Facilitando la interpretación de la inferencia en redes Bayesianas

Trabajo Fin de Máster presentado por

Marta Alonso Tubía

Universidad Politécnica de Madrid.
Máster en Ciencia de Datos

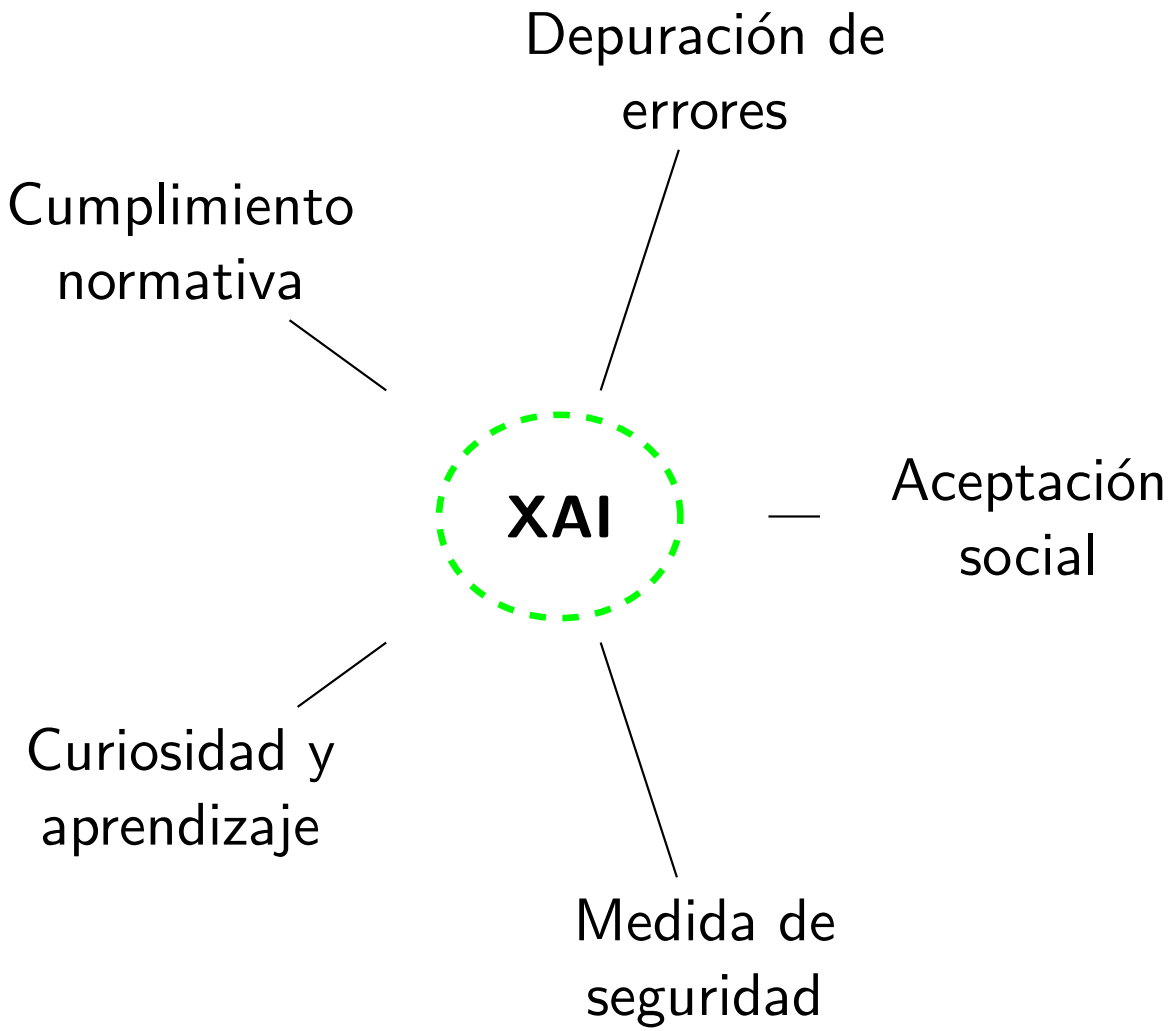
20 Julio 2023

Tutores: Concha Bielza & Pedro Larrañaga

Índice

Índice

Motivación


Depuración de errores
Cumplimiento normativa
XAI
Aceptación social
Curiosidad y aprendizaje
Medida de seguridad

Definición

Una red Bayesiana (RB), es una tupla $\mathcal{B} = (\mathcal{G}, \theta)$*, donde* $\mathcal{G} = (V, A)$ *es un grafo dirigido y acíclico con un conjunto de nodos* $V = \{X_1, \ldots, X_n\}$ *y un conjunto de arcos* $A \subseteq V \times V$*. El conjunto* $\theta = \{P(x_i|\mathbf{X}_{Pa_{X_i}})\}$ *define una CPD para cada nodo del grafo. Una RB representa una distribución conjunta de probabilidad* $P(\mathbf{x})$ *del vector aleatorio* $\mathbf{X} = (X_1, ..., X_n)$

$$P(\mathbf{x}) = \prod_{i=1}^{n} P(x_i|\mathbf{X}_{Pa_{X_i}}).$$

Adecuadas para interpretabilidad:

- Representa visualmente las dependencias entre variables.
- Modela la incertidumbre usando probabilidades.
- Permite aprendizaje incremental.
- Permite incorporar conocimientos previos.

