es factible explorar todas las posibles estructuras para su representación, por lo que se debe recurrir a métodos que reduzcan o penalizen algunas propiedades en la estructura (tal como interacciones entre los nodos). Al igual, la falta de herramientas matemáticas y computacionales son necesarias para tratar este problema de interés biológico.

Las DBNs en conjunto con alguna función de puntuación, permiten simular el comportamiento del sistema dinámico de una GRN mediante la prueba de información mutua entre las variables que intervienen en el modelo. La bacteria E.coli ha sido una de las bacterias más estudiadas en los últimos años dentro de las investigaciones biológicas debido a la estructura que posee y sus componentes, siendo tratada desde diversas perspectivas (por ejemplo los TFs). Las herramientas para evaluar la calidad de redes dinámicas y funcionales del flagelo de la bacteria son escasas. El flagelo de la bacteria E.coli es sólo un pequeño ejemplo de una GRN que será usado en este trabajo para estudiar el aprendizaje de los modelos gráficos en GRNs.

1.5. Alcances y limitaciones

Los alcances de este trabajo son:

- Usar una medida de calidad (MIT) para la red bayesiana dinámica, basada en la estructuras propuestas.
- Obtener un producto de software (script en Matlab) del algoritmo que permita la aplicación de la técnica de medida de calidad de red bayesiana dinámica.
- Verificar en cuál de todas las estructuras propuestas se obtuvo el puntaje más alto, debido a que esta nos va a indicar cual fue el aprendizaje de la estructura propuesta.
- Obtener un producto de software (programa en HTML5) que permita observar las interacciones

entre cada gen que posee el flagelo.

Las limitaciones de este trabajo son:

- Dentro de los datos experimentales probados existen sólo diez factores de transcripción (TFs), para realizar las pruebas pertinentes.
- Los métodos de aprendizaje se basan en técnicas estadísticas, como nosotros estamos utilizando la prueba estadística de X^2 , entonces lo mejor sería que tuvieramos más de 30 observaciones de los datos experimentales. Sin embargo la cantidad de datos disponibles para cada factor de transcripción son pocos ya que estos son los únicos que se obtuvieron del laboratorio y los cuales arrojaron que para cada TF sólo existe 9 observaciones. Por lo tanto nuestra aproximación del aprendizaje de la DBN sólo será para darnos una idea de cuales variables son independientes y cuales variables no son independientes.

Capítulo 2

Marco de referencia

2.1. Antecedentes

Cada organismo posee miles de decenas de genes, la unidad básica que representa a los organismos es la célula, la cual puede ser de dos tipos: eucarionte y procarionte. Desde el año de 1886 se descubrió la bacteria E.coli y años más tarde se empezó a observar la secuencia de su genoma. Actualmente es estudiada debido a la resistencia de ciertas cepas a antibióticos [SMB]. La bacteria E.coli pertenece a los organismos procariontes unicelulares que son capaces de moverse por medio del flagelo, con deslizamientos sobre superficies sólidas, por simples sacudidas o contracciones. El caso de estudio de este trabajo, es el flagelo de la bacteria mencionada representado por una GRN [Neapolitan2009] [MICRO].

La ingeniería inversa de las GRNs ha sido empleada en biología durante varias décadas y se utiliza para descubrir automáticamente la estructura de algún sistema biológico dinámico mediante un modelo (como ecuaciones diferenciales, redes booleanas, redes neuronales o BN por citar algunos). El representar una GRN a partir de la cantidad de datos experimentales obtenidos de la medición de los niveles de expresión de un gran número de genes, genera la necesidad de descubrir patrones en tales datos. Los patrones sirven para conocer la dependencia o independencia condicional entre los datos.

Un sistema biológico se puede representar con el modelo de una BN, ésta también es llamada red causal. En ella se analizan los patrones de la expresión de genes a partir de los datos observados en el sistema, ya que la estructura de la BN es conocida. Sin embargo, cuando sólo se cuentan con pocos datos del sistema biológico entonces es más difícil encontrar su estructura, representarla en un modelo y analizar los patrones de la expresión de sus genes. Para encontrar patrones desconocidos entre los datos, se realiza la inferencia dentro del modelo con el propósito de obtener nuevos datos. A este proceso se le conoce como aprendizaje del modelo.

Una BN representa la estructura de una GRN, modelada mediante un grafo dirigido acíclico, refleja las relaciones de dependencia o independencia contenidas entre los genes de la red con una distribución de probabilidad. Una BN estática tiene el problema de que no posee algún mecanismo para modelar datos en series de tiempo. Ciertas GRNs poseen ciclos. En los experimentos biológicos la información de la expresión de los genes es obtenida en series de tiempo que se pueden discretizar. El modelo que se puede utilizar para resolver ese problema es una DBN [Gardner2005], [Murphy1999], [BIO131].

Es preciso mencionar que las BNs al igual que las DBNs, son capaces de modelar las relaciones de causalidad entre cada elemento presente en un sistema biológico y utilizan los mismos métodos para el aprendizaje del sistema.

2.2. Redes regulatorias genéticas

Los organismos forman parte del medio ambiente, por ejemplo una planta, animal, bacteria, protista o fungusida e incluso el ser humano. Estos habitan en lugares distintos pero en ocasiones comparten alguna característica en especial [Neapolitan2009]. Los organismos cuyas células no contienen un núcleo son llamados procariontes. Las bacterias pueden estar clasificadas dentro de los organismos procariontes unicelulares, por ejemplo la bacteria *Escherichia coli*.

El cromosoma de un organismo sirve para transportar biológicamente características hereditarias. Es una macromolécula consistente de cadenas de ácido desoxirribonucleico (ADN). Las cadenas de ADN contienen información codificada para producir cada proteína o molécula de ácido ribonucleico

(ARN) presente en un organismo. Cuando una cadena está dividida en pequeños fragmentos, estos son conocidos como genes, su función dentro de un organismo consiste en obtener la información necesaria para construir una molécula de ARN o una proteína. Una proteína contiene una región codificadora de ADN, la cual puede funcionar como un factor de transcripción para algunos sitios de ADN.

Cuando la célula de un organismo reconoce el inicio de un gen o varios genes para construir una molécula de ARN o proteína, éste proceso se realiza mediante la sección de ADN llamado *promotor*. El promotor de un gen es el que indica que se debe iniciar una transcripción de ADN a ARN, éste proceso se conoce como *regulación de transcripción de ADN*. Posteriormente cuando se reconoce el inicio de un gen éste se copia en una molécula de ARN, lo que resulta de este proceso de regulación del ARN y transporte se conoce como *ARNm*, el ARNm también es conocido como ARN mensajero. El proceso de regulación de translación de ARN convierte el código de ARN en una proteína. Para que se pueda tener una proteína modificada es necesario pasar por el proceso de regulación de modificación de proteína. Todo este conjunto de procesos es importante dentro de la biología molecular [Hidde2000] (ver Figura 2.1).

Un sistema regulatorio genético está estructurado por interacciones regulatorias entre ADN, ARN, proteínas y otras moléculas [Hidde2000]. Una GRN puede ser vista como un modelo abstracto de las interacciones que tiene el genoma humano. En computación, la información del modelo de un sistema biológico es útil para formar una GRN.

La GRN sirve para representar las interacciones de los genes en una estructura, la puede modelar un grafo dirigido compuesto por nodos (genes) y arcos (interacciones). Cada interacción entre los genes se comporta como una activación (induce la transcripción con otros genes) o inhibición (reprime la actividad de transcripción) [Cho2007].

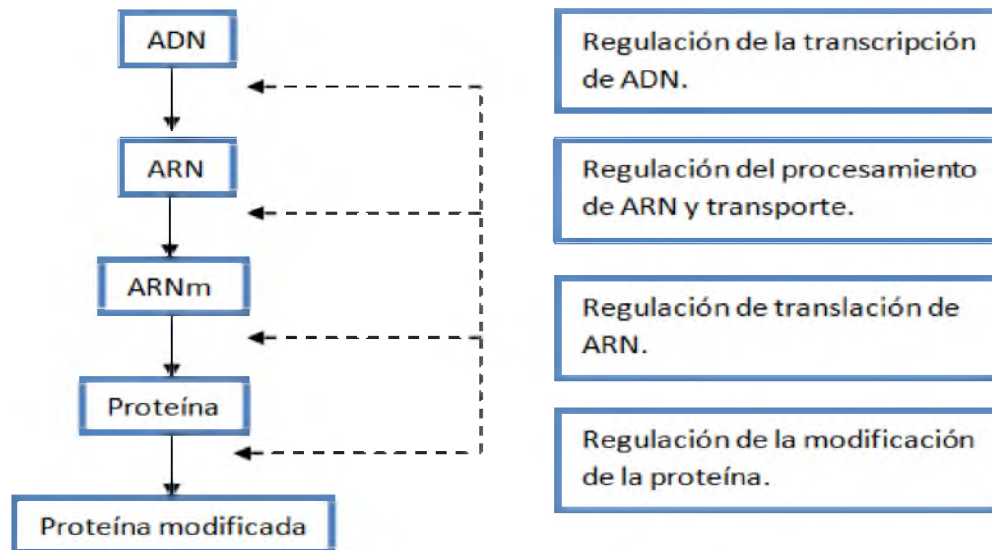


Figura 2.1: Regulación de la expresión de un gen en diferentes etapas de la síntesis de una proteína. Imagen tomada de [Hidde2000].

2.3. Bacteria Escherichia Coli

Las bacterias son capaces de producir proteínas para el beneficio de algún organismo, por ejemplo el ser humano. *Escherichia coli* es una bacteria que vive dentro del intestino de animales y humanos. Está conformada de cepas, las cuales en su mayoría son inofensivas, sin embargo las cepas que son dañinas como es el caso de O157:H7, pueden producir una seria intoxicación alimentaria en los seres humanos [Martinez2011] (ver Figura 2.2).



Figura 2.2: Escherichia coli. Imagen tomada de la foto librería de ciencia de [ECDC].

Estudios han demostrado que la cepa K-12 MG1655 es inofensiva en el intestino del ser humano debido a que hace más de 100 años fue aislada de éste para ser utilizada en investigaciones, con la fi-

nalidad de obtener conocimiento sobre sus componentes moleculares y observar su comportamiento. Se usa como modelo de investigación de la genética y fisiología, e incluso en aplicaciones industriales [EPA]. Para nuestro propósito, realizaremos el análisis de la estructura de una GRN que representa las interacciones de los factores de transcripción que producen proteínas relacionadas con la formación del flagelo de la bacteria E.coli K-12 MG1655. Del número total de genes de E.coli, aproximadamente 300 han sido predecidos para codificar los factores de transcripción (TFs) y 7 para los factores σ . Esta información está contenida en algunas bases de datos, por ejemplo RegulonDB [Martinez2012].

La bacteria E.coli recibe el nombre de peritrica porque cuenta con flagelos repartidos por toda la superficie, existen más bacterias pertenecientes a la misma clasificación. *Los **flagelos** permiten que las bacterias se desplacen en medios líquidos. La longitud del flagelo en la bacteria E.coli es de 20 nm, su diámetro es constante y uniforme para cada especie de bacterias.* El movimiento flagelar de la bacteria tiene un sistema dinámico el cual le permite modificar su mecanismo básico. En la bacteria E.coli, la síntesis del flagelo es el producto de la actividad de al menos 50 genes manipulados por precisión respecto al movimiento o interacción que ejercen y como resultado se obtiene un proceso rápido de regulación [MICRO] [Chilcott2000]. La actividad de los genes requiere de uno de los 7 factores σ en E.coli (FliA) y un regulador maestro (FlhDC). Este proceso es altamente dependiente del resto de la actividad de los genes de la bacteria, la evidencia que existe es que el regulador maestro (FlhDC) es uno de los genes más regulados en la bacteria E.coli [Valdez] [Martinez2012].

Los aspectos dinámicos y funcionales del flagelo se basan en el movimiento que realizan sobre algún tipo de célula bacteriana. Los flagelos están compuestos aproximadamente por 20 proteínas base, pero para su regulación y coordinación requieren de más proteínas [MICRO].

Las actividades promotoras de los TFs para la formación del flagelo de la bacteria Escherichia coli K-12 MG1655 (ver Cuadro 2.1) han sido reportadas por algunos investigadores como [Martinez2012].

Los experimentos relacionados con la formación del flagelo de la bacteria E.coli K-12 MG1655 miden los niveles de expresión de los factores de transcripción implicados en este proceso, estos niveles de expresión también pueden interpretarse como la cantidad de proteína sintetizada.

TFs	Descripción
lrhA	DNA-binding transcriptional repressor of flagellar, motility and chemotaxis genes
ihfA	integration host factor (IHF), DNA-binding protein, alpha subunit
flhDC	DNA-binding transcriptional dual regulator with FlhD y FlhC
qseB	quorum sensing DNA-binding response regulator in two-component regulatory system with QseC
ompR	DNA-binding response regulator in two-component regulatory system with EnvZ
fis	global DNA-binding transcriptional dual regulator
hns	global DNA-binding transcriptional dual regulator H-NS
hdfR	DNA-binding transcriptional regulator
fliA	RNA polymerase, sigma 28 (sigma F) factor
CRP	DNA-binding transcriptional dual regulator

Cuadro 2.1: Escherichia coli K12. Descripción tomada de [IBIS]

Como la información del movimiento flagelar de la bacteria E.coli esta compuesto de un sistema dinámico, entonces está puede ser modelada por medio de una red bayesiana dinámica. El objetivo de realizar este modelo, es para observar algunos instantes de tiempo en los cuales se obtienen patrones de la expresión de los genes a partir de los datos observados en el sistema y con ello, poder dar una explicación acerca del comportamiento del movimiento.

