# Métodos de Inteligencia Artificial

L. Enrique Sucar (INAOE)

esucar@inaoep.mx

ccc.inaoep.mx/esucar

Tecnologías de Información

UPAEP

# Redes Bayesianas: Parte II

- Inferencia
  - Propagación en redes multiconectadas
- Aprendizaje
  - Parámetros
  - Estructura

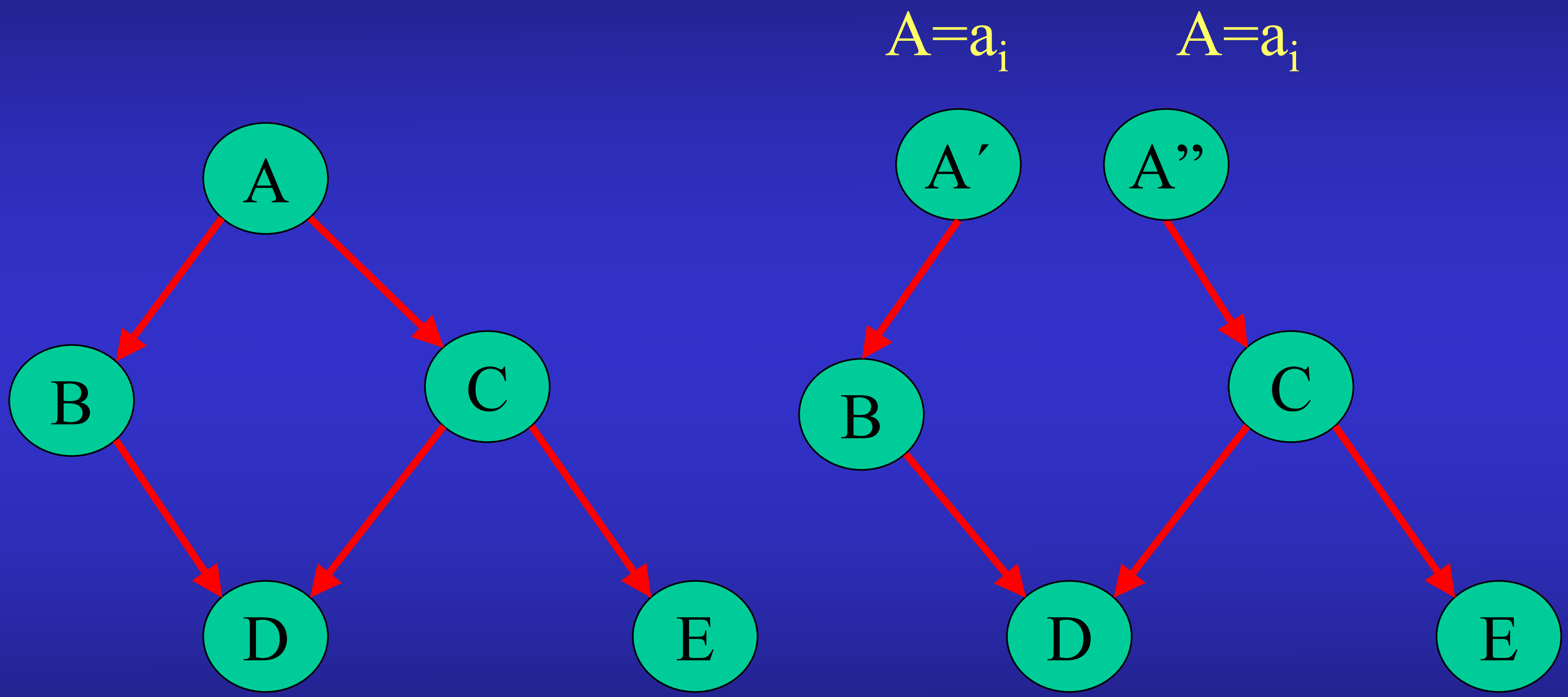
# Inferencia en RB multiconectadas

- Hay 3 tipos de métodos para calcular las probabilidades posteriores de todas las variables no conocidas en redes multi-conectadas:
  - **Condicionamiento**
  - **Simulación estocástica**
  - **Agrupamiento**

# Condicionamiento

- Si *instanciamos* (asignamos un valor) a una variable, ésta ***bloquea*** las trayectorias de propagación.
- Entonces, asumiendo valores para un grupo seleccionado de variables podemos descomponer la gráfica en un conjunto de redes conectadas en forma sencilla.
- Propagamos para cada valor posible de dichas variables y luego promediamos las probabilidades ponderadas.

# Condicionamiento



# Procedimiento

- Al "cortar" en A, la probabilidad de cualquier variable (b) la podemos obtener mediante la regla de probabilidad total:

$$P(b|E) = \Sigma_i \, P(b|a_i, E) \, P(a_i|E)$$

- Donde:
  - $P(b|a_i, E)$ : probabilidad posterior por propagación para cada valor de A
  - $P(a_i|E)$ : "peso"

# Simulación estocástica

- Se asignan valores aleatorios a las variables no asignadas, se calcula la distribución de probabilidad, y se obtienen valores de cada variable dando una muestra.
- Se repite el procedimiento para obtener un número apreciable de muestras y en base al número de ocurrencias de cada valor se determina la probabilidad de dicha variable.

# Muestreo Lógico

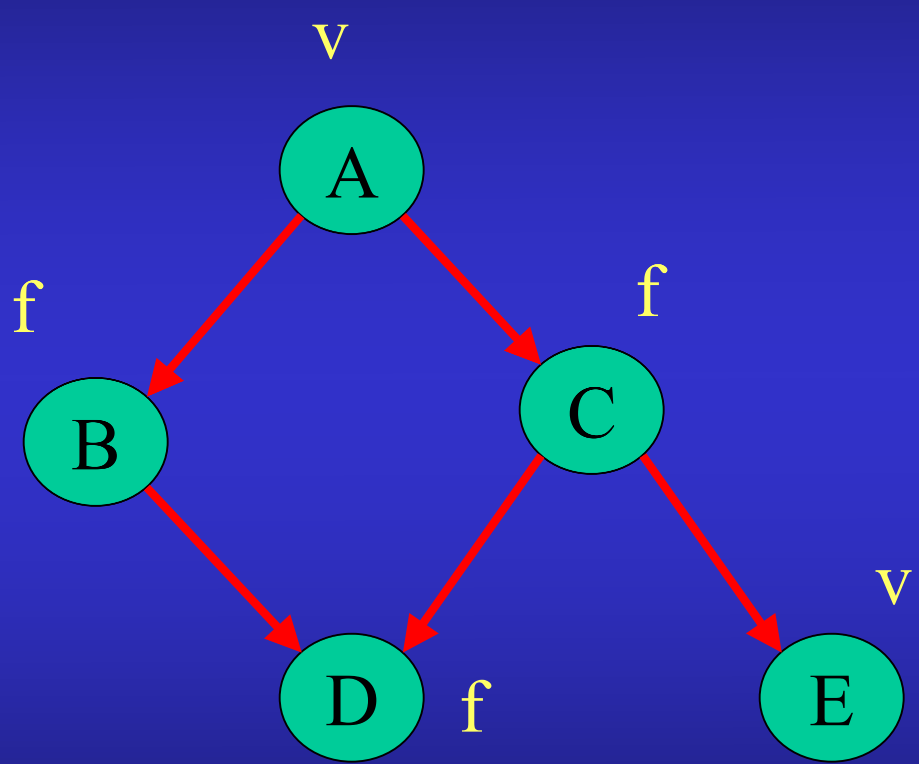
Para “N” muestras, repetir:

1. Dar valores aleatorios a los nodos raíz de acuerdo a sus probabilidades
2. En base a los valores anteriores, dar valores aleatorios a las siguientes variables (hijos de los nodos raíz) en función de la probabilidad condicional
3. Repetir (2) hasta llegar a los nodos hoja

Obtener probabilidades posteriores como frecuencias

- Si hay nodos evidencia, sólo considerar las muestras que correspondan a dichos valores

# Muestreo Lógico: ejemplo




| |
|---|
| vfffv |
| fvvff |
| vffvf |
| ffvfv |
| vfvvf |
| ffffv |
| fvvvf |
| fffff |
| fffvf |
| vvvvf |

# Ejemplo

- Sin evidencia:
  - P(A=V) = 4/10 = 0.4
  - P(B=V) = 3/10 = 0.3
  - P(C=V) = 5/10 = 0.5
  - P(D=V) = 5/10 = 0.5
  - P(E=V) = 3/10 = 0.3
- Con evidencia: D=V (aplican 5 muestras):
  - P(A=V) = 3/5 = 0.6
  - P(B=V) = 2/5 = 0.4
  - P(C=V) = 3/5 = 0.6
  - P(E=V) = 1/5 = 0.2

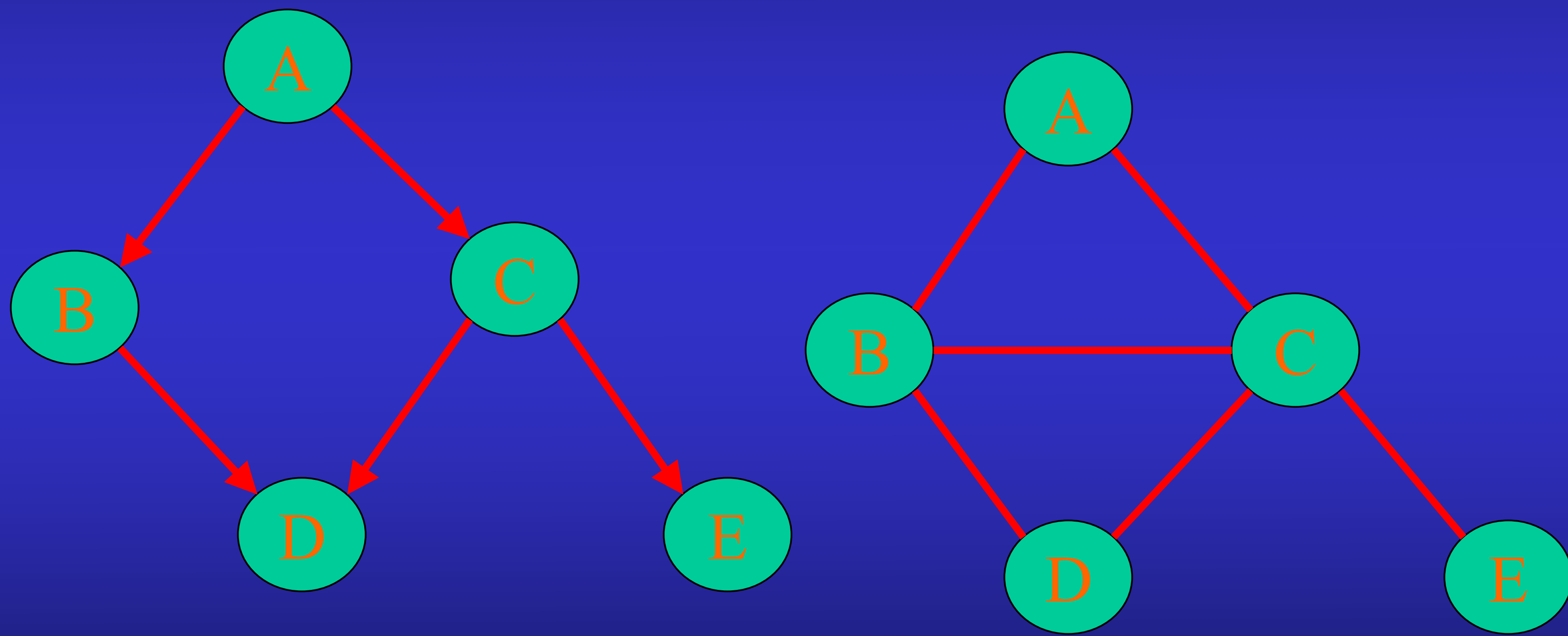
# Agrupamiento

- El método de agrupamiento consiste en transformar la estructura de la red para obtener un árbol, mediante agrupación de nodos usando la teoría de grafos.
- La propagación se realiza sobre el árbol de macro-nodos obtenido, donde cada macro-nodo corresponde a un clique o *unión* de la RB original (*junction tree*)

