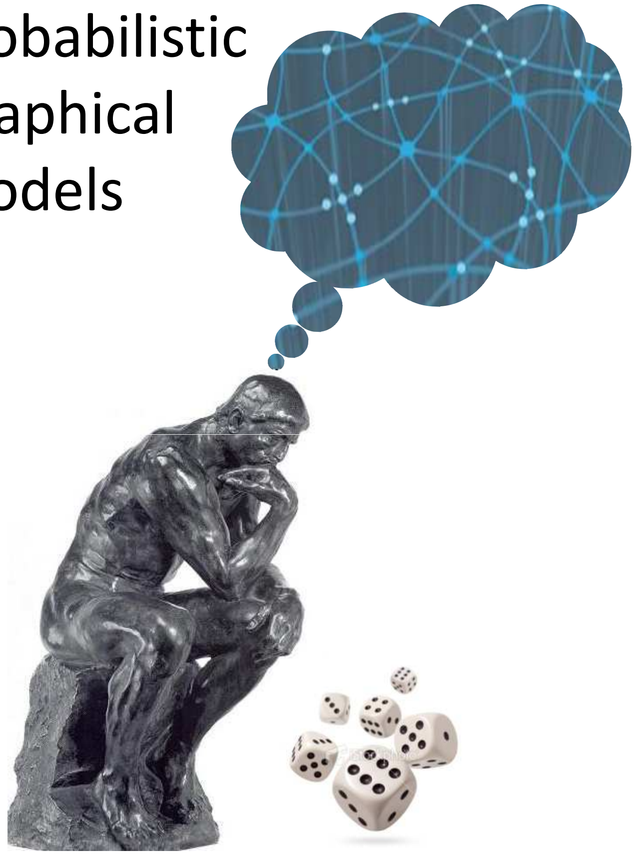
Probabilistic
Graphical
Models

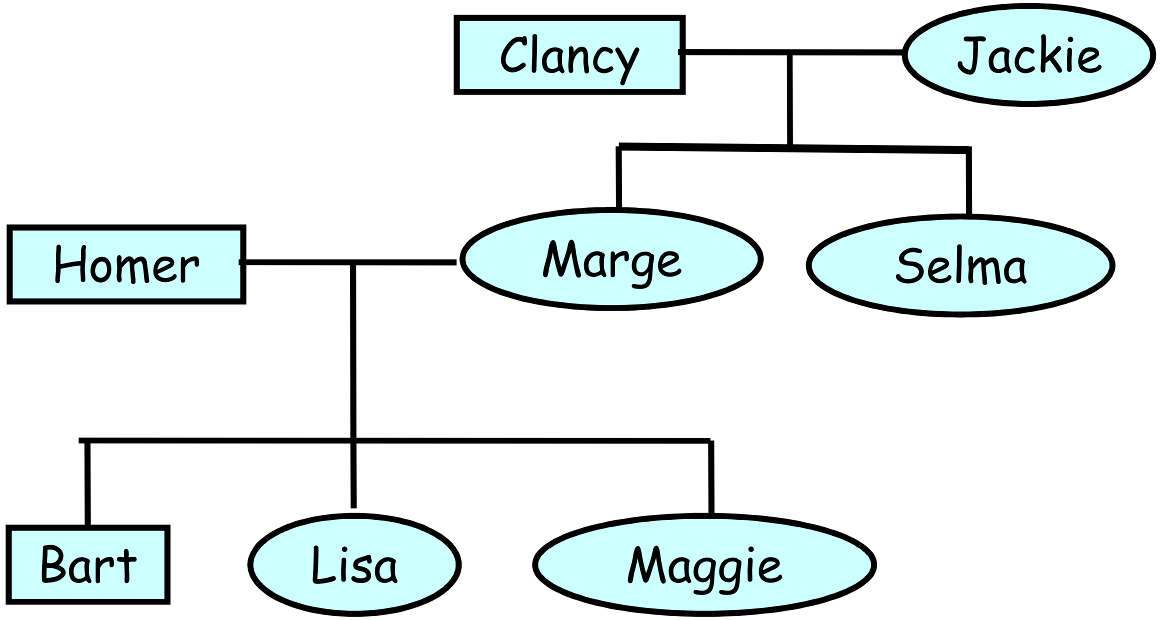


Representation

Template Models

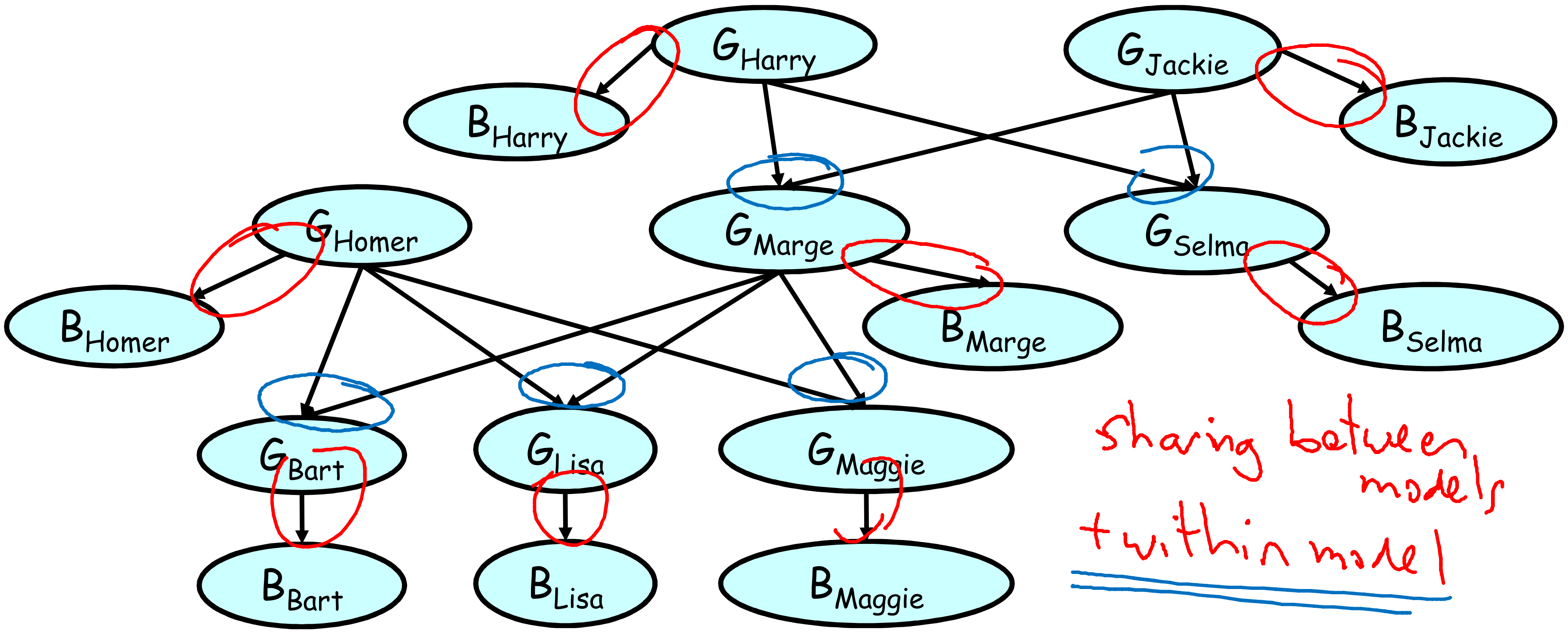
Overview

Genetic Inheritance