En nuestro trabajo, presentamos un modelo dinámico dado por medio de una DBN, el cual representa las actividades promotoras generadas a partir de los factores de transcripción, además incluimos la inferencia de los parámetros encontrados en cada uno de estos [Martinez2012].

2.4. Modelos gráficos

Los modelos gráficos son utilizados en diversas áreas, por ejemplo biología, matemáticas y computación. Para algunos autores es necesario relacionar los modelos gráficos con probabilidad, independientemente del área en la que se esté trabajando.

*Un **modelo gráfico** está ligado a la teoría de probabilidad y la teoría de grafos. Proporciona una herramienta para resolver problemas en matemáticas aplicadas e ingeniería, haciendo uso de la incertidumbre, complejidad para el diseño y análisis del aprendizaje de algoritmos [CSBN]. La probabilidad es utilizada en algunas situaciones donde se conocen los posibles resultados pero no se sabe cual de ellos ocurrirá o en que orden ocurrirá.*

La representación de los modelos gráficos probabilísticos se realiza por medio de grafos los cuales contienen nodos (representan las variables aleatorias del modelo) y los arcos (la independencia condicional entre los nodos). Los modelos gráficos dirigidos a comparación de los no dirigidos, cuentan con la capacidad de interpretación causal entre arcos y nodos. Por ejemplo la interacción de $A \rightarrow B$, significa que existe un mecanismo causal que hace que la presencia del nodo A produzca al nodo B.

Un modelo gráfico dirigido está conformado de los siguientes aspectos según [Neapolitan2009] y [Diez2012]:

- **Arco:** Se refiere a la representación gráfica que viene dada por una flecha, por ejemplo un arco (X,Y) indica $X \rightarrow Y$.
- **Grafo dirigido:** Se refiere a $G=(N,A)$, donde N es el conjunto de nodos y A un conjunto de arcos definidos sobre los nodos.
- **Padre:** Por ejemplo, si X es el nodo padre del nodo Y entonces debe de existir un arco de $X \rightarrow Y$ y por lo tanto se puede decir que Y es hijo X. Al conjunto de los padres de X se representa como $pa(X)$ y al conjunto de hijos de X se puede representar por $S(X)$.
- Dos variables X e Y son **independientes** si la probabilidad de $P(X | Y)=P(X)$.
- Dos variables X e Y son **independientes** dado una tercer variable Z si la probabilidad que se tiene es $P(X | Y,Z)=P(X | Y)$.
- **Hallazgo:** Se define como la determinación del valor de una variable, $H=h$, a partir de un dato (una observación, una medida, etc.) [Diez2012].
- **Evidencia:** Se define como el conjunto de todos los "hallazgos" disponibles en un determinado momento o situación. Por ejemplo $e=H_1=h_1, \dots, H_r=h_r$.
- **Probabilidad a priori:** Es la probabilidad de una variable o un conjunto de variables que se calcula cuando no hay ningún hallazgo.
- **Probabilidad a posteriori:** Es la probabilidad de una variable o conjunto de variables dada una evidencia $e:P(x | e)$.

Para un modelo dirigido, es preciso especificar la distribución de probabilidad condicional (en inglés conocido como CPD, Conditional Probability Distribution) en cada variable continua. Sin embargo si las variables son discretas, se debe utilizar una tabla de probabilidad condicional (en inglés conocida como CPT, Conditional Probability Table) la cual sirve para enumerar la probabilidad de que el nodo hijo tome en cada uno de sus diferentes valores, el valor de su padre [Murphy1999]. Ejemplo de algún modelo gráfico dirigido es una red bayesiana.

2.4.1. Redes bayesianas

Una **red bayesiana** también conocida como red de creencia consta de tres elementos: un conjunto de *variables aleatorias* X (por ejemplo , x_1, \dots, x_n), un *grafo dirigido acíclico* (en inglés conocido como DAG, Directed Acyclic Graph) $G=(X,A)$ (donde G representa el DAG, X son las variables aleatorias y A son los arcos (o aristas) entre cada variable X_i) y una *distribución de probabilidad* sobre X , representada en la red bayesiana como $P(X)$ [Murphy1999], la cual puede ser factorizada de la siguiente forma:

$$P(x) = \prod_i P(x_i | pa(X_i)),$$

donde las $P(x_i | pa(X_i))$ son las probabilidades condicionales de $P(X)$ y $pa(X_i)$ representa todos los padres que puede tener la variable x_i .

A continuación se describe un ejemplo ilustrativo para construir una red bayesiana estática: Supongamos que estoy en el trabajo, mi vecino Juan llama para decirme que mi alarma está sonando, pero la vecina Mary aún no me llama. La alarma de mi casa se puede encender por medio de dos causas, una por temblores ligeros y la otra porque es posible que haya un ladrón. Por consiguiente los dos aspectos importantes para el razonamiento probabilístico de este problema son:

1. **Las variables aleatorias** para representar la red:

x_1 =ladrón de casas, x_2 =temblor, x_3 =alarma, x_4 =llamada de Juan, x_5 =llamada de Mary.

2. **La topología de la red** (ver Figura 2.3) que refleja el conocimiento "causal":

- Un ladrón puede encender la alarma.
- Un temblor puede encender la alarma.

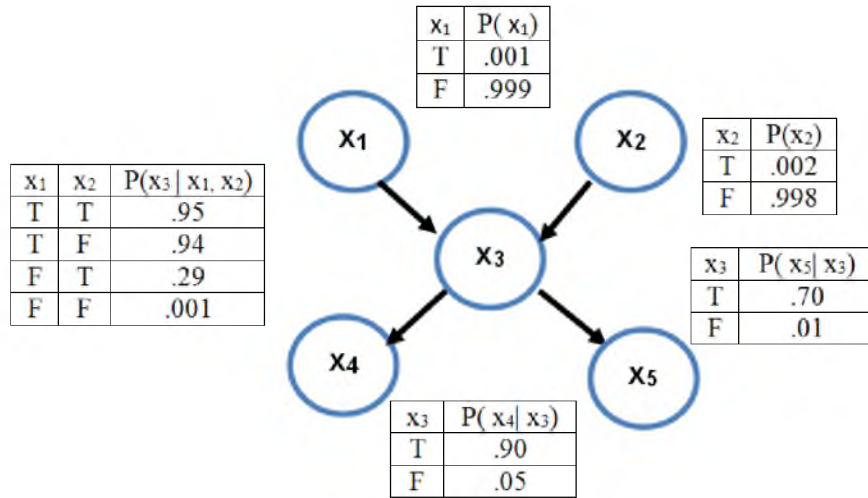


Figura 2.3: Ejemplo de una red bayesiana estática de una alarma y los parámetros asociados a ella. Imagen tomada de [Valenzuela2008].

- La alarma puede causar que Mary llame.
- La alarma puede causar que Juan llame.

Las CPTs de las variables aleatorias $X=(x_1, x_2, x_3, x_4, x_5)$ de la figura (ver Figura 2.3) están dadas por las siguientes probabilidades:

1. Probabilidad de x_1 es $P(x_1)$
2. Probabilidad de x_2 es $P(x_2)$
3. Probabilidad de x_3 dado x_1 y x_2 es $P(x_3 | x_1, x_2)$
4. Probabilidad de x_4 dado x_3 es $P(x_4 | x_3)$
5. Probabilidad de x_5 dado x_3 es $P(x_5 | x_3)$

En ocasiones el resultado de la probabilidad de alguna (as) variable (s) no puede (n) ser tan satisfactorio (s) ya que el valor que se obtiene es muy pequeño. Por ejemplo si la x_4 y x_5 se realizó dado que la x_3 se activó y x_1 o x_2 no tuvieron nada que ver con que la alarma se haya activado entonces el resultado de la probabilidad de 0.001 es muy pequeño. Esto se ve reflejado en la siguiente distribución de probabilidad:

$$\begin{aligned}
& P(x_4 \wedge x_5 \wedge x_3 \wedge \neg x_1 \wedge \neg x_2) \\
& = P(x_4 | x_3)P(x_5 | x_3)P(x_3 | \neg x_1, \neg x_2) P(\neg x_1)P(\neg x_2) \\
& = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\
& = 0.00063
\end{aligned}$$

También puede existir el caso en que los datos u observaciones de una BN provienen de una CPT en la cual existe uno o varios parámetros desconocidos, los cuales se pueden estimar por medio de un número pequeño diferente de cero (por ejemplo 0.001), como se muestra en la CPT cuando los valores x_1, x_2 son falsos dada la probabilidad de $P(x_3 | x_1, x_2)$, entonces su valor es 0.001.

Las BNs cuentan con una parte cualitativa y otra cuantitativa, la parte cualitativa es un DAG, representa causalidad, relevancia o relaciones (de dependencia o independencia) entre variables aleatorias y la parte cuantitativa permite introducir incertidumbre en el DAG mediante las tablas de probabilidad las cuales cuantifican estas relaciones [Campos2006].

Las BNs poseen algunas ventajas: como las relaciones de dependencia y causalidad entre cada variable aleatoria que manejen, también pueden combinar el conocimiento de un experto humano con los datos que se proponen en la red. Sin embargo también poseen dos desventajas, la primera esta ligada a sustentar aspectos temporales de datos en series de tiempo y la segunda es que no son capaces de modelar fenómenos cíclicos, a comparación de las DBNs.

2.4.2. Redes bayesianas dinámicas

En el ejemplo de la figura (ver Figura 2.3) de la BN, existen algunos sucesos que no han sido detectados en cierto tiempo. Hoy en día hay diversos métodos que han sido desarrollados por medio de estadística para el razonamiento sobre relaciones temporales de las diferentes entidades que conforman un modelo, generalmente ese razonamiento es conocido como análisis de series de tiempo.

Una red bayesiana (o de creencia) dinámica es un modelo gráfico debido a que describe un sistema que cambia dinámicamente o evoluciona a través del tiempo, prediciendo el comportamiento futuro del sistema [Mihajlovic]. Para algunos autores la naturaleza del modelo de una BN estática puede ser

adaptado a un modelo dinámico como lo es la DBN (ver Figura 2.4).

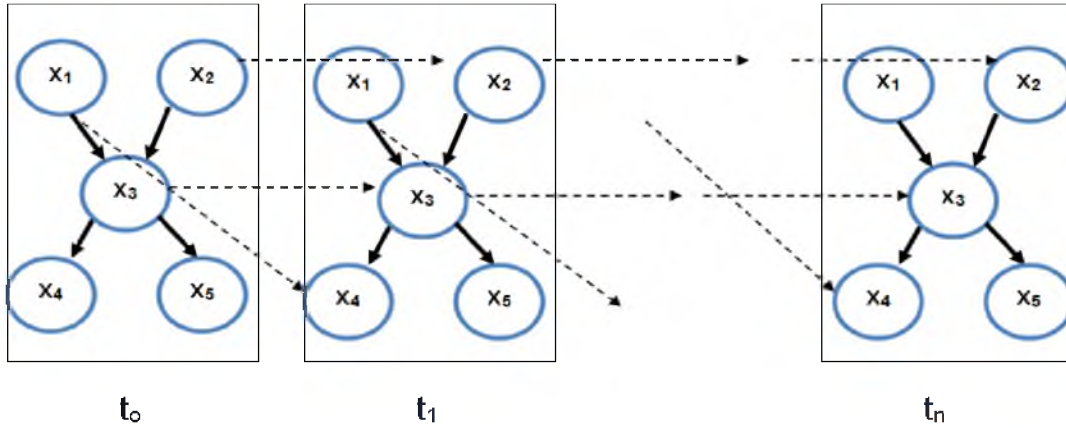


Figura 2.4: Representa el ejemplo de la BN de una alarma pero utilizando procesos temporales. Imagen tomada de [Valenzuela2008] para el inicio de los procesos temporales.

Los espacios de tiempo de la figura (ver Figura 2.4) son interconectados con relaciones temporales representados por arcos punteados entre variables aleatorias particulares en los cuales se observa la dinámica de la red. Las variables del modelo son denotadas como los estados de tiempo de la DBN, los estados del tiempo t dependen tanto de su estado anterior conocido como el tiempo t_1 , como de los tiempos anteriores t_2, t_3, \dots, t_n . Cabe mencionar que los nodos, arcos y la probabilidad de la interpretación estática de la BN también se ven reflejados en la DBN. Como el modelo de la BN es "expandido" mediante una DBN, nosotros sólo podemos conocer algunos valores de probabilidad condicional de las dependencias e independencias entre las variables para el primer tiempo siempre y cuando tengamos algunas observaciones del modelo estático, pero para encontrar las relaciones de independencia entre las demás variables se deben medir o calcular los estados anteriores. Las relaciones temporales son incorporadas en tablas de probabilidad condicional (CPTs) entre cada variable para cada espacio de tiempo.

2.4.3. Aprendizaje en redes bayesianas

El aprendizaje es utilizado en diversos campos de investigación para la adquisición de conocimientos de alguna tarea difícil e incluso imposible. Los sistemas de aprendizaje poseen la capacidad de generar conocimiento nuevo de alguna tarea y ajustar el conocimiento al existente.

En la actualidad se siguen construyendo modelos para representar fenómenos biológicos realizados a mano con la ayuda de expertos humanos, los cuales comparten entre ellos lo aprendido respecto a la experiencia que han tenido con la finalidad de crear una base de conocimientos. En los últimos años distintas personas también se han encargado de investigar y desarrollar técnicas para aprender a partir de los datos, tanto de la estructura del fenómeno físico como de los parámetros asociados a un modelo gráfico. Aunque se han creado diversas técnicas para el aprendizaje, también se puede realizar la mezcla de la base de conocimientos para inferir sobre los datos y que se modifique el modelo; es decir, para que aprenda el modelo sobre los nuevos datos.

