Práctica 4

Redes Bayesianas

Inteligencia Artificial

3º curso, Grado de Ingeniería en Informática

9 de diciembre de 2021

Leticia Sánchez Romero, 760549

Índice de contenidos

1. Estudio previo - ejercicio 1

PRÁCTICA 4: Redes Bayesianas

Ejercicio 1

a)

$\boxed{1.}$ $I \perp H \mid S$?

2 caminos:

- $\overbrace{\underbrace{I\ C}\ \underbrace{P\ H}}^{\text{activo}}$ → activo → no independientes (ya que todos los tripletes del camino son activos)
- $I\ C\ S\ H$ (activo, activo, inactivo) → inactivo → independientes

→ como hay un camino activo no son independientes

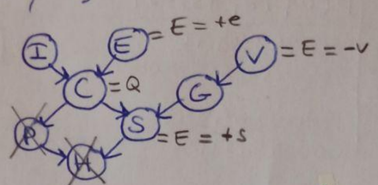
$\boxed{2}$ $V \perp H \mid S$?

2 caminos:

- $V\ G\ S\ H$ (activo, inactivo) → inactivo → independientes
- $V\ G\ S\ C\ P\ H$ (activo, activo, activo, activo) → activo → no independientes

→ como hay un camino activo no son independientes

b) $P(C \mid +e, +s, -v)$ → $Q = C$; $E_1 = +e$; $E_2 = +s$; $E_3 = -v$



Algoritmo de eliminación de variables

$\boxed{1.}$ Ignorar variables que no sean ancestro de Q o E.

- se ignoran P y H

$\boxed{2}$ Instanciar las tablas de probabilidad con la evidencia

I	P(i)
+i	0'10
−i	0'90

E	P(+E)
+e	0'05

I E C	P(C|i, +e)
+i +e +c	0'01
−i +e +c	0'20
+i +e −c	0'99
−i +e −c	0'80

C G S	P(+S|c, g)
+c +g +s	0'90
+c −g +s	0'60
−c +g +s	0'80
−c −g +s	0'02

V	P(−V)
−v	0'80

V G	P(G|−v)
−v +g	0'08
−v −g	0'92

$\boxed{3.}$ Eliminar para cada variable H (ni $\underset{\text{query}}{Q}$ ni $\underset{\text{Evidencia}}{E}$) → las variables H son: I y G

- Eliminar variable I (paso 1: unión factores ; paso 2: marginalización)

I	P(i)
+i	0'10
−i	0'90

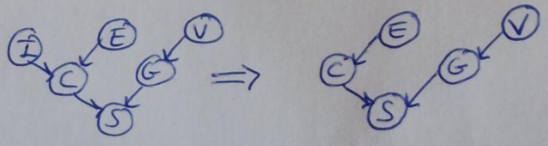
I E C	P(C|i, +e)
+i +e +c	0'01
−i +e +c	0'20
+i +e −c	0'99
−i +e −c	0'80

unión ⟹

E I C		P(C, I|+e)
+e +i +c	0'1 · 0'01	0'001
+e −i +c	0'9 · 0'2	0'18
+e +i −c	0'1 · 0'99	0'099
+e −i −c	0'9 · 0'8	0'72

marginalización ⟹

E C	P(C|+e)
+e +c	0'181
+e −c	0'819



3. continuación.

- Eliminar variable G

V	G	P(G|−v)
−v	+g	0'08
−v	−g	0'92

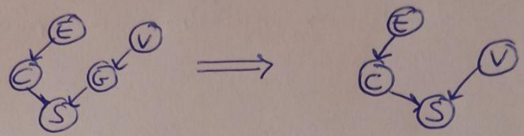
C	G	S	P(+s|c,g)
+c	+g	+s	0'90
+c	−g	+s	0'60
−c	+g	+s	0'80
−c	−g	+s	0'02

$\xRightarrow{\text{UNIÓN}}$

	C	G	S	P(+s,g|c)
0'08·0'9	+c	+g	+s	0'072
0'92·0'6	+c	−g	+s	0'552
0'08·0'8	−c	+g	+s	0'064
0'92·0'02	−c	−g	+s	0'0184

