

Modelos Gráficos Probabilistas

L. Enrique Sucar

INAOE

Sesión 8:

Campos de Markov

“...la verdadera lógica de este mundo es el cálculo de probabilidades, que toma en cuenta la magnitud de las probabilidades, las cuales deben estar en la mente de todo hombre razonable”

James C. Maxwell, 1850

Campos de Markov

- Introducción
 - Modelo de Ising
- Representación
 - Tipos de Modelos
 - Redes de Markov
 - Vecindades
- Algoritmos
 - ICM, Metrópolis, Recocido Simulado
- Aplicaciones
 - Filtrado y recuperación de imágenes

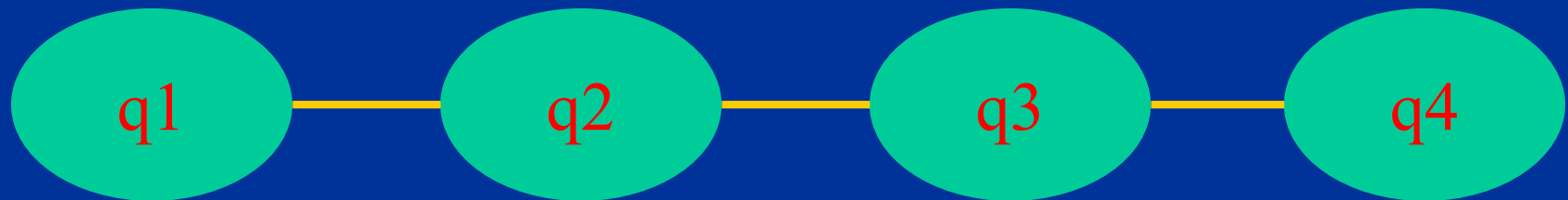
Campos de Markov (CAM)

- Ciertos procesos, como un campo magnético o una imagen, se pueden ver como una serie de estados en el espacio o una malla de estados.
- A este tipo de modelos se les conoce como *Campos de Markov*.
- Estos procesos se pueden considerar como una extensión de las cadenas de Markov en los cuales el índice de tiempo se substituye por un índice espacial

Modelo de Ising

- Surgen del problema de modelar materiales ferromagnéticos en lo que se conoce como el Modelo *Ising*.
- Se tiene una serie de dipolos en una línea que pueden estar orientados hacia “arriba” (+) o hacia “abajo” (-).
- El estado de cada dipolo se ve influenciado por los dipolos cercanos - probabilidad para cada estado depende de los estado de los puntos vecinos.

Modelo de Ising



Posibles configuraciones:

+ + + +

+ + + -

+ + - +

....

Modelo de Ising

- Un campo de Markov asigna probabilidad a cada configuración en el espacio de posibles configuraciones.
- Se considera que la probabilidad del estado de una variable es independiente de los demás dados sus 2 vecinos (para una cadena), es decir que tiene la propiedad Markoviana

$$P(S_n = q_i | S_1 = q_j, S_2 = q_k, \dots) = P(S_n = q_i | S_{n-1} = q_j, S_{n+1} = q_k)$$

Configuración más probable

- Dadas las probabilidades locales, el problema central es encontrar la probabilidad de cada una de las posibles configuraciones, y en particular cual es la configuración más probable.

- + + + +

- + + + -

- + + - +

- ...

- - - + +

- ...