Para poder saber cuales son los genes que interactúan en la representación del modelo de una GRN, estos dependen de los datos que se tienen. Naturalmente el aprendizaje en BNs puede ser una alternativa para el modelado de la GRN. El proceso de aprendizaje para obtener una red bayesiana a partir de los datos, conlleva elegir uno o ambos de los siguientes aspectos:

- **Aprendizaje estructural:** Para llegar a este tipo de aprendizaje se debe obtener la estructura o topología de la red.
- **Aprendizaje paramétrico:** El tipo de aprendizaje que se maneja es que, dada una estructura, se deben obtener las probabilidades asociadas.

Si lo que requerimos es utilizar el aprendizaje paramétrico en BN, entonces se deben especificar los parámetros del modelo. Los parámetros consisten en las probabilidades a priori de los nodos raíz y las probabilidades condicionales de las otras variables, dado sus padres. Para obtener los nodos raíz, se debe de estimar la probabilidad marginal, por ejemplo:

$P(A_i) \sim NA_i/N$, donde NA_i es el número de ocurrencias del valor i de la variable A y N el número total de casos o registros donde se busca el valor i [Sucar2006].

Para describir una BN es necesario conocer su topología (estructura) y sus parámetros. Para que una BN aprenda, puede hacerlo mediante esas dos formas. Sin embargo el aprendizaje de la estructura es mucho más difícil a comparación del aprendizaje de parámetros. Esto se complica más cuando algunos nodos están ocultos, o hacen falta datos debido a que no se realizaron las observaciones en

un experimento de forma adecuada en comparación a cuando si se realizaron las observaciones del experimento.

Existen cuatro problemas que se presentan cuando se maneja el aprendizaje en BN descritos a continuación (ver Cuadro 2.2):

Estructura/Observabilidad	Método
Conocida, Total	Muestra estadística
Conocida, Parcial	EM (Expectativa de maximización) o gradiente de ascenso
Desconocida, Total	Búsqueda en el espacio del modelo
Desconocida, Parcial	Estructura EM

Cuadro 2.2: Problemas de aprendizaje de BN. Tabla tomada del autor [Murphy1999]

1. El primer caso es más sencillo para el aprendizaje, ya que si se realiza un experimento y a partir de este se conoce su estructura y todas sus observaciones, entonces se pueden construir las tablas de probabilidad condicional ya que el aprendizaje sólo equivale al conteo de las observaciones. Sin embargo al haber muy pocas muestras u observaciones de la estructura, el aprendizaje no es tan confiable y por lo tanto se obtiene una aproximación del aprendizaje de la estructura.
2. El segundo caso se refiere a que si la estructura del modelo de un experimento es completamente conocida y si sólo tenemos algunas de sus observaciones, entonces los valores de ciertas variables que influyen en la estructura no son conocidos. Esta situación podría deberse a que en un principio no fueron medidas las variables. Este tipo de variables también son conocidas como variables ocultas.

La forma básica para calcular los parámetros faltantes en el modelo es utilizando el método Máxima Verosimilitud (en inglés conocido como EM, Expectation Maximization), el cual se basa principalmente en estimar la independencia entre cada variable del modelo y estimar cuáles son las posibles variables ocultas mediante la inferencia del modelo consiguiendo para ello el proceso de aprendizaje.

3. El tercer caso indica que si se desconoce cómo modelar todas las observaciones de un experimento, entonces se dice que la estructura del modelo es desconocida. Incluso al tratar de

interpretar los datos que tenemos de las observaciones no es una tarea fácil, ya que si no se tiene experiencia de lo que se está intentando realizar, se pueden llegar a suponer algunas relaciones entre los nodos que no son relevantes para el modelo, provocando algunos errores al momento de predecirlo, por ejemplo decidir el número de padres que un nodo puede tener o las relaciones que existirán entre los nodos.

Una cuestión importante que se debe tomar en cuenta es lo siguiente: dado que el modelo no existe, es complicado saber si está bien lo que se ha realizado ya que no se cuenta con qué comparar el modelo, así que lo único que queda es recurrir a algoritmos de búsqueda que no exploren todas las posibles soluciones para que sus búsquedas no sean de tipo exhaustivas, debido a que no se sabe hasta qué punto dejar de buscar. Existen algunas técnicas que evitan búsquedas extensas siempre y cuando existan observaciones evidentes entre las variables y una estimación inicial de la estructura del modelo para encontrar los valores de los nodos ocultos, con la finalidad de tener resultados favorables. Los dos enfoques más utilizados para construir este tipo de estructuras son Branch and Bound (el cual restringe un número máximo de padres para cada nodo del modelo) y el Análisis de Dependencia Extenso (en inglés conocido como EDA, Extended Dependency Analysis), este utiliza la distribución X^2 para decidir la información mutua entre cada nodo del modelo, denotando así una medida de confianza en las conexiones que se aprendan al momento de construir el modelo.

4. Finalmente llegamos al cuarto caso, este es el más difícil porque si se realizó un experimento de manera errónea, no podremos conocer la estructura del modelo que queremos representar, ya que no se tienen todas las observaciones y esto implica que nuestro modelo tenga datos faltantes y variables ocultas. Una función de puntuación bayesiana muy utilizada en este caso es el Criterio de Información Bayesiana (en inglés conocido como BIC, Bayesian Information Criterion), el cual realiza una aproximación de los nodos ocultos y los parámetros que debe contener la estructura del modelo.

En este caso si podemos decir que hace uso de búsquedas exhaustivas y el proceso de aprendizaje es más extenso, ya que no siempre al proponer nodos ocultos quiere decir que el modelo a obtener sea más compacto.

A grandes rasgos existen dos razones por las cuales nuestro trabajo está situado dentro de los primeros dos casos del aprendizaje de una DBN. La primera se debe a que planteamos una DBN para flagelo de la bacteria E.coli y en la segunda obtenemos las probabilidades asociadas entre los datos del modelo, a través de su conteo de observaciones.

2.4.3.1. Prueba de información mutua (MIT)

La prueba de información mutua $MI_D(X_i, Pa_G(X_i))$ sirve para medir el grado de interacción entre cada variable aleatoria X_i y sus padres $Pa_G(X_i)$ si y sólo si ellos son independientes, pero penalizando su valor por medio de un término relativo a la distribución X^2 .

El proceso del MIT se basa en el tratamiento de la información mutua proporcionada por cada nodo de la red, con la finalidad de indagar que información comparte un nodo con otro. Verifica si la información extraída de la red es relevante para el problema que se este modelando [Campos2006], [BIO131]. Las funciones de puntaje representan una opción para medir los grados de conveniencia de un DAG con un conjunto de datos; es decir, se encargan de reducir el número de variables necesarias para representar una DBN dependiendo de la probabilidad de las variables que influyen en el DAG. El MIT ha sido considerado para el aprendizaje de BNs pero no en su totalidad para el aprendizaje de DBNs [Campos2006] , [Nguyen2001] y [Xuan2012]. La siguiente ecuación representa el MIT para BN estáticas:

$$SS_{MIT}(G : D) = \sum_{i=1, Pa_i \neq \emptyset}^n \{2N, I(X_i, Pa_i) - \sum_{j=1}^{s_i} X_{\alpha_j} l_i \sigma^i(j)\}$$

A continuación se da una breve explicación de los términos empleados en la formula del MIT:

- X_i : variable aleatoria.
- n : total de variables aleatorias.
- D : conjunto de datos de N observaciones.
- G : un DAG.
- $Pa_i = X_{i1}, \dots, X_{is_i}$: conjunto de padres de X_i .

- $s_i = |Pa_i|$: se refiere a la cardinalidad de padres (es decir a la cantidad de padres que tiene un nodo hijo).
- $I(X_i, Pa_i)$: significa la información mutua entre X_i y sus padres la cual es estimada a partir de D .
- X_{α}, l_{ij} : se refiere a la distribución X^2 .
- α : significa el nivel de confianza de la prueba de independencia para la distribución X^2 .
- El conjunto de $X = X_1, \dots, X_n$ indica el conjunto de variables n correspondiente a los estados discretos r_1, \dots, r_n .

El término $l_{i\sigma_i(j)}$ está dado por la siguiente función:

$$l_{i\sigma_i(j)} = \begin{cases} (r_i - 1)(r_{i\sigma_i(j)} - 1) \prod_{k=1}^{j-1} r_{i\sigma_i(k)}, & j = 2, \dots, s_i \\ (r_i - 1)(r_{i\sigma_i(j)} - 1), & j = 1 \end{cases}$$

Los elementos que componen a esa función se denotan como:

- $\sigma_i = \sigma_i(1), \dots, \sigma_i(s_i)$ el cual corresponde a alguna permutación del conjunto de padres que tiene un nodo hijo.
- Y $j=2, \dots, s_i$ significa que pueden existir 2 ó más nodos padres que interactúan con un nodo hijo, en el caso de que $j=1$ entonces indica que existe un nodo padre que interactúa al menos con un nodo hijo.

La prueba de la información mutua indica que tan "bueno" o "malo" es el modelo que nosotros proponamos respecto a los datos con los que contamos. La forma en que el MIT trabaja el aprendizaje del modelo de una DBN, consiste en discretizar el tiempo y crear una réplica de cada variable aleatoria para cada punto temporal de ese modelo. Para realizar la discretización del tiempo se utiliza un método de discretización para las observaciones que se tienen de éste. El modelo dinámico debe tener al menos dos instantes de tiempo, por ejemplo para dos instantes debe de ser t_0 y t_1 y para más de dos instantes debe de ser t_0 y t_1, \dots, t_n . Para pruebas de éste trabajo utilizaremos sólo dos instantes de tiempo los cuales son el tiempo t_0 y el tiempo t_1 .

Si se maneja una DBN en conjunto con el MIT se utilizan valores temporales dentro de las observaciones. Esto se ve reflejado cuando el instante de tiempo t_1 es "posterior" al instante de tiempo t_0 . Las series de tiempo para crear la DBN son denotadas de la siguiente forma:

$$SS_{MIT}(G : D) = \sum_{i=1, Pa_i \neq \emptyset}^n \{2N, I(X_i^{1 \rightarrow}, Pa_i) - \sum_{j=1}^{s_i} X_{\alpha, l_i} \sigma_i(j)\}$$

Cuando los datos están compuestos de series de tiempo separadas por N_t , el número efectivo de las observaciones se basa en $N_e = N \cdot N_t$. La figura (ver Figura 2.5) muestra el comportamiento de una DBN respecto a dos variables X_j y X_i , las cuales se relacionan en N instantes de tiempo. Algo muy importante de este proceso es que en la figura (ver Figura 2.5) podemos observar la replicación de la variable aleatoria X_j contenida en X_i en cada instante de tiempo (donde los instantes se denotan de 2 hasta N).

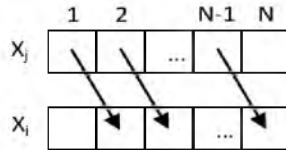


Figura 2.5: Alineamiento de datos para una DBN con un arco de $X_j \rightarrow X_i$ y el número efectivo de observaciones está conformado como $N-1$. Imagen tomada de [Nguyen2001].

La información mutua calculada en una DBN entre un nodo padre y sus nodos hijos para cada instante de tiempo hace uso de la lógica del modelo de primer orden de suposición de Markov (en inglés conocido como first order Markov Assumption), el cual introduce el cálculo de las variables aleatorias del modelo en el tiempo t y está denotado como $X_i^{1 \rightarrow} = \{X_{i2}, X_{i3}, \dots, X_{iN}\}$ [Nguyen2001] y [BIO131].

2.4.3.2. Ji cuadrada o X^2

Existen diversas definiciones para X^2 dependiendo del contexto en el que se este trabajando, por ejemplo:

- Es una prueba estadística para evaluar hipótesis acerca de dos variables categóricas.
- Consiste en comprobar si dos características cualitativas están relacionadas entre si (por ejemplo, el color de los ojos de Mary está relacionado con el color de su cabello).

- Se utiliza para probar si una serie de datos observados concuerda con un modelo, es decir con una serie de datos esperados.
- Se utiliza para probar las diferencias entre las porciones de varios grupos (en tablas de contingencia).

La prueba normalmente es usada en problemas estadísticos, se puede aplicar tanto a variables aleatorias (o estocásticas) discretas o continuas. Si se utiliza para medir porciones de varios grupos, entonces se utilizan las tablas de contingencia. Las tablas de contingencia sirven para clasificar la frecuencia con la que aparecen las variables categóricas [TCONT].

Cuando se hace uso de esta prueba estadística se requiere utilizar un nivel de significación, para el MIT, indica que el resultado de éste va ser satisfactorio o no satisfactorio dependiendo del número de redes que proponamos. El nivel de significación es comúnmente representado por el símbolo griego α (alfa). Algunas referencias citadas mencionan cual valor de alfa es pertinente utilizar en la prueba estadística de X^2 . A continuación se mencionan tres cuestiones:

1. Si conocemos como son los datos del laboratorio e investigamos acerca de ellos, podemos suponer que el valor de alfa es de 0.999, por lo tanto el resultado de la prueba estadística (X^2) contendrá menos error.
2. Si sólo nos dedicamos a realizar análisis estadísticos el valor de alfa puede ser de 0.05 ya que es un valor aceptable en algunas pruebas, por lo tanto el resultado de la prueba estadística (X^2) contendrá un nivel aceptable de errores.
3. Pero si suponemos valores de alfa mucho menores a 0.05, el resultado de la prueba estadística (X^2) contendrá un nivel elevado de errores. Sin embargo esto no quiere decir que esté mal lo que se propone sino que indica que la prueba que se está realizando no ha sido tan investigada y por ello no hay más datos o información que respalde lo que se está proponiendo.