$\xRightarrow{\text{MARGINALIZACIÓN}}$

V	C	S	P(+s|c,−v)
−v	+c	+s	0'624
−v	−c	+s	0'0824



4. Juntar todos los factores restantes y normalizar

+e	c	+s	−v	P(+e,c,+s,−v)
+e	+c	+s	−v	0'05 · 0'181 · 0'624 · 0'8 = 0'00451776
+e	−c	+s	−v	0'05 · 0'819 · 0'0824 · 0'8 = 0'002699424

⇓ NORMALIZAR

c	P(c|+e,+s,−v)
+c	$\frac{0'00451776}{(0'00451776+0'002699424)} = \boxed{0'625}$ ← Probabilidad de que tenga covid.
−c	$\frac{0'002699424}{(0'00451776+0'002699424)} = 0'375$

2. Estudio previo - ejercicio 2

Ejercicio 2

a) C ≡ carrera

G ≡ género

A ≡ admitidos

$$\begin{tikzcd} C \arrow[rr] \arrow[dr] & & G \arrow[dl] \\ & A & \end{tikzcd}$$

- La carrera influye en el género que solicita
- La carrera influye en el número de admitidos (según las solicitudes)
- El género influye en el número de admitidos (según las solicitudes)

b)

C	P(C) = solicitudes carrera / total solicitudes
A	(825+108)/3228 = 0'290
B	(560+25)/3228 = 0'181
C	(325+593/3228 = 0'284
D	(417+375)/3228 = 0'245

(*) Solicitudes totales =
= 825 + 108 + 560 + 25 +
+ 325 + 593 + 417 + 375 =
= 3228

C	G	P(G|C)
A	F	108/(825+108) = 0'116
A	M	825/(825+108) = 0'884
B	F	25/(560+25) = 0'0427
B	M	560/(560+25) = 0'9573
C	F	593/(325+593) = 0'646
C	M	325/(325+593) = 0'354
D	F	375/(417+375) = 0'473
D	M	417/(417+375) = 0'527

C	G	A	P(A|C, G)
A	F	a	89/108 = 0'824
A	F	-a	1-0'824 = 0'176
A	M	a	512/825 = 0'62
A	M	-a	1-0'620 = 0'38
B	F	a	17/25 = 0'68
B	F	-a	1-0'68 = 0'32
B	M	a	353/560 = 0'63
B	M	-a	1-0'63 = 0'37
C	F	a	202/593 = 0'34
C	F	-a	1-0'34 = 0'66
C	M	a	120/325 = 0'369
C	M	-a	1-0'369 = 0'631
D	F	a	131/375 = 0'349
D	F	-a	1-0'349 = 0'651
D	M	a	138/417 = 0'33
D	M	-a	1-0'33 = 0'67

c) Observando la tabla P(A|C, G) se observa que la probabilidad de ser admitido/a en las carreras A, B y D es superior para el género femenino que para el masculino. Por el contrario en la carrera C es superior la probabilidad de ser admitido si se tiene género masculino.

→ Por estas dos razones, SÍ que puede haberse dado una discriminación de género a la hora de realizar las admisiones.

d)

[1.] Ignorar variables

• No se puede ya que C es ancestro de Q y de E.

[2] Instanciar tablas de probabilidad con la evidencia

C	P(C)
A	0'290
B	0'181
C	0'284
D	0'245

C	G	P(f|C)
A	f	0'116
B	f	0'0427
C	f	0'646
D	f	0'473

C	G	A	P(A|C,f)
A	f	a	0'824
A	f	-a	0'176
B	f	a	0'68
B	f	-a	0'32
C	f	a	0'34
C	f	-a	0'66
D	f	a	0'349
D	f	-a	0'651

[3] Eliminar variable H (ni Q ni E) $\longrightarrow$ H = C = Carrera

P(C) ∪ P(f|C) ∪ P(A|C,f)

C	G	A	P(A,C,f)
A	f	a	0'29·0'116·0'824 = 0'0277
A	f	-a	0'29·0'116·0'176 = 0'00592
B	f	a	0'181·0'0427·0'68 = 0'00525
B	f	-a	0'181·0'0427·0'32 = 0'002473
C	f	a	0'284·0'646·0'34 = 0'06237
C	f	-a	0'284·0'646·0'66 = 0'121
D	f	a	0'245·0'473·0'349 = 0'04044
D	f	-a	0'245·0'473·0'651 = 0'0692

marginalización $\Longrightarrow$

G	A	P(A,f)
f	a	0'13576
f	-a	0'198593

[4] Juntar todos los factores restantes y normalizar.