- - - - -

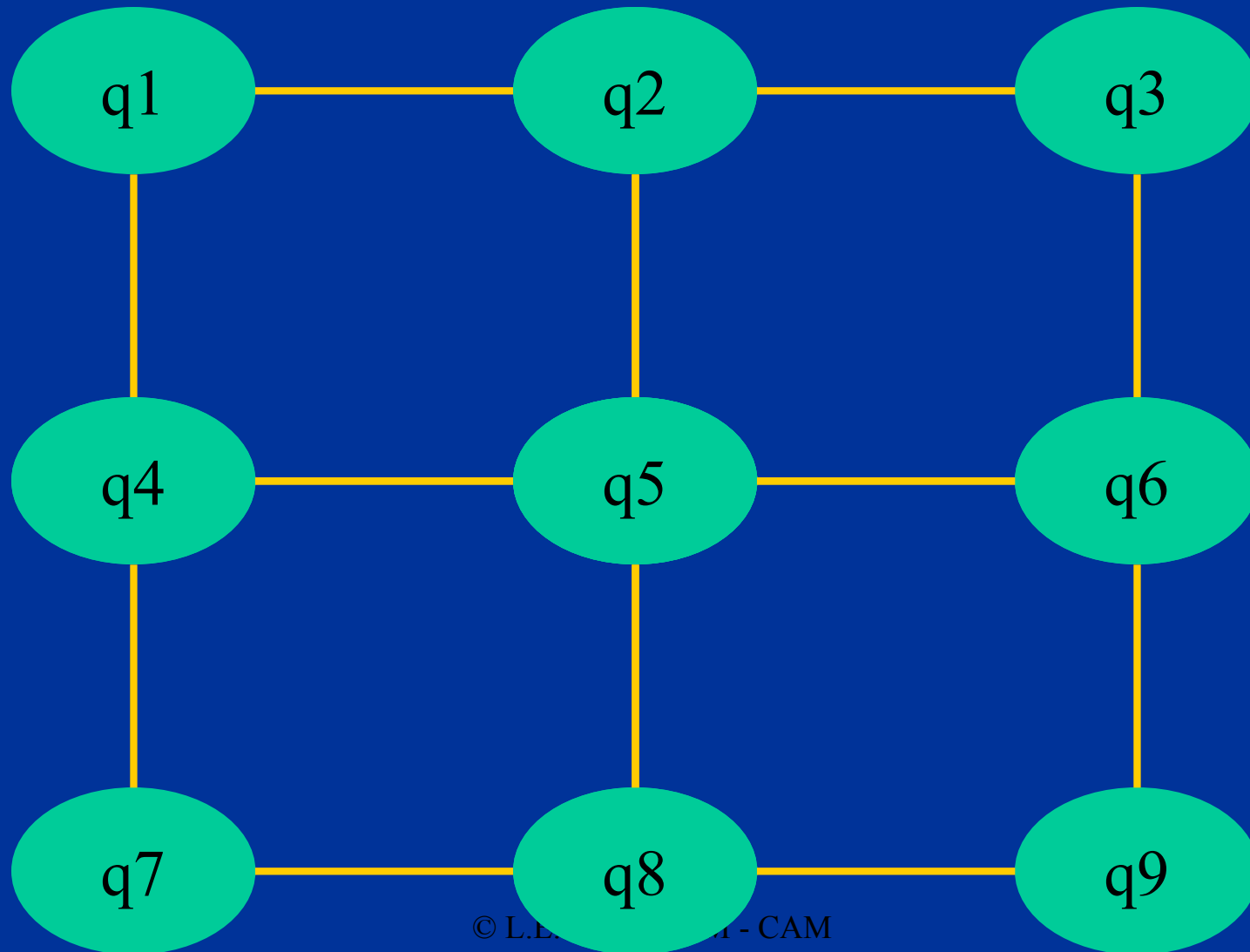
Probabilidades

- Podemos distinguir dos factores que determinan la probabilidad de una configuración:
 - la P a *priori* de cada estado,
 - la P conjunta con sus vecinos.
- En el modelo de *Ising*, estos corresponden a la influencia de un campo magnético externo, y a las interacciones entre los dipolos vecinos.

Campos en 2-D

- Esto lo podemos extender de una dimensión a dos dimensiones. En este caso tenemos una malla de puntos, donde el estado de cada punto depende del estado de sus vecinos (4, 8, etc).

Ejemplo



Representación

- Un campo aleatorio es una colección de variables aleatorias indexadas por sitios.
- Se considera un conjunto de variables aleatorias $F = \{F_1, \dots, F_M\}$, asociadas a cada sitio del sistema de sitios S . Cada variable toma un valor f_i de un conjunto de posibles valores L . Entonces F es un *campo aleatorio*.
- Un *campo aleatorio de Markov (CAM)* es un campo aleatorio con la propiedad de “localidad”.

Propiedades

Un CAM debe satisfacer las siguientes propiedades:

- $P(f) > 0$ – *positivo*
- $P(f_i | f_{s-i}) = P(f_i | \text{vec}(f_i))$

Donde $\text{vec}(f_i)$ son los vecinos de f_i

Vecindad

Un sistema de vecindad para S se define como:

$$V = \{V_i \mid \forall_i \in S\}$$

- Cumple con las siguientes dos propiedades:
 1. Un sitio no es vecino de si mismo.
 2. La relación de vecindad es mutua.
- Se pueden tener diferentes “vecindades” (primer orden, segundo orden, etc.).

Vecindad

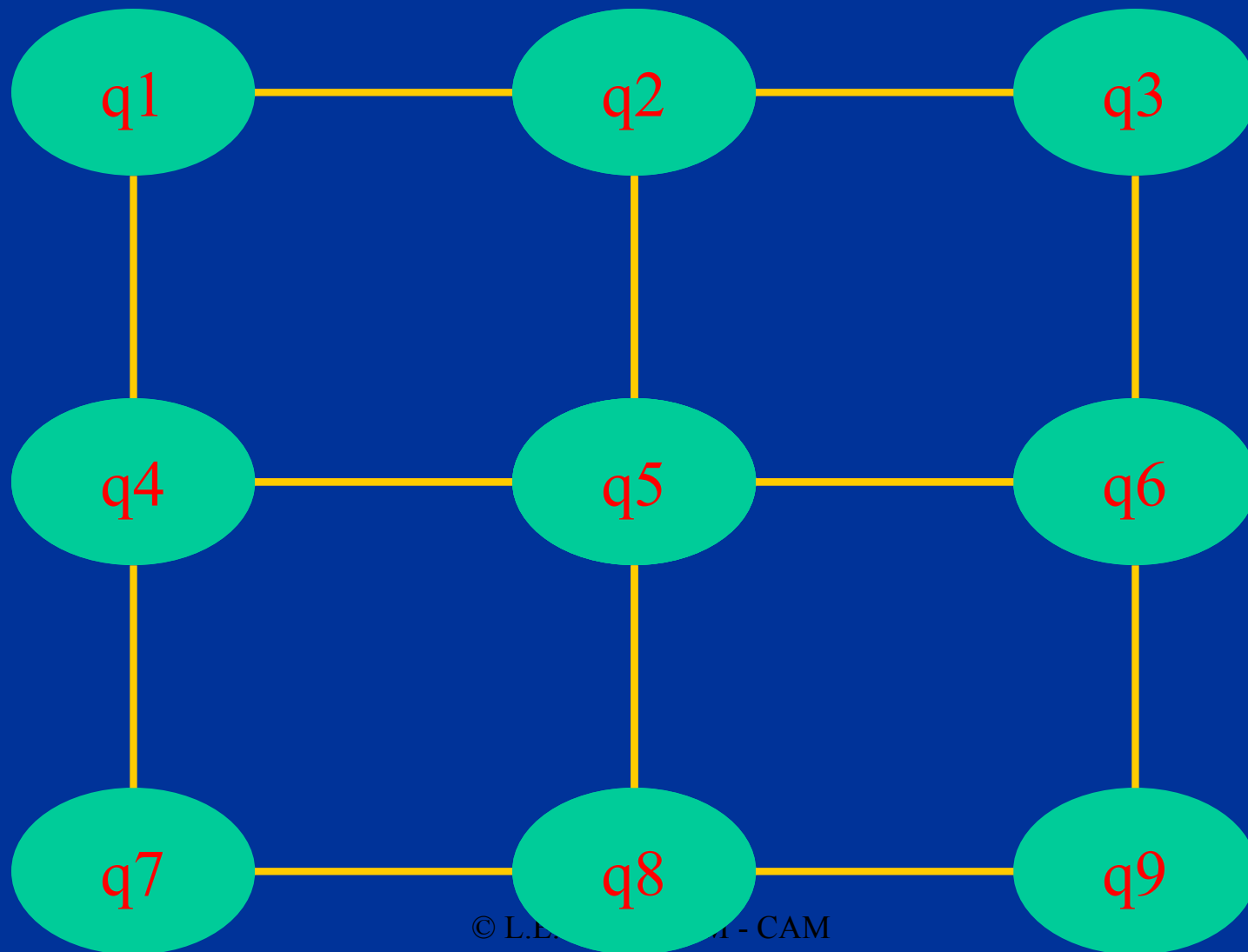
Para una malla regular, la vecindad de orden i con radio r se define como:

$$V_i = \left\{ v_{(i)} \in S \mid \text{dist}(i, v(i)) \leq r \right\}$$

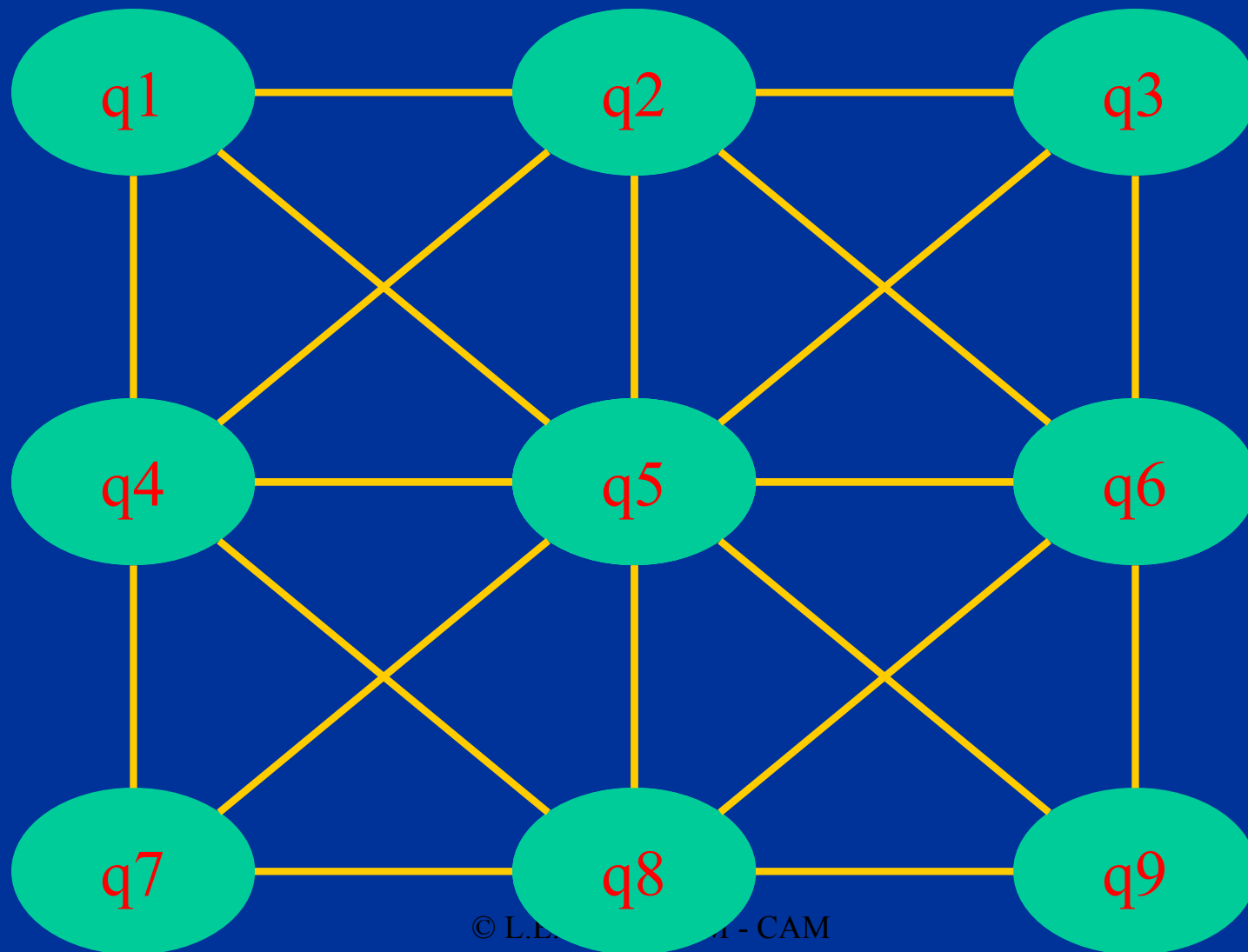
Donde $\text{dist}(x,y)$ es la distancia euclidiana entre x y y .

En un sistema de vecindad de primer orden, cada sitio (interior) tiene 4 vecinos; en uno de segundo orden, 8 vecinos: en uno de tercer orden, 12 vecinos, etc.

Ejemplo – 1er orden



Ejemplo – 2do orden



Conjuntos Completos

El conjunto de sitios junto con las vecindades nos definen un grafo no-dirigido: $G=(S, V)$.

Un *conjunto completo* (C) se define como un subconjunto de sitios de forma que están todos conectado contra todos

Un *conjunto completo* puede ser un solo sitio, C_1 , un par de sitios vecinos, C_2 , una tripleta de vecinos, C_3 , y así sucesivamente. De forma que la colección de C para un grafo esta dada por:

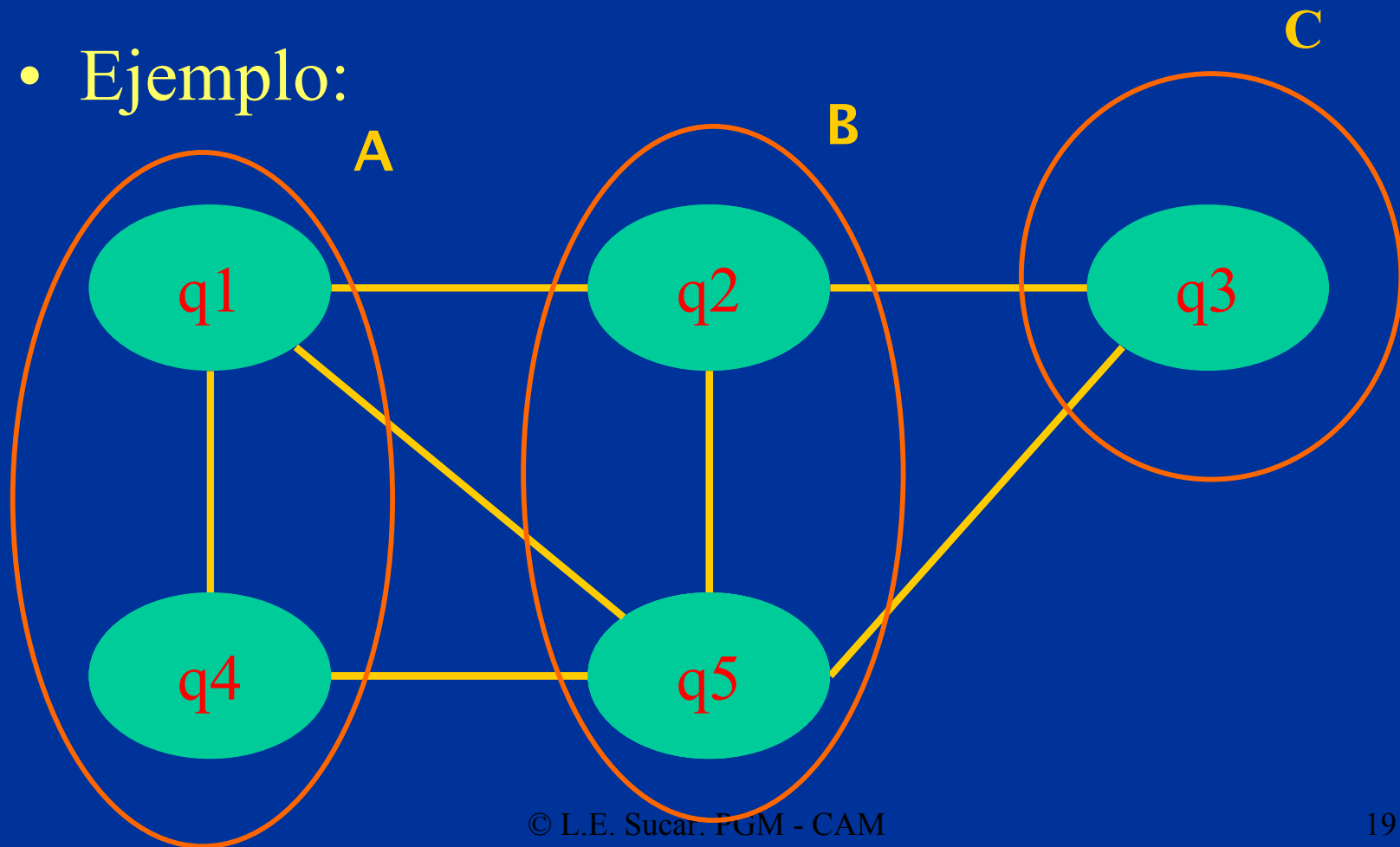
$$C = C_1 \cup C_2 \cup C_3 \cup \dots$$

Redes de Markov

- Podemos considerar los CAM como un caso especial de un modelo más general – que son las *Redes de Markov*
- Las redes de Markov (RM) son modelos gráficos probabilísticos no-dirigidos, que representa las dependencias e independencias de un conjunto de variables
- Un subconjunto A de variables es independiente de otro subconjunto C dado un tercer subconjunto B , si los nodos en B separan en el grafo A de B

Redes de Markov

- Ejemplo:



Redes de Markov

Parametrización:

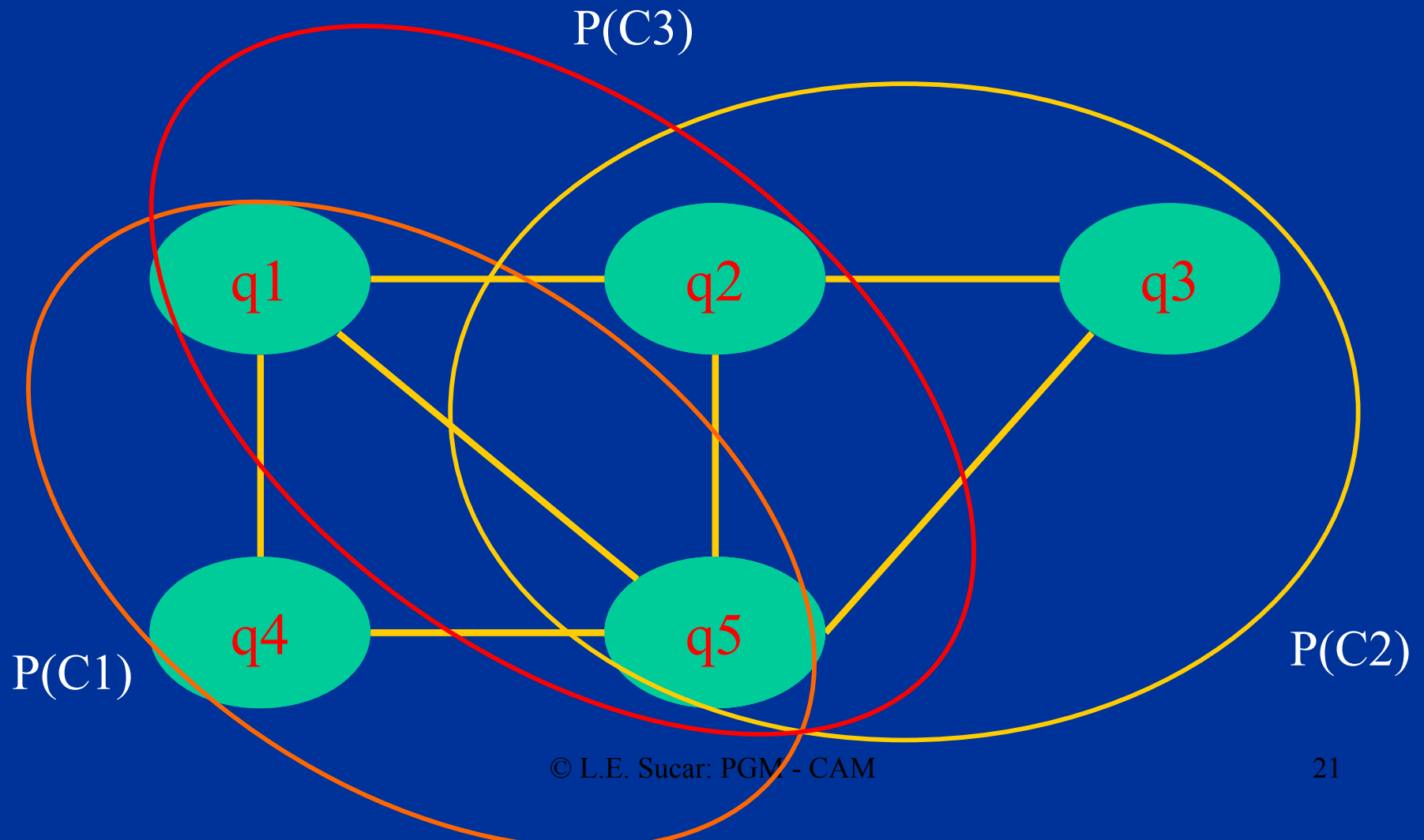
- Podemos en general representar la probabilidad conjunta de una RM como el producto de las probabilidades locales (potenciales) de conjuntos completos con un factor de normalización:

$$P(X_1, X_2, \dots, X_n) = 1/z P(C_1) P(C_2) \dots P(C_m)$$

- Es suficiente considerar sólo los *cliques*, pero también se pueden considerar otros conjuntos (esto se hace porque es conveniente en la práctica)

Redes de Markov

- Parámetros:



Configuración más probable

Una aplicación típica de CAMs es para obtener la configuración más probable dadas ciertas restricciones representadas por las probabilidades locales (potenciales)

Podemos expresar la probabilidad conjunta, como el producto de las probabilidades de las vecindades:

$$P_{F/G}(f) = k \prod_c P_c$$

Configuración más probable

Dichas probabilidades de las vecindades (potenciales) se pueden ver como “restricciones” que van a *favorecer* o *desfavorecer* ciertas configuraciones.