Para las pruebas de nuestro trabajo utilizaremos un $\alpha=0.05$. Cuando se utiliza la prueba de X^2 para algún modelo gráfico, esta requiere de una hipótesis inicial. La hipótesis sirve para escoger entre dos o más variables propuestas en el modelo, las que hipotetizan una relación causal, en otras palabras esto significa cuál es la independencia existente entre las variables del modelo.

2.5. Tecnologías de software utilizadas

La tecnología utilizada en la implementación del algoritmo para el aprendizaje de la estructura en modelos gráficos es Matlab. Las razones por las cuales se decidió implementar Matlab son las siguientes:

1. Es una herramienta que nos permite codificar de forma más rápida dicho algoritmo y por tal motivo nos ayuda a reducir el tiempo de desarrollo de este proyecto.
2. Existen paquetes de herramientas matemáticas, comúnmente denominadas toolboxes las cuales incluyen funciones destinadas para algunas áreas específicas tales como: Control, procesamiento digital de Señales, Algoritmos Genéticos, Redes Neuronales, entre otras [Cubillos2007].
3. Es utilizado en diversas universidades y comunidades científicas.
4. Cuenta con una toolbox realizada por [Nguyen2001] que es aplicable al problema inverso de redes regulatorias genéticas. Ésta toolbox agilizó la programación de algunas funciones que eran necesarias para la tesis.
5. Es posible generar excelentes gráficas que sirven para ilustrar la red bayesiana dinámica del flagelo de la bacteria *Escherichia coli*.
6. Es un potente lenguaje de programación para el caso de vectores y matrices a comparación de C o C++.
7. Su única desventaja a comparación de otros lenguajes o herramientas de programación es que su velocidad es relativamente menor.

Cabe mencionar que para hacer más visual la estructura original del flagelo de la bacteria *Escherichia coli* se decidió realizar una aplicación para visualizar el comportamiento de esa estructura en HTML5. Este es un lenguaje de marcado de hipertextos, el cual sirve para agrupar nuevas tecnologías de desarrollo para aplicaciones web. HTML5 cuenta con nuevas funciones para mejorar la creación de aplicaciones modernas [Vega2011].

Las razones por las cuales se decidió implementar la tecnología de HTML5 son las siguientes:

1. Es una propuesta de una aplicación para la web, en la cual se puede mostrar en una gráfica el comportamiento de cada factor transcripción respecto a los datos experimentales obtenidos del flagelo de la bacteria *Escherichia coli* .
2. Se puede visualizar la red regulatoria de la formación del flagelo, con la finalidad de representar la codificación de los factores de transcripción denotando para ello su activación, inhibición o forma dual dada.

Capítulo 3

Protocolo experimental

3.1. Métodos y técnicas

A continuación se mostrarán las razones por las cuales proponemos un modelo gráfico para el flagelo de la bacteria E.coli a través de una red bayesiana dinámica, también se explicará el porqué seleccionamos dentro de las funciones de puntuación de información teórica la prueba de información mutua para ser aplicado al modelo gráfico construido a partir de datos y observaciones conocidas del flagelo.

Anexamos dos scripts desarrollados en Matlab, los cuales son utilizados para el aprendizaje y comprobación de la calidad de la red bayesiana dinámica que proponemos. Dentro de este documento se explicará de forma general el algoritmo implementado dentro de todos los scripts, pero dentro de cada script existe una leyenda en donde se muestra la funcionalidad de estos. También anexamos el programa desarrollado bajo el lenguaje de HTML5 el cual tiene la finalidad de mostrar visualmente el comportamiento que tienen los factores de transcripción extraídos de [Martinez2012] y de [Valdez].

3.1.1. Estructura del modelo

Partimos de la estructura de la formación del flagelo de la bacteria Escherichia coli (el cual es un fenómeno dinámico) (ver Figura 3.1) y de las observaciones de cada gen que lo conforman (ver

Figura 3.2) con la finalidad de obtener datos a posteriori, tomamos la red reportada en [Martinez2012] y [Valdez].

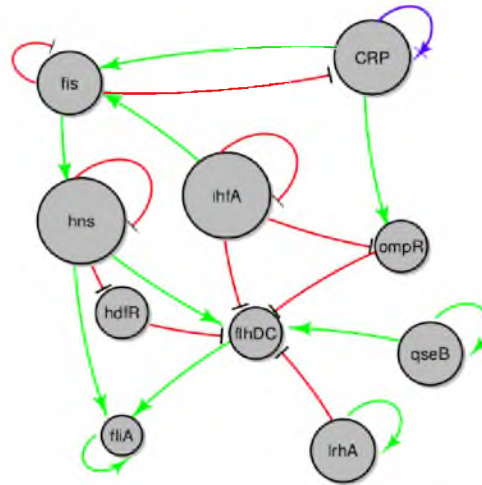


Figura 3.1: Red regulatoria de la formación del flagelo. Imagen tomada de [Valdez].

r(t)					
Time	lrhA	ihfA	flhDC	qseB	ompR
0	0.151457418023	0.162081906963	0.000000000000	0.139792160601	0.162341145011
1	0.138010516406	0.153360095696	0.000000000000	0.122254454804	0.152306948682
2	0.118647257516	0.145283365488	0.000000000000	0.103223449008	0.138359798168
3	0.095332506832	0.142596490980	0.000000000000	0.080445976075	0.118144641928
4	0.075681247697	0.141442034034	0.000000000000	0.062205989863	0.101424909059
5	0.067464741157	0.154132304912	0.000000000000	0.055277561913	0.091511892568
6	0.065981140222	0.154879846561	0.000000000000	0.053973900752	0.092141824102
7	0.068931230266	0.145002811898	0.000000000000	0.056982136237	0.095146933205
8	0.069463596878	0.132309743180	0.000000000000	0.059452170754	0.094972693670
Time	fis	hns	hdfR	fliA	CRP
0	0.166345468722	0.594940827600	0.143558644405	0.144390050887	0.151916143639
1	0.167063182047	0.617380141943	0.132174451782	0.132521391723	0.137372982549
2	0.152973128072	0.651794227383	0.113161124208	0.112455694647	0.122430040909
3	0.134924040729	0.727359571106	0.090011434649	0.088771753276	0.108702034482
4	0.124439363535	0.801502606204	0.073262396397	0.068856177259	0.096833681030
5	0.122171021028	0.860450063908	0.067525206050	0.058728904749	0.089404294159
6	0.117292234228	0.928709497784	0.066909368590	0.057895001098	0.086421337403
7	0.114108539054	1.000000000000	0.067799006585	0.061287892113	0.086322098893
8	0.104649695425	0.953162637868	0.067843713181	0.062963973870	0.085820117759

Figura 3.2: Normalización de los valores del GFP/OD. Imagen tomada de [Valdez].

Además tomamos 20 variables para representar los promotores del flagelo de la bacteria para cada factor de transcripción y se mencionan a continuación:

$$x_1=lrhA, x_2=ihfA, x_3=flhDC, x_4=qseB, x_5=ompR, x_6=fis, x_7=hns, x_8=hdfR, x_9=fliA, x_{10}=CRP.$$

x_{p1} =lrhA, x_{p2} =ihfA, x_{p3} =flhDC, x_{p4} =qseB, x_{p5} =ompR, x_{p6} =fis, x_{p7} =hns, x_{p8} =hdfR, x_{p9} =fliA, x_{p10} =CRP.

Una parte importante del análisis de la estructura (ver Figura 3.1) fue tomar algunas decisiones que nos permitieran modelar el conocimiento a posteriori de la DBN. La primera observación que se realizó dentro de la estructura se debió a que era posible obtener dos DBNs las cuales no diferían mucho. La razón por la que sólo se tomaron dos DBNs dependió de las interacciones de los promotores obtenidos de la red reportada, como sólo se observó un ciclo entre los factores de transcripción entonces por eso sólo obtuvimos dos redes las cuales incluyen la interacción de inhibición de fis a CRP para la primera DBN y la interacción de la activación de CRP a fis para la segunda DBN. La figura (ver Figura 3.3) muestra la arista que existe del nodo x_{10} al nodo x_6 equivalente a la activación y la figura (ver Figura 3.4) muestra la arista que existe del nodo x_6 al nodo x_{10} equivalente a la inhibición.

Para establecer la forma de introducir los datos tanto de la estructura de la formación del flagelo como sus respectivas observaciones para modelar el conocimiento a posteriori de una DBN es muy importante, ya que si se cambia un dato de entrada entonces cambia la salida del conocimiento a posteriori de la DBN.

La construcción de las estructuras gráficas se realizó de la siguiente forma: En la figura (ver Figura 3.1) se observa que sólo existen 10 nodos (los cuales son equivalentes a los 10 TFs del flagelo) que interactúan entre sí. La representación utilizada para los nodos equivale a las variables que van desde x_1 hasta x_{10} , pero esa parte sólo simulaba una red bayesiana estática. Para que ésta estructura funcionará como una red bayesiana dinámica, se tenía que representar otra red bayesiana estática, la cual tuviera exactamente el mismo número de nodos e interacciones de la otra red que se había formado. Para representar los nodos de la segunda red, se utilizaron las variables que van desde x_{p1} hasta x_{p10} y finalmente para la parte dinámica representamos a las variables $x_1, x_2, x_4, x_6, x_7, x_9$ y x_{10} por medio de la unión de las dos redes bayesianas estáticas como se muestra a continuación: $x_{p1} - x_1, x_{p2} - x_2, x_{p4} - x_4, x_{p6} - x_6, x_{p7} - x_7, x_{p9} - x_9$ y $x_{p10} - x_{10}$. En la figura (ver Figura 3.5) o la figura (ver Figura 3.3) y la figura (ver Figura 3.6) o (ver Figura 3.4) se observan las dos propuestas que realizamos.

La construcción de las tablas o matrices de adyacencia sirven para conocer cuales son los factores de transcripción que pueden estar relacionados de acuerdo al análisis que realizamos. Por lo que si existe una relación o interacción entre un factor de transcripción respecto a otro, entonces en la tabla se indica como un 1, en caso de no haber una interacción entonces en la tabla se indica como un 0. El cuadro (ver Cuadro 3.1) representa la figura (ver Figura 3.3) y el cuadro (ver Cuadro 3.2) representa la figura (ver Figura 3.4).

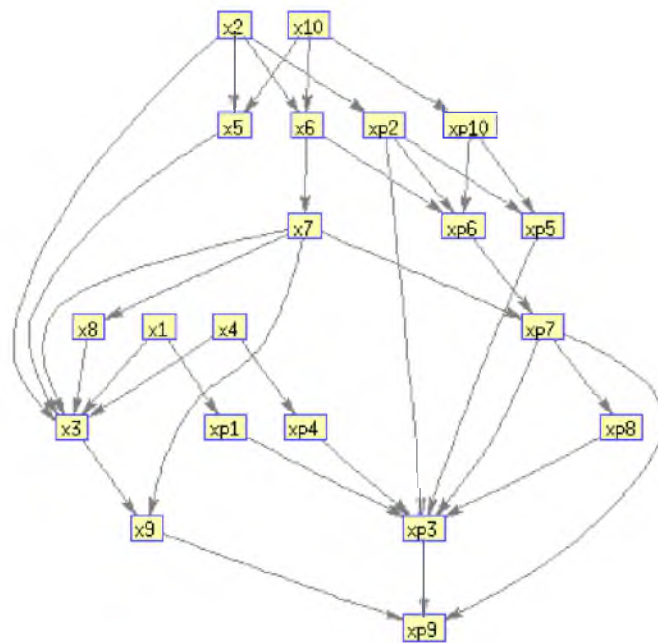


Figura 3.3: Red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_6 depende del nodo x_{10} .

En nuestro trabajo, el DAG está compuesto de nodos (representados como rectángulos con un nombre en su interior) y arcos (que sirven para conectar a los nodos). Los nodos representan variables aleatorias (que van desde x_1 hasta x_{p10}) y los arcos representan las dependencias entre las variables aleatorias. Las variables aleatorias pueden ser continuas o discretas, pero las discretas a comparación de las continuas se representan por el conjunto de estados deterministas que puede tomar la variable aleatoria, las continuas también pueden representarse por un conjunto de estados, sólo que sus estados suelen tender a infinito. Como los valores obtenidos de las observaciones de análisis del flagelo eran continuos, entonces aplicamos un método para discretizar los valores.

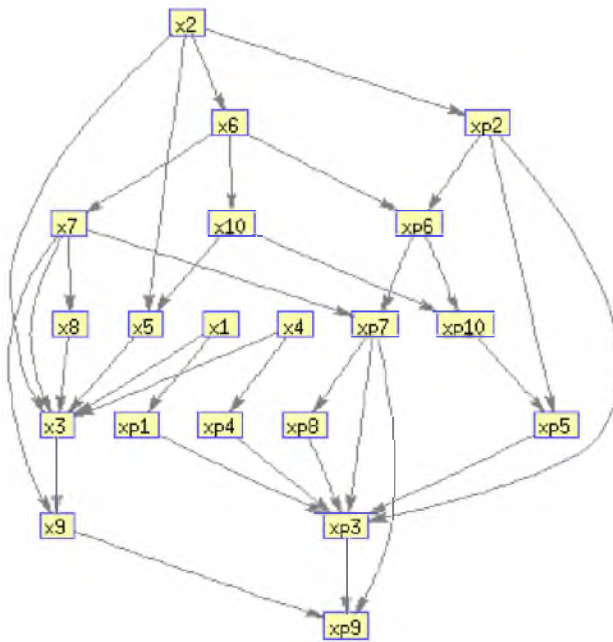


Figura 3.4: Red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_{10} depende del nodo x_6 .