• No hay factores restantes.

• Se normaliza respecto a f para obtener P(A,f)

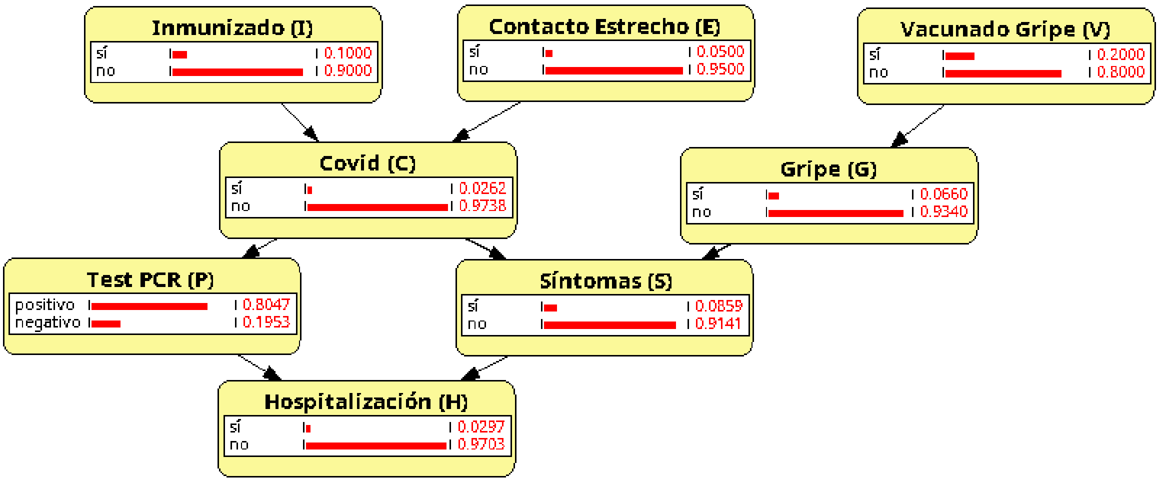
A	P(A|f)
a	$\frac{0'13576}{0'13576+0'198593} = \boxed{0'406}$
-a	$\frac{0'198593}{0'13576+0'198593} = 0'594$

→ probabilidad ser admitida si es mujer.

3. OpenMarkov - ejercicio 1

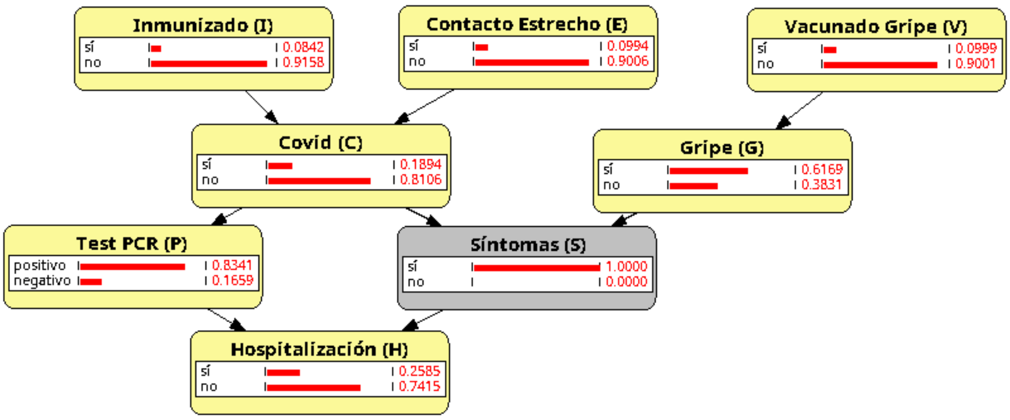
<u>Red Bayesiana para el Covid</u>

La red junto con las probabilidades de sus variables se puede ver a continuación:



<u>Ejercicio 1.a)</u>

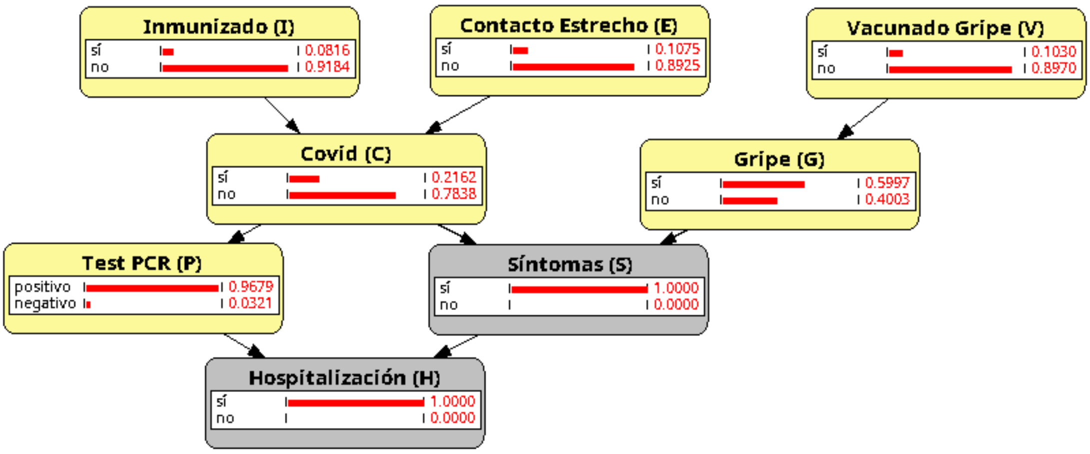
Red Bayesiana con evidencia en Síntomas siendo esta +s añadida:



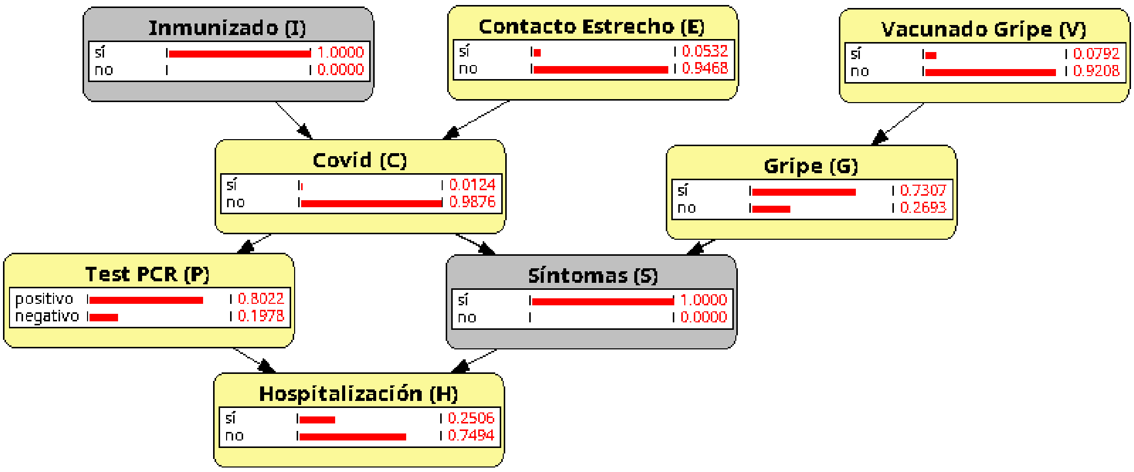
I y H son independientes dado S

X ⊥ Y si ∀ x , y : P (x , y) = P (x) P (y)
P (x | y) = P (x)
P (y | x) = P (y)

Si se calcula P (I | +h, +s) se observa que P (I) cambia su valor respecto a la red que se tiene inicialmente con solo la evidencia de S, por lo que **no son independientes**:

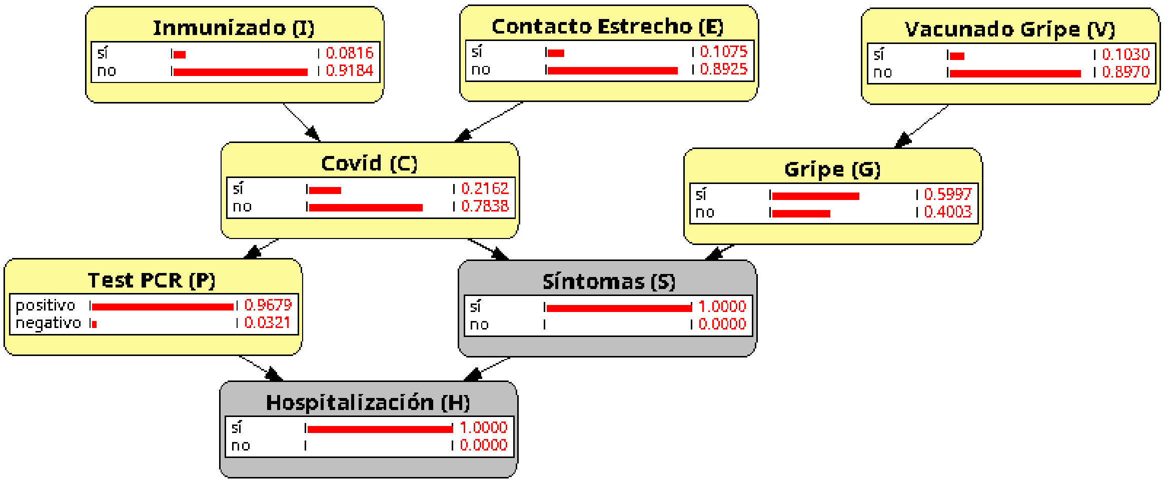


Si se calcula P (H | +i, +s) se observa que P (H) cambia su valor respecto a la red que se tiene inicialmente con solo la evidencia de S, por lo que **no son independientes**:

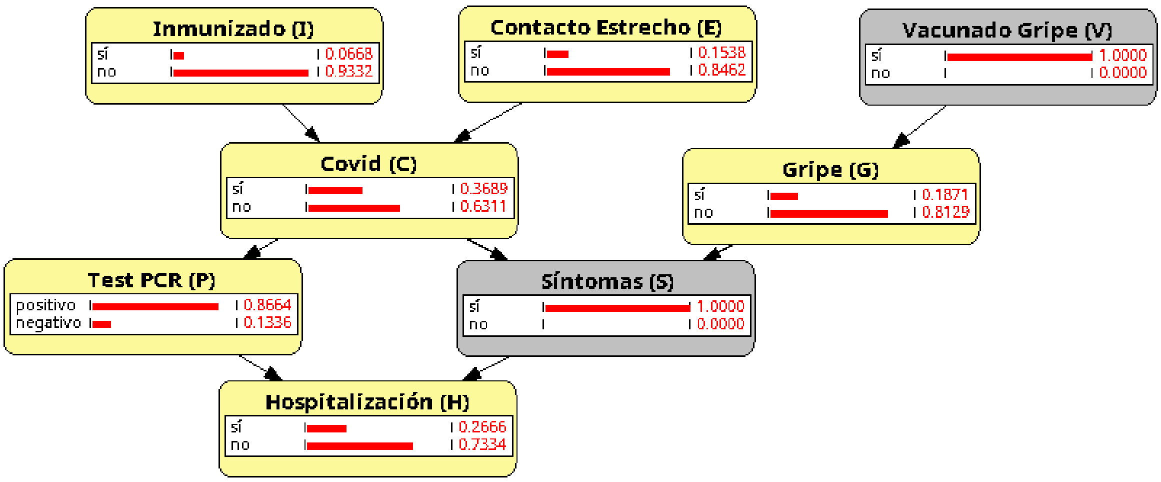


V y H son independientes dado S

Si se calcula P (V | +h, +s) se observa que P (V) cambia su valor respecto a la red que se tiene inicialmente con solo la evidencia de S, por lo que **no son independientes**:

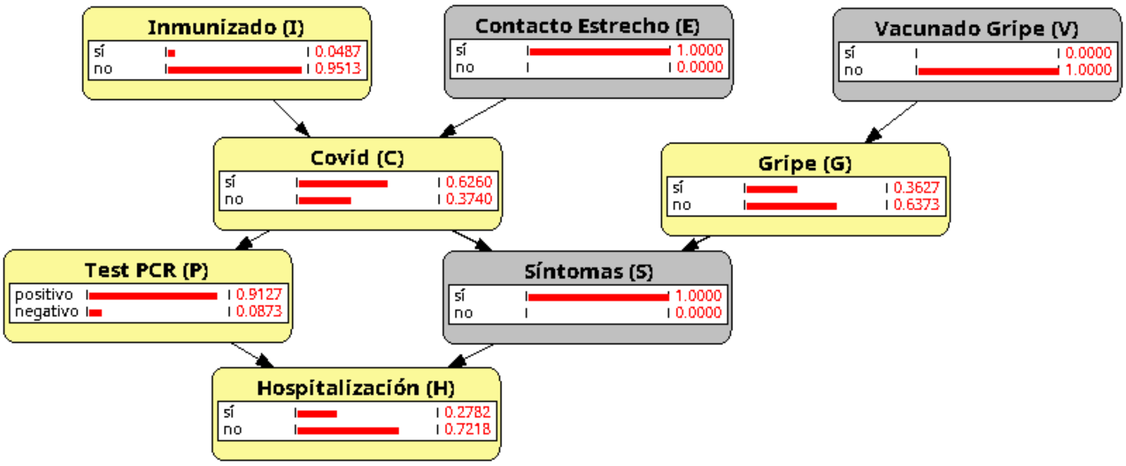


Si se calcula P (H | +v, +s) se observa que P (H) cambia su valor respecto a la red que se tiene inicialmente con solo la evidencia de S, por lo que **no son independientes**:



Ejercicio 1.b)

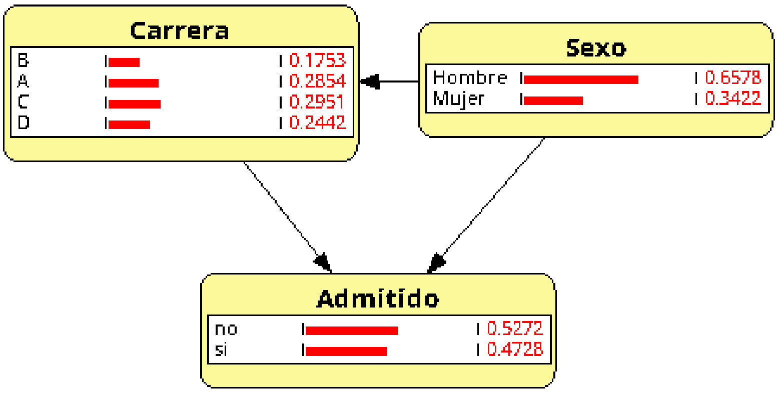
El resultado de la probabilidad pedida en el estudio previo se observa en la imagen de la red, en el nodo Covid, donde tener covid tiene un 0.6260 de probabilidad.
Se ha conseguido añadiendo los hallazgos a las variables evidencia, que son tener contacto estrecho, no estar vacunado y tener síntomas.



4. OpenMarkov - ejercicio 2

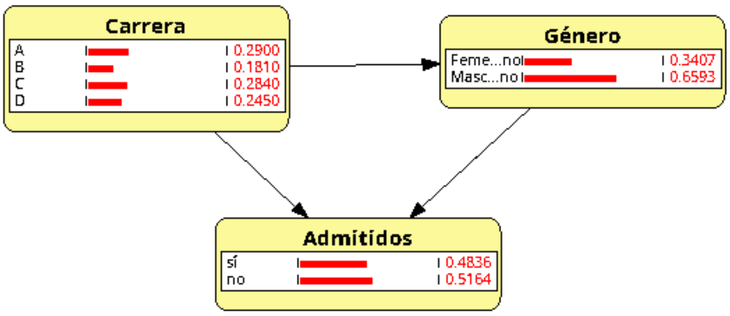
<u>Ejercicio 2.a)</u>

Aprendiendo la Red Bayesinaa con datos del fichero *DatosBerkeley.xls* se obtiene la siguiente red:



La red aprendida es distinta a la que se ha obtenido en el estudio previo, ya que en el estudio previo se ha considerado que la causalidad entre carrera y sexo es en la dirección contraria, es decir, se ha considerado que la carrera es la que influye en el género que solicita la admisión.