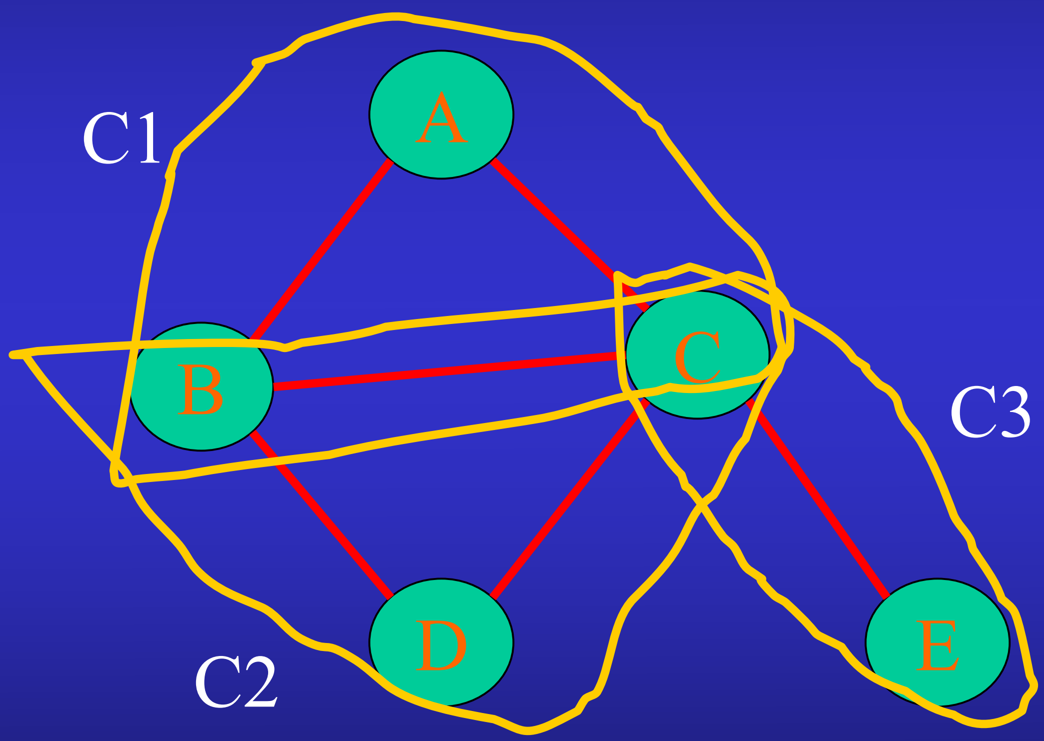
# Agrupamiento

- Transformación:
  - Eliminar direccionalidad de los arcos
  - Ordenamiento de los nodos por máxima cardinalidad
  - Moralizar el grafo (arco entre nodos con hijos comunes)
  - Triangular el grafo
  - Obtener los *cliques* y ordenar
  - Construir árbol de *cliques*

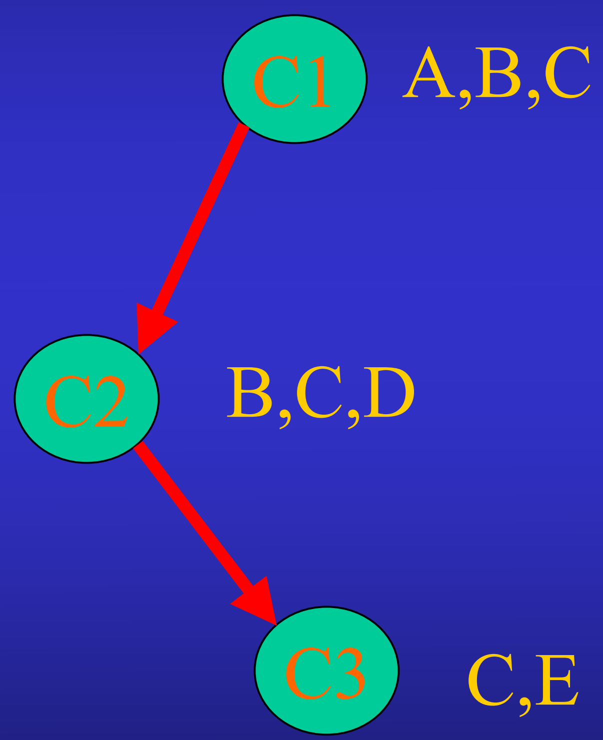
# Ejemplo



# Ordenamiento de Cliques



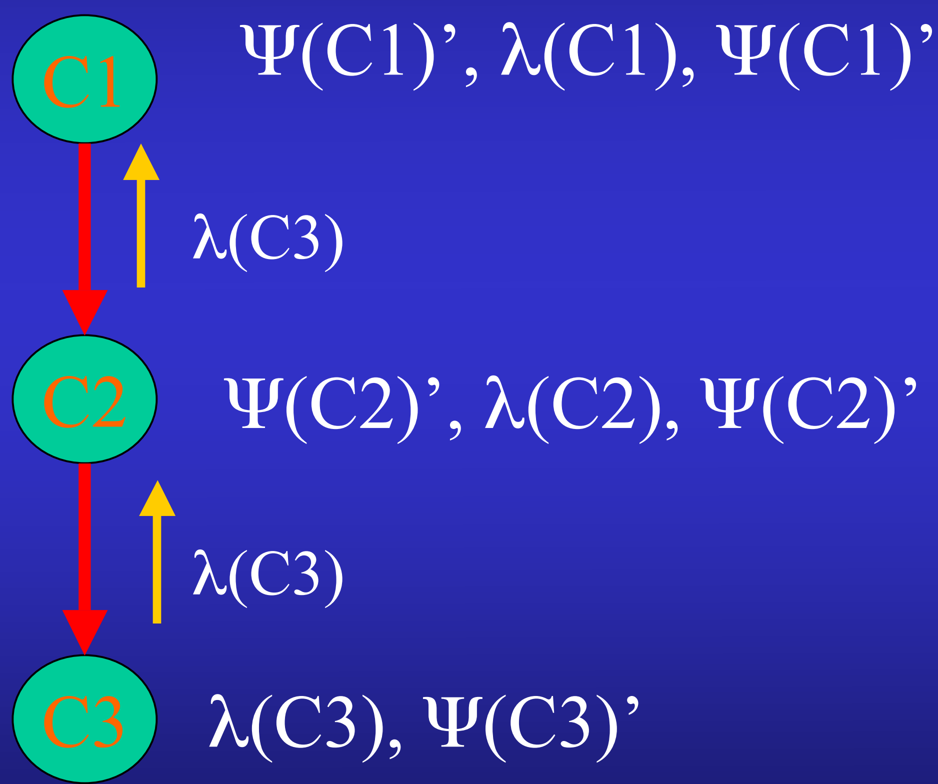
# Árbol de Cliques



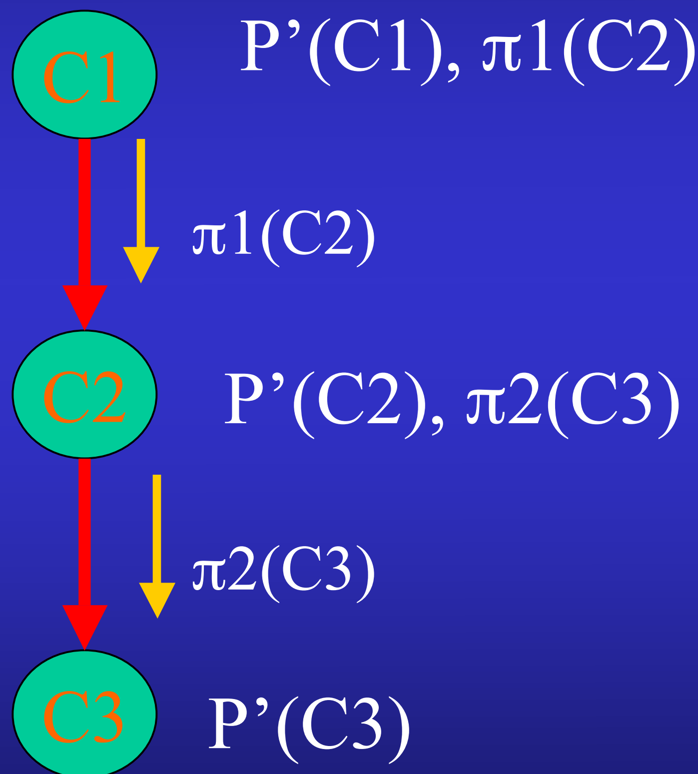
# Propagación

- La propagación es mediante el envío de mensajes en el árbol de *cliques* (en forma similar a árboles)
- Inicialmente se calcula la probabilidad conjunta (potencial) de cada *clique*, y la condicional dado el padre
- Dada cierta evidencia se recalculan las probabilidades de cada *clique*
- La probabilidad individual de cada variable se obtiene de la del *clique* por marginalización

# Ejemplo – propagación λ



# Ejemplo – propagación π



# Complejidad

- En el peor caso, la propagación en redes bayesianas es un problema NP-duro
- En la práctica, en muchas aplicaciones se tienen redes no muy densamente conectadas y la propagación es eficiente aún para redes muy grandes (función del *clique* mayor)
- Para redes muy complejas (muchas conexiones), la mejor alternativa son técnicas de simulación estocástica o técnicas aproximadas

# Aprendizaje

El aprendizaje inductivo consiste en obtener conocimiento a partir de datos.

En redes bayesianas se divide en 2 aspectos:

- Obtener la estructura de la red – aprendizaje estructural
- Obtener las probabilidades asociadas – aprendizaje paramétrico

# Aprendizaje Paramétrico

- Datos completos - se estiman las probabilidades a partir de frecuencias

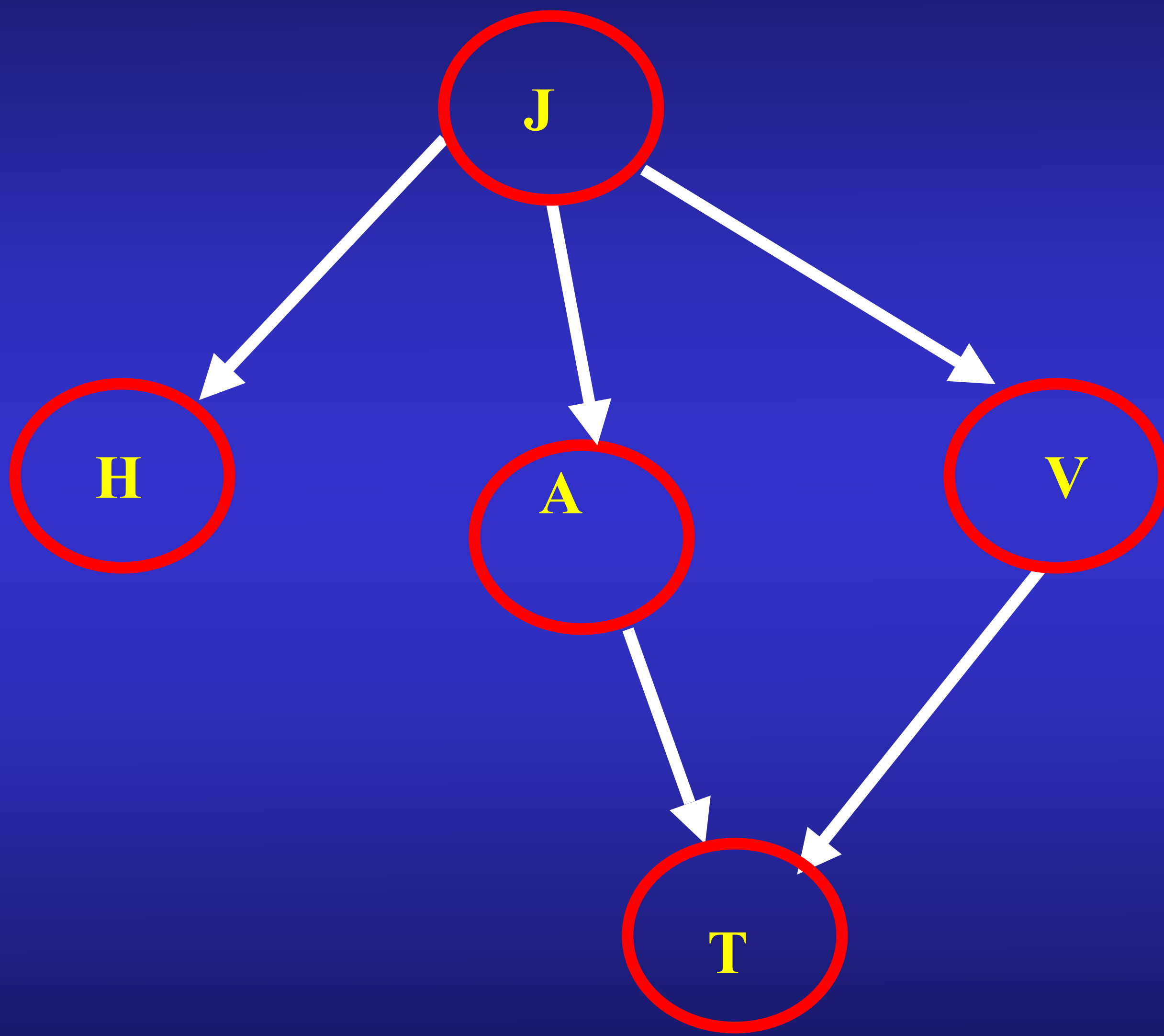
$$P(A) \sim N_a / N_t$$

$$P(B|A1, ..., An) \sim N_{a1, ..., an, b} / N_{a1, ..., an}$$

# Ejemplo - ¿Cuándo jugar golf?

| Ambiente | Temp. | Humedad | Viento | Jugar |
|---|---|---|---|---|
| soleado | alta | alta | no | N |
| soleado | alta | alta | si | N |
| nublado | alta | alta | no | P |
| lluvia | media | alta | no | P |
| lluvia | baja | normal | no | P |
| lluvia | baja | normal | si | N |
| nublado | baja | normal | si | P |
| soleado | media | alta | no | N |
| soleado | baja | normal | no | P |
| lluvia | media | normal | no | P |
| soleado | media | normal | si | P |
| nublado | media | alta | si | P |
| nublado | alta | normal | no | P |
| lluvia | media | alta | si | N |

# Ejemplo – estructura



# Ejemplo

- P(J)
  - P(N) = 5/14
  - P(P) = 9/14
- P(V|J)
  - P(si|N)=3/5,  P(si|P)=3/9
  - P(no|N)=2/5,  P(no|P)=6/9
- Etc.