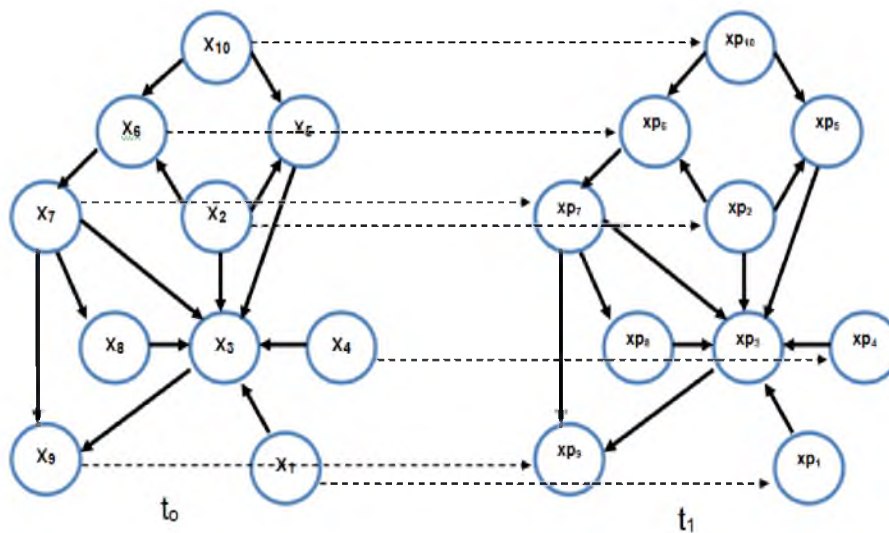


Figura 3.5: Primera propuesta de la red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_6 depende del nodo x_{10} .

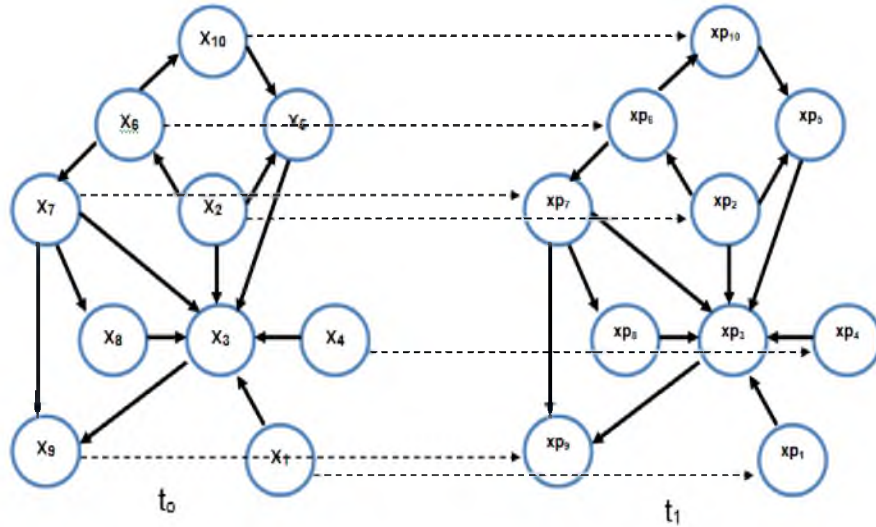


Figura 3.6: Segunda propuesta de la red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_{10} depende del nodo x_6 .

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	0	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_7	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0
x_8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
x_{p1}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_{p2}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0
x_{p3}	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
x_{p4}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_{p5}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_{p6}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
x_{p7}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
x_{p8}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_{p9}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_{p10}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0

Cuadro 3.1: Primera matriz de adyacencia de la red bayesiana dinámica elaborada. Tabla tomada a partir de la figura (ver Figura 3.3).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	0	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0
x_7	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0
x_8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
x_{p1}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_{p2}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0
x_{p3}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
x_{p4}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_{p5}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_{p6}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
x_{p7}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
x_{p8}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_{p9}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_{p10}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Cuadro 3.2: Segunda matriz de adyacencia de la red bayesiana dinámica elaborada. Tabla tomada a partir de la figura (ver Figura 3.4).

La finalidad de realizar la discretización se debe a que es más práctico trabajar la inferencia y comprobar la calidad de la DBN creada con datos discretizados. Nosotros discretizamos los datos de la figura (ver Figura 3.2), utilizando el método de la mediana para cada factor de transcripción. Estos datos representan las observaciones del flagelo de la bacteria E.coli.

El resultado que obtuvimos se puede observar en el cuadro (ver Cuadro 3.3), en donde utilizamos 1's y 2's para representar los valores discretos.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
2	2	2	2	2	2	1	2	2	2
2	1	2	2	2	2	1	2	2	2
2	1	2	2	2	2	1	2	2	2
2	2	2	2	2	2	1	2	2	2
1	1	2	1	1	1	1	1	1	1
1	1	2	1	1	1	2	1	1	1
1	2	2	1	1	1	2	1	1	1
1	2	2	1	1	1	2	1	1	1
1	1	2	1	1	1	2	1	1	1

Cuadro 3.3: Resultados de la discretización obtenidos a partir de la figura (ver Figura 3.2)

Cabe mencionar que el número de observaciones obtenidas del laboratorio son sólo nueve para cada factor de transcripción como se muestra en la figura (ver Figura 3.2), ahí se puede observar

que los valores son continuos. Estos valores equivalen al conocimiento a posteriori de la DBN que nosotros proponemos a partir de la figura (ver Figura 3.3). Del conocimiento obtenido nosotros podemos realizar inferencias sobre cada relación existente entre los factores de transcripción de la red para obtener el aprendizaje de la DBN.

En tanto, para la representación de los tiempos utilizados en la red bayesiana dinámica podemos decir que la red que tiene las variables de x_1 a x_{10} , se refiere al instante de tiempo 0 (t_0) y la red que tiene las variables de x_{p1} a x_{p10} se refiere al instante de tiempo 1 (t_1). Es evidente que en ésta representación dinámica aún no se tiene introducida la prueba de información mutua, por lo que no se puede concluir todavía nada.

Antes de aplicar el MIT a la DBN es necesario mencionar que utilizamos dos pruebas para verificar como trabaja el MIT la dinámica de una red. La primera prueba consiste en el decremento de variables y la segunda prueba en el incremento de variables. Las variables que se utilizan para estos procesos son $x_1, x_2, x_4, x_6, x_7, x_9$ y x_{10} , las cuales son analizadas para cada red propuesta, debido a que en nuestro trabajo es esencial trabajar sobre ellas.

El decremento de variables consiste en proponer siete matrices de adyacencia para cada una de las redes propuestas e ir eliminando la interacción de dos variables para cada una de las matrices, conservando los valores de cada variable dinámica que no se elimine. Por ejemplo: el cuadro (ver Cuadro 3.4) elimina la interacción $x_1 - x_{p1}$, el cuadro (ver Cuadro 3.5) elimina la interacción $x_2 - x_{p2}$, el cuadro (ver Cuadro 3.6) elimina la interacción $x_4 - x_{p4}$, el cuadro (ver Cuadro 3.7) elimina la interacción $x_6 - x_{p6}$, el cuadro (ver Cuadro 3.8) elimina la interacción $x_7 - x_{p7}$, el cuadro (ver Cuadro 3.9) elimina la interacción $x_9 - x_{p9}$ y el cuadro (ver Cuadro 3.10) elimina la interacción $x_{10} - x_{p10}$.

El incremento de variables consiste en proponer siete matrices de adyacencia a partir de las dos redes propuestas. Nuestro razonamiento consistió en buscar cual eran las interacciones que formaban ciclos dentro de la red estática y sugerimos que al encontrar esos ciclos, podía haber la posibilidad de que esto fuera aplicado a la dinámica de la red y obtuvieramos con ello puntajes positivos, también realizamos algunas pruebas colocando interacciones aleatorias entre las variables y obtuvimos pun-

tajes negativos.

Para la primera propuesta, los ciclos que encontramos son los siguientes:

- El cuadro (ver Cuadro 3.4) no aumenta ninguna interacción pero elimina la interacción $x_1 - xp_1$, el cuadro (ver Cuadro 3.5) no aumenta ninguna interacción pero elimina la interacción $x_2 - xp_2$.
- El cuadro (ver Cuadro 3.6) no aumenta ninguna interacción pero elimina la interacción $x_4 - xp_4$, el cuadro (ver Cuadro 3.7) aumenta la interacción $x_6 - xp_2, x_6 - xp_{10}$ y elimina la interacción $x_6 - xp_6$
- El cuadro (ver Cuadro 3.8) aumenta la interacción $x_7 - xp_6, x_7 - xp_2$ y elimina la interacción $x_7 - xp_7$
- El cuadro (ver Cuadro 3.9) aumenta la interacción $x_9 - xp_1, x_9 - xp_2, x_9 - xp_3, x_9 - xp_4, x_9 - xp_5, x_9 - xp_7, x_9 - xp_8$ y elimina la interacción $x_9 - xp_9$.
- El cuadro (ver Cuadro 3.10) no aumenta ninguna interacción, pero elimina la interacción $x_{10} - xp_{10}$.

Para la segunda propuesta, los ciclos que encontramos son los siguientes:

- El cuadro (ver Cuadro 3.11) no aumenta ninguna interacción pero elimina la interacción $x_1 - xp_1$.
- El cuadro (ver Cuadro 3.12) no aumenta ninguna interacción pero elimina la interacción $x_2 - xp_2$.
- El cuadro (ver Cuadro 3.13) no aumenta ninguna interacción pero elimina la interacción $x_4 - xp_4$
- El cuadro (ver Cuadro 3.14) aumenta la interacción $x_6 - xp_2$ y elimina la interacción $x_6 - xp_6$.
- El cuadro (ver Cuadro 3.15) aumenta la interacción $x_7 - xp_6, x_7 - xp_2$ y elimina la interacción $x_7 - xp_7$.
- El cuadro (ver Cuadro 3.16) aumenta la interacción $x_9 - xp_1, x_9 - xp_2, x_9 - xp_3, x_9 - xp_4, x_9 - xp_5, x_9 - xp_7, x_9 - xp_8$ y elimina la interacción $x_9 - xp_9$.
- El cuadro (ver Cuadro 3.17) aumenta la interacción $x_{10} - xp_6, x_{10} - xp_2$ y elimina la interacción $x_{10} - xp_{10}$.

	DECREMENTO RED 1										DECREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.4: Decremento del nodo x_1

	DECREMENTO RED 1										DECREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.5: Decremento del nodo x_2

	DECREMENTO RED 1										DECREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.6: Decremento del nodo x_4

	DECREMENTO RED 1										DECREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.7: Decremento del nodo x_6

	DECREMENTO RED 1										DECREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.8: Decremento del nodo x_7

	DECREMENTO RED 1										DECREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.9: Decremento del nodo x_9

	DECREMENTO RED 1										DECREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Cuadro 3.10: Decremento del nodo x_{10}

	INCREMENTO RED 1										INCREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.11: Incremento de las aristas de x_1 no tiene.

	INCREMENTO RED 1										INCREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.12: Incremento de las aristas de x_2 no tiene.

	INCREMENTO RED 1										INCREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.13: Incremento de las aristas de x_4 no tiene.

	INCREMENTO RED 1										INCREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.14: Incremento de las aristas de x_6 dado x_2 y x_{10} . Incremento de la arista de x_6 dado x_2 .

	INCREMENTO RED 1										INCREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.15: Incremento de las aristas de x_7 dado x_2 y x_6 . Incremento de las aristas de x_7 dado x_2 y x_6 .

	INCREMENTO RED 1										INCREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	0	1	1	0	0
x_{10}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Cuadro 3.16: Incremento de las aristas de x_9 dado $x_1, x_2, x_3, x_4, x_5, x_7$ y x_8 . Incremento de las aristas de x_9 dado $x_1, x_2, x_3, x_4, x_5, x_7$ y x_8 .

	INCREMENTO RED 1										INCREMENTO RED 2									
	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}	x_{p1}	x_{p2}	x_{p3}	x_{p4}	x_{p5}	x_{p6}	x_{p7}	x_{p8}	x_{p9}	x_{p10}
x_1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x_2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_4	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x_7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
x_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
x_{10}	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0

Cuadro 3.17: Incremento de las aristas de x_{10} no tiene. Incremento de las aristas de x_{10} dado x_2 y x_6 .

En el apartado de resultados se introducirá el puntaje de cada una de las matrices de adyacencia de acuerdo al decremento o incremento de interacciones entre variables y se comparará con el puntaje de las dos matrices de adyacencia propuestas.

3.2. Prueba de información mutua (MIT)

En capítulos anteriores se mencionó que actualmente es posible comprobar la calidad de una DBN a través de alguna función de puntuación de información teórica. Para nuestro trabajo, empleamos un algoritmo para el aprendizaje de DBN. Cabe mencionar que utilizamos una toolbox en Matlab para realizar la prueba de información mutua con la finalidad de agilizar la elaboración de éste trabajo, haciéndole algunas adaptaciones conforme a nuestro problema. La toolbox fue extraída de [CODEMIT] y [Nguyen2001].