Aplicación

Inferencia
- Probabilidad marginal: $P(\mathbf{Y})$
- Probabilidad condicionada: $P(\mathbf{Y}|\mathbf{e})$
- Inferencia abductiva: $\mathbf{y} = argmax_{\mathbf{Y}} P(\mathbf{Y}|\mathbf{e}), \quad \mathbf{Y} \subseteq \mathbf{X}$
- Explicación más relevante (MRE)

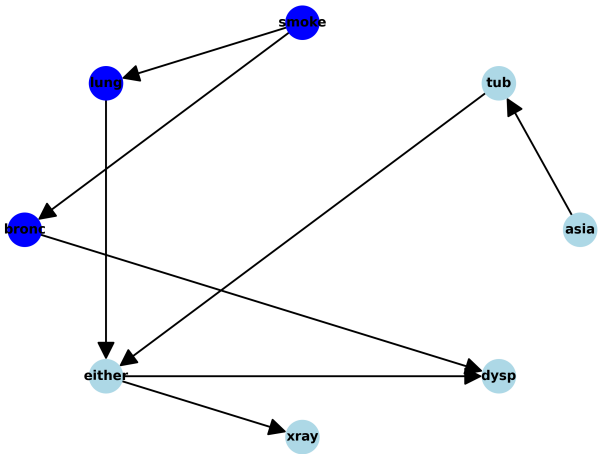
Índice

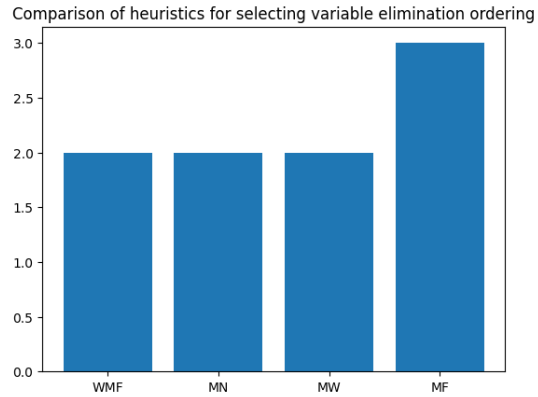
Algoritmos exactos

- Eliminación de variables y Árbol de cliqués.
- **Objetivo**: generar visualizaciones que ilustren pasos intermedios y justifiquen decisiones del razonamiento.
- Librerías: networkX y matplotlib.

Eliminación de variables

- Eliminación sistemática y ordenada de variables en una lista de factores.
- Es posible una optimización inicial: ***pruning***.
- Ordenación clave en la eficiencia del procedimiento.
  - → Comparación (heurísticas) usando *induced-width*.

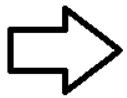




Variable a eliminar Earthquake
Factores que participan

Alarm	Burglary	Earthquake	phi(Alarm,Burglary,Earthquake)
Alarm(0)	Burglary(0)	Earthquake(0)	0.9990
Alarm(0)	Burglary(0)	Earthquake(1)	0.7100
Alarm(0)	Burglary(1)	Earthquake(0)	0.0600
Alarm(0)	Burglary(1)	Earthquake(1)	0.0500
Alarm(1)	Burglary(0)	Earthquake(0)	0.0010
Alarm(1)	Burglary(0)	Earthquake(1)	0.2900
Alarm(1)	Burglary(1)	Earthquake(0)	0.9400
Alarm(1)	Burglary(1)	Earthquake(1)	0.9500