# Aprendizaje Estructural

Diversos métodos:

- Aprendizaje de árboles
- Aprendizaje de poliárboles
- Aprendizaje de redes multiconectadas
  - Métodos basados en medidas
  - Métodos basados en relaciones de dependencia

# Aprendizaje de árboles

- Algoritmo desarrollado por Chow y Liu para aproximar una distribución de probabilidad por un producto de probabilidades de segundo orden (árbol).
- La probabilidad conjunta de n variables se puede representar como:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid X_{j(i)}\right)$$

- donde $X_{j(i)}$ es la causa o padre de $X_i$.

# Aprendizaje de árboles

- Se plantea el problema como uno de optimización - obtener la estructura que más se aproxime a la distribución "real".
- Medida de la diferencia de información entre la distribución real (P) y la aproximada (P*):

$$I\left(P, P^{*}\right) = \sum_{x} P\left(X\right) \log \frac{P(X)}{P^{*}(X)}$$

- El objetivo es minimizar I.

# Aprendizaje de árboles

- Se puede definir dicha diferencia en función de la información mutua entre pares de variables, que se define como:

$$I(X_i, X_j) = \sum_{x_i, x_j} P(X_i, X_j) \log \frac{P(X_i, X_j)}{P(X_i) P(X_j)}$$

- Se puede demostrar (Chow 68) que la diferencia de información es una función del negativo de la suma de las informaciones mutuas (pesos) de todos los pares de variables que constituyen el árbol
- Encontrar el árbol más próximo equivale a encontrar el árbol con mayor peso.

# Aprendizaje de árboles - algoritmo

1. Calcular la información mutua entre todos los pares de variables (n(n - 1)/2).
2. Ordenar las informaciones mutuas de mayor a menor.
3. Seleccionar la rama de mayor valor como árbol inicial.
4. Agregar la siguiente rama mientras no forme un ciclo, si es así, desechar.
5. Repetir (3-4) hasta que se cubran todas las variables (n -1 ramas).

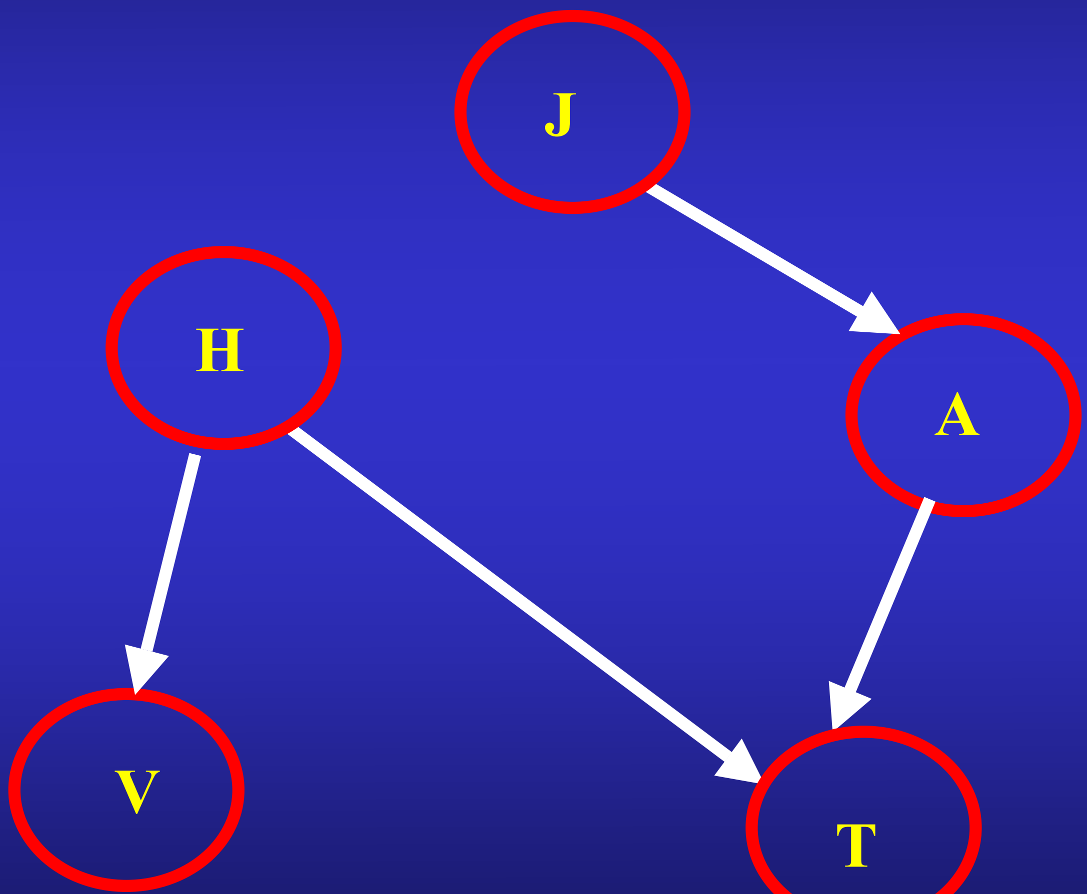
- El algoritmo NO provee la dirección de los arcos, por lo que ésta se puede asignar en forma arbitraria o utilizando semántica externa (experto).

# Ejemplo (golf)

- Informaciones mutuas ordenadas

| No. | Var 1 | Var 2 | I.M. |
|---|---|---|---|
| 1 | temp. | humedad | .1128 |
| 2 | humedad | viento | .0860 |
| 3 | ambiente | juega | .0745 |
| 4 | ambiente | temp. | .0074 |
| 5 | humedad | juega | .0457 |
| 6 | viento | juega. | .0145 |
| 7 | temp. | juega | … |
| 8 | viento | ambiente | … |
| 9 | humedad | viento | … |
| 10 | viento | temp. | … |

# Ejemplo (golf)



# Aprendizaje de poliárboles

- Parte del esqueleto (estructura sin direcciones) obtenido con el algoritmo anterior
- Determina la dirección de los arcos utilizando pruebas de dependencia entre tripletas de variables.
- Dadas 3 variables, existen 3 casos posibles:
  - Arcos divergentes
  - Arcos secuenciales
  - Arcos convergentes
- Los primeros dos casos son indistinguibles, pero el tercero es diferente, ya que las dos variables "padre" son marginalmente independientes.

# Prueba de Tripletas

- Tripleta de variables:

$$X – Z – Y$$

- Si X – Y son independientes dado Z, entonces pueden ser secuenciales o divergentes

$$X \rightarrow Z \rightarrow Y;\ X \leftarrow Z \rightarrow Y$$

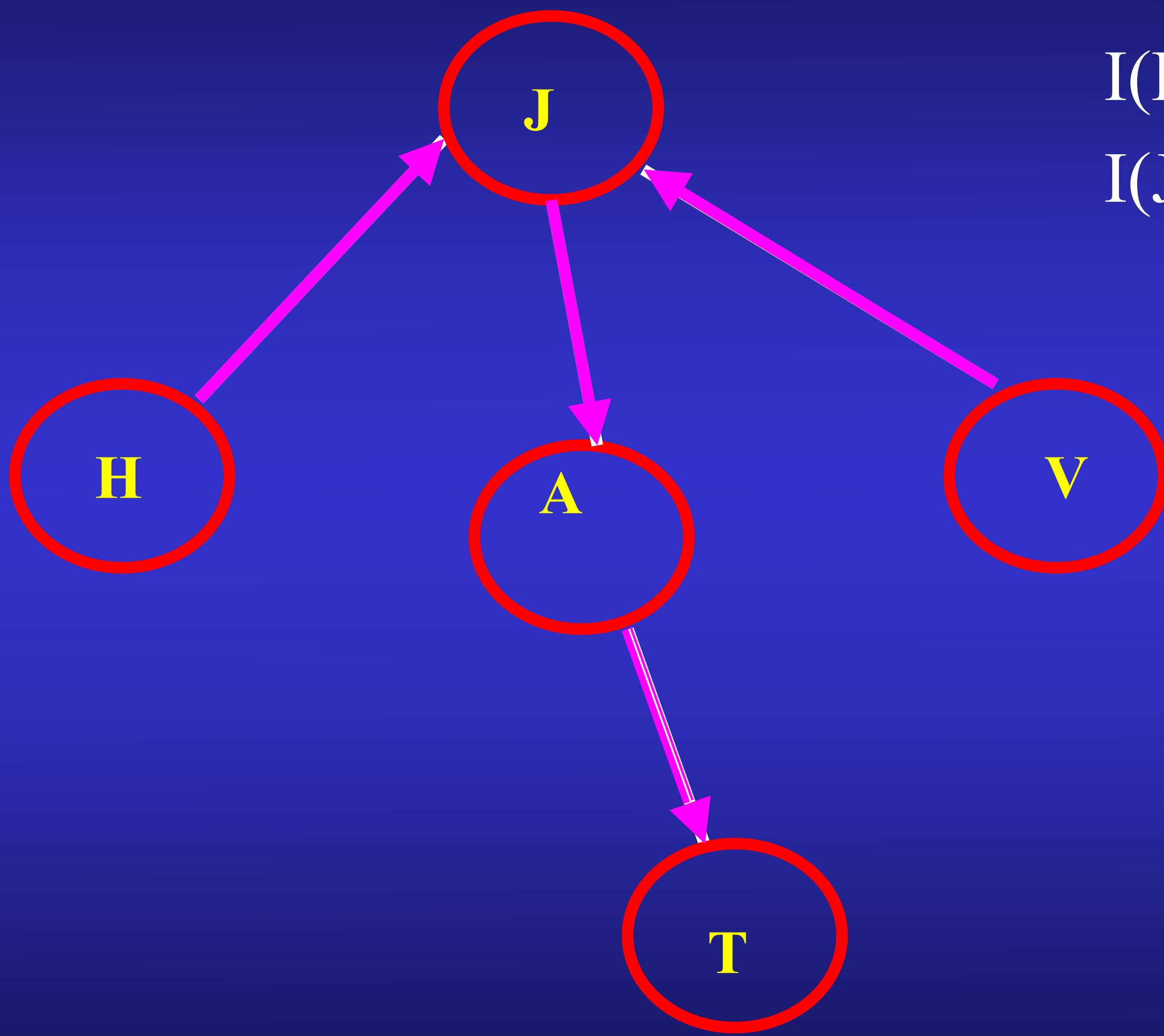
- Si X – Y no son independientes dado Z, entonces son arcos convergentes

$$X \rightarrow Z \leftarrow Y$$

# Aprendizaje de poliárboles - algoritmo

1. Obtener esqueleto utilizando el algoritmo de Chow y Liu
2. Recorrer la red hasta encontrar una tripleta de nodos que sean convergentes (tercer caso) - nodo multipadre-
3. A partir de un nodo multipadre determinar las direcciones de los arcos utilizando la prueba de tripletas hasta donde sea posible (base causal).
4. Repetir 2-3 hasta que ya no se puedan descubrir más direcciones.
5. Si quedan arcos sin dirección, utilizar semántica externa para obtener su dirección (o fijar direcciones).

# Ejemplo




~I(H,J,V)

I(H,J,A)

I(J,A,T)

# Aprendizaje de redes multiconectadas

Existen dos tipos de métodos para el aprendizaje genérico de redes bayesianas:

1. Métodos basados en medidas de ajuste y búsqueda
2. Métodos basados en pruebas de independencia

# Métodos basados en medidas

Se generan diferentes estructuras y se evalúan respecto a los datos utilizando alguna medida

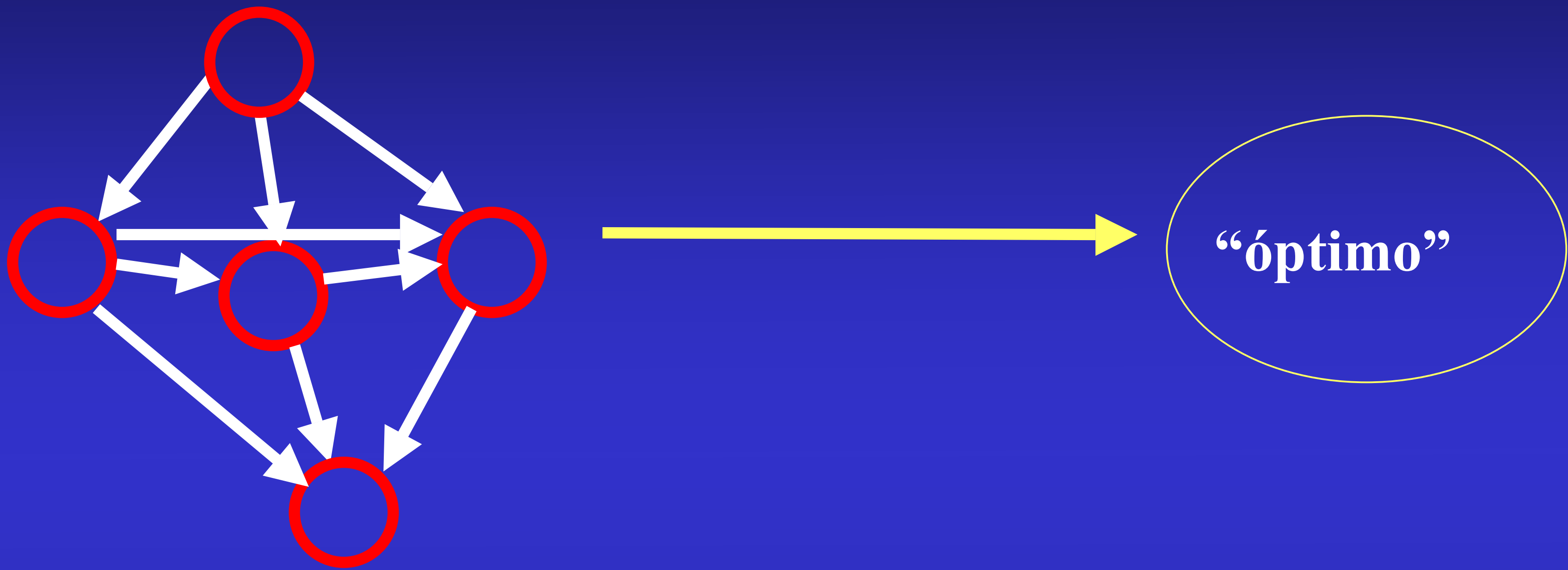
Dos aspectos principales:

- Medida de “ajuste” de la estructura a los datos
- Búsqueda de la “mejor” estructura

# Buscando la mejor estructura

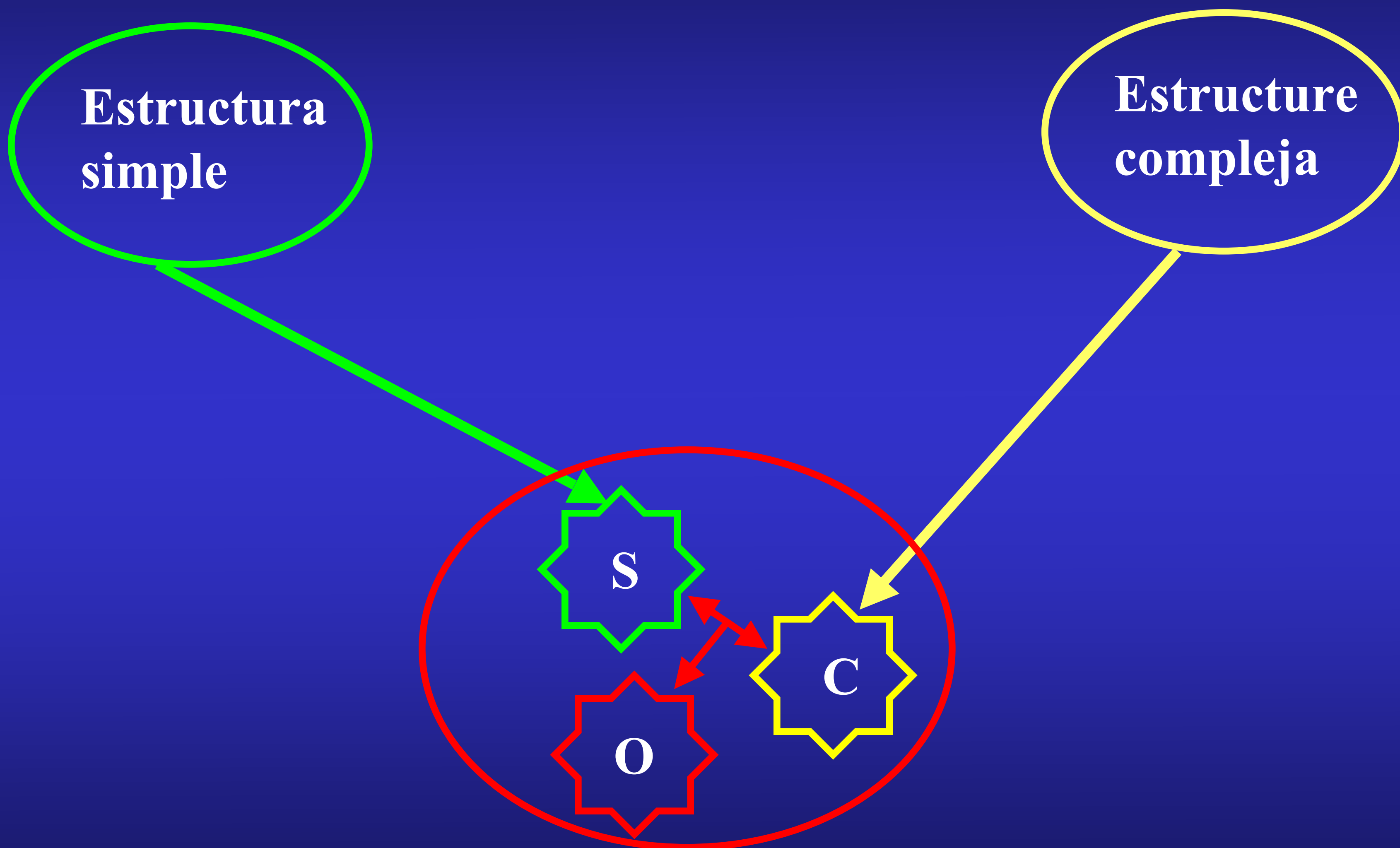

- Búsqueda de ascenso de colinas (*hill climbing*)
- Se inicia con una estructura simple (árbol) y se van agregando arcos hasta llegar a un *mínimo local*

# Buscando la mejor estructura




- Se puede iniciar con una estructura compleja (máximo número de arcos) y se van eliminando arcos hasta llegar a un *mínimo local*

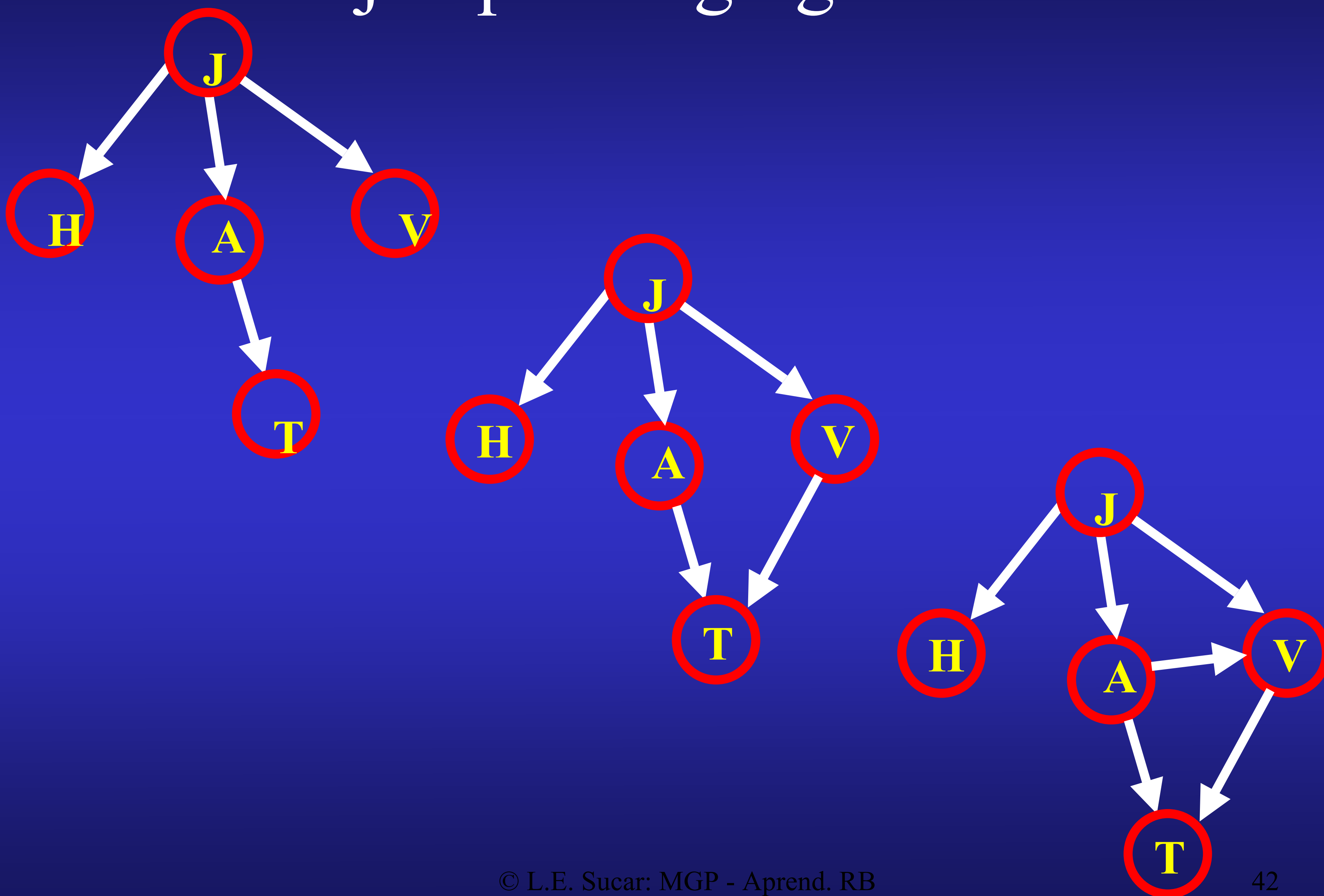
# Búsqueda bidirecional



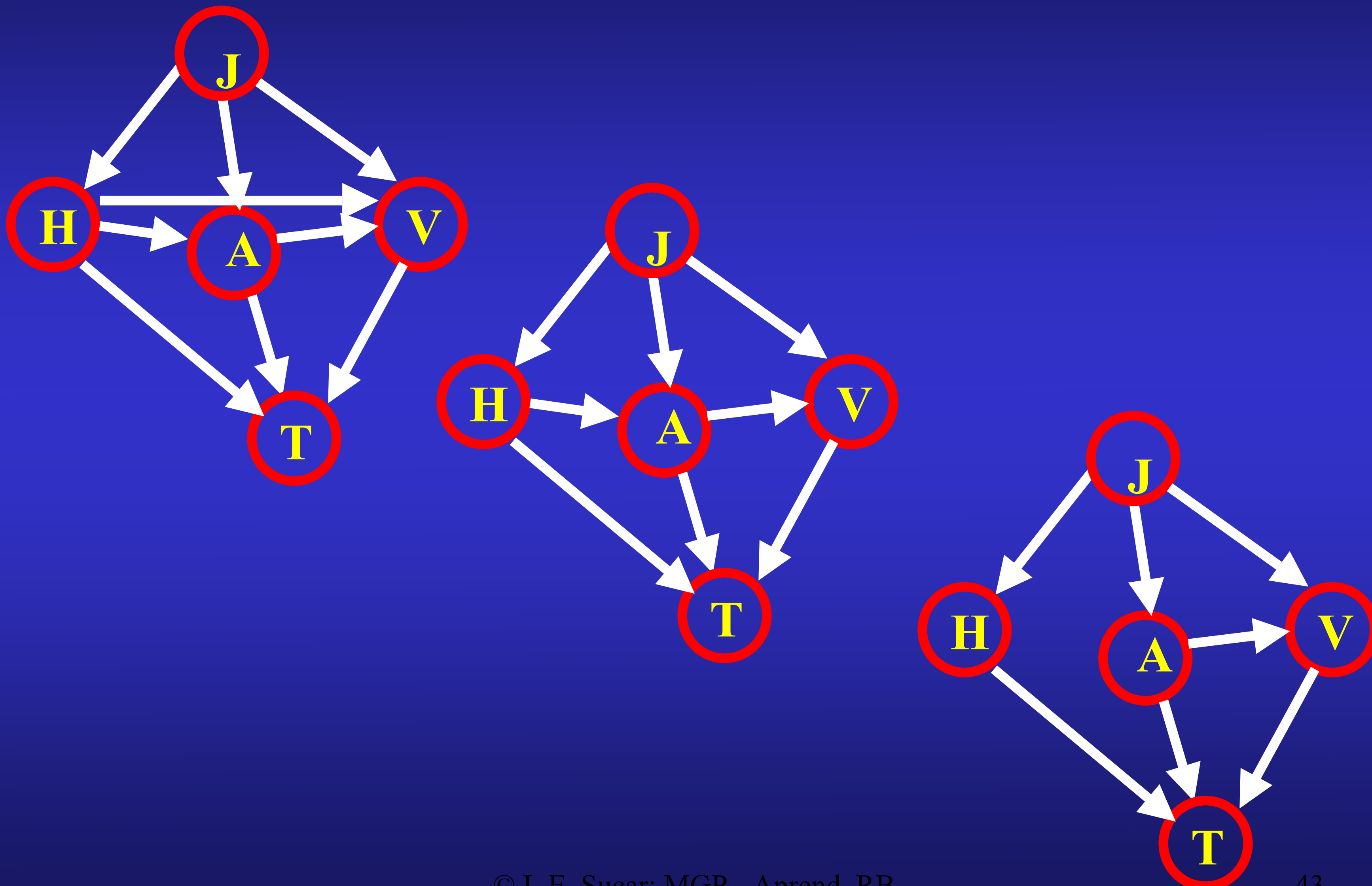
# Ejemplo - ¿Cuándo jugar golf?