Por lo tanto lo que realizamos fue el aprendizaje de las interacciones regulatorias entre genes del

flagelo de la bacteria E.coli dadas mediante su estructura. La prueba fue aplicada a las dos redes propuestas, las redes conformadas de decrementos y las redes que tenían incrementos y como resultado obtuvimos 30 puntajes. A continuación se establece el código del algoritmo para la prueba que se desarrolló (el cual se extrajo de la toolbox) y se explica más adelante cual es su funcionalidad:

La función "funcion.MIT" recibe como entrada los valores discretizados del cuadro (ver Cuadro 3.3) y la figura (ver Figura 4.4). Durante el desarrollo de la función se buscó utilizar el factor de penalización respecto al cuadro mencionado así que el resultado obtenido para cada TF fue de 2 y a partir de esto se buscó la información mutua entre cada variable propuesta para el modelo, es decir entre los padres e hijos, con el objetivo de probar la hipótesis de independencia entre X_i y X_j dado el valor de $\alpha=0.05$.

```

function score=funcion_MIT(a,net)
contador=1;
n_state=max(max(a)); %factor de penalización
if nargin<3
    alpha=0.05;
end;
[n dim]=size(a); % 9 10
vectorPuntaje=zeros(1,dim);
chi=zeros(1,dim); %máximo número de padres que tiene un nodo
for i=1:dim
    valor=n_state^(i-1)*(n_state-1)^2;
    chi(i)= chi2inv(alpha,n_state^(i-1)*(n_state-1)^2);
end
score=0;
%la función de puntaje es descompuesta, se obtiene el puntaje de cada variable o nodo
for i=1:dim %columnas de la matriz de la red que representa los padres de un nodo
    nPa=sum(net(:,i)); %obtenemos el número de padres xi
    Pa=1:dim; fprintf('Padres de: %d',i);
    Pa=Pa(boolean(net(:,i))) %lista de padres
    score_i=0;

    if nPa>=1 %si el nodo tiene algún padre entonces se calcula el puntaje de la siguiente manera
        fprintf('Pa(a)=%d, i=%d',Pa(j),i); %Observamos la información mutua entre los padres del tiempo 0 al tiempo 1
        score_i=2*(n-1)*informacion_mutua_condicional(a,Pa(1), i, 0,n_state)-chi(1)
    end
    if nPa>=2 %sino entonces
        for j=2:nPa
            fprintf('Pa(a)=%d, i=%d',Pa(j),i); %Observamos la información mutua entre los padres del tiempo 0 al tiempo 1
            score_i=score_i+2*(n-1)*informacion_mutua_condicional(a,Pa(j), i, Pa(1:j-1),n_state)-chi(j) %numero total de observaciones=n
        end
    end
    vectorPuntaje(contador)=score_i;
    score=score+score_i;
    contador=contador+1;
end

```

Figura 3.7: Algoritmo de la prueba de independencia mutua. Imagen tomada de [CODEMIT]

```

function CMI=informacion_mutua_condicional(a,Xi, Xj, Pa,n_state)
if nargin<5 n_state=max(max(a));end;
[n dim]=size(a); scanned=zeros(1,n); prob_Pa=[]; %probabilidad de cada padre Pa
MI_arr=[]; %el valor correspondiente a la información mutua
Ne=n-1; %número de observaciones efectivas
nPa=length(Pa);
if Pa==0
    Mem1=a(1:end-1,Xi); Mem2=a(2:end,Xj); Cont=Contingency(Mem1,Mem2,n_state);
    CMI=Mutu_info(Cont,n_state); %calcula la información mutua condicional
    return;
end
for i=1:n-1 %construye una tabla de contingencia correspondiente a la combinación de Pa
    if scanned(i)==0 %sirve para encontrar una nueva combinación de Pa en la matriz discretizada (a(i,Pa))
        count=1; %es para encontrar el número de veces que aparece la combinación de la matriz discretizada a
        scanned(i)=1; T=zeros(n_state,n_state); T(a(i,Xi),a(i+1,Xj))=T(a(i,Xi),a(i+1,Xj))+1;
        for j=i+1:n-1
            if (scanned(j)==0 && sum(a(i,Pa)==a(j,Pa))==nPa)
                scanned(j)=1; T(a(j,Xi),a(j+1,Xj))=T(a(j,Xi),a(j+1,Xj))+1; count=count+1;
            end
        end MI_arr=[MI_arr Mutu_info(T,n_state)]; prob_Pa=[prob_Pa count];
    end
end
prob_Pa=prob_Pa/sum(prob_Pa); %normaliza la probabilidad
CMI=sum(prob_Pa.*MI_arr);
function MI=Mutu_info(T,n_state)
%calcula la información mutua
a=sum(T'); b=sum(T); N=sum(a); MI=0;
for i=1:n_state
    for j=1:n_state
        if T(i,j)>0
            MI=MI+T(i,j)*log(T(i,j)*N/(a(i)*b(j)));
        end;
    end
end
MI=MI/N;
function Cont=Contingency(Mem1,Mem2,n_state) %Busca la contingencia del número de Pa
Cont=zeros(n_state,n_state);
for i = 1:length(Mem1);
    Cont(Mem1(i),Mem2(i))=Cont(Mem1(i),Mem2(i))+1;cap
end

```

Figura 3.8: Algoritmo de la independencia mutua condicional. Imagen tomada de [CODEMIT]

Capítulo 4

Resultados

4.1. Prueba de información mutua

A continuación se dará una breve explicación de los resultados obtenidos en la compilación del script elaborado en Matlab:

A partir de la red regulatoria del flagelo (ver Figura 3.1), se tomaron las variables aleatorias que intervinieron en el proceso de aprendizaje de la DBN. Como se puede observar, la figura cuenta con ciclos o enlaces hacia las variables, debido a ello se decidió proponer dos redes bayesianas dinámicas y probarlas con el algoritmo implementado del MIT, con la finalidad de corroborar la fiabilidad de las estructuras y a partir del resultado obtenido poder colocar las interacciones faltantes entre cada variable de la red bayesiana dinámica. En el cuadro (ver Cuadro 4.1) se observan los resultados de las dos redes bayesianas propuestas. El puntaje de la red1 está dado por la figura (ver Figura 4.4), el cual es equivalente a 24.520086 y el puntaje de la red2 está dado por la figura (ver Figura 3.4), el cual es equivalente a 24.041938.

De las primeras dos redes propuestas, podemos decir que al momento de calcular la información mutua entre cada nodo de la red, el puntaje más alto entre los nodos padres y sus hijos (24.520086) posee un incremento de información mutua, es decir cuando ésta red posee mayores relaciones de independencia condicional entonces se obtiene un mejor puntaje.

De la comparación de puntaje de la red1 equivalente a 24.520086 contra los decrementos que se observan en el cuadro (ver Cuadro 4.2) pertenecientes a la segunda columna podemos decir que sólo obtuvimos una aproximación de la red, aunque existan los valores de 25.230809 (ver Figura 4.1) y 24.622673. El resultado de nuestro modelo está dentro de un rango aceptable e incluso los demás valores no están tan alejados del puntaje de 24.520086. A comparación del puntaje de la red2 equivalente a 24.041938 contra los decrementos que se observan en el cuadro (ver Cuadro 4.2) perteneciente a la tercera columna, podemos inferir que es muy poco probable considerar este modelo ya que los valores de esa columna están muy alejados del puntaje de 24.041938.

La comparación del puntaje de la red1 24.520086 contra los incrementos observados en el cuadro (ver Cuadro 4.3) pertenecientes a la segunda columna y la comparación del puntaje de la red2 24.041938 contra los incrementos observados en el cuadro (ver Cuadro 4.3) pertenecientes a la cuarta columna, podemos decir de forma general que los puntajes obtenidos son más altos que en los decrementos y eso no es conveniente en el modelo que proponemos, ya que no buscamos un modelo complicado que contenga demasiados parámetros. En la figura (ver Figura 4.2) se puede observar que la variable x_{p_6} ya no tiene una independencia condicional de la variable x_6 , sin embargo x_{p_2} y $x_{p_{10}}$ depende condicionalmente de x_6 en el t_0 , pero como surgió esta nueva dependencia entonces el modelo requiere más parámetros y se vuelve más complejo. Por ese motivo llegamos a la conclusión que el modelo que mejor se adaptaba es la red1, esta contiene un puntaje de 24.520086.

Durante el análisis de la toolbox, se observó que en la figura (ver Figura 3.7) en conjunto con la figura (ver Figura 2.5) \mathbf{Pa} representa X_i el cual es equivalente al instante de tiempo 1 e \mathbf{i} representa X_j el cual es equivalente al instante de tiempo 0. Por consiguiente el resultado que obtuvimos de la figura (ver Figura 4.3) muestra la DBN en el tiempo 0 y el tiempo 1 a partir de la estructura con el puntaje más alto 24.520086 (ver Figura 4.4). En la figura (ver Figura 4.3) las líneas rojas punteadas indican las interacciones dinámicas de las variables $x_{p_1} - x_1$, $x_{p_2} - x_2$, $x_{p_4} - x_4$, $x_{p_6} - x_6$, $x_{p_7} - x_7$, $x_{p_9} - x_9$ y $x_{p_{10}} - x_{10}$ y las líneas negras punteadas indican la información mutua entre los nodos padres y los hijos.

Nodos (X_i)	Padres del nodo (P_a) de la red1	Padres de nodo (P_a) de la red2	Puntaje red1	Puntaje red2
x_1	0	0	0.000000	0.000000
x_2	0	0	0.000000	0.000000
x_3	$x_1, x_2, x_4, x_5, x_7, \text{ y } x_8$	$x_1, x_2, x_4, x_5, x_7, \text{ y } x_8$	-31.583438	-31.583438
x_4	0	0	0.000000	0.000000
x_5	$x_2 \text{ y } x_{10}$	$x_2 \text{ y } x_{10}$	7.705904	7.705904
x_6	$x_2 \text{ y } x_{10}$	x_2	7.705904	0.537221
x_7	x_6	x_6	11.086423	11.086423
x_8	x_7	x_7	3.850963	3.850963
x_9	$x_3 \text{ y } x_7$	$x_3 \text{ y } x_7$	3.748376	3.748376
x_{10}	0	x_6	0.000000	6.082399
x_{p1}	x_1	x_1	6.082399	6.082399
x_{p2}	x_2	x_2	0.537221	0.537221
x_{p3}	$x_{p1}, x_{p2}, x_{p4}, x_{p5}, x_{p7} \text{ y } x_{p8}$	$x_{p1}, x_{p2}, x_{p4}, x_{p5}, x_{p7} \text{ y } x_{p8}$	-31.583438	-31.583438
x_{p4}	x_4	x_4	6.082399	6.082399
x_{p5}	$x_{p2} \text{ y } x_{p10}$	$x_{p2} \text{ y } x_{p10}$	7.705904	7.705904
x_{p6}	$x_6, x_{p2} \text{ y } x_{p10}$	$x_6 \text{ y } x_{p2}$	6.995181	7.705904
x_{p7}	$x_7 \text{ y } x_{p6}$	$x_7 \text{ y } x_{p6}$	10.983836	10.983836
x_{p8}	x_{p7}	x_{p7}	3.850963	3.850963
x_{p9}	$x_9, x_{p3} \text{ y } x_{p7}$	$x_9, x_{p3} \text{ y } x_{p7}$	5.269089	5.269089
x_{p10}	x_{10}	x_{10}, x_{p6}	6.082399	5.979812
Puntaje total			24.520086	24.041938

Cuadro 4.1: Resultados del decremento de cada nodo

DECREMENTO		
Eliminación de aristas	Puntaje red1	Puntaje red2
x_1-x_{p1}	18.437688	17.959539
x_2-x_{p2}	23.982865	23.504717
x_4-x_{p4}	18.437688	17.959539
x_6-x_{p6}	25.230809	16.873254
x_7-x_{p7}	24.622673	24.144524
x_9-x_{p9}	22.999374	22.521225
$x_{10}-x_{p10}$	18.437688	24.144524

Cuadro 4.2: Resultados del decremento de cada nodo

INCREMENTO			
Incremento de aristas	Puntaje red1	Incremento de aristas	Puntaje red2
x_1-0	18.437688	x_1-0	17.959539
x_2-0	23.982865	x_2-0	23.504717
x_4-0	18.437688	x_4-0	17.959539
$x_6-x_{p2} \text{ y } x_6-x_{p10}$	26.751728	x_6-x_{p2}	18.496760
$x_7-x_{p2} \text{ y } x_7-x_{p6}$	25.239634	$x_7-x_{p2} \text{ y } x_7-x_{p6}$	26.783400
$x_9-x_{p1}, x_9-x_{p2}, x_9-x_{p3}, x_9-x_{p4}, x_9-x_{p5}, x_9-x_{p7} \text{ y } x_9-x_{p8}$	-21.469796	$x_9-x_{p1}, x_9-x_{p2}, x_9-x_{p3}, x_9-x_{p4}, x_9-x_{p5}, x_9-x_{p7} \text{ y } x_9-x_{p8}$	-21.947944
$x_{10}-0$	18.437688	$x_{10}-x_{p2} \text{ y } x_{10}-x_{p6}$	25.057307

Cuadro 4.3: Resultados del incremento de cada nodo

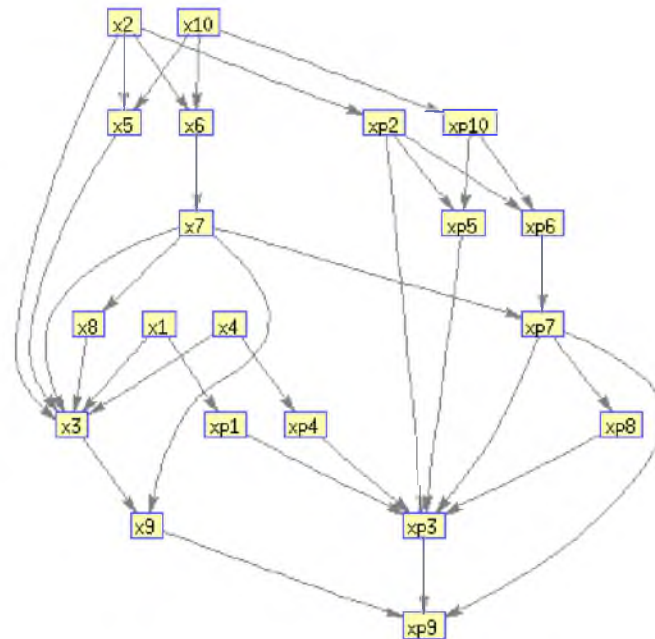


Figura 4.1: Decremento de información mutua de la red regulatoria.