Earthquake	phi(Earthquake)
Earthquake(0)	0.9980
Earthquake(1)	0.0020

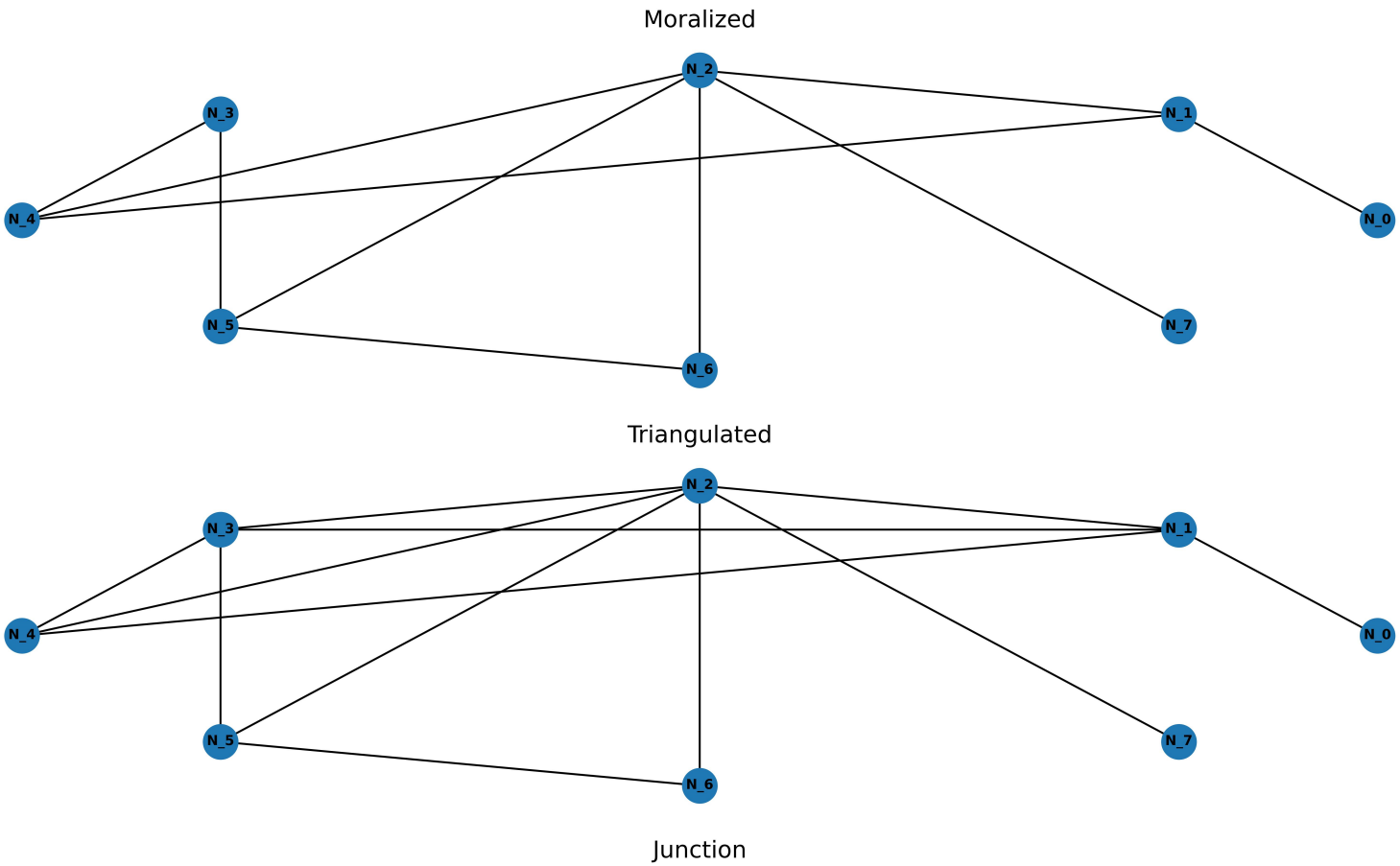


Phi

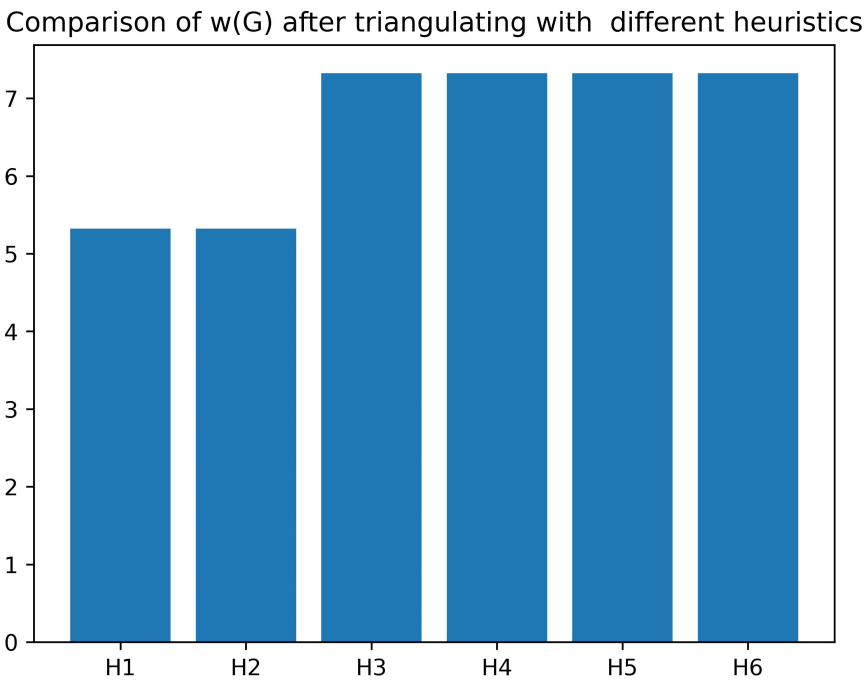
Burglary	Alarm	phi(Burglary,Alarm)
Burglary(0)	Alarm(0)	0.9984
Burglary(0)	Alarm(1)	0.0016
Burglary(1)	Alarm(0)	0.0600
Burglary(1)	Alarm(1)	0.9400