| Ambiente | Temp. | Humedad | Viento | Jugar |
|---|---|---|---|---|
| soleado | alta | alta | no | N |
| soleado | alta | alta | si | N |
| nublado | alta | alta | no | P |
| lluvia | media | alta | no | P |
| lluvia | baja | normal | no | P |
| lluvia | baja | normal | si | N |
| nublado | baja | normal | si | P |
| soleado | media | alta | no | N |
| soleado | baja | normal | no | P |
| lluvia | media | normal | no | P |
| soleado | media | normal | si | P |
| nublado | media | alta | si | P |
| nublado | alta | normal | no | P |
| lluvia | media | alta | si | N |

# Ejemplo – agregando arcos



# Ejemplo – eliminando arcos



# Métodos basados en medidas

- Se genera la estructura en base a ir agregando/eliminando arcos de acuerdo a medidas de dependencia entre variables
- Ejemplos:
  - Árboles – método de Chow y Liu
  - Poliárboles – método de Rebane y Pearl
  - Multiconectadas – existen varios algoritmos basados en diferentes medidas

# Algoritmo PC

- Se basa en pruebas de independencia entre variables:

$$I\ (Xi,\ Xj \mid A)$$

- Donde A es un subconjunto de variables
- Asume que:
  - Se tienen suficientes datos
  - Las pruebas estadísticas no tienen errores

# Prueba de Independencia

- Para probar si X, Y son independientes dado A se utiliza la entropía cruzada condicional:

CE(X,Y | Z) =

$$\Sigma_z P(z)\, \Sigma_{x,y} P(x,y|z) \log [P(x,y|z) / P(x|z)\, P(y|z)]$$

- Si es cero o cercana a cero, quiere decir que son independientes (se puede usar un umbral o una prueba estadística con cierto nivel de significancia)

# Algoritmo

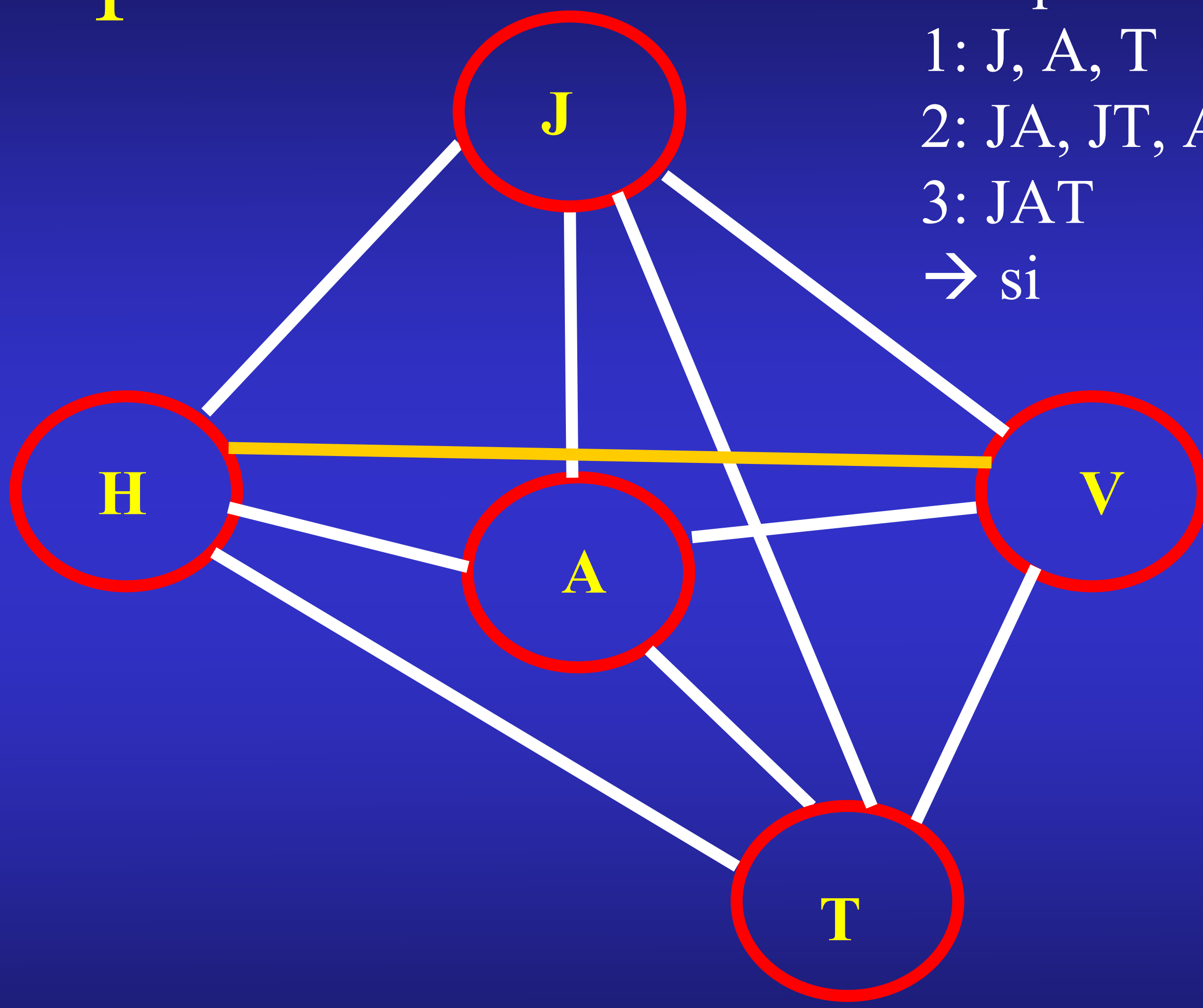
1. Encontrar un “esqueleto” (grafo no dirigido)
2. Encontrar arcos convergentes en tripletas de variables por pruebas de independencia
3. Orientar el resto de las ligas de forma que no se produzcan ciclos

# Esqueleto

- La idea básica para determinar el esqueleto es iniciar con un grafo completo (conectando todos vs. todos los nodos) y eliminar el arco entre X – Y si hay un subconjunto de nodos en G (excepto X, Y) que los hace independientes
- En principio se consideran todos los posibles subconjuntos de variables, de tamaño 1 hasta de tamaño N-1 (N es el número de nodos adyacentes a X)
- El considerar todos los posibles subconjuntos es muy ineficiente, y normalmente se limita a considerar sólo subconjuntos de 1, 2, …, k nodos