De esta forma, la configuración más probable se puede ver como aquella que tiene una mayor *compatibilidad* con las probabilidades locales.

Configuración más probable

Podemos expresar los potenciales en forma de exponenciales (de esta forma pueden ser cualquier número real):

$$\text{Pot}(\mathbf{x}_c) = \exp\{-U_c(\mathbf{x}_c)\}$$

Así que el producto se vuelve una suma:

$$U_p(f) = \sum U_c(\mathbf{x}_c)$$

Por lo que la probabilidad conjunta se puede expresar como:

$$P_{F/G}(f) = \left(\frac{1}{Z}\right) \exp[-U_p(f)]$$

Campo de Gibbs

Lo anterior también se puede obtener mediante una analogía entre los CAM y los Campo *Aleatorio de Gibbs* (CAG).

Una distribución de Gibbs tiene la siguiente forma (Z es una constante de normalización):

$$P(f) = \left(\frac{1}{Z} \right) \exp \left[\frac{-U(f)}{T} \right]$$

Donde:

$$Z = \sum_{f \in F} \exp \left[\frac{-U(f)}{T} \right]$$

Campo de Gibbs

$U(f)$ se conoce como la función de energía y se obtiene como la suma de los *potenciales* de todos los C:

$$U(f) = \sum c V_c(f)$$

La configuración más probable corresponde a la de energía mínima.

La función de energía se puede expresar en términos de los C de cada tamaño:

$$U(f) = \sum_{c1} V_1(f_i) + \sum_{c2} V_2(f_i, f_j) + \sum_{c3} V_3(f_i, f_j, f_k) + \dots$$

Campo de Gibbs

Se puede demostrar que un CAM y el correspondiente CAG son equivalentes.

Esto permite que se pueda expresar la probabilidad conjunta especificando los potenciales de los C. En la práctica se seleccionan los potenciales de acuerdo al comportamiento deseado – con los potenciales se codifica el conocimiento *a priori* del problema.

Configuración más probable

- Entonces, para especificar un CAM se requiere:
 - Definir el esquema de vecindad
 - Especificar las probabilidades (potenciales) para cada uno de los conjuntos completos de nodos

Para el caso de vecindad de primer orden:

$$U_p(f) = \sum_c V_c(f) + \lambda \sum_o V_o(f)$$

Configuración más probable

- V_C corresponde a P_F o la información del dominio dada por los vecinos y V_O corresponde a $P_{G/F}$ o la información de las observaciones; λ es una constante que da el peso relativo entre ambas.
- Bajo este enfoque, la solución a un problema particular corresponde en encontrar la configuración del CAM de mayor probabilidad o de “energía” (U_P) mínima. La función que se logre depende de la forma de las funciones para V_C y V_O .

Ejemplo

- Por ejemplo, podemos querer “suavizar” una imagen; es decir, minimizar la “distancia” de cada pixel a sus vecinos, pero también mantenerlo cercano a su valor en la imagen (observación):

