

# Introduction to Markov Random Fields

Sarah Michele Rajtmajer

Applied Research Lab, Penn State University

21 February 2012

### *Stochastic processes as Dynamic Bayesian Networks*

A **Dynamic Bayesian Network** is a probabilistic **graphical model** that represents a sequence of random variables and their conditional dependencies.

A **Markov Chain** is a simple Dynamic Bayesian Network with the Markov property.

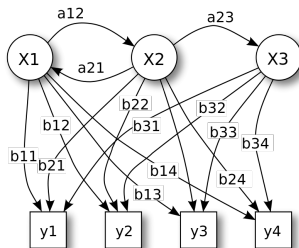
$$p(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p(X_{n+1} = x \mid X_n = x_n)$$



Random walks (Graph Theory), Thermodynamics, Enzyme activity (Chemistry), Data compression and pattern recognition (Information Science), Google's PageRank, Asset pricing (Finance), Population processes (Biology), Algorithmic music composition, Baseball statistics, Text generation...

## Latent Variables and Hidden Markov Models

A **Hidden Markov Model** is another example of a Dynamic Bayesian Network.

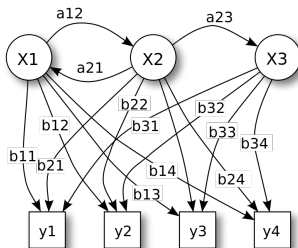


### Inference tasks for HMMs

- *Filtering*: Given model parameters and sequence of observations, compute distribution over hidden states,  $p(X_t | y(1), \dots, y(t))$ .
- *Smoothing*: Given model parameters and sequence of observations, compute distribution over hidden states for point in time in the past,  $p(X_k | y(1), \dots, y(t))$ , for  $k < t$ .
- *Probability of an observed sequence*
- *Most likely explanation*

## Latent Variables and Hidden Markov Models

A **Hidden Markov Model** is another example of a Dynamic Bayesian Network.

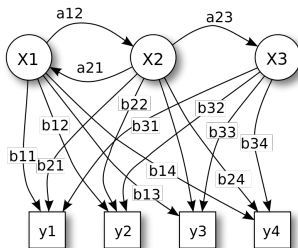


### Inference tasks for HMMs

- *Filtering*: Given model parameters and sequence of observations, compute distribution over hidden states,  $p(X_t | y(1), \dots, y(t))$ .
- *Smoothing*: Given model parameters and sequence of observations, compute distribution over hidden states for point in time in the past,  $p(X_k | y(1), \dots, y(t))$ , for  $k < t$ .
- *Probability of an observed sequence*
- *Most likely explanation*

## Latent Variables and Hidden Markov Models

A **Hidden Markov Model** is another example of a Dynamic Bayesian Network.

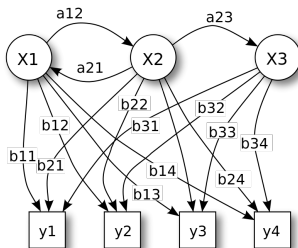


### Inference tasks for HMMs

- *Filtering*: Given model parameters and sequence of observations, compute distribution over hidden states,  $p(X_t | y(1), \dots, y(t))$ .
- *Smoothing*: Given model parameters and sequence of observations, compute distribution over hidden states for point in time in the past,  $p(X_k | y(1), \dots, y(t))$ , for  $k < t$ .
- *Probability of an observed sequence*
- *Most likely explanation*

## Latent Variables and Hidden Markov Models

A **Hidden Markov Model** is another example of a Dynamic Bayesian Network.



### Inference tasks for HMMs

- *Filtering*: Given model parameters and sequence of observations, compute distribution over hidden states,  $p(X_t | y(1), \dots, y(t))$ .
- *Smoothing*: Given model parameters and sequence of observations, compute distribution over hidden states for point in time in the past,  $p(X_k | y(1), \dots, y(t))$ , for  $k < t$ .
- *Probability of an observed sequence*
- *Most likely explanation*

## Properties of Bayesian Networks

- Given  $G$  a directed, acyclic graph over random variables  $X_1, \dots, X_n$ .
- Let  $X_{\pi_i}$  be the set of parents for node  $X_i$ .
- We associate with each node the conditional probability distribution of  $X_i$  given its parents:  $p(X_i | X_{\pi_i})$ .
- Joint probability distribution  $p$  **factorises according to**  $G$  if  $p$  can be expressed as

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{\pi_i}).$$

- Individual factors are  $p(x_i | x_{\pi_i})$  are called **conditional probability distributions**.