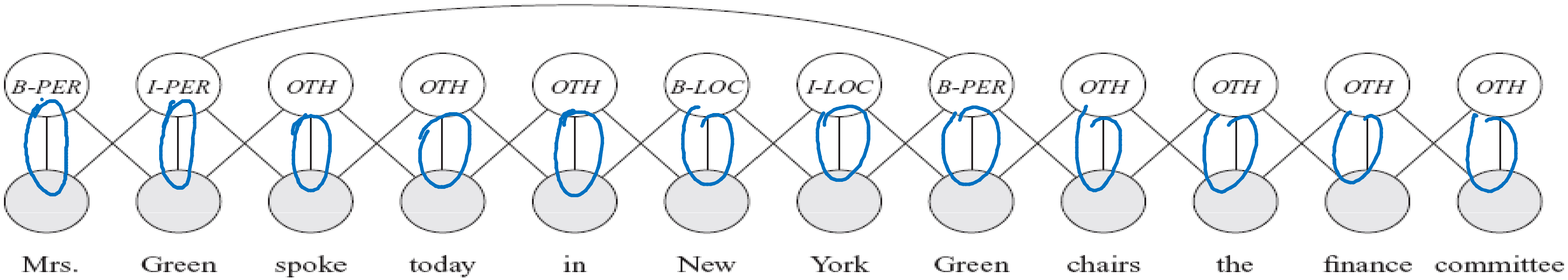


pedigree
trait

Genetic Inheritance



NLP Sequence Models



KEY

B-PER	Begin person name	*I-LOC*	Within location name
I-PER	Within person name	*OTH*	Not an entitiy
B-LOC	Begin location name		