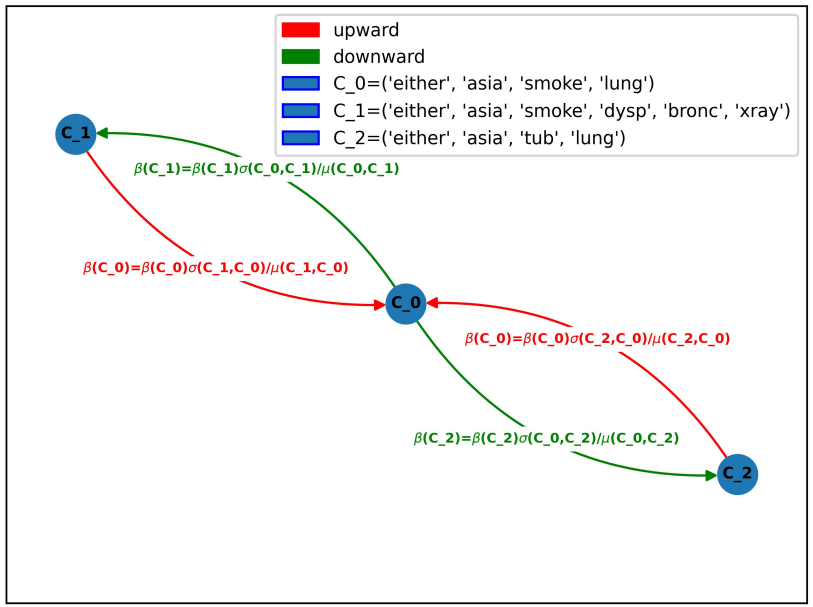
Árbol de cliqués

Construcción del árbol de cliqués



Coste de triangulación y message passing





$$w(\mathcal{G}) = log_2 \sum_C \prod_{v_i \in C} n_i$$

Algoritmos aproximados: Muestreo de Gibbs

Repetir muestreo hasta que la distribución conjunta esté más cerca de la distribución posterior $P(\mathbf{X}|\mathbf{e})$.

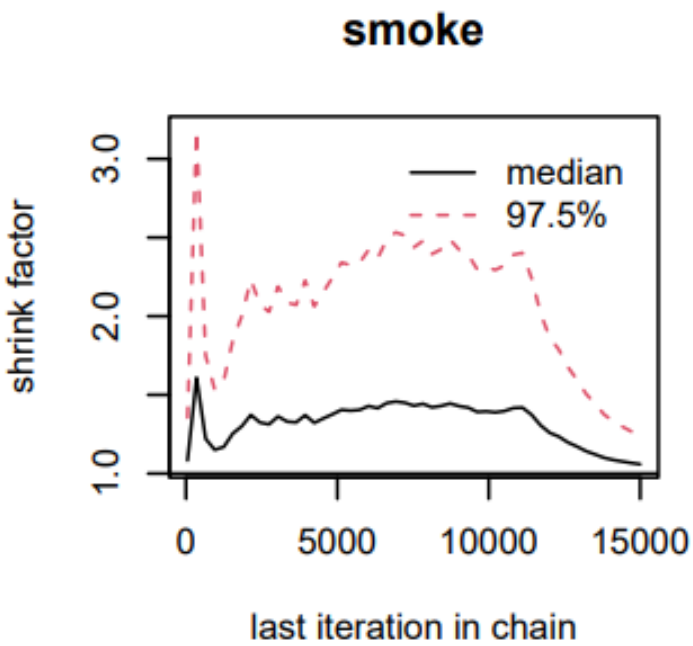
¿CONVERGENCIA?

1. Opciones para la convergencia del muestreo de Gibbs $\rightarrow$ métodos estadísticos de diagnóstico de convergencia.
2. Limitaciones de librerías que hacen inferencia aproximada (*bnlearn*, *pgmpy*, *pomegranate*...).
   $\rightarrow$ extender *pgmpy* con otra librería especializada en test de diagnóstico de convergencia de cadenas de Markov (*coda*(R)).
3. Comunicar python-R y transformar los objetos de una a otra librería. $\rightarrow$ *rpy2*. Implementar.