Cryptanalysis (Mathematics), Speech recognition, Part-of-speech tagging (Natural Language Processing), Gene prediction, Protein folding, Bio-sequencing (Biology)...

## Properties of Bayesian Networks

- Given  $G$  a directed, acyclic graph over random variables  $X_1, \dots, X_n$ .
- Let  $X_{\pi_i}$  be the set of parents for node  $X_i$ .
- We associate with each node the conditional probability distribution of  $X_i$  given its parents:  $p(X_i | X_{\pi_i})$ .
- Joint probability distribution  $p$  **factorises according to**  $G$  if  $p$  can be expressed as

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{\pi_i}).$$

- Individual factors are  $p(x_i | x_{\pi_i})$  are called **conditional probability distributions**.

Cryptanalysis (Mathematics), Speech recognition, Part-of-speech tagging (Natural Language Processing), Gene prediction, Protein folding, Bio-sequencing (Biology)...



## Properties of Bayesian Networks

- Given  $G$  a directed, acyclic graph over random variables  $X_1, \dots, X_n$ .
- Let  $X_{\pi_i}$  be the set of parents for node  $X_i$ .
- We associate with each node the conditional probability distribution of  $X_i$  given its parents:  $p(X_i | X_{\pi_i})$ .
- Joint probability distribution  $p$  **factorises according to**  $G$  if  $p$  can be expressed as

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{\pi_i}).$$

- Individual factors are  $p(x_i | x_{\pi_i})$  are called **conditional probability distributions**.

Cryptanalysis (Mathematics), Speech recognition, Part-of-speech tagging (Natural Language Processing), Gene prediction, Protein folding, Bio-sequencing (Biology)...

## Properties of Bayesian Networks

- Given  $G$  a directed, acyclic graph over random variables  $X_1, \dots, X_n$ .
- Let  $X_{\pi_i}$  be the set of parents for node  $X_i$ .
- We associate with each node the conditional probability distribution of  $X_i$  given its parents:  $p(X_i | X_{\pi_i})$ .
- Joint probability distribution  $p$  **factorises according to**  $G$  if  $p$  can be expressed as

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{\pi_i}).$$

- Individual factors are  $p(x_i | x_{\pi_i})$  are called **conditional probability distributions**.

Cryptanalysis (Mathematics), Speech recognition, Part-of-speech tagging (Natural Language Processing), Gene prediction, Protein folding, Bio-sequencing (Biology)...

## Properties of Bayesian Networks

- Given  $G$  a directed, acyclic graph over random variables  $X_1, \dots, X_n$ .
- Let  $X_{\pi_i}$  be the set of parents for node  $X_i$ .
- We associate with each node the conditional probability distribution of  $X_i$  given its parents:  $p(X_i | X_{\pi_i})$ .
- Joint probability distribution  $p$  **factorises according to**  $G$  if  $p$  can be expressed as

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{\pi_i}).$$

- Individual factors are  $p(x_i | x_{\pi_i})$  are called **conditional probability distributions**.

Cryptanalysis (Mathematics), Speech recognition, Part-of-speech tagging (Natural Language Processing), Gene prediction, Protein folding, Bio-sequencing (Biology)...

## Properties of Bayesian Networks

- Given  $G$  a directed, acyclic graph over random variables  $X_1, \dots, X_n$ .
- Let  $X_{\pi_i}$  be the set of parents for node  $X_i$ .
- We associate with each node the conditional probability distribution of  $X_i$  given its parents:  $p(X_i | X_{\pi_i})$ .
- Joint probability distribution  $p$  **factorises according to**  $G$  if  $p$  can be expressed as

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{\pi_i}).$$

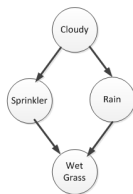
- Individual factors are  $p(x_i | x_{\pi_i})$  are called **conditional probability distributions**.