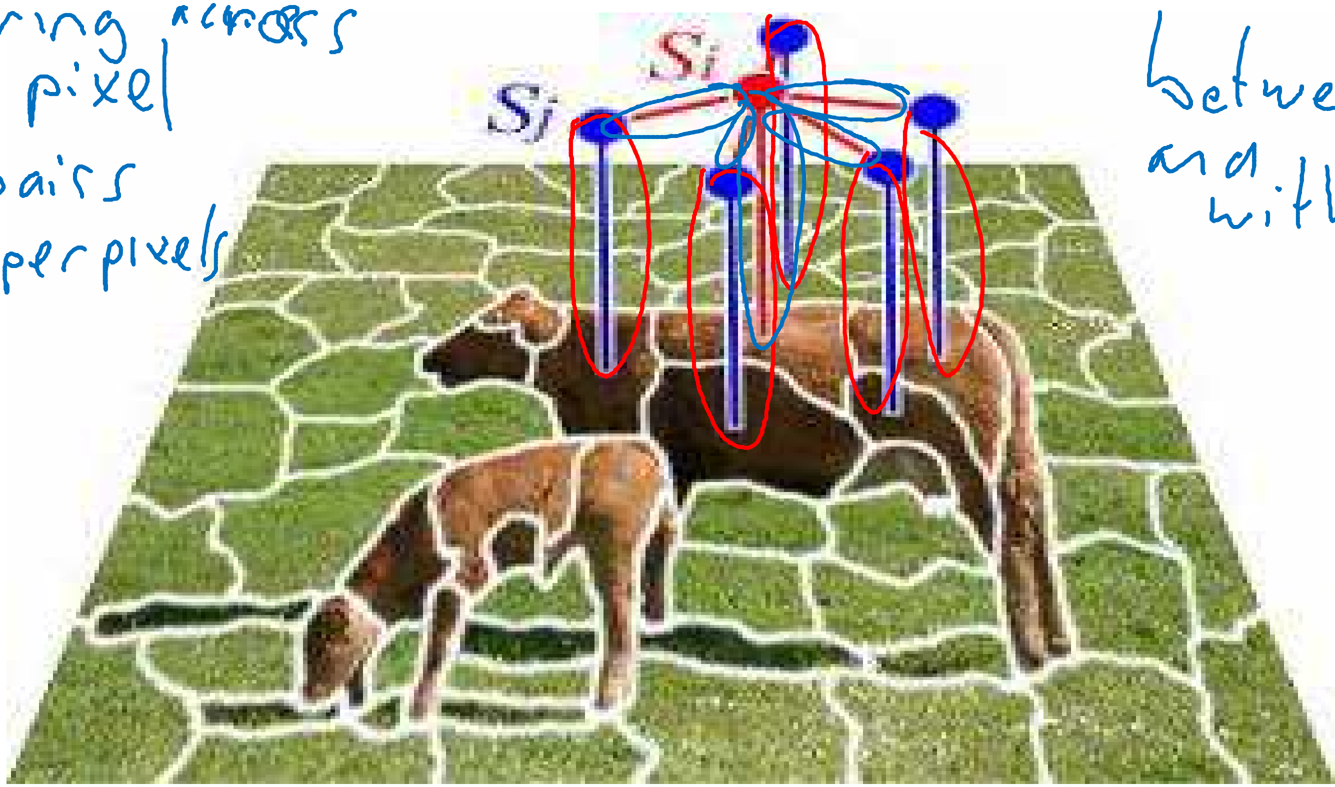
Named entity recognition

Daphne Koller

Image Segmentation

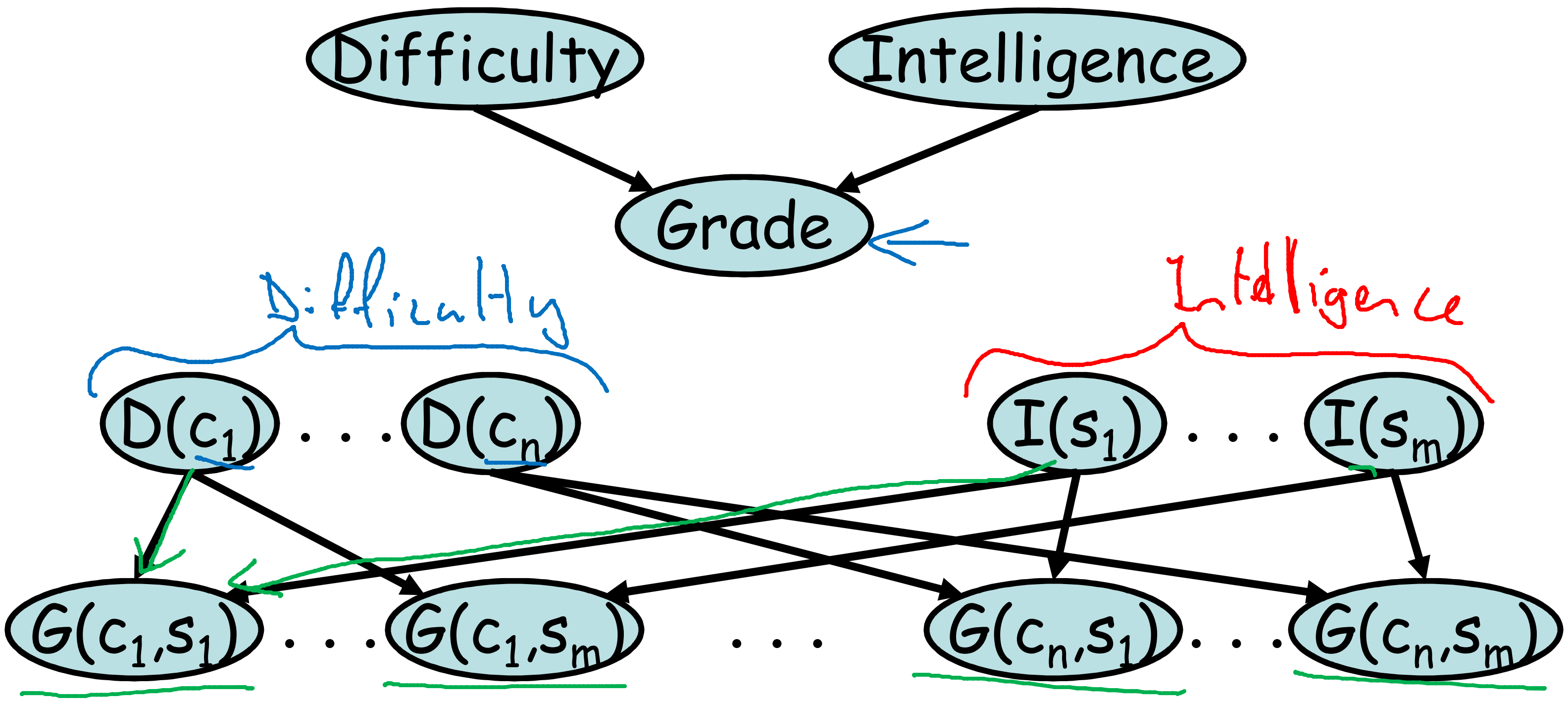
sharing across pixel

and pairs of superpixels



between and within

The University Example

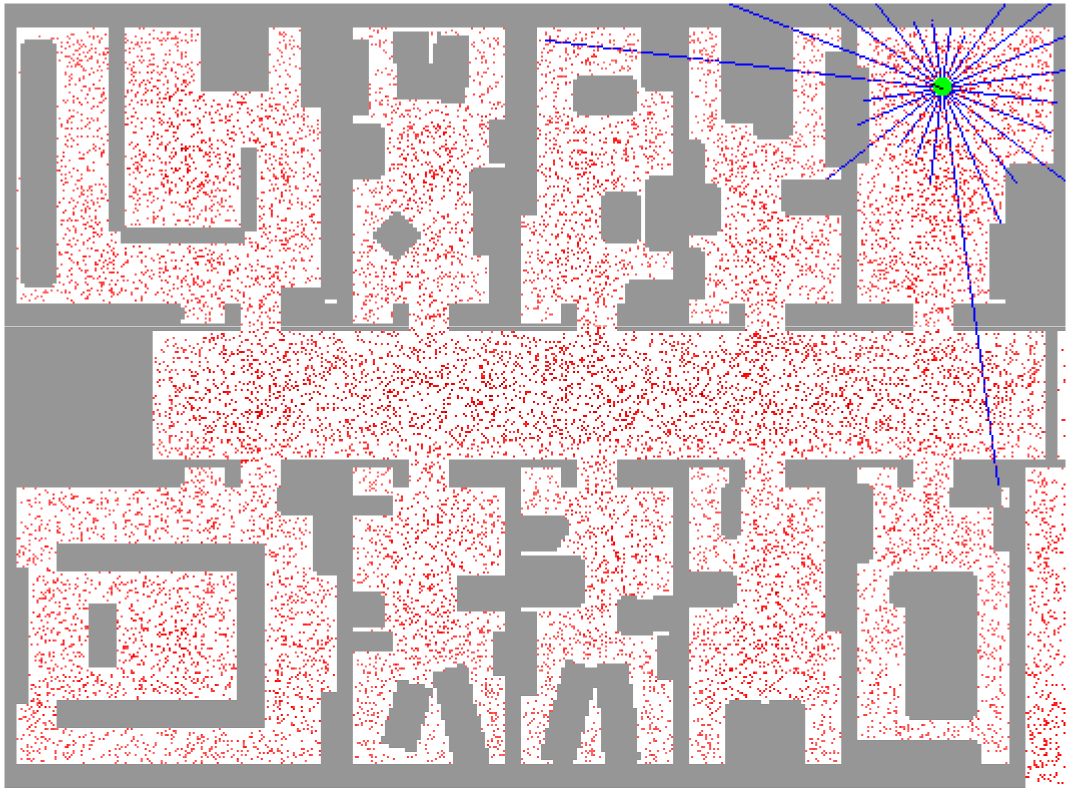


Robot Localization

time series

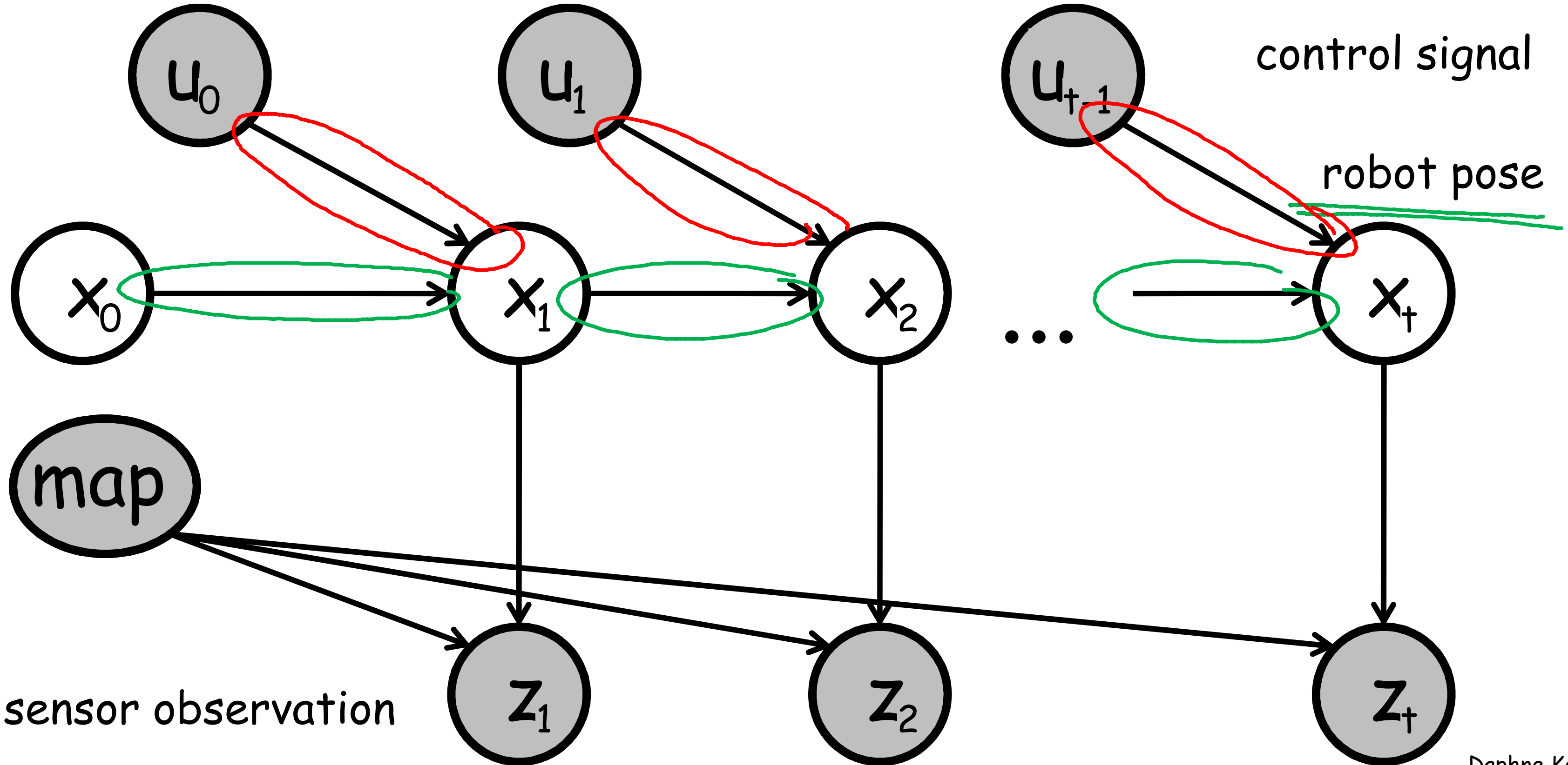
position at time t changes over time