Gelman-Rubin

	Point est
asia	1.00
tub	1.32
either	1.85
dysp	1.04
smoke	1.03
bronc	1.00
lung	1.73
xray	1.55

Multivariate psrf: 1.44

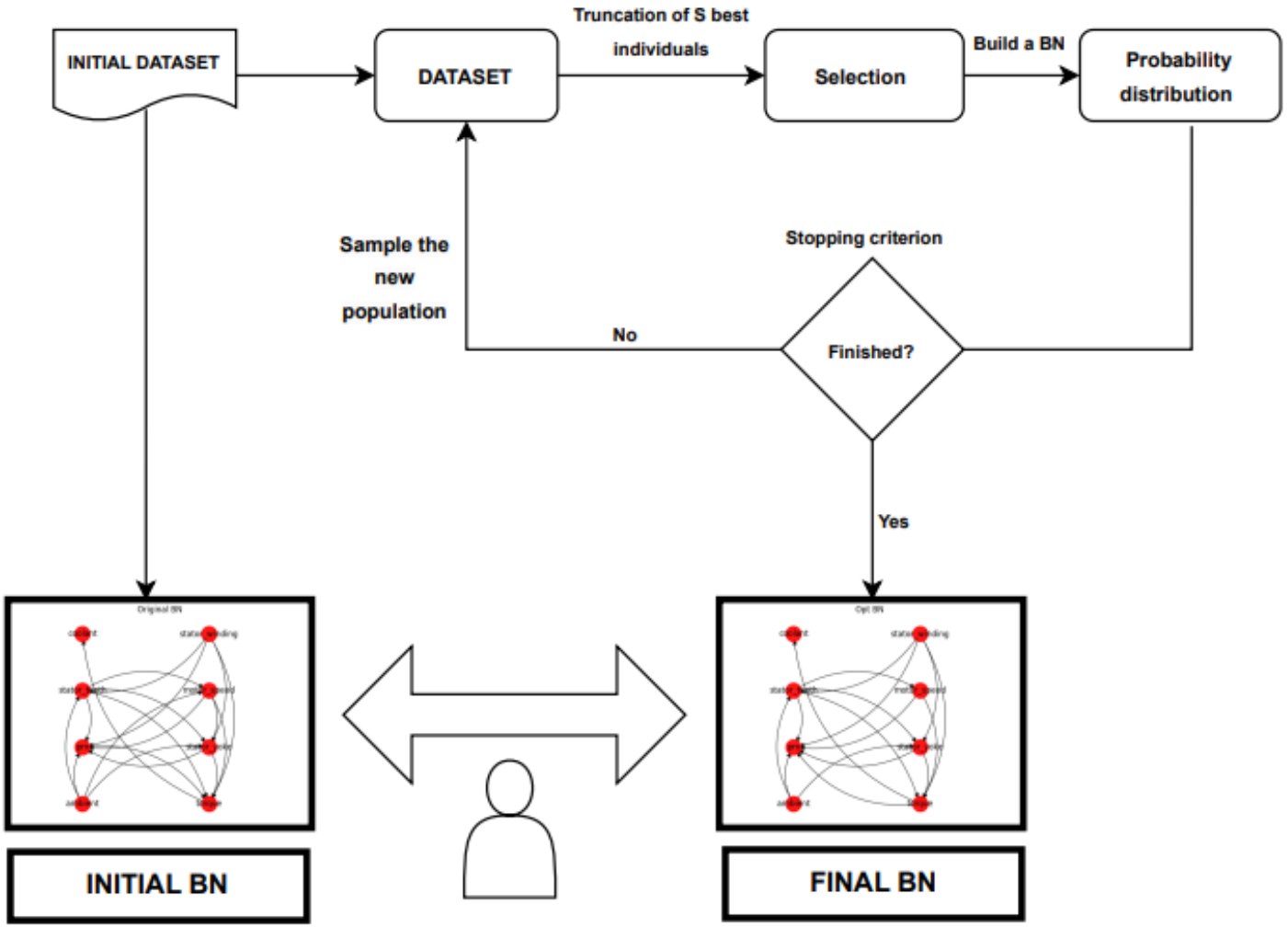


Índice

Propuesta

1. Identificar propuestas existentes.
2. Identificar preguntas interesantes que quedan aún por responder y pueden añadir valor al conocimiento del usuario.
3. **Objetivo**: Encontrar una realidad hipotética modelada por una RB en la que la situación más probable tenga una probabilidad mayor que un umbral inferior fijado por el usuario.
   Scope: dataset Gaussiano multivariante.
4. Definir una metodología basada en el uso de EDAs y el análisis de las diferencias (cualitativas y cuantitativas) entre la RB original y la RB obtenida tras ejecutar el EDA.
   Librería: **EDAspy**. Implementación: extensión de **EGNA**.