Cryptanalysis (Mathematics), Speech recognition, Part-of-speech tagging (Natural Language Processing), Gene prediction, Protein folding, Bio-sequencing (Biology)...

## Goals of graphical models

- Provide *compact factorisations* of large, joint probability distributions.
- Use *local functions* which exploit *conditional independencies* in the models.

Consider the example.



By the chain rule, the joint probability is given by

$$P(C, S, R, W) = P(C) \times P(S | C) \times P(R | C, S) \times P(W | C, S, R).$$

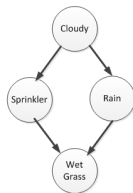
By using conditional independence relationships we can rewrite this as

$$P(C, S, R, W) = P(C) \times P(S | C) \times P(R | C) \times P(W | S, R).$$

## Goals of graphical models

- Provide *compact factorisations* of large, joint probability distributions.
- Use *local functions* which exploit *conditional independencies* in the models.

Consider the example.



By the chain rule, the joint probability is given by

$$P(C, S, R, W) = P(C) \times P(S \mid C) \times P(R \mid C, S) \times P(W \mid C, S, R).$$

By using conditional independence relationships we can rewrite this as

$$P(C, S, R, W) = P(C) \times P(S \mid C) \times P(R \mid C) \times P(W \mid S, R).$$

## Markov Random Fields as Undirected Graphical Models

A **Markov Random Field** is an *undirected* probabilistic graphical model representing random variables and their conditional dependencies.

Given  $G = (V, E)$  undirected graph over random variables  $(X_v)_{v \in V}$ .

### Pairwise Markov Property

$$X_u \perp X_v \mid X_{V \setminus \{u, v\}}, \text{ for } \{u, v\} \notin E$$

Any two non-adjacent variables are conditionally independent given all other variables.

### Local Markov Property

$$X_v \perp X_{V \setminus v \cup N(v)} \mid X_{N(v)}, \text{ for } N(v) = \text{neighbors of } v$$

A variable is conditionally independent of all other variables given its neighbours.

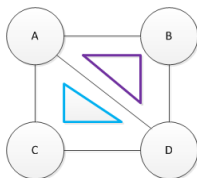
### Global Markov Property

$$X_A \perp X_B \mid X_S, \text{ where every path from } A \text{ to } B \text{ passes through } S$$

Any two subsets of variables are conditionally independent given a separating subset.

- In the directed model, we had a local conditional probability distributions at each node, depending on nodes' parents. These served as the factors of the joint probability distribution.
- What is our equivalent for the undirected model?

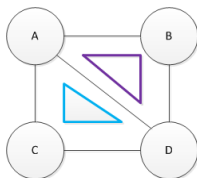
→ Define local factors on (maximal) **cliques**.





- In the directed model, we had a local conditional probability distributions at each node, depending on nodes' parents. These served as the factors of the joint probability distribution.
- What is our equivalent for the undirected model?

→ Define local factors on (maximal) **cliques**.



Represent the joint probability distribution as a **product of clique potentials**

$$p(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \prod_{c_i \in C} \psi_i(c_i),$$

where  $c_i$  is a clique in the set of all cliques  $C$  in the graph and  $\psi_i(c_i)$  is the  $i$ th clique potential, a function of only the values of the clique members in  $c_i$ . Each potential function  $\psi_i$  must be positive, but unlike probability distribution functions they need not sum to 1. Normalization constant  $Z$  is therefore required in order to create a valid probability distribution

$$Z = \sum_x \prod_{c_i \in C} \psi_i(c_i).$$

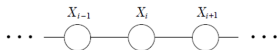
(Hammersley-Clifford Theorem) Every MRF can be specified via clique potentials.