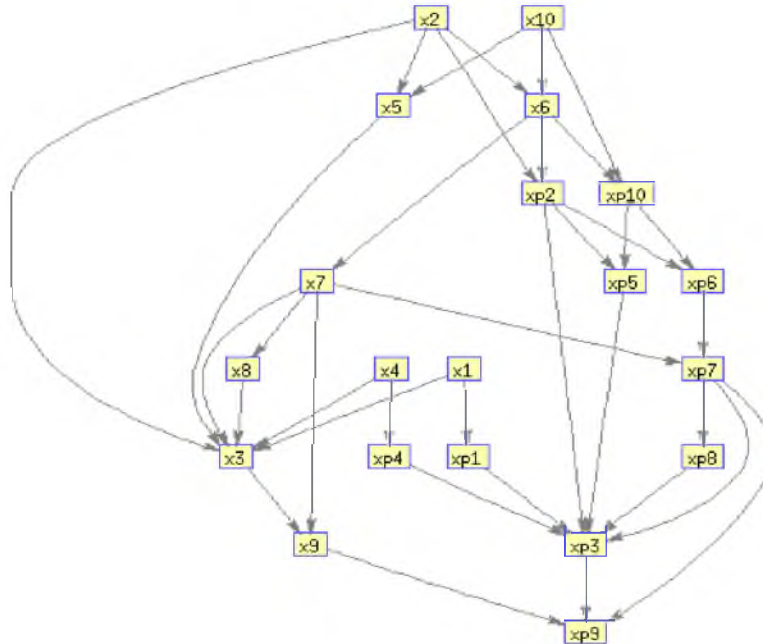


Figura 4.2: Incremento de información mutua de la red regulatoria.

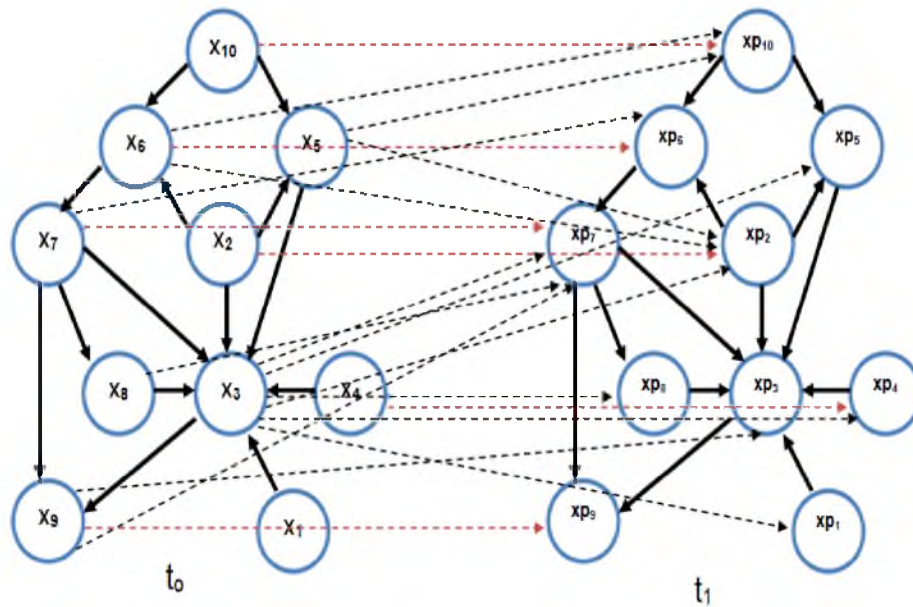


Figura 4.3: DBN obtenida a partir del MIT.

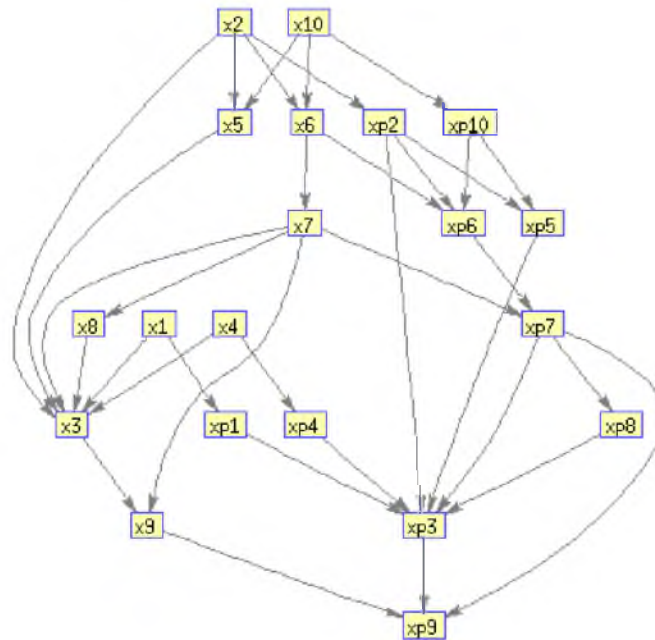


Figura 4.4: Red bayesiana dinámica elaborada a partir de la red regulatoria de la formación del flagelo. Donde el nodo x_6 depende del nodo x_{10} .

4.2. Aprendizaje de la DBN

De la red regulatoria obtenida con el más alto puntaje (ver Figura 4.4) de 24.520086 se realizó el conteo de parámetros o frecuencia absoluta de parámetros para conocer el aprendizaje dentro de la estructura, ya que se tenían algunas de sus observaciones. La probabilidad a posteriori de la estructura está dada por las siguientes CPTs:

	p(A)
a=0	5/9
a=1	4/9

Cuadro 4.4: Probabilidad del nodo A.

	p(D)
d=0	5/9
d=1	4/9

Cuadro 4.5: Probabilidad del nodo D.

F	G	p(G F)
0	0	1/5
0	1	4/5
1	0	4/4
1	1	0.001

Cuadro 4.6: Probabilidad del nodo G dado el nodo F.

G	H	p(H G)
0	0	1/5
0	1	4/5
1	0	4/4
1	1	0.001

Cuadro 4.7: Probabilidad del nodo H dado el nodo G.

Cuando no se cuentan con todas las observaciones en algunas CPTs se puede suponer que el valor de la observación es de 0.001. Esto indica que no se descarta la posibilidad de que pueda haber una relación de independencia entre los nodos.

Podemos observar que los resultados de las CPTs de las figuras (ver Figura 4.9) y (ver Figura 4.10) son similares. Para las variables B y J decimos que son condicionalmente dependientes dado E o F. También podemos observar algunas independencias condicionales como:

C	G	I	$p(I C,G)$
1	0	1	4/5
1	0	1	
1	0	1	
1	0	1	
1	0	0	1/5
1	1	0	1
1	1	0	
1	1	0	
1	1	0	

Cuadro 4.8: Probabilidad del nodo I dado los nodos C y nodo G.

B	J	C	$p(E B,J)$
1	1	1	1
1	1	1	
0	1	1	1
0	1	1	
0	0	0	1
0	0	0	
0	0	0	
1	0	0	0.001
1	0	0	

Cuadro 4.9: Probabilidad del nodo E dado el nodo B y el nodo J.

B	J	F	$p(F B,J)$
1	1	1	1
1	1	1	
0	1	1	1
0	1	1	
0	0	0	1
0	0	0	
0	0	0	
1	0	0	0.001
1	0	0	

Cuadro 4.10: Probabilidad del nodo F dado los nodos B y nodo J.

A	B	D	E	G	H	C	$p(C A,B,D,E,G,H)$
1	1	1	1	0	1	1	1
1	1	1	1	0	1	1	
1	0	1	1	0	1	1	1
1	0	1	1	0	1	1	
0	0	0	0	0	0	1	1
0	0	0	0	1	0	1	1
0	0	0	0	1	0	1	
0	1	0	0	1	0	1	1
0	1	0	0	1	0	1	

Cuadro 4.11: Probabilidad del nodo F dado el nodo A, el nodo B, el nodo D, el nodo E, el nodo G y el nodo H.

1. C, E y F son condicionalmente independientes dado B; o

2. las variables F y E son condicionalmente independientes dado J.

Esto es necesario conocer en nuestra red bayesiana dinámica debido a que todas las independencias condicionales en el grafo propuesto corresponden a las relaciones de independencia colocadas en las CPTs.

4.3. Graficación del flagelo de la bacteria *Escherichia coli*

En la (ver Figura 4.5) se muestra la red regulatoria del flagelo de la bacteria *Escherichia coli*. Esta figura generaliza la representación de la dinámica de la red respecto a cada variable aleatoria utilizada en el flagelo.

Podemos observar que en la figura (ver Figura 4.6) es posible seleccionar de una lista cada factor de transcripción. Existen dos finalidades de la sección "MANEJO DE DATOS", la primera es que dentro de los "Resultados obtenidos del modelo optimizado (puntos azules)", se muestran los datos obtenidos del laboratorio vistos para cada factor de transcripción y los cuales son correspondientes a 9 muestras que se obtuvieron en [Martinez2012]. La segunda finalidad se refiere a los "Valores normalizados GFP/OD (puntos negros)", ahí se muestran los datos que obtuvieron algunos investigadores como [Valdez] respecto al modelo matemático que propusieron.

La sección donde se muestran las gráficas de cada TF se encuentran ubicadas en (ver Figura 4.7). Esta sección se va a estar actualizando conforme a cada TF que seleccionemos de la lista. En la figura (ver Figura 4.8) se muestra la ecuación que utilizó [Valdez] en su modelo matemático para cada TF que se seleccione de la lista. Además en la figura (ver Figura 4.8) existe un botón, el cual al oprimirlo podemos observar en otra figura (ver Figura 4.10) todas las interacciones con las que cuenta la formación del flagelo la bacteria *Escherichia coli*.

Finalmente se puede apreciar que cuando seleccionemos cada TF de la lista, en la figura (ver Figura 4.9) se puede obtener un subgrafo de las ecuaciones que [Valdez] propone para su modelo, en donde se visualizan las interacciones de los TFs. Ya sea que se comporte cada interacción como una inhibición, activación o forma dual.

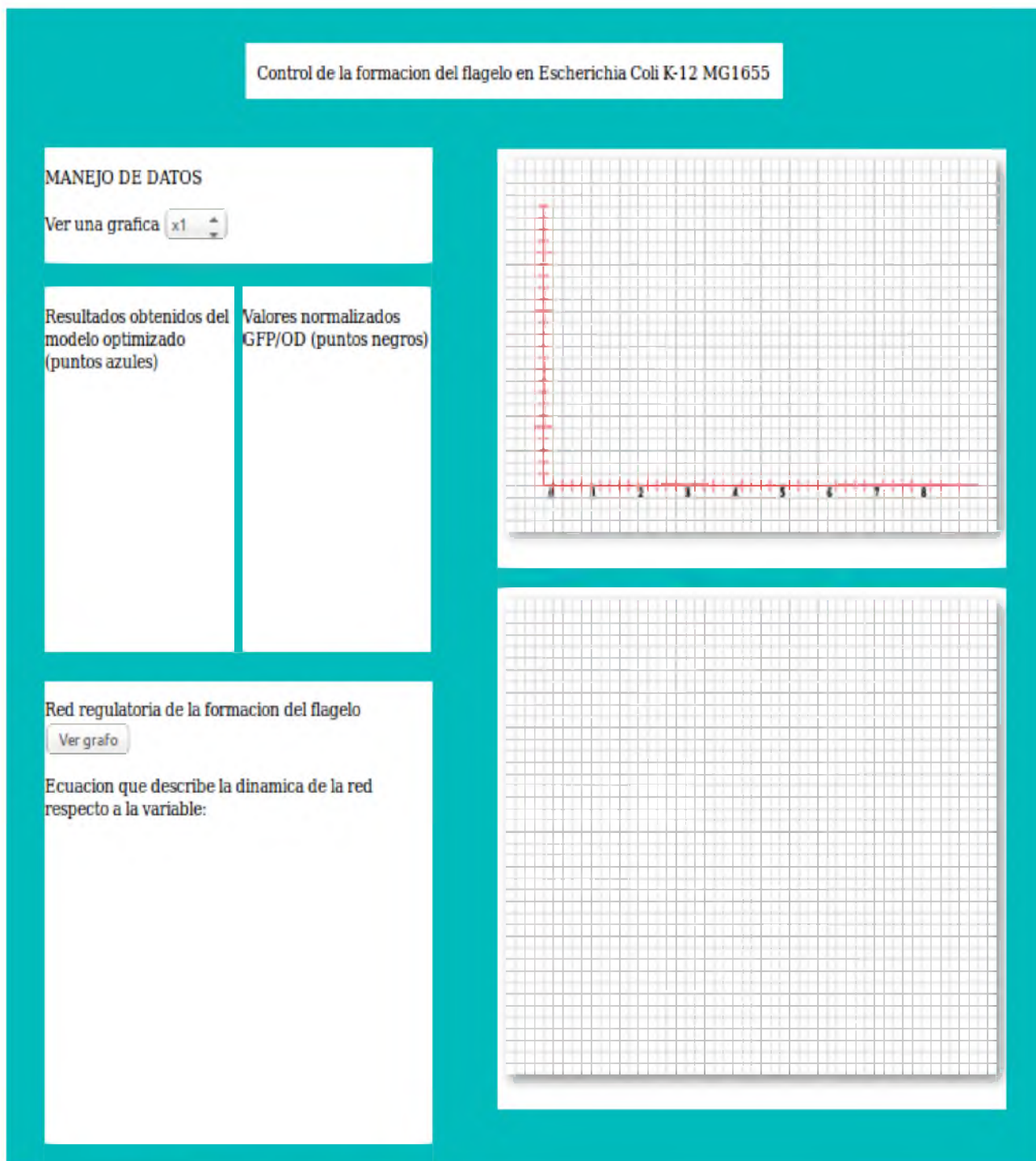


Figura 4.5: Diseño del modelo gráfico del flagelo de la bacteria Escherichia coli.