Metodología



Implementación

- **Procedimiento**: minimizar la función de coste C_l, $l = 1, 2, 3 \ldots$

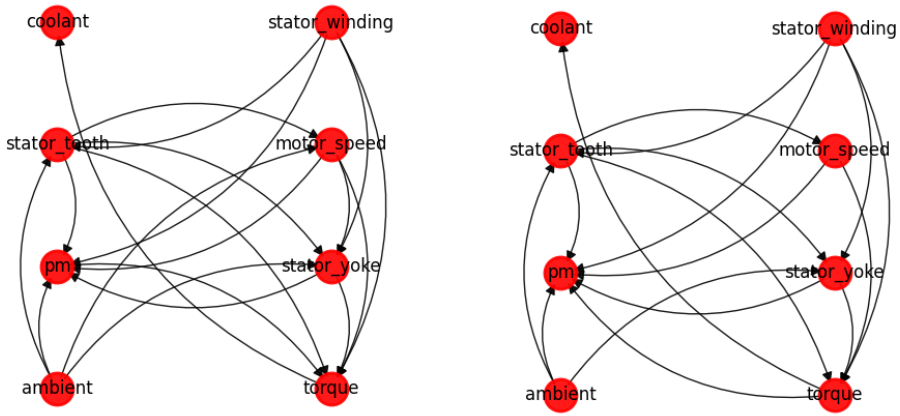
$$C_l(\mathbf{x}) = \begin{cases} log(f_l(\boldsymbol{\mu}_l)) - log(f_l(\mathbf{x})) & \|\tilde{\mathbf{e}} - \mathbf{e}\|_\infty \leq \epsilon \\ 99999 & \text{otro caso} \end{cases}$$

donde $\mathbf{x} = (\mathbf{y}, \tilde{\mathbf{e}})$, con criterio de parada:

$$\text{death iter} \geq N_{iter} \quad \text{or} \quad \frac{f_l(\boldsymbol{\mu}_l)}{f_0(\boldsymbol{\mu}_0)} \geq \alpha > 1$$

Resultados (dataset motor eléctrico)

Objetivo: mantenimiento del motor más eficiente.



- Incremento en la función de densidad 1.638
- Las dependencias directas velocidad motor - temperatura estator y velocidad motor - temperatura ambiente, desaparecen.
- La temperatura del motor decrece su correlación notablemente con la temperatura de diferentes partes del estator.

Índice

Explicación más relevante (MRE)

- Una buena explicación es **concisa** y **precisa** $\rightarrow$ necesitamos una métrica.

$$GBF(\mathbf{x};\mathbf{e}) \equiv \frac{P(\mathbf{e}|\mathbf{x})}{P(\mathbf{e}|\bar{\mathbf{x}})}$$

El GBF presenta características teóricas que permiten identificar automáticamente la explicación más relevante.

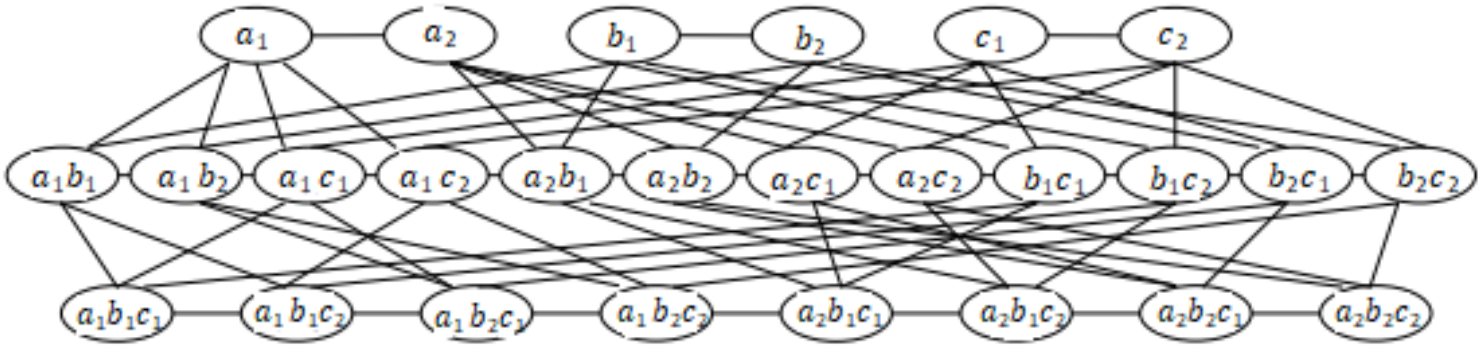
Definición

Sea $\mathbf{M}$ *un conjunto de variables target, y* $\mathbf{e}$ *la evidencia parcial en el resto de variables en una RB. La explicación más relevante es el problema de encontrar una explicación* $\mathbf{x}$ *para* $\mathbf{e}$ *de forma que tenga el máximo factor generalizado de Bayes GBF* $(\mathbf{x};\mathbf{e})$*, i.e.,*

$$MRE(\mathbf{M};\mathbf{e}) \equiv argmax_{\mathbf{x},\emptyset\subset\mathbf{X}\subseteq\mathbf{M}} \quad GBF(\mathbf{x};\mathbf{e}).$$