## Interpretation of clique potentials

Often, clique potentials take the form  $\psi_i(c_i) = \exp(-f_i(c_i))$  with  $f_i(c_i)$  an energy function over values  $c_i$ . The energy is an indicator of the likelihood of the corresponding relationships within the clique, with higher energy configuration having lower probability. The joint p.d. becomes

$$p(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \exp \left[ - \sum_{c_i \in C} f_i(c_i) \right].$$

Example: *Spin Glass model*



		$x_i$	
		-1	1
-1	1.5	0.2	
1	0.2	1.5	

		$x_{i+1}$	
		-1	1
-1	1.5	0.2	
1	0.2	1.5	

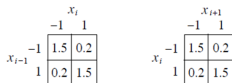
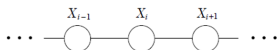
- A spin glass is a collection of magnetic moments (spins) whose low temperature state is disordered.
- Spins can be in one of two states,  $(+1, -1)$ .
- There is competition among interactions between moments (so-called *frustration*).
- Interactions are at least partially random.

## Interpretation of clique potentials

Often, clique potentials take the form  $\psi_i(c_i) = \exp(-f_i(c_i))$  with  $f_i(c_i)$  an energy function over values  $c_i$ . The energy is an indicator of the likelihood of the corresponding relationships within the clique, with higher energy configuration having lower probability. The joint p.d. becomes

$$p(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \exp \left[ - \sum_{c_i \in C} f_i(c_i) \right].$$

Example: *Spin Glass model*



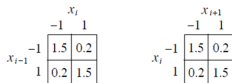
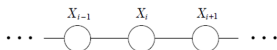
- A spin glass is a collection of magnetic moments (spins) whose low temperature state is disordered.
- Spins can be in one of two states,  $(+1, -1)$ .
- There is competition among interactions between moments (so-called *frustration*).
- Interactions are at least partially random.

## Interpretation of clique potentials

Often, clique potentials take the form  $\psi_i(c_i) = \exp(-f_i(c_i))$  with  $f_i(c_i)$  an energy function over values  $c_i$ . The energy is an indicator of the likelihood of the corresponding relationships within the clique, with higher energy configuration having lower probability. The joint p.d. becomes

$$p(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \exp \left[ - \sum_{c_i \in C} f_i(c_i) \right].$$

Example: *Spin Glass model*



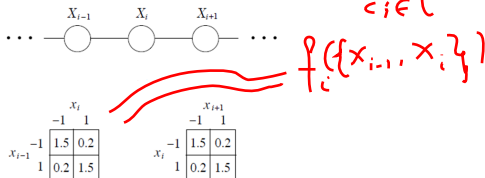
- A spin glass is a collection of magnetic moments (spins) whose low temperature state is disordered.
- Spins can be in one of two states,  $(+1, -1)$ .
- There is competition among interactions between moments (so-called *frustration*).
- Interactions are at least partially random.

## Interpretation of clique potentials

Often, clique potentials take the form  $\psi_i(c_i) = \exp(-f_i(c_i))$  with  $f_i(c_i)$  an energy function over values  $c_i$ . The energy is an indicator of the likelihood of the corresponding relationships within the clique, with higher energy configuration having lower probability. The joint p.d. becomes

$$p(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \exp \left[ - \sum_{c_i \in C} f_i(c_i) \right].$$

Example: Spin Glass model  $E(X_1 = x_1, \dots, X_n = x_n) = \sum_{c_i \in C} f_i(c_i)$



- A spin glass is a collection of magnetic moments (spins) whose low temperature state is disordered.
- Spins can be in one of two states,  $(+1, -1)$ .
- There is competition among interactions between moments (so-called *frustration*).
- Interactions are at least partially random.

## The Ising Model (2-D Markov Random Field)

+	+	+	+	+	+	+	+
+	-	+	-	-	-	-	+
+	-	-	+	+	+	+	+
+	+	-	+	-	-	-	+
+	+	+	+	+	+	+	+

- Spins are arranged in a ( $d$ -dimensional) lattice, and each spin interacts with its nearest neighbors.
- Energy is given by  $H(\mathbf{x}) = \sum_{ij} \beta_{ij} x_i x_j + \sum_i \alpha_i x_i$ .
- Goal: Find phase transitions.