Se ha realizado la red con las probabilidades del estudio previo en el programa y se ha obtenido la siguiente red:



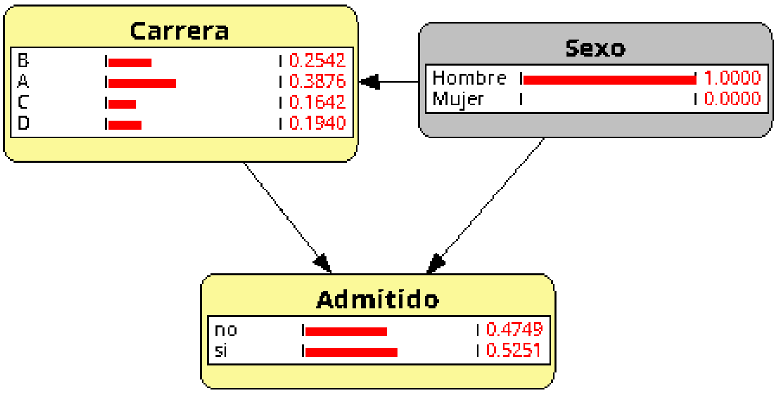
Las probabilidades de admitidos se observa que aun con esta diferencia en la causalidad si se redondea a las décimas quedan mismos resultados, ya que solo se diferencia en las centésimas. A continuación la tabla de probabilidades de la red calculada (se puede comparar con la tabla del estudio previo).

Sexo	Mujer	Mujer	Mujer	Mujer	Hombre	Hombre	Hombre	Hombre
Carrera	D	C	A	B	D	C	A	B
no	0.678938	0.680363	0.186275	0.329268	0.678404	0.620148	0.375784	0.376941
si	0.321062	0.319637	0.813725	0.670732	0.321596	0.379852	0.624216	0.623059

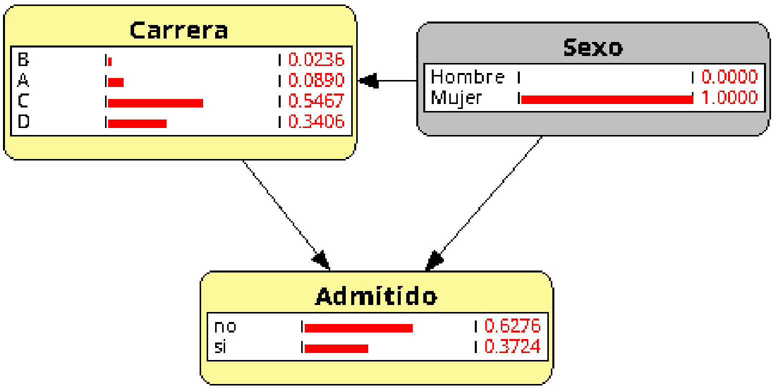
Ejercicio 2.b)

En este apartado se calculan las probabilidades con la red aprendida.

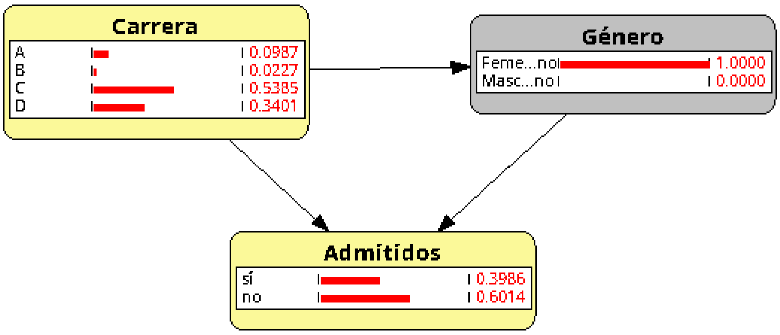
La **probabilidad de admisión para un hombre** es 0.5251 como se observa en la red al aplicar el hallazgo (evidencia sexo = hombre).



La **probabilidad de admisión para una mujer** es 0.3724 como se observa en la red al aplicar el hallazgo (evidencia sexo = mujer).



Con la red modelada en el estudio previo **la probabilidad de admisión para una mujer** es 0.3986 como se observa en la red al aplicar el hallazgo (evidencia Género = Femenino).



Con la red anterior se comprueba que los cálculos en el estudio previo son correctos.

Ejercicio 2.c)

Puede ser que haya discriminación de género pero si se ven las probabilidades del nodo Carrera en el que afecta el Sexo, se observa que en la carrera A y B hay más probabilidad de que sean hombre, y en la C y D más probabilidad de que sean mujeres. Por lo que puede ser que también influya el número de solicitudes y de qué sexo haga esas solicitudes, ya que si hay más solicitudes de mujeres que de hombres, es normal que haya más admisiones de mujeres que de hombres y viceversa.

Sexo	Mujer	Hombre
B	0.023643	0.254178
A	0.089025	0.387572
C	0.546702	0.164236
D	0.34063	0.194014