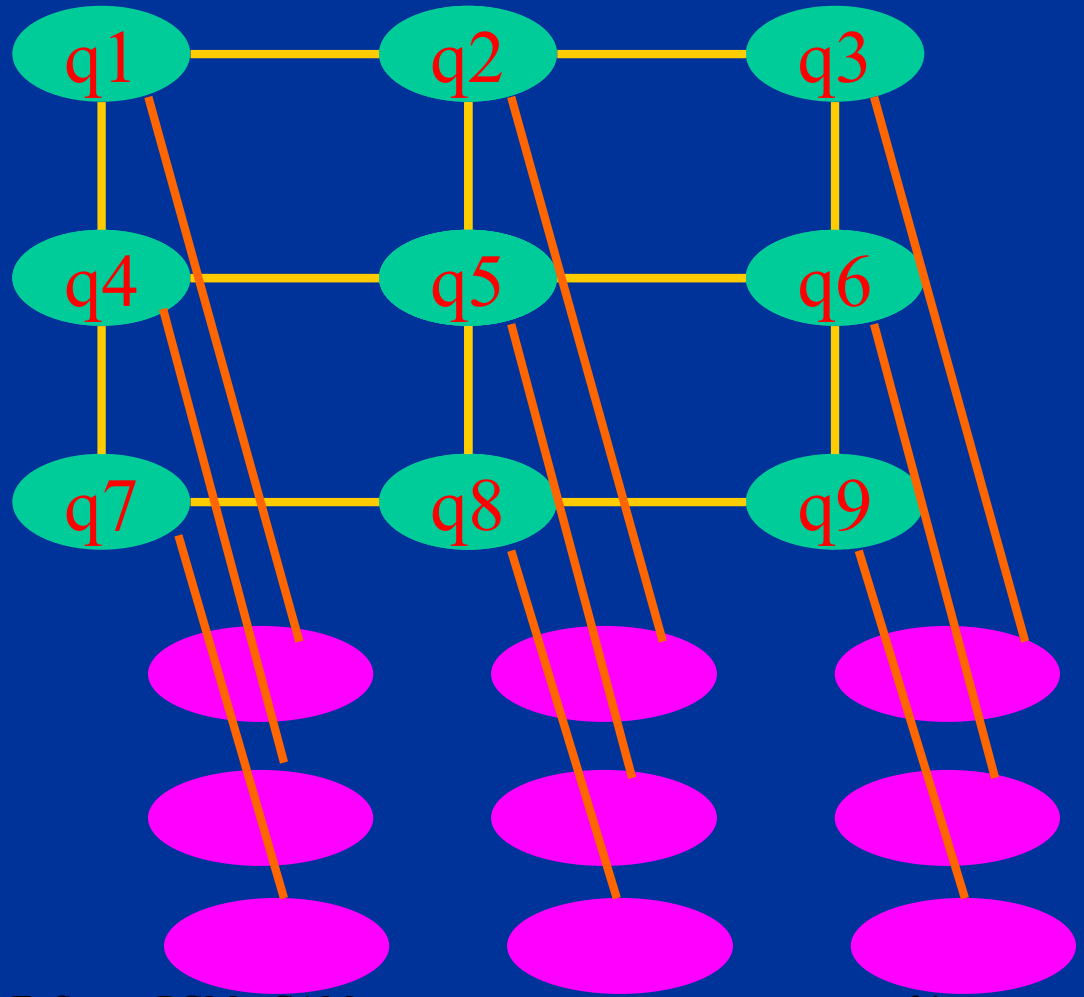
$$V_c(f) = (f - u)^2$$

$$V_o(f) = (f - g)^2$$

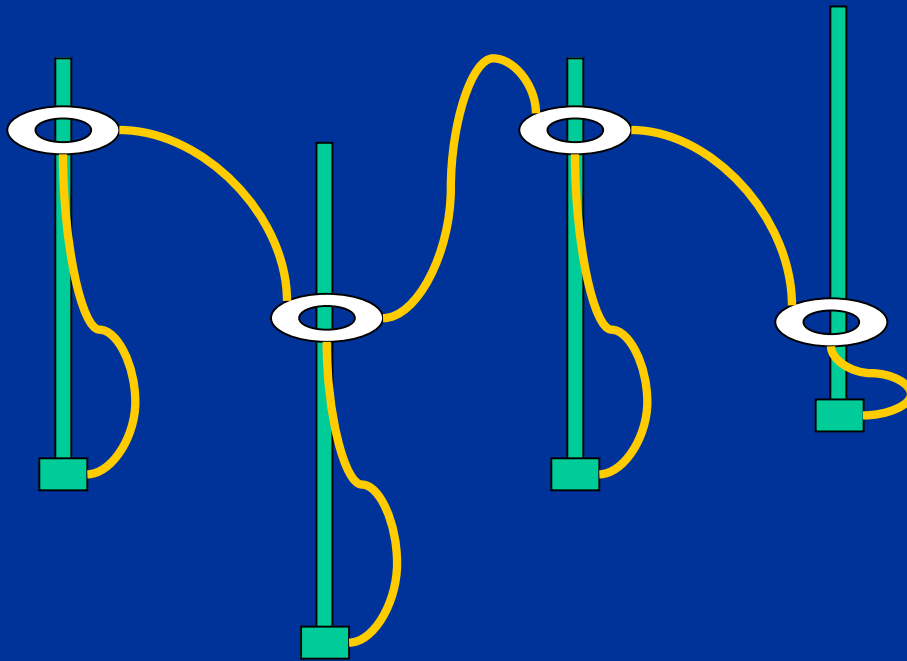
Ejemplo

F
imagen “suavizada”

G
imagen “original”