The Ising model is very popular for explaining the effect of "society" or "environment" on a "component" or "individual". Example applications from flocking behaviour, behaviour of neural networks, sociology...

## The Ising Model (2-D Markov Random Field)

+	+	+	+	+	+	+	+
+	-	+	-	-	-	-	+
+	-	-	+	+	+	+	+
+	+	-	+	-	-	-	+
+	+	+	+	+	+	+	+

- Spins are arranged in a ( $d$ -dimensional) lattice, and each spin interacts with its nearest neighbors.
- Energy is given by  $H(\mathbf{x}) = \sum_{ij} \beta_{ij} x_i x_j + \sum_i \alpha_i x_i$ .
- Goal: Find phase transitions.

The Ising model is very popular for explaining the effect of "society" or "environment" on a "component" or "individual". Example applications from flocking behaviour, behaviour of neural networks, sociology...



## The Ising Model (2-D Markov Random Field)

+	+	+	+	+	+	+	+
+	-	+	-	-	-	-	+
+	-	-	+	+	+	+	+
+	+	-	+	-	-	-	+
+	+	+	+	+	+	+	+

- Spins are arranged in a ( $d$ -dimensional) lattice, and each spin interacts with its nearest neighbors.
- Energy is given by  $H(\mathbf{x}) = \sum_{ij} \beta_{ij} x_i x_j + \sum_i \alpha_i x_i$ .
- Goal: Find phase transitions.

The Ising model is very popular for explaining the effect of "society" or "environment" on a "component" or "individual". Example applications from flocking behaviour, behaviour of neural networks, sociology...

## The Ising Model (2-D Markov Random Field)

+	+	+	+	+	+	+	+
+	-	+	-	-	-	-	+
+	-	-	+	+	+	+	+
+	+	-	+	-	-	-	+
+	+	+	+	+	+	+	+

- Spins are arranged in a ( $d$ -dimensional) lattice, and each spin interacts with its nearest neighbors.
- Energy is given by  $H(\mathbf{x}) = \sum_{ij} \beta_{ij} x_i x_j + \sum_i \alpha_i x_i$ .
- Goal: Find phase transitions.

The Ising model is very popular for explaining the effect of "society" or "environment" on a "component" or "individual". Example applications from flocking behaviour, behaviour of neural networks, sociology...

### Conditional Random Fields

A special type of a Markov Random Field, a **Conditional Random Field** allows that each random variable may also be conditioned upon a set of global observations.

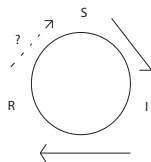
Define  $G = (V, E)$  to be an undirected graphical Markov Random Field model over random variable  $(X_v)_{v \in V}$  and  $\mathbf{O}$  a random variable representing observation sequences.  $(\mathbf{X}, \mathbf{O})$  is a conditional random field. The probability of a particular label sequence  $\mathbf{x}$  given observation sequence  $\mathbf{o}$  is a **normalized product of potential functions**, each of the form

$$\exp \left( \sum_j \lambda_j t_j(x_{i-1}, x_i, \mathbf{o}, i) + \sum_k \mu_k s_k(x_i, \mathbf{o}, i) \right),$$

where  $t_j(x_{i-1}, x_i, \mathbf{o}, i)$  is a transition feature function of the entire observation sequence and the labels at positions  $i$  and  $i - 1$  in the label sequence;  $s_k(x_i, \mathbf{o}, i)$  is a state feature function of the label at position  $i$  and the observation sequence; and  $\lambda_j$  and  $\mu_k$  are parameters (to be estimated from training data).

## Conditional Random Fields, an example

### The SIR epidemic model



Let  $p$  be the probability that a susceptible node becomes infected by a sick neighbor in a given time interval.

Let  $q$  be the probability that an infected node recovers (or dies) in a given time interval.