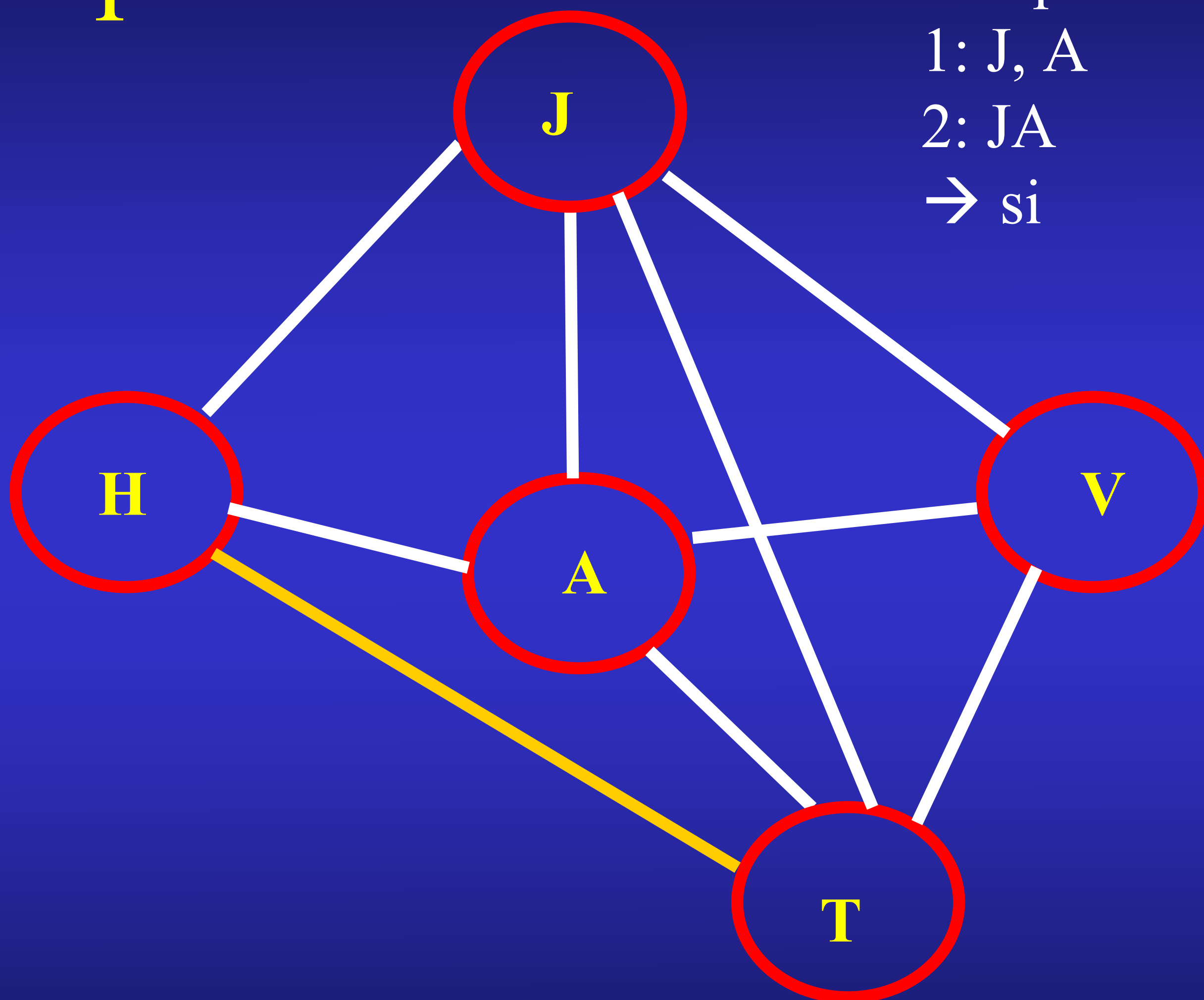
# Ejemplo

Probar si H,V son Independientes dados:
1: J, A, T
2: JA, JT, AT
3: JAT
→ si



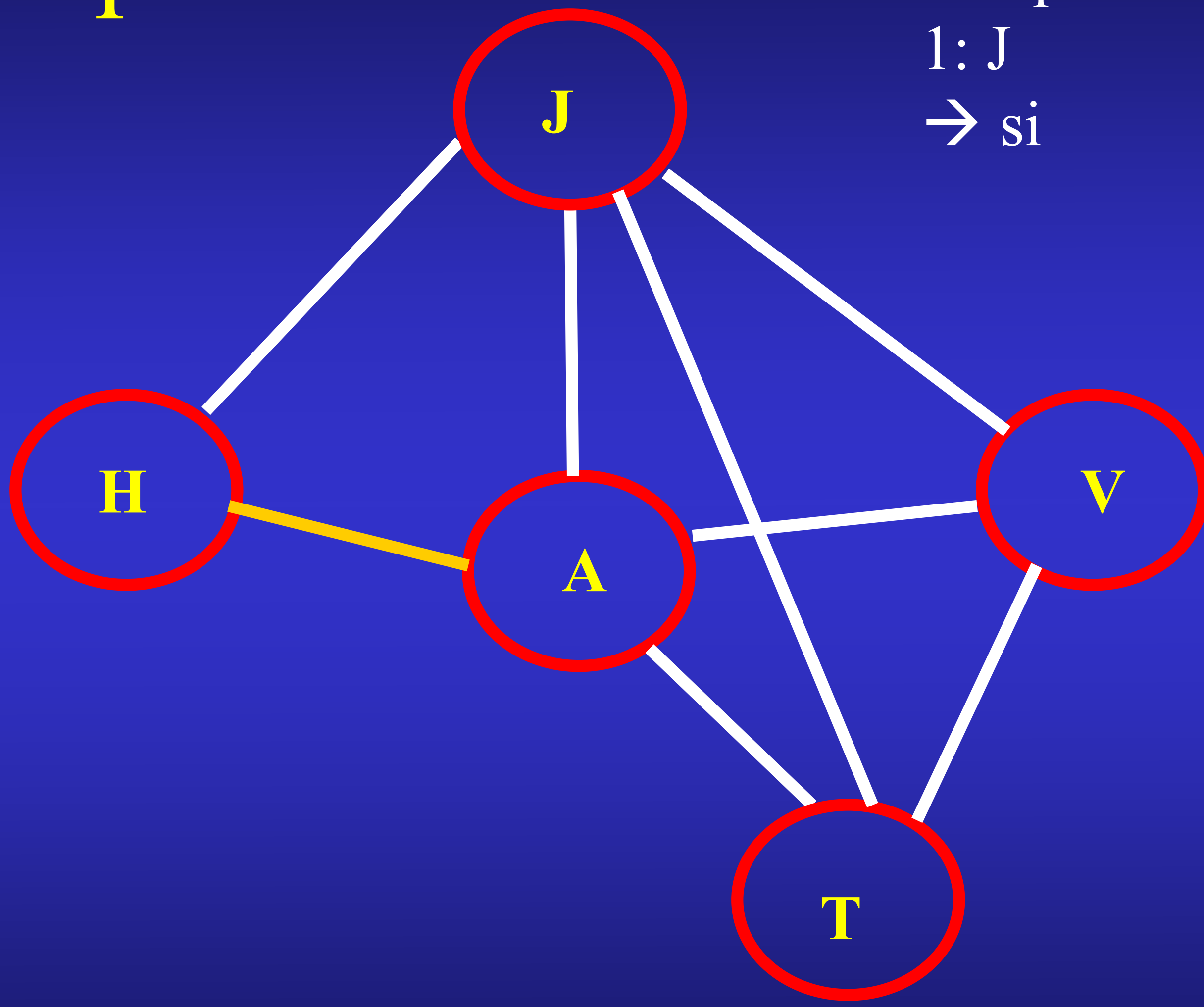
# Ejemplo

Probar si H,T son Independientes dados:
1: J, A
2: JA
→ si



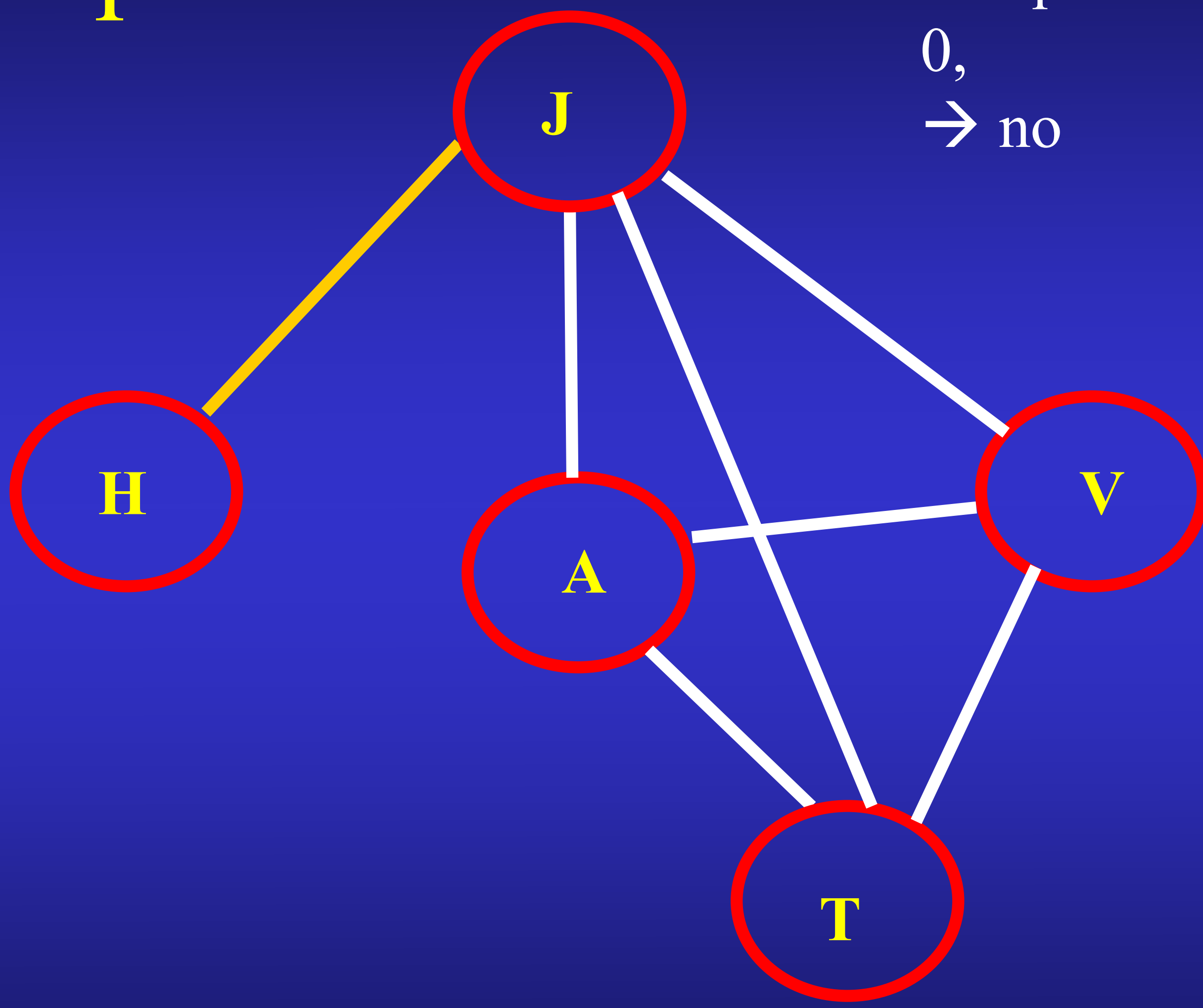
# Ejemplo

Probar si H,A son Independientes dados:
1: J
→ si



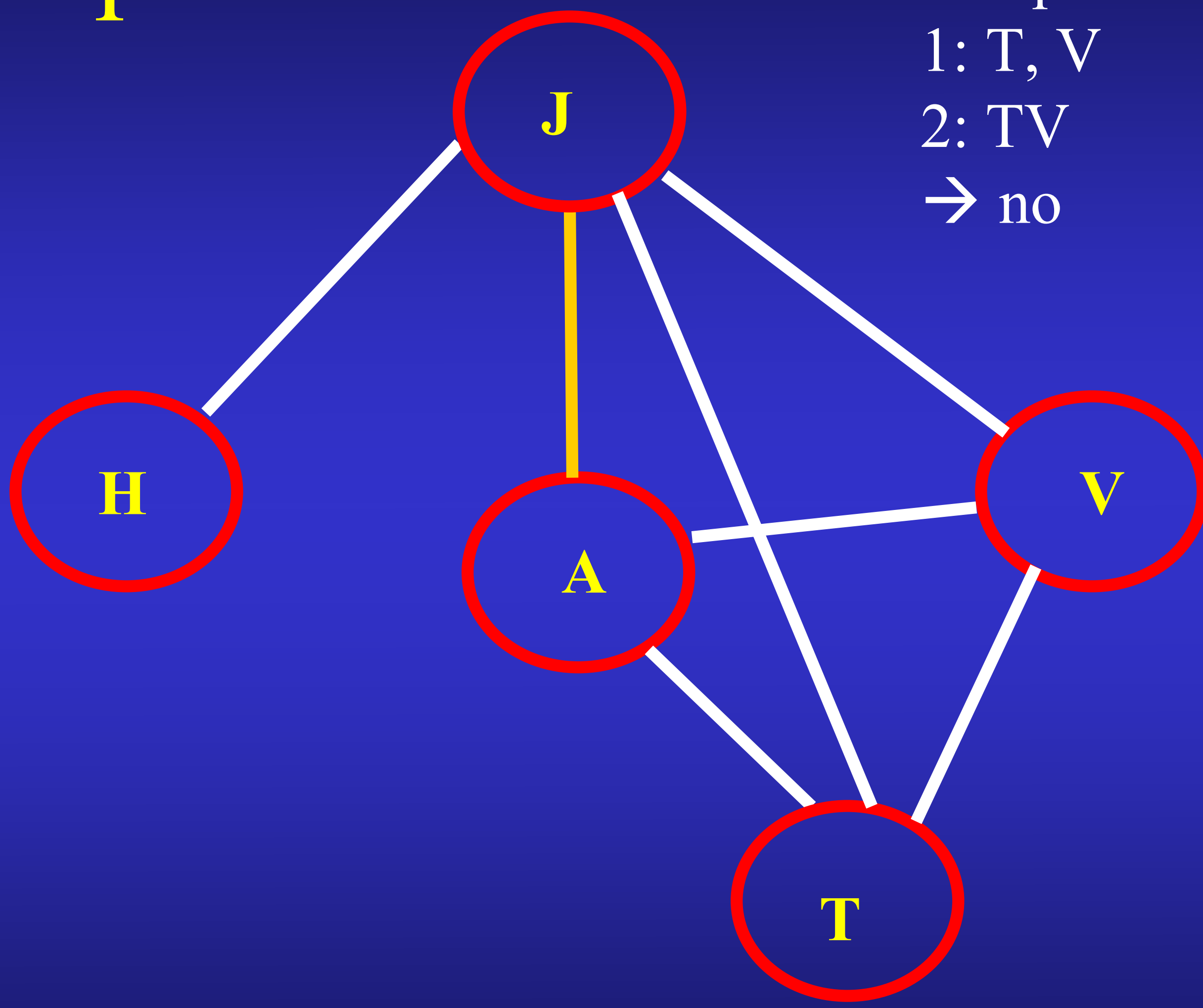
# Ejemplo

Probar si H,J son Independientes dados:
0,
→ no



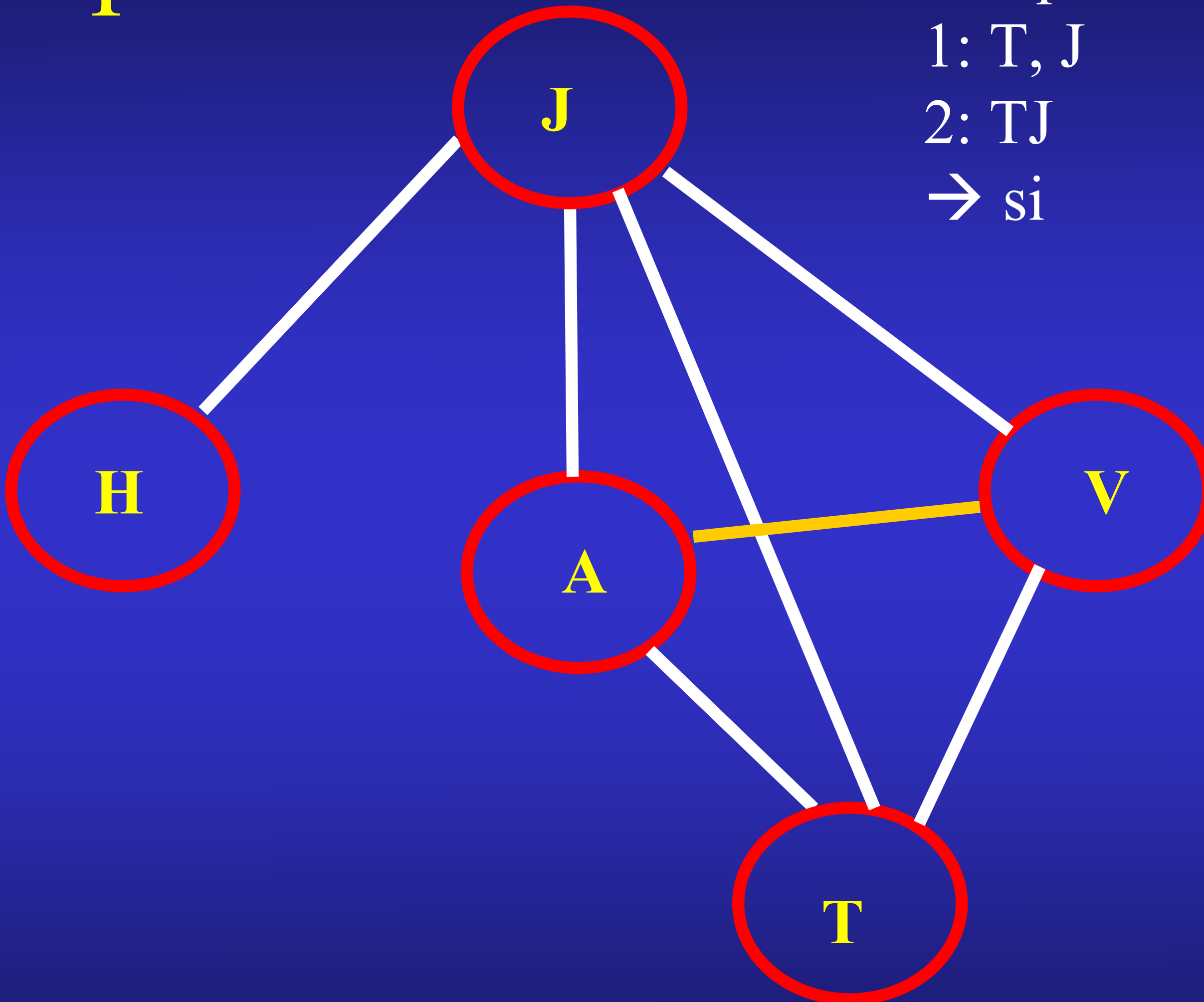
# Ejemplo

Probar si A,J son Independientes dados:
1: T, V
2: TV
→ no



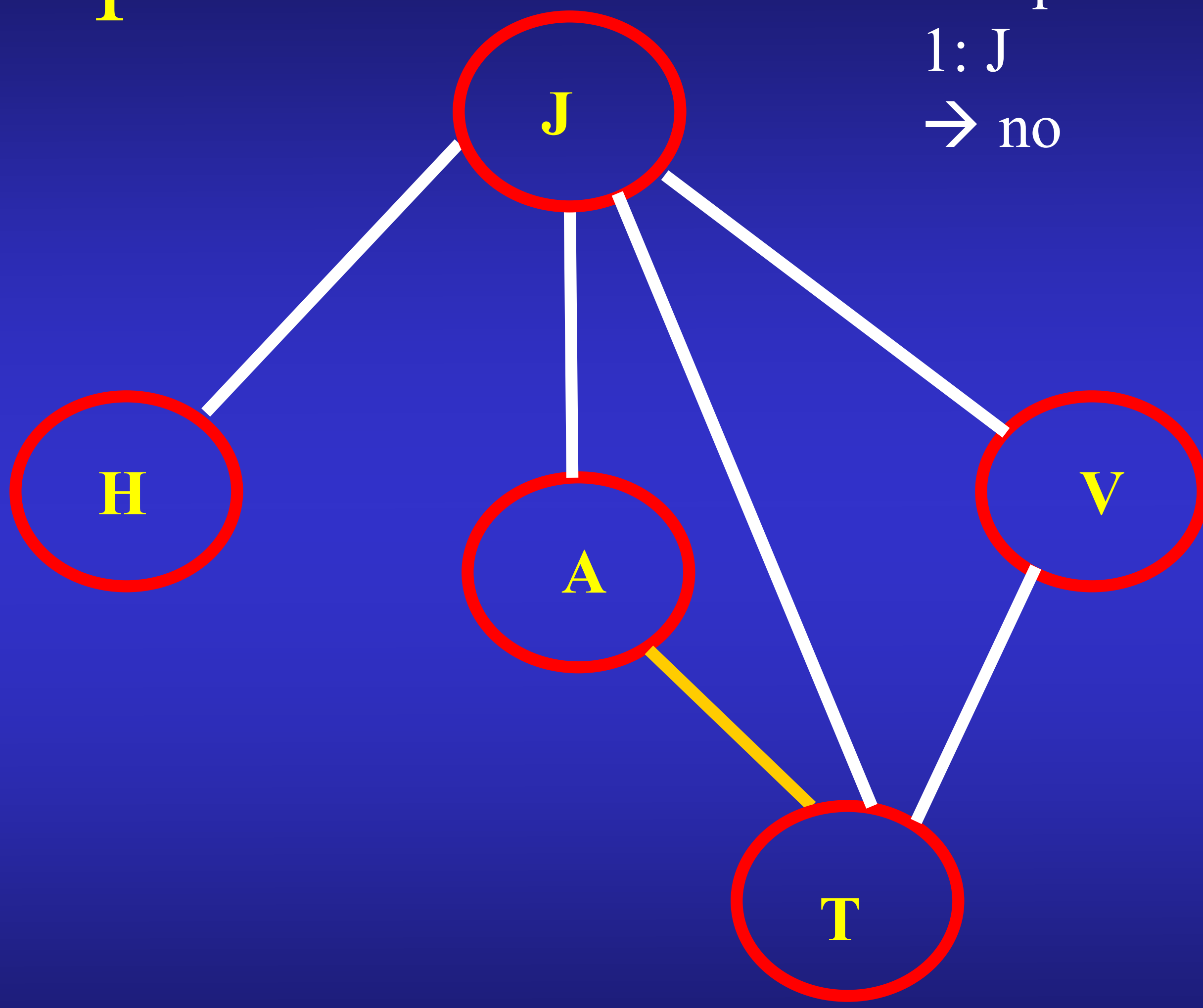