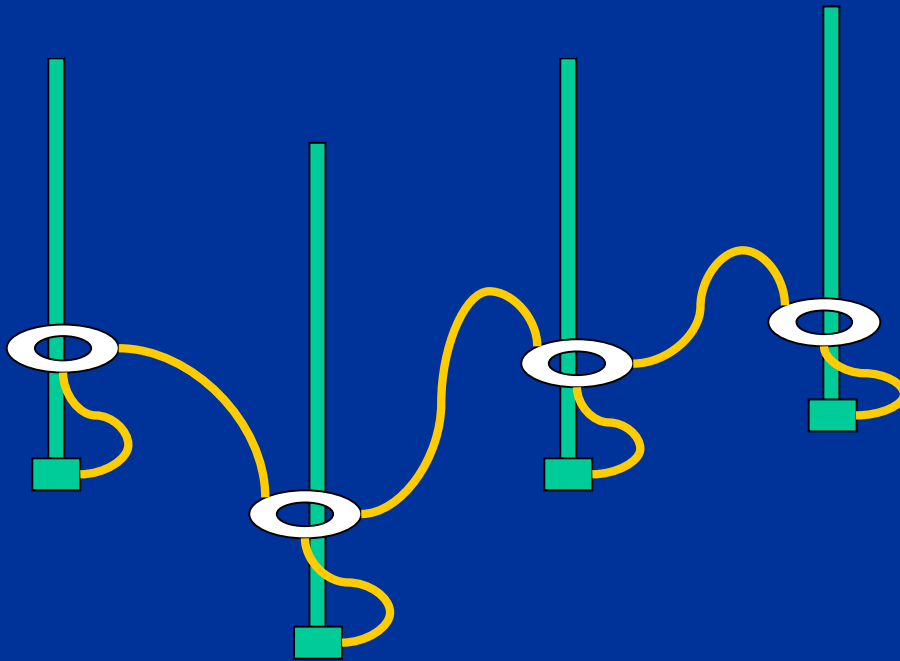


Analogía Física



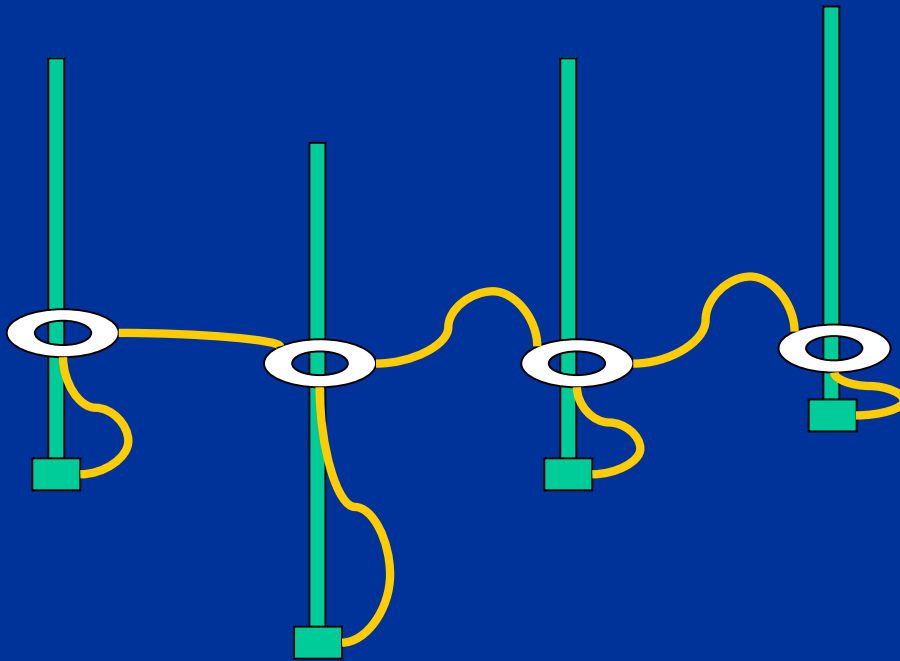
Analogía Física

- más peso a las observaciones -



Analogía Física

- más peso a los vecinos -



Algoritmos

- El obtener la configuración de mayor probabilidad (mínima energía) en forma directa (exhaustiva) prácticamente imposible (excepto para problemas muy pequeños), por lo que se plantea como un problema de *búsqueda*. Se busca la configuración de mayor probabilidad, sin tener que calcular directamente las probabilidades de cada configuración.

Algoritmos

El problema incluye 3 aspectos:

1. Representación: CAM con un esquema de vecindad y los potenciales asociados
2. Función objetivo: Función de energía a minimizar.
3. Algoritmo de optimización: simulación estocástica (Metropolis)

Algoritmo básico

- Inicializar con un valor aleatorio cada variable.
- Repetir para cada variable en el campo:
 - Calcular el valor de energía (potencial) de cada variable en base a la función deseada y los valores de los vecinos.
 - Si el valor de energía es menor al anterior cambiar de valor.
 - Si no, con cierta probabilidad también cambiar de valor.
- Hasta que se cumplan N iteraciones o ya no haya cambios (convergencia)
- Obtener configuración “óptima”

Variantes

Cálculo del óptimo:

- MAP: se toma el valor para cada variable al final de las iteraciones.
- MPM: se toma el valor de mayor frecuencia de acuerdo a su ocurrencia durante la simulación.

Variantes

Forma de optimización:

- **Metrópolis:** con probabilidad fija se puede pasar a estados de mayor energía.
- **Recocido simulado:** se va disminuyendo la probabilidad de pasar a estados de mayor energía (temperatura).
- **ICM:** tomar siempre el estado de menor energía.

Probabilidades de transición

- Probabilidad de transición a estado de mayor energía:

$$P = e^{-\delta V / T}$$

- Donde δV es la diferencia de energía y T es la *temperatura*

Ejemplo

- Dada la siguiente imagen:
- Consideramos:
 - Configuración inicial de ceros
 - Vecindad de primer orden
 - Potenciales de suavizamiento con $\lambda=4$

| | | | |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 0 | 1 | 0 | 1 |
| 0 | 1 | 1 | 1 |

Ejemplo

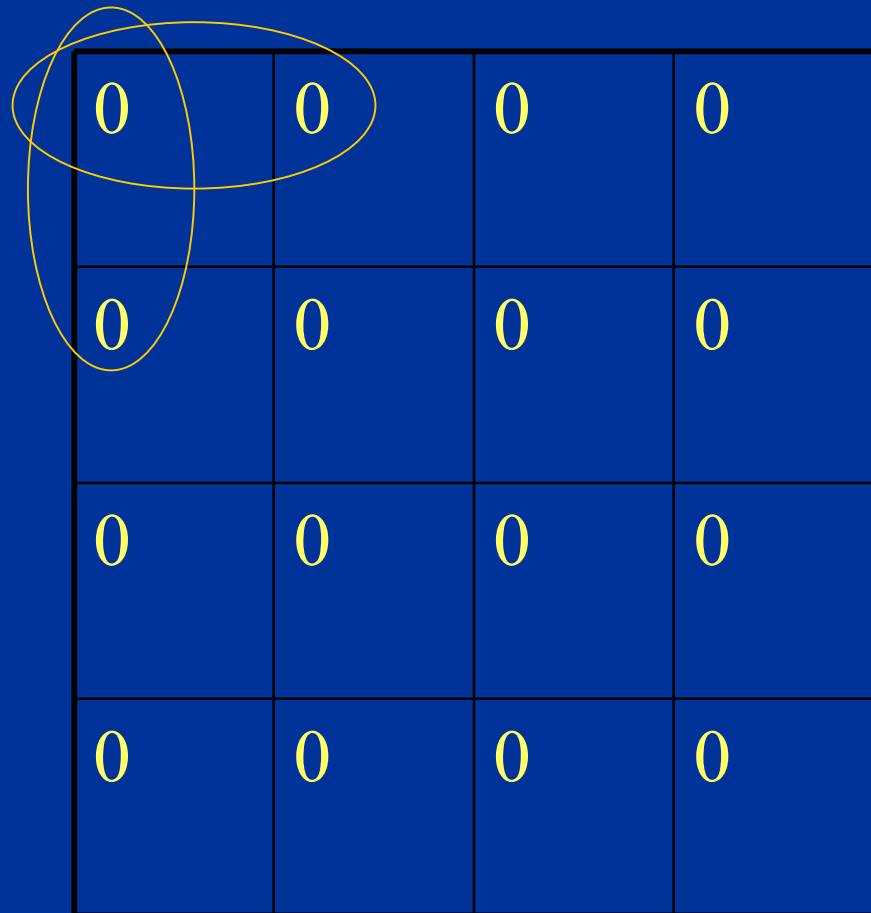
1era iteración:

- 1er sitio:

$$V1(0) = 0$$

$$V1(1) = 2 + 4(1) = 6$$

- ...
- ...



| | | | |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |

Ejemplo

- ...
- ...
- 11vo sitio:
 $V_{11}(0) = 2$
 $V_{11}(1) = 6$
- ...

| | | | |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 |

Ejemplo

2da iteración

- ...
- 11ovo sitio:
 $V_{11}(0) = 4$
 $V_{11}(1) = 4$
- ...

| | | | |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 |

Aplicaciones

- Procesamiento de imágenes
 - Eliminación de ruido
 - Filtrado de una imagen
 - Segmentación de texturas
 - Visión estereoscópica
 - Recuperación de forma (3-D)
 - Reconocimiento de objetos
 - Etiquetado de imágenes