Implementación y k-MRE

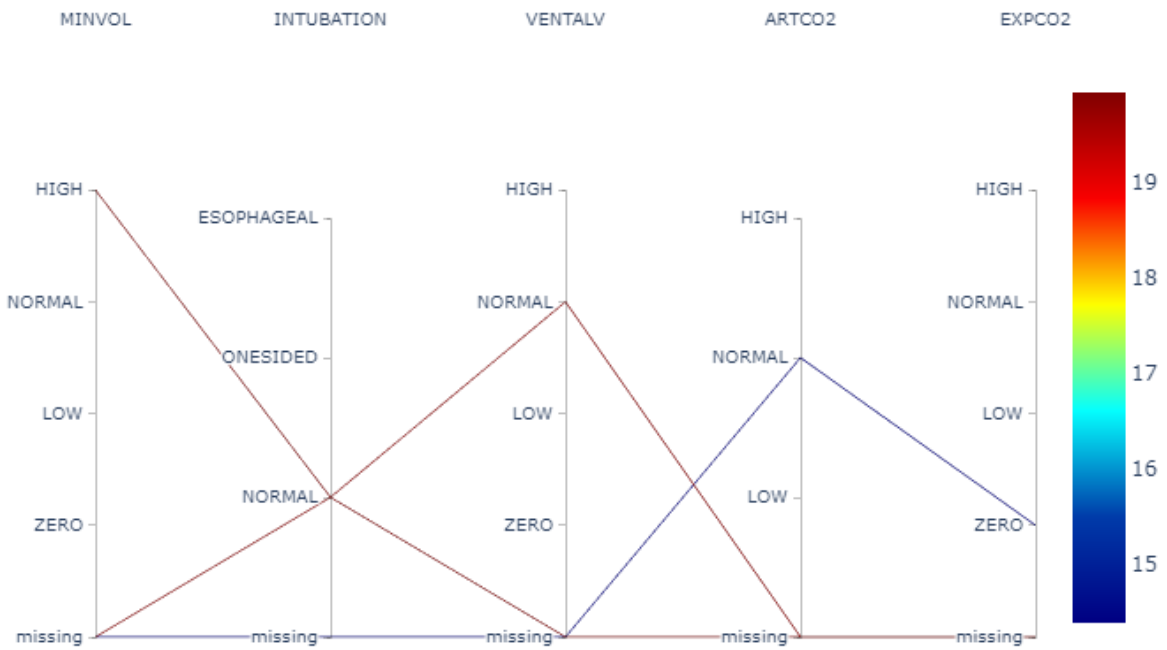
- **Implementación**: Forward search



- **k-MRE**: poder escoger entre posibles explicaciones y comparar la calidad de las mismas. → Diversidad y GBF

Resultados (dataset ALARM)

Evidencia: baja ventilación pulmonar.



MINVOL	INTUBATION	VENTALV	ARTCO2	EXPCO2	GBF	Hamming distance
HIGH	NORMAL	*missing*	*missing*	*missing*	19.93	0 (MRE)
missing	NORMAL	NORMAL	*missing*	*missing*	19.79	2
missing	*missing*	*missing*	NORMAL	ZERO	14.39	4

Índice

Conclusiones

• Visualizaciones de los diferentes pasos de los algoritmos de inferencia exacta pueden ayudar en la toma de decisiones:

1. Arrojando luz sobre cómo se hacen los cálculos en cada etapa.
2. Detectando resultados inesperados que pueden indicar problemas en el proceso de inferencia $\rightarrow$ Validación de resultados.
3. Las representaciones visuales son más accesibles que las descripciones puramente matemáticas.

• **Inferencia aproximada**: herramientas de diagnóstico de convergencia para el algoritmo de sampleo de Gibbs:

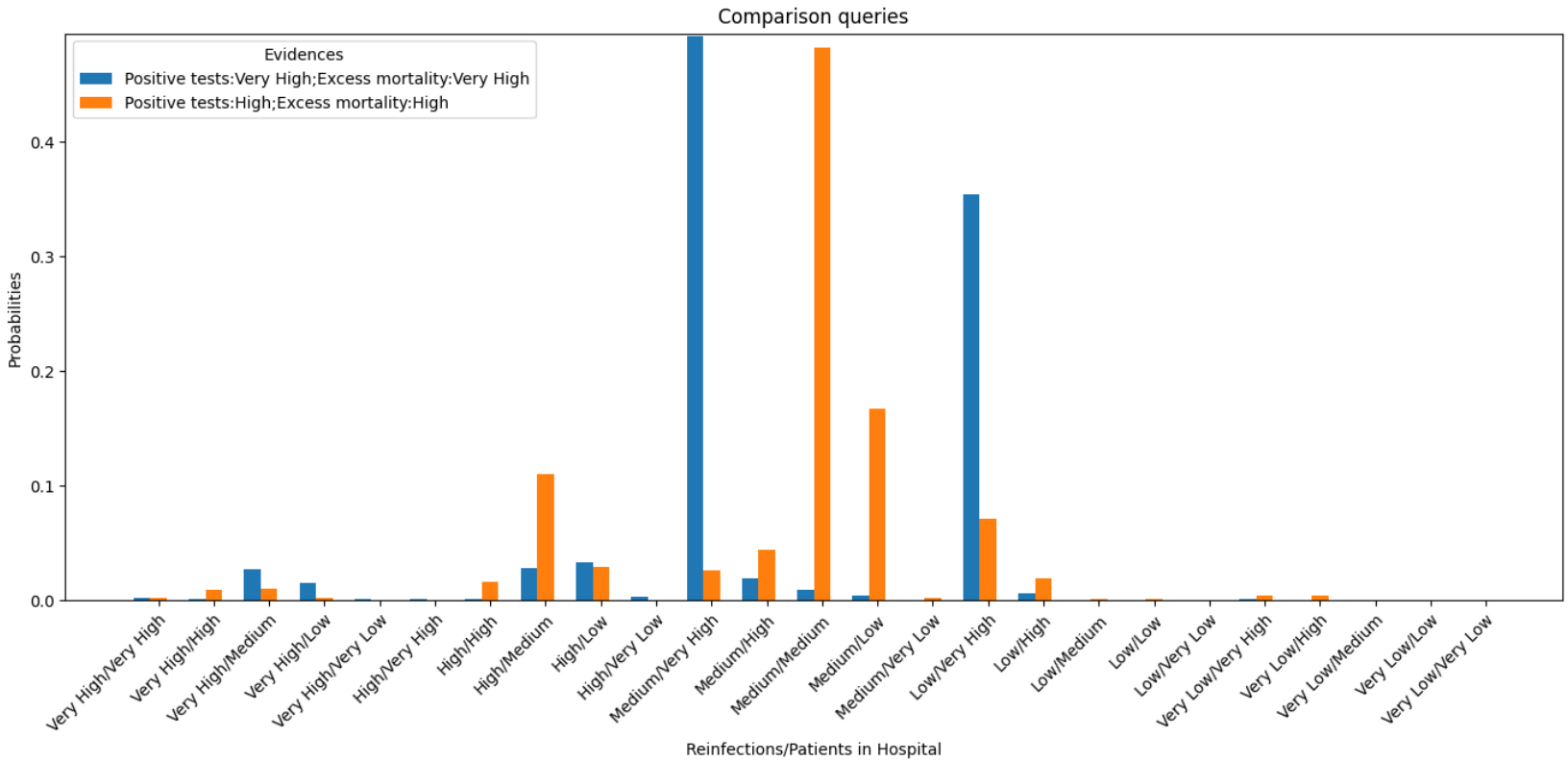
1. Si el algoritmo no ha convergido, los resultados pueden no reflejar la verdadera distribución posterior, disminuyendo la precisión de los resultados.
2. Visualizar la evolución del factor de convergencia puede ayudar a **mejorar la eficiencia**: aproximar el número de muestras necesarias para converger.

• **Inferencia abductiva** $\rightarrow$ *¿cómo sería la red bayesiana de una realidad hipotética en la que tenemos una mayor confianza sobre la situación más probable?*

• **MRE** $\rightarrow$ Algoritmo aproximado de búsqueda local + comparación de explicaciones con representación de coordenadas paralelas y distancia de hamming.

Líneas futuras

- Aumentar el scope de aplicación de nuestra metodología de inferencia abductiva a dataset híbridos o dataset categóricos sin ninguna restricción.
- Generalización de GBF, definida para variables discretas
  No se conocen librerías de diseño extensible que soporten dataset continuos.
- Poder comparar consultas de probabilidad de varias evidencias.


Comparison queries
Evidences
Positive tests:Very High;Excess mortality:Very High
Positive tests:High;Excess mortality:High
Probabilities
0.0
0.1
0.2
0.3
0.4
Very High/Very High
Very High/High
Very High/Medium
Very High/Low
Very High/Very Low
High/Very High
High/High
High/Medium
High/Low
High/Very Low
Medium/Very High
Medium/High
Medium/Medium
Medium/Low
Medium/Very Low
Low/Very High
Low/High
Low/Medium
Low/Low
Low/Very Low
Very Low/Very High
Very Low/High
Very Low/Medium
Very Low/Low
Very Low/Very Low
Reinfections/Patients in Hospital