# Ejemplo

Probar si A,V son Independientes dados:
1: T, J
2: TJ
→ si



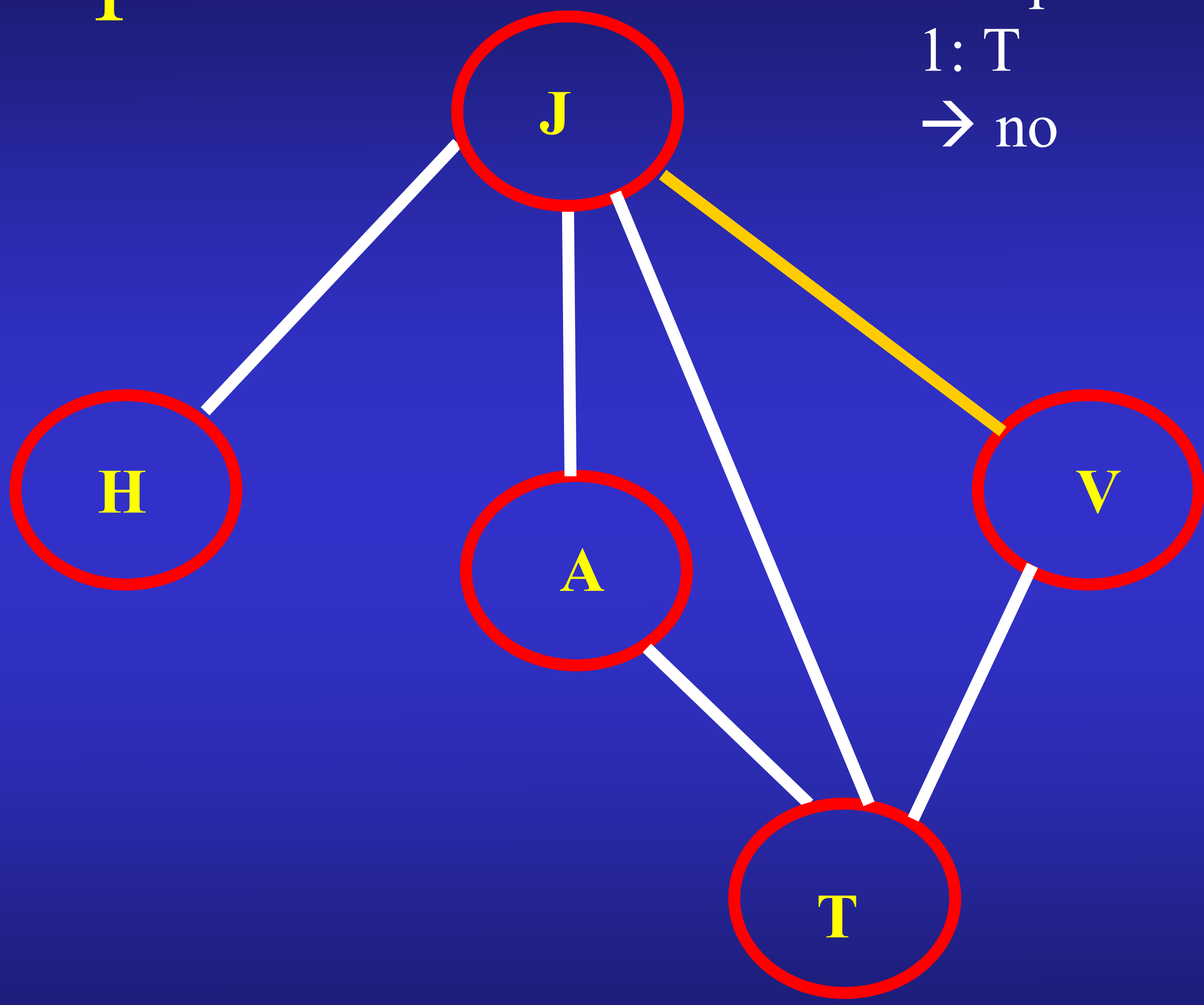
# Ejemplo

Probar si A,T son Independientes dados:
1: J
→ no



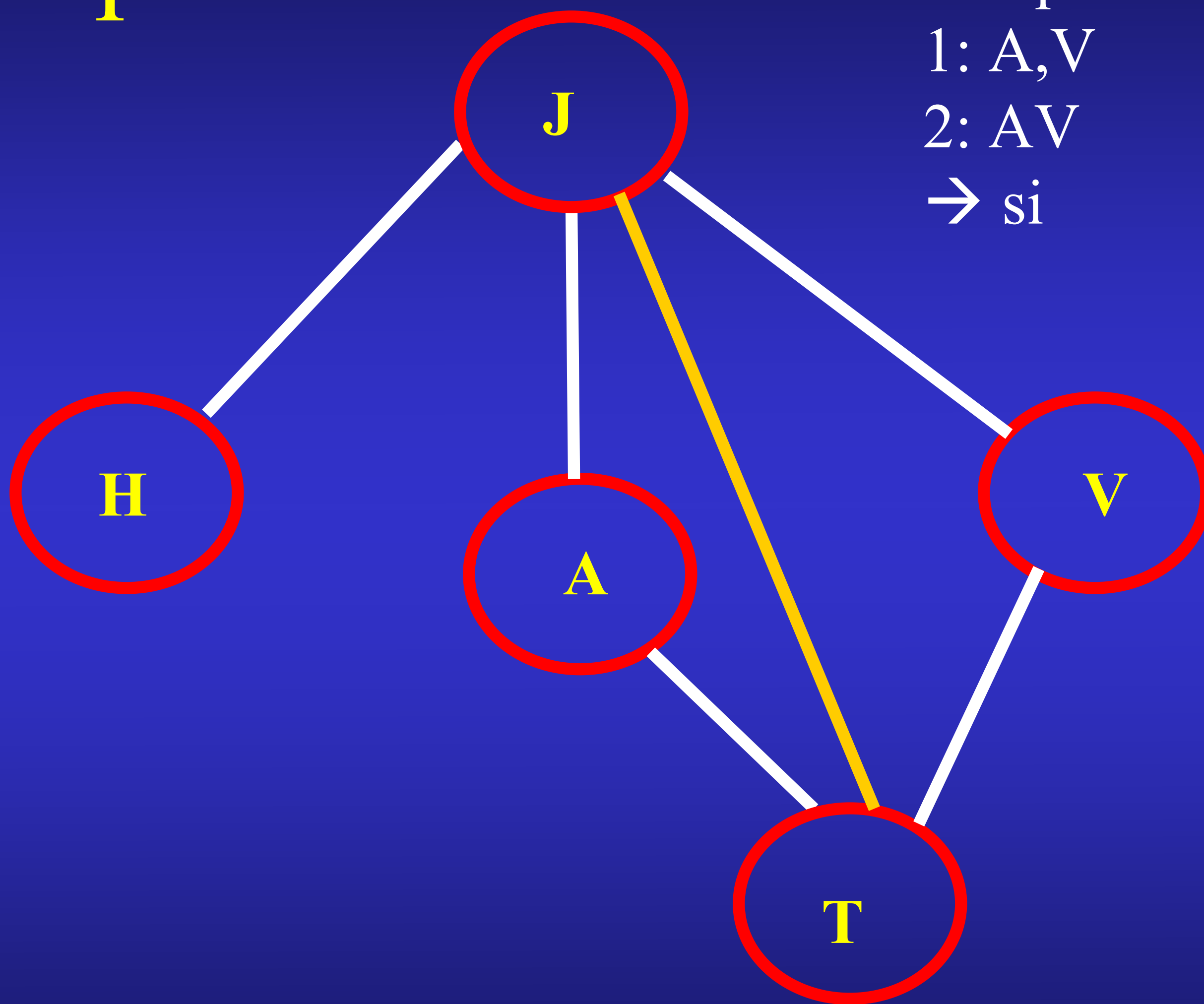
# Ejemplo

Probar si J,V son Independientes dados:
1: T
→ no



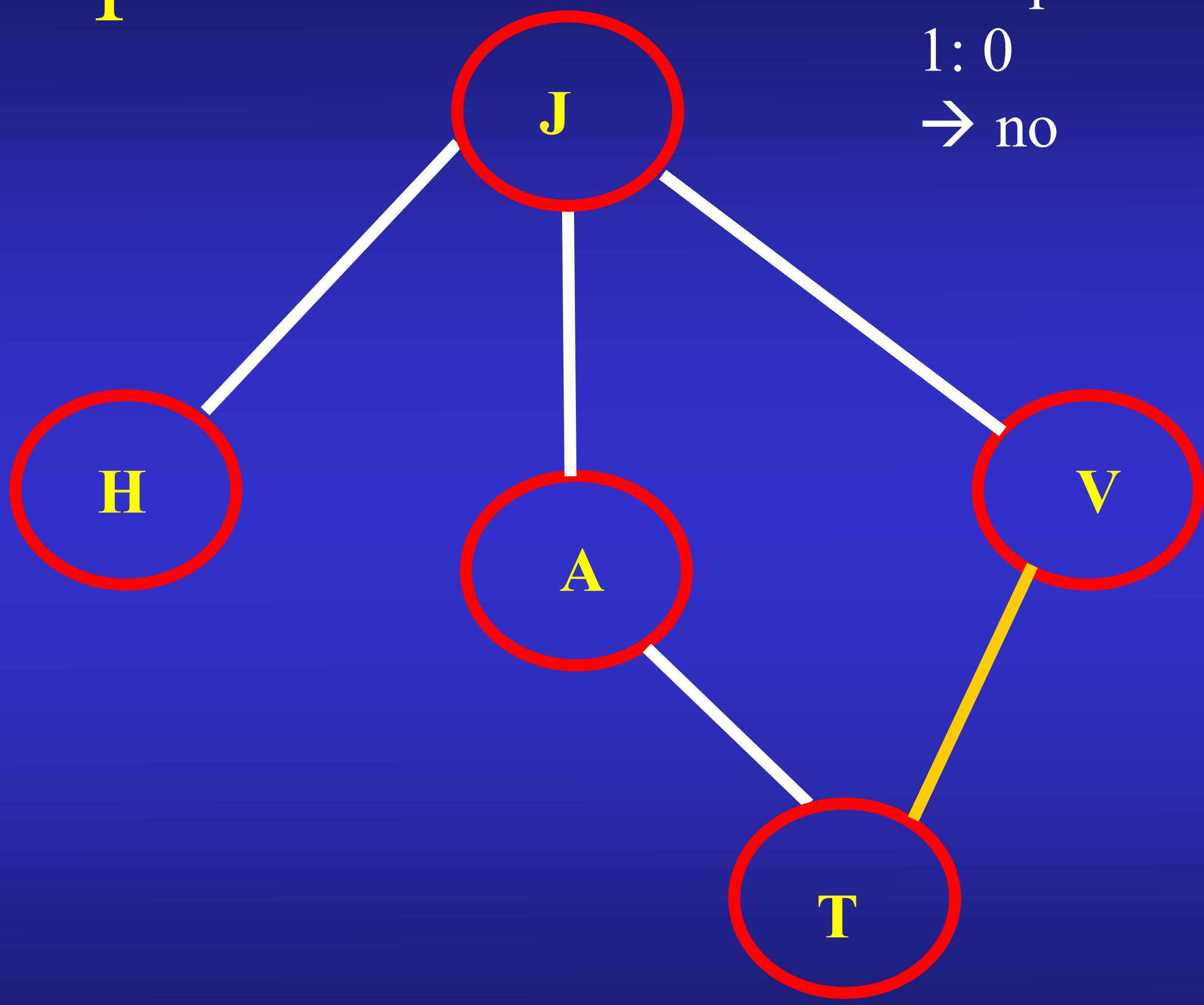
# Ejemplo

Probar si J,T son
Independientes dados:
1: A,V
2: AV
→ si



# Ejemplo

Probar si V,T son Independientes dados:
1: 0
→ no



# Arcos convergentes

- Se verifica cada tripleta de variables para encontrar arcos convergentes mediante pruebas de independencia:

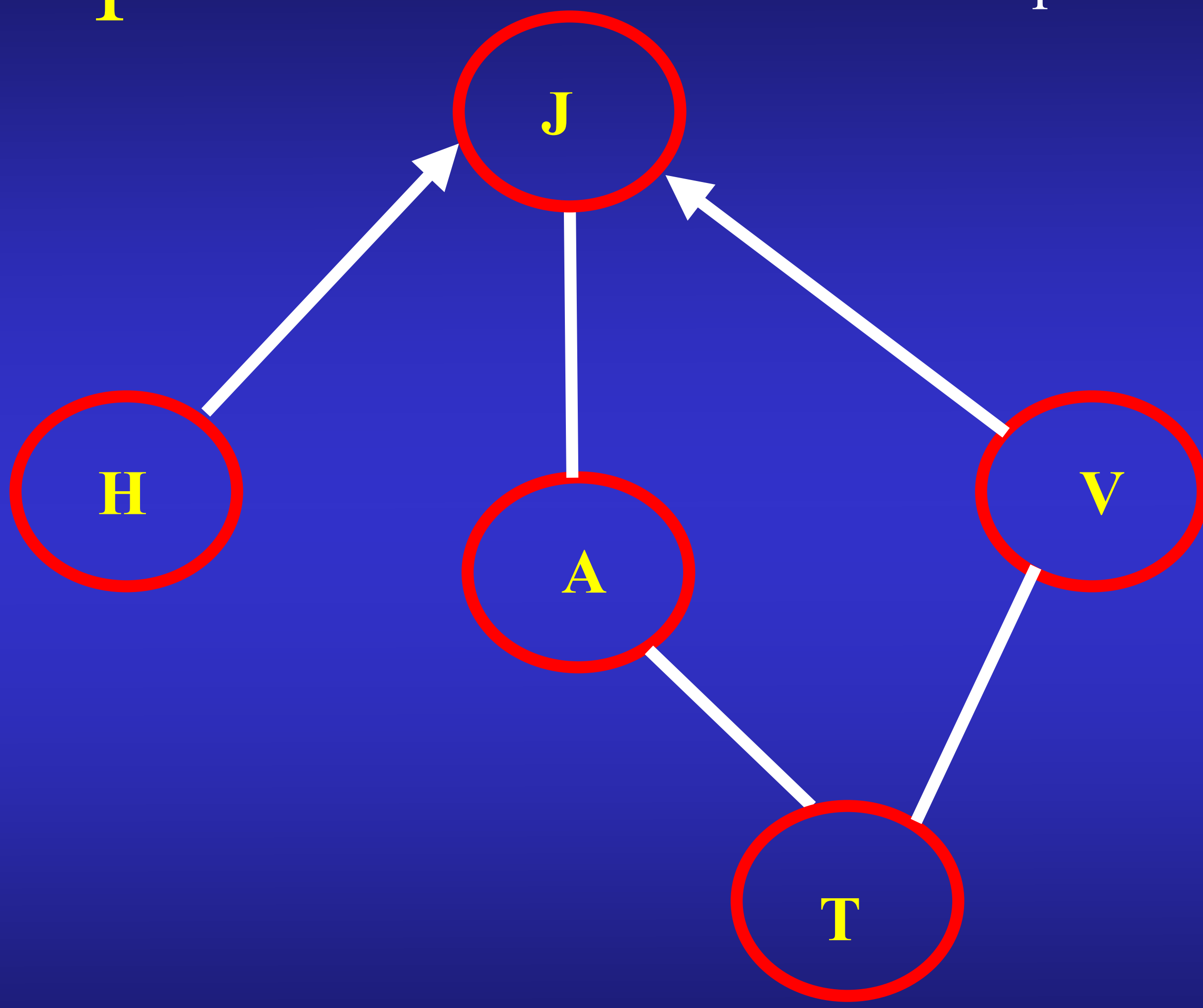
$$X - Z - Y$$

- Si X – Y no son independientes dado Z, entonces son arcos convergentes

$$X \rightarrow Z \leftarrow Y$$

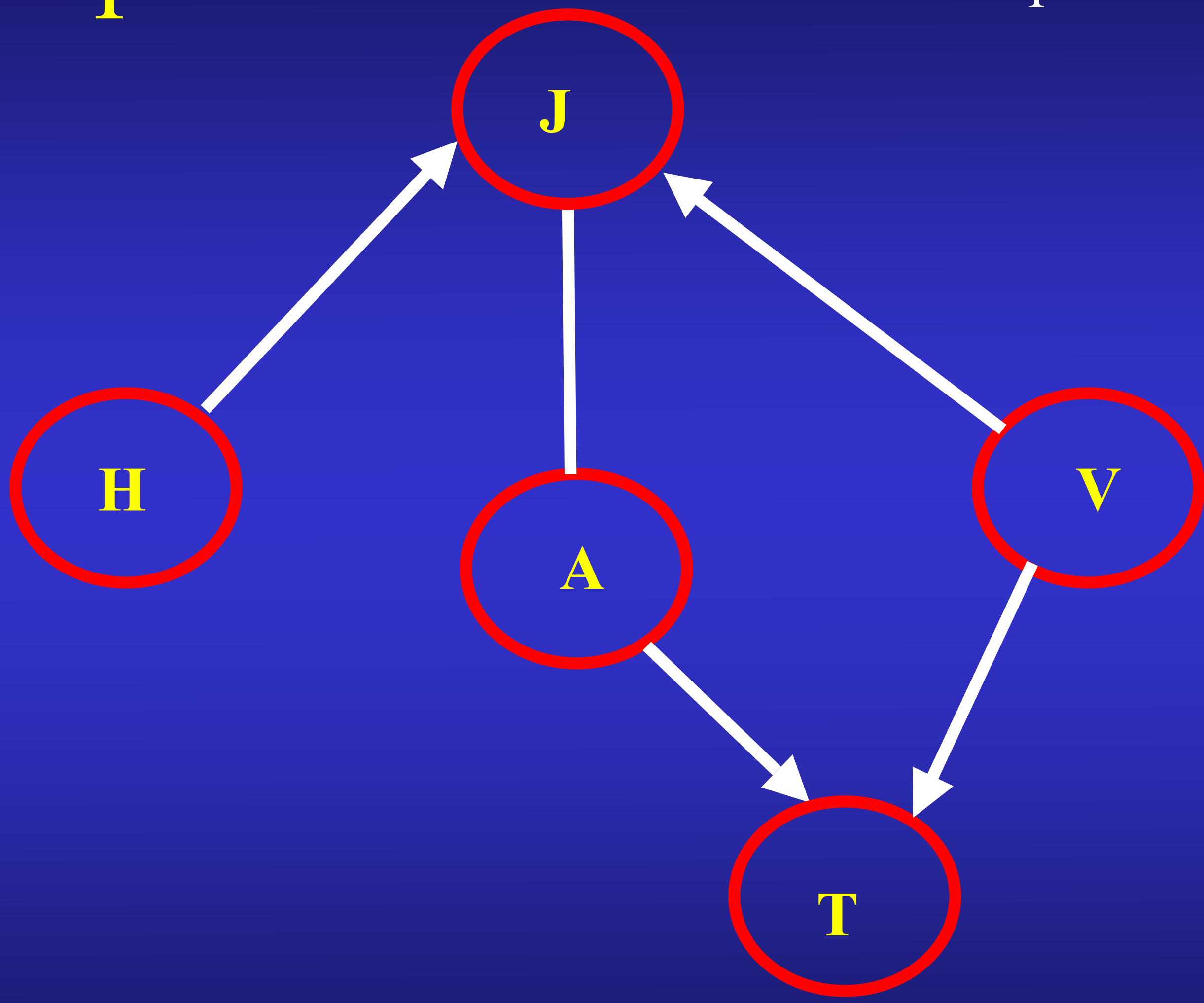
# Ejemplo

H,V no son
Independientes dado J



# Ejemplo

A,V no son
Independientes dado T

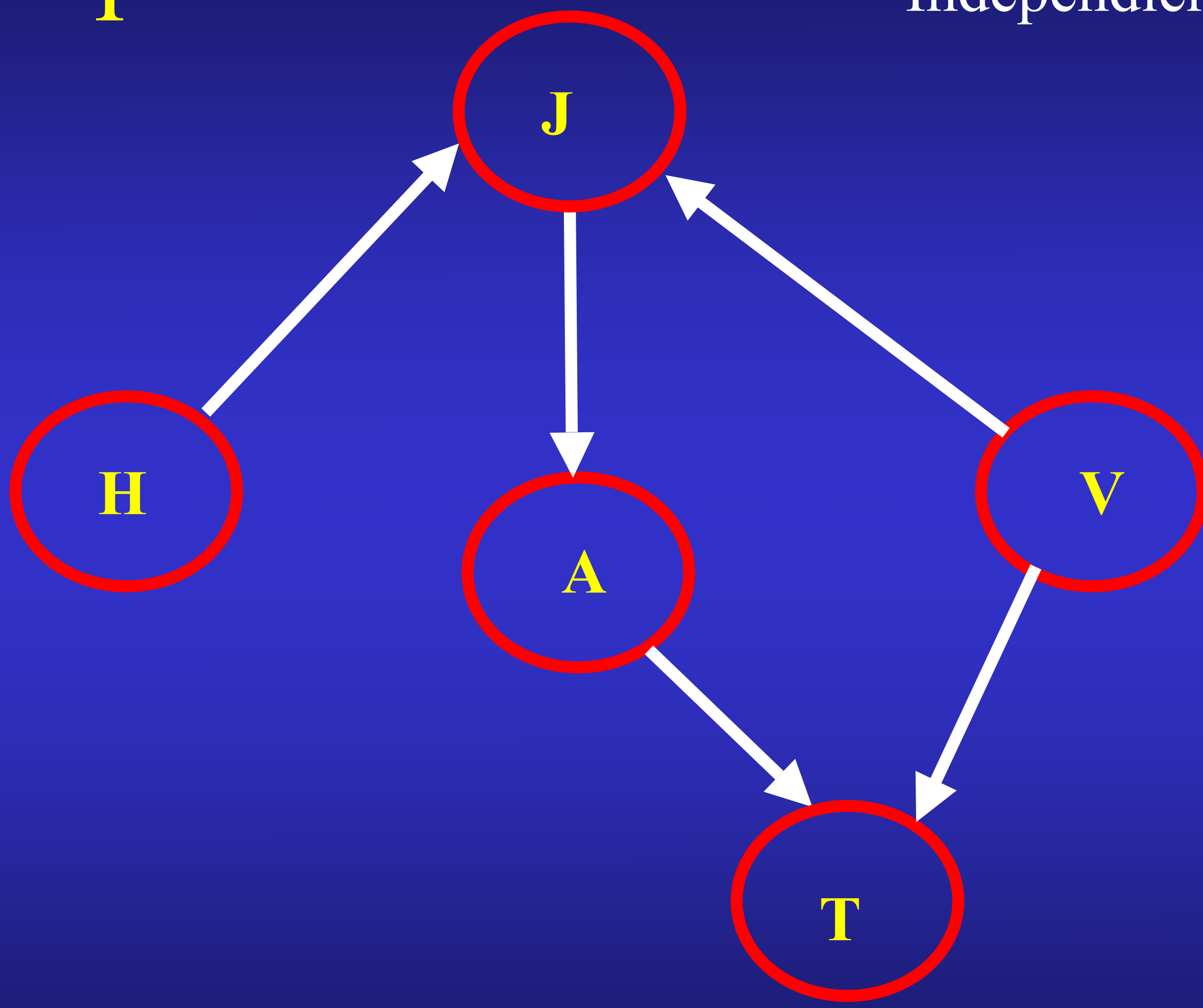


# Otras orientaciones

- En base a los arcos existentes, se orientan los demás con pruebas de independencia, evitando crear ciclos
- Si quedan al final arcos sin orientar, se direccionan en forma aleatoria, evitando ciclos

# Ejemplo

H, A son
Independientes dado J



# Tarea

- Leer sobre inferencia y aprendizaje de redes bayesianas (capítulo en la página)
- Hacer ejercicios para el jueves 10 de marzo