	<i>Susceptible</i>	<i>Infectious</i>	<i>Recovered</i>
Susceptible	$(1 - p)^{in(v)}$	$1 - (1 - p)^{in(v)}$	0
Infectious	0	$1 - q$	$q$
Recovered	0	0	1

**Table 1:** Transition probabilities at each time interval during the lifespan of the disease

Rajtmajer and Vukičević, 2011

$\Gamma$  - (weighted, oriented) graph of network states and transitions

$v \in \Gamma$  - one of the possible  $3^n$  states

$uv \in \Gamma$  - if  $u$  can be transformed to  $v$  in one step, with  $\text{weight}(uv) = \text{probability of transition}$

For example:

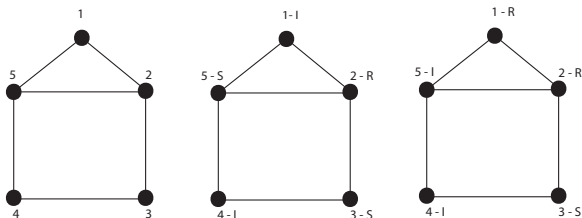


Figure 1: graph  $G$ , state  $u$  (S,I and R denote susceptible, infectious or recovered), state  $v$

The weight of  $uv \in \Gamma$  is calculated as

$$q \cdot (1 - p) \cdot 1 \cdot (1 - q) \cdot (1 - (1 - p)^2).$$

$I$  - initial probability vector

(Given fixed known starting state  $x \in V(\Gamma)$ ,  $I_x = 1$  and  $I_x = 0$  elsewhere.)

$M$  - transition matrix size  $3^n \times 3^n$

$M^k \cdot I$  - progression from  $I$  in  $k$  steps

### Theorem

*Let  $p, q \in (0, 1)$  and let  $I$  be any initial probability vector. There is a limit vector  $W = M^k \cdot I$ . Moreover, in the limit vector probabilities corresponding to all acute states are equal to zero.*

→ We can model the (entire) course of the epidemic.

Real world problems should be attacked using a Monte Carlo approach.

## Classification of network hubs:

- *outhub* - node which causes the most harm to the network as a whole (as measured by the sum of other nodes' chances to get sick) if it first to become infected and introduce the disease to rest of the population
- *inhub* - node which is most susceptible to get the disease, given that the disease may be start at any other node in the graph (measured as the average probability to becoming infected, over all  $n - 1$  introduction points)
- *transition hub* - node which would most alleviate the harm to the network, should it become immune (by vaccination), and therefore unable to catch or spread the disease

Should all three hubs should be given as the nodes with highest degree?

## Classification of network hubs:

- *outhub* - node which causes the most harm to the network as a whole (as measured by the sum of other nodes' chances to get sick) if it first to become infected and introduce the disease to rest of the population
- *inhub* - node which is most susceptible to get the disease, given that the disease may be start at any other node in the graph (measured as the average probability to becoming infected, over all  $n - 1$  introduction points)
- *transition hub* - node which would most alleviate the harm to the network, should it become immune (by vaccination), and therefore unable to catch or spread the disease

Should all three hubs should be given as the nodes with highest degree?



## Classification of network hubs:

- *outhub* - node which causes the most harm to the network as a whole (as measured by the sum of other nodes' chances to get sick) if it first to become infected and introduce the disease to rest of the population
- *inhub* - node which is most susceptible to get the disease, given that the disease may be start at any other node in the graph (measured as the average probability to becoming infected, over all  $n - 1$  introduction points)
- *transition hub* - node which would most alleviate the harm to the network, should it become immune (by vaccination), and therefore unable to catch or spread the disease

Should all three hubs should be given as the nodes with highest degree?

Consider the network.