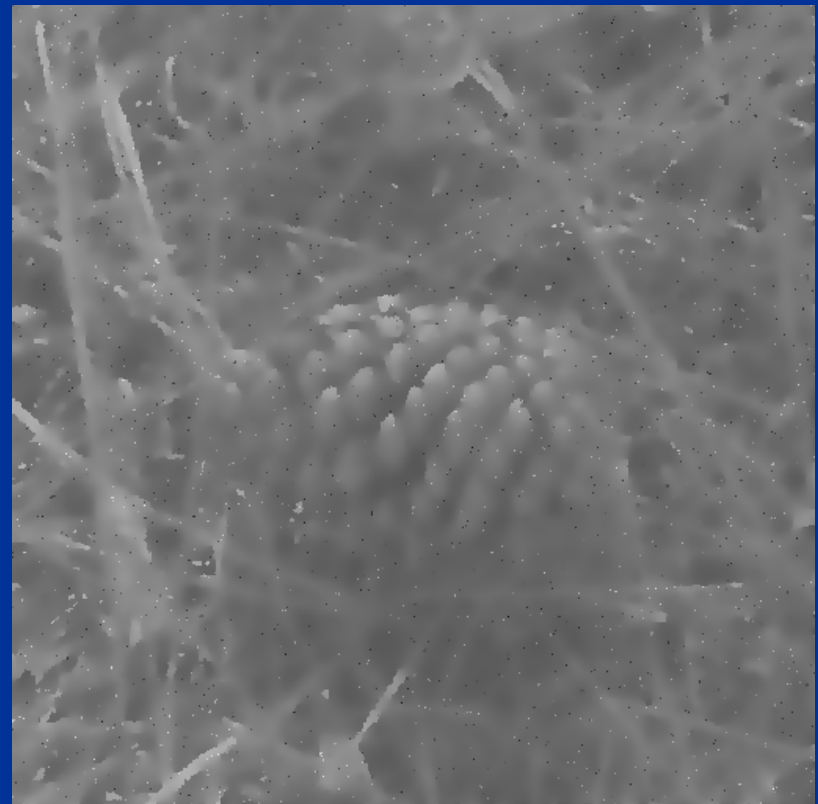
Ejemplo – eliminación de ruido



Ejemplo – eliminación de ruido



Ejemplo – eliminación de ruido



Otro ejemplo: anotación de imágenes

| | | | |
|----------|-----|----------|-----|
| Grass | 0.6 | People | 0.4 |
| Tree | 0.2 | Tree | 0.3 |
| rock | 0.1 | Mountain | 0.2 |
| building | 0.1 | Jet | 0.1 |



| | |
|----------|-----|
| Church | 0.3 |
| Grass | 0.3 |
| Sky | 0.2 |
| Elephant | 0.2 |

| | |
|-------|-----|
| Tree | 0.5 |
| Grass | 0.3 |
| Sky | 0.1 |
| Jet | 0.1 |

Considerar relaciones espaciales entre regiones

| | R_1 | R_2 | R_3 | R_4 |
|-----------|----------|--------|-------|----------|
| C_1 | grass | people | tree | church |
| C_2 | grass | tree | tree | church |
| $C_{..}$ | | | | |
| C_i | grass | tree | tree | elephant |
| $C_{..}$ | | | | |
| C_j | tree | people | tree | church |
| C_{256} | building | Jet | jet | elephant |

Usar CAMs para obtener mejor "configuración"

Campos de Markov

- Combinar probabilidad previa (clasificador) con la probabilidad de la etiqueta dados sus vecinos (relaciones espaciales)

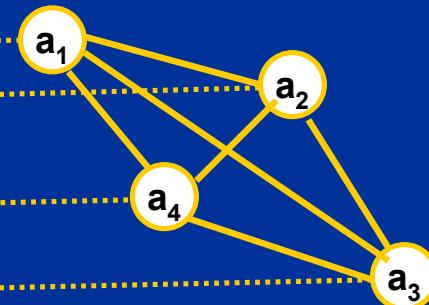
| | |
|----------|-----|
| Grass | 0.6 |
| Tree | 0.2 |
| rock | 0.1 |
| building | 0.1 |

| | |
|----------|-----|
| Church | 0.3 |
| Grass | 0.3 |
| Sky | 0.2 |
| Elephant | 0.2 |



| | |
|----------|-----|
| People | 0.4 |
| Tree | 0.3 |
| Mountain | 0.2 |
| Jet | 0.1 |

| | |
|-------|-----|
| Tree | 0.5 |
| Grass | 0.3 |
| Sky | 0.1 |
| Jet | 0.1 |



Función de energía

- Considera el potencial inicial de acuerdo a las características de la región (V_O) y los potenciales dados por sus vecinos para 3 tipos de relaciones espaciales:
 - Topológicas (V_T)
 - Horizontales (V_H)
 - Verticales (V_V)

$$U_p(f) = \alpha_1 V_T(f) + \alpha_2 V_H(f) + \alpha_3 V_V(f) + \lambda \sum_o V_o(f)$$

Potenciales

- Los potenciales son inversamente proporcionales a las probabilidades de cada relación:

$$V_T(f) = \frac{1}{\sum_c P_{1c}(f) \oplus P_{2c}(f)}$$
$$V_H(f) = \frac{1}{\sum_c P_{3c}(f) \oplus P_{4c}(f)}$$
$$V_V(f) = \frac{1}{\sum_c P_{5c}(f) \oplus P_{6c}(f) \oplus P_{7c}(f)}$$

Resultados

| Algorithm | Relation group | Smoothing | Accuracy |
|-----------|----------------|---------------------|---------------|
| kNN | None | None | 36.81% |
| MRFs | Topological | None | 42.72% |
| | | Laplacian | 43.51% |
| | | Expert info. | 43.25% |
| | Horizontal | None | 41.72% |
| | | Laplacian | 43.08% |
| | | Expert info. | 43.58% |
| | Vertical | None | 43.73% |
| | | Laplacian | 44.93% |
| | | Expert info. | 44.88% |
| | All | None | 43.29% |
| | | Laplacian | 45.41% |
| | | Expert info. | 45.64% |

Referencias

- Li, “Markov Random Fields Models in Computer Vision”, Springer-Verlag
- Chellapa, Jain, “Markov Random Fields: Theory and Models”, Academic Press.
- C. Hernández, L.E. Sucar, “Markov Random Fields and Spatial Information to Improve Automatic Image Annotation”, Advances in Image and Video Technology, Lecture Notes in Computer Science 4872, Springer-Verlag , pp. 879--892, 2007.
- [Koller & Friedman]: Capítulo 4

Actividades

- Hacer ejercicio de Campos de Markov