MANEJO DE DATOS	
Ver una grafica <input type="button" value="x3"/>	
Resultados obtenidos del modelo optimizado (puntos azules)	Valores normalizados GFP/OD (puntos negros)
x3	x3
0.000000000000	0.000000000000
0.000000000000	0.066480082482
0.000000000000	0.124448786483
0.000000000000	0.174995815281
0.000000000000	0.219071358528
0.000000000000	0.257503954049
0.000000000000	0.291016062798
0.000000000000	0.320237649776
0.000000000000	0.345718026195
0.000000000000	

Figura 4.6: Datos pertenecientes a los resultados de laboratorio y los resultados de [Valdez].

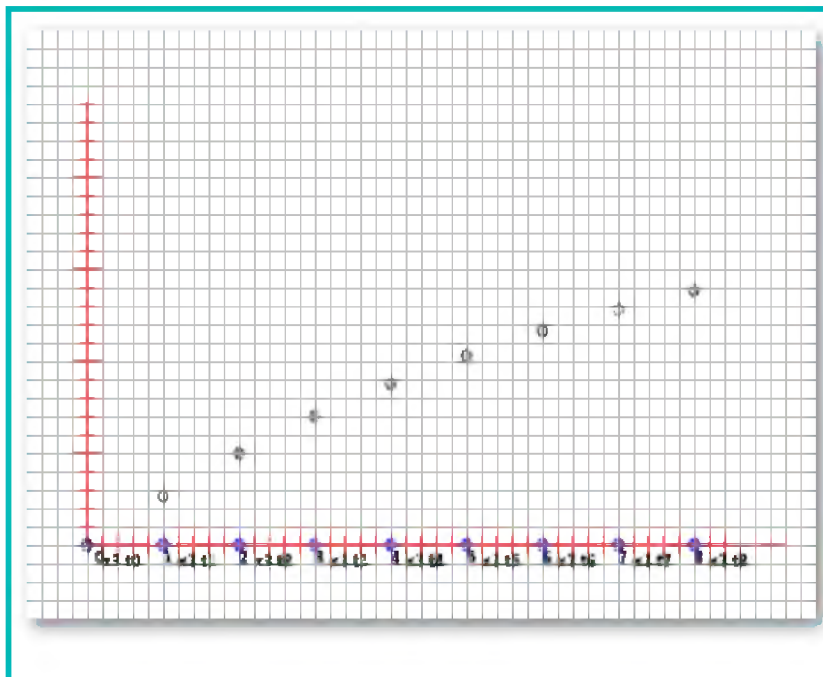


Figura 4.7: Gráfica perteneciente a la variable aleatoria seleccionada.

Red regulatoria de la formación del flagelo

[Ver grafo](#)

Ecuación que describe la dinamica de la red respecto a la variable:

x_3

$$x_3(t+1) = a_3 x_3(t) + (1-a_3) [k_4,3H(x_4(t)-T_{4,3}) + k_1,3H(-(x_1(t)-T_{1,3})) + k_2,3H(-(x_2(t)-T_{2,3})) + k_5,3H(-(x_5(t)-T_{5,3})) + k_8,3H(-(x_8(t)-T_{8,3})) + k_7,3H(x_7(t)-T_{7,3})]$$

Figura 4.8: Ecuación obtenida a partir de la variable aleatoria seleccionada de acuerdo al sistema de ecuaciones de [Valdez]

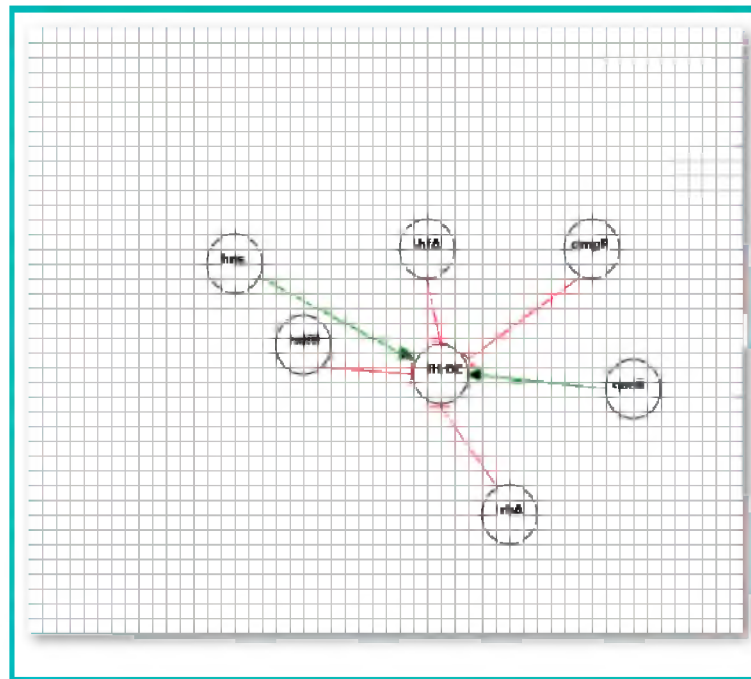


Figura 4.9: Subgrafo obtenido a partir de la ecuación de [Valdez].

Capítulo 5

Conclusiones

Como primer punto para desarrollar nuestro trabajo, nos dimos a la tarea de investigar los modelos gráficos utilizados en la ingeniería inversa de redes regulatorias genéticas, obteniendo con ello las redes booleanas o las redes bayesianas por citar algunos, que fueran aplicados a sistemas biológicos. Nuestro siguiente paso de investigación se basó en pensar que el modelo a emplear, necesitaría utilizar las relaciones de independencia condicional en instantes de tiempo, el aprendizaje y comprobación de éste. La prueba de información mutua en conjunto a las redes bayesianas permiten utilizar la causalidad en el modelo gráfico, adecuando la estructura de la BN estática a la DBN para obtener su aprendizaje. Parte de lo que realizamos fue proponer una estructura dinámica de la formación del flagelo de la bacteria *Escherichia coli*, por medio de una red bayesiana dinámica, del resultado de ella podemos decir que es una reconstrucción relativa del comportamiento del flagelo. Esta red fue elaborada a partir de las observaciones que se tenían de [Valdez] y [Martinez2012]. Encontramos que la relación de causalidad de CRP a fis tiene mayores relaciones de independencia condicional (del t_0 al t_1) entre las variables relacionadas por la inhibición de la red que proponemos con el puntaje de 24.520086.

Al revisar trabajos sobre métodos de aprendizaje de los modelos gráficos para algún sistema biológico nos dimos cuenta que, si ya se contaban con algunas observaciones de los datos denotados por los factores de transcripción y la propuesta de algún modelo, entonces nos era más fácil dar una aproximación de las relaciones de independencia condicional entre ellos para el aprendizaje de la

red bayesiana dinámica. Mediante el razonamiento mencionado, utilizamos el método de conteo de parámetros, ya que fue el que mejor se adecuaba para aprender de las DBNs.

En las pruebas realizadas sobre la toolbox en Matlab para la prueba de información mutua, utilizando los datos de [Martinez2012], encontramos que a medida que la cantidad de observaciones aumenta, es posible obtener mejores resultados e incluso esto nos ayudaría a aumentar el número de tiempos en la DBN y con ello podríamos conocer mejor las relaciones de independencia entre cada variable utilizada en el modelo. Esto abre la posibilidad de utilizar la toolbox propuesta para futuras investigaciones que contengan mayores observaciones de laboratorio.

Un punto importante a tomar en cuenta es la forma de discretizar los datos continuos con los que contamos, ya que estos pueden modificar la salida de la matriz de adyacencia. Recurrimos a utilizar la mediana debido a que puede ser calculada aunque sólo se tuviera una observación por cada factor de transcripción. La búsqueda de otros métodos de discretización puede abrir el paso hacia nuevas investigaciones.

Para la representación gráfica de la dinámica del flagelo se usó el lenguaje HTML, donde la comparación que se observa de las gráficas se debe a las ecuaciones que propone el autor [Valdez] y a los datos de laboratorio. En éste apartado se puede visualizar las interacciones de los factores de transcripción respecto a sus tipos de comportamientos ya sea como activación, inhibición o dual. Sólo nos queda decir que es una propuesta de un modelo gráfico planteado para una página web y que puede llegar a ser mejorado con datos dinámicos e incluso con series de tiempo en otro trabajo de investigación.

Bibliografía

[BIO131] **Bio131**. <http://www.biomedcentral.com/1471-2105/13/131>

[Campos2006] De Campos Luis M. **A scoring function for learning bayesian networks based on mutual information and conditional independence tests**. Machine Learning. Vol.7.

[Carvalho2009] Carvalho A. 2009. **Scoring functions for learning Bayesian networks**.

[Chickering1996] Chickering D.M. 1996. **Learning Bayesian networks is NP-complete**. In **Learning form Data**. Artificial Intelligence Statistics. Springer Verlag.

[Chilcott2000] Chilcott G.S and Hughes K.T. **Coupling of Flagellar Gene Expression to Flagellar Assembly in Salmonella enterica Serovar Typhimurium and Escherichia coli**. Microbiology and molecular biology reviews 2000, 64(4):694-708.

[Cho2007] Cho K. et al. 2007. **Reverse engineering of gene regulatory networks**. IET Syst. Biol. Pág. 149–157.

[CODEMIT] **Codemit**. <http://code.google.com/p/globalmit>.

[Cooper1990] Cooper G.F. 1990. **The computational complexity of a probabilistic inference using bayesian belief networks**. Artificial Intelligence. Vol.42: 393-405.

[CSBN] **Csbn**. <http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html>

[Cubillos2007] Cubillos Varela A. 2007. **Guía de MATLAB para Ing. Mecánica**. Universidad de Ibagué.

[Diez2012] Díez F.J. 2012. **Introducción a los Modelos Gráficos Probabilistas**. Departamento de Inteligencia Artificial.

[ECDC] **Ecdc.** http://ecdc.europa.eu/en/healthtopics/escherichia_coli

[EPA] **Epa.** <http://epa.gov/oppt/biotech/pubs/fra/fra004.htm>

[Gardner2005] Gardner and Faith. 2005. **Reverse-engineering transcription control networks.** Department of Biomedical Engineering, Boston University. Pág: 66-82.

[Heckerman1995] Heckerman et al. 1995. **Learning Bayesian networks : the combinations knowledge and statical data.** Machine learning. Vol. 20.

[Hidde2000] Hidde de Jong. 2000. **Modeling and Simulation of Genetic Regulatory.** Institute National de Recherche en informatique et en automatique.

[IBIS] **Ibis.** <http://ibis.tau.ac.il/perluser/2013/lessons/Ecoli.ptt>

[Ideker2001] Ideker T.et al. 2001. **A new approach to decoding life: systems biology.** Vol.2:343-372

[Martinez2011] Martínez A. 2011. **Escherichia coli transcriptional regulatory network.** Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. Feb-May. Pág. 21-31.

[Martinez2012] Martinez et al. 2012. **Hierarchical dynamics of a transcription factors network in E. coli.** MolecularBioSystems.

[Mihajlovic] Mihajlovic V. and Petkovic M. **Dynamic Bayesian Networks: A State of the Art.** Computer Science Departament. University of Twente The Netherlands.

[MICRO] **Micro:** http://www.biologia.edu.ar/microgeneral/micro-ianez/11_micro.htm

[Murphy1999] Murphy K. y Mian S. 1999. **Modelling Gene Expression Data using Dynamic Bayesian Networks.**

[Neapolitan2009] Neapolitan R. 2009. **Probabilistic Methods for Bioinformatics.** Bioinformatics. Ed. Elsevier. Pág. 79-102.

[Nguyen2001] Vinh Nguyen.2011. **The GlobalMIT Toolkit for Learning Optimal Dynamic Bayesian Network User Guide.** Monash University, Victoria, Australia.

- [Rau2011] Rau A. 2011. **Reverse Engineering Gene Regulatory Networks Using Approximate Bayesian Computation**. Institut National de la Recherche Agronomique. Pág. 1-4.
- [Robinson1977] Robinson. 1977. **Counting unlabeled acyclic digraphs. Lecture notes in mathematics 622: Combinatorial mathematics V**. New York.
- [SMB] **Smb**. <http://www.seresmodelicos.csic.es/bacteri.html>
- [Sucar2006] Sucar L. **Redes bayesianas**. Aprendizaje Automático: conceptos básicos y avanzados, Pearson Educación. INAOE. Pág. 1-25.
- [TCONT] **Tcont**. <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/GuiaSPSS/12contin.pdf>
- [Tiessen2012] Tiessen A. 2012. **Fundamentos de mejoramiento genético vegetal. Conceptos básicos de genética, biología molecular, bioquímica y fisiología vegetal**. Ed. EAE. Pág.191-193.
- [Valdez] Valdez et al. **Parameter fitting of a discrete time model of the regulatory network which controls flagella formation in Escherichia coli K-12 MG1655**. Pág.14.
- [Valenzuela2008] Valenzuela R.M. 2008. **Redes Bayesianas**. Centro de Sistemas Inteligentes Tecnológico de Monterrey, Campus Monterrey.
- [Vega2011] Vega and Van Der Henst. 2011. **Guía HTML5. El presente de la web. HTML5, css3 y javascript**.
- [Xuan2012] Xuan N. et al. 2012. **Gene regulatory network modeling via global optimization of high-order dynamic Bayesian network**.