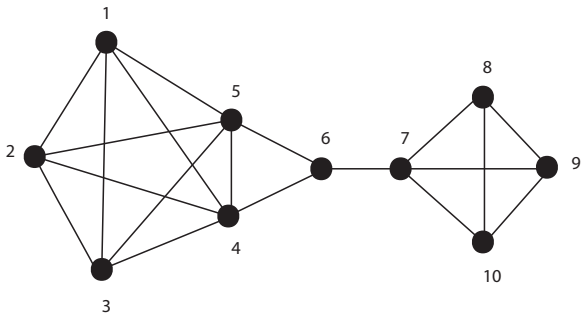


Figure 2: Sample network

We fix  $p$  and  $q$ . Using both the complete algorithm with complexity  $6^n$  to determine all possible courses of the epidemic on the network and the Monte Carlo method (for comparison), we determine the following probabilities.

*Outhub*: probability that a randomly chosen vertex will become infected given vertex  $i$  is first to be infected

*Inhub*: probability that  $i$  will become infected given a randomly chosen vertex is first to be infected

*Trans hub*: probability that a randomly chosen vertex will become infected given  $i$  is vaccinated

Vertex ( $i$ )	<i>Outhub</i>	<i>Inhub</i>	<i>Trans hub</i>	Vertex ( $i$ )	<i>Outhub</i>	<i>Inhub</i>	<i>Trans hub</i>
1	0.593589	0.591562	0.478138	1	0.59406	0.59147	0.47786
2	0.593589	0.591562	0.478138	2	0.59153	0.59174	0.47825
3	0.593589	0.591562	0.478138	3	0.59163	0.59136	0.47774
4	<b>0.619712</b>	<b>0.611778</b>	0.426127	4	<b>0.62041</b>	<b>0.61165</b>	0.42626
5	<b>0.619712</b>	<b>0.611778</b>	0.426127	5	<b>0.61917</b>	<b>0.61152</b>	0.42623
6	0.585350	0.594192	<b>0.376840</b>	6	0.58460	0.59377	<b>0.37710</b>
7	0.535528	0.539843	0.397074	7	0.53791	0.53956	0.39708
8	0.475483	0.478576	0.507061	8	0.47471	0.47838	0.50675
9	0.475483	0.478576	0.507061	9	0.47584	0.47824	0.50686
10	0.475483	0.478576	0.507061	10	0.47613	0.47830	0.50725

Table 2: Figure 12 hubs with the complete calculation and Monte Carlo approximation, respectively

Vertex 6 is the transition hub, although it has lowest degree(!)

## *Applications of Random Field models*

- Labelling and parsing sequential data
  - Natural language text
  - Biological sequences
- Computer vision
  - Image segmentation
  - Object recognition

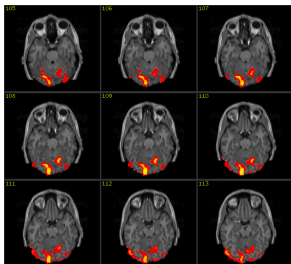


Figure 3: fMRI

## *Applications of Random Field models*

- Labelling and parsing sequential data
  - Natural language text
  - Biological sequences
- Computer vision
  - Image segmentation
  - Object recognition

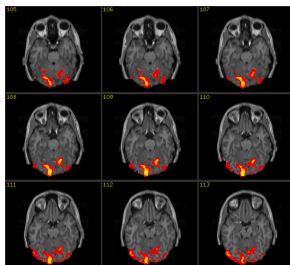


Figure 3: fMRI

## *Applications of Random Field models*

- Labelling and parsing sequential data
  - Natural language text
  - Biological sequences
- Computer vision
  - Image segmentation
  - Object recognition

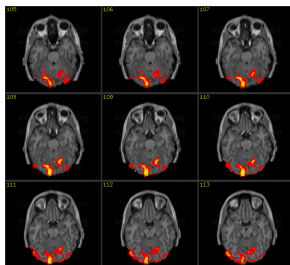


Figure 3: fMRI

## *Applications of Random Field models*

- Labelling and parsing sequential data
  - Natural language text
  - Biological sequences
- Computer vision
  - Image segmentation
  - Object recognition

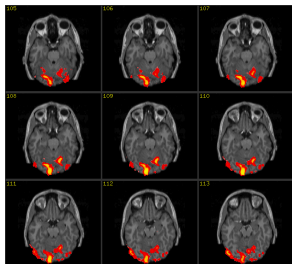


Figure 3: fMRI

## *Applications of Random Field models*

- Labelling and parsing sequential data
  - Natural language text
  - Biological sequences
- Computer vision
  - Image segmentation
  - Object recognition

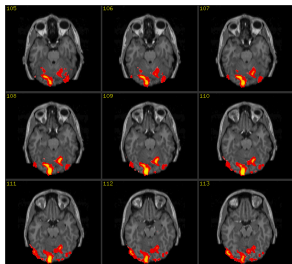


Figure 3: fMRI



## *Applications of Random Field models*

- Labelling and parsing sequential data
  - Natural language text
  - Biological sequences
- Computer vision
  - Image segmentation
  - Object recognition

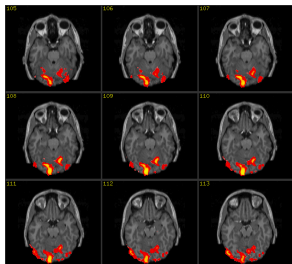


Figure 3: fMRI

## References

- H. Wallach, *Conditional Random Fields: An Introduction*, U. Penn Technical Report MS-CIS-04-21, 2004.
- S. Roweis, *Undirected Graphical Models*, Uncertainty and Learning in AI, NYU, 2004.
- C. Sutton and A. McCallum, *An Introduction to Conditional Random Fields for Relational Learning*, MIT Press, 2006.
- D. Precup, *Graphical Models*, Probabilistic Reasoning in AI, McGill University, 2008.
- J. Cao and K.J. Worsley, *Applications of Random Fields in Human Brain Mapping*, American Mathematical Society, 1991.

## References

- H. Wallach, *Conditional Random Fields: An Introduction*, U. Penn Technical Report MS-CIS-04-21, 2004.
- S. Roweis, *Undirected Graphical Models, Uncertainty and Learning in AI*, NYU, 2004.
- C. Sutton and A. McCallum, *An Introduction to Conditional Random Fields for Relational Learning*, MIT Press, 2006.
- D. Precup, *Graphical Models*, Probabilistic Reasoning in AI, McGill University, 2008.
- J. Cao and K.J. Worsley, *Applications of Random Fields in Human Brain Mapping*, American Mathematical Society, 1991.

## References

- H. Wallach, *Conditional Random Fields: An Introduction*, U. Penn Technical Report MS-CIS-04-21, 2004.
- S. Roweis, *Undirected Graphical Models*, Uncertainty and Learning in AI, NYU, 2004.
- C. Sutton and A. McCallum, *An Introduction to Conditional Random Fields for Relational Learning*, MIT Press, 2006.
- D. Precup, *Graphical Models*, Probabilistic Reasoning in AI, McGill University, 2008.
- J. Cao and K.J. Worsley, *Applications of Random Fields in Human Brain Mapping*, American Mathematical Society, 1991.

## References

- H. Wallach, *Conditional Random Fields: An Introduction*, U. Penn Technical Report MS-CIS-04-21, 2004.
- S. Roweis, *Undirected Graphical Models*, Uncertainty and Learning in AI, NYU, 2004.
- C. Sutton and A. McCallum, *An Introduction to Conditional Random Fields for Relational Learning*, MIT Press, 2006.
- D. Precup, *Graphical Models*, Probabilistic Reasoning in AI, McGill University, 2008.
- J. Cao and K.J. Worsley, *Applications of Random Fields in Human Brain Mapping*, American Mathematical Society, 1991.

## References

- H. Wallach, *Conditional Random Fields: An Introduction*, U. Penn Technical Report MS-CIS-04-21, 2004.
- S. Roweis, *Undirected Graphical Models*, Uncertainty and Learning in AI, NYU, 2004.
- C. Sutton and A. McCallum, *An Introduction to Conditional Random Fields for Relational Learning*, MIT Press, 2006.
- D. Precup, *Graphical Models*, Probabilistic Reasoning in AI, McGill University, 2008.
- J. Cao and K.J. Worsley, *Applications of Random Fields in Human Brain Mapping*, American Mathematical Society, 1991.