robot dynamics are fixed



Daphne Koller

Robot Localization



Daphne Koller

Template Variables

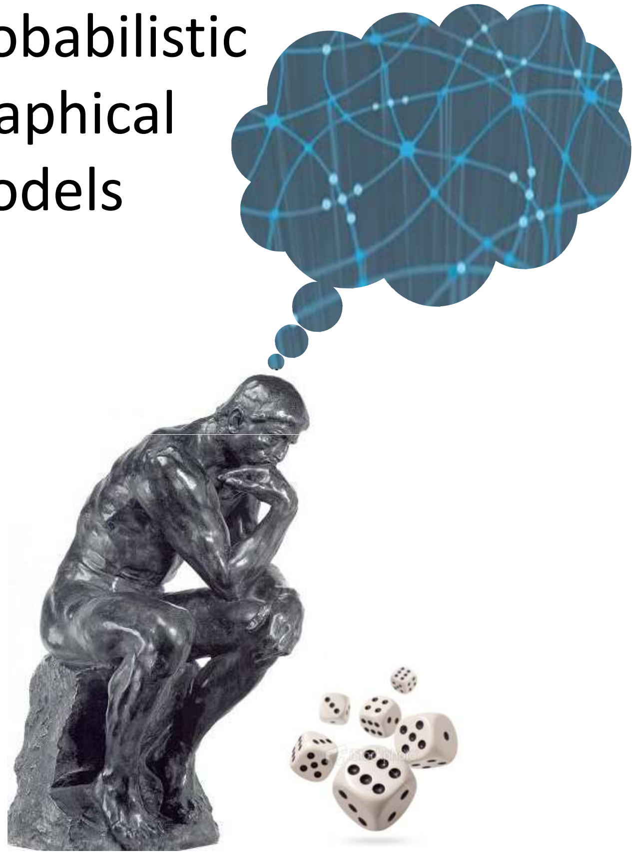
- Template variable $X(U_1,\ldots,U_k)$ is instantiated (duplicated) multiple times
  - Location(t), Sonar(t)
  - Genotype(person), Phenotype(person)
  - Label(pixel)
  - Difficulty(course), Intelligence(student), Grade(course, student)

Template Models

- Languages that specify how ground variables inherit dependency model from template
- Dynamic Bayesian networks ← temporal
- Object-relational models people, courses, pixels, …
  - Directed
    - Plate models
  - Undirected

Daphne Koller

Probabilistic
Graphical
Models



Representation

Template Models

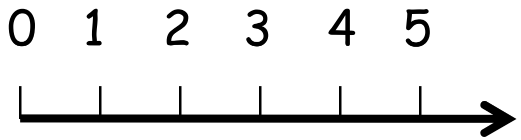
Temporal Models

evolves over time

Distributions over Trajectories

discretize time

sensor



- Pick time granularity Δ
- $X^{(t)}$ – variable X at time $t\Delta$
- $X^{(t:t')} = \{X^{(t)}, \ldots, X^{(t')}\}$ $(t \leq t')$

- Want to represent $P(X^{(t:t')})$ for any t, t'

Markov Assumption

$$\underbrace{P(\boldsymbol{X}^{(0:T)})}_{\text{time flows forward}} = P(\boldsymbol{X}^{(0)}) \prod_{t=0}^{T-1} P(\underbrace{\boldsymbol{X}^{(t+1)}}_{\text{state at t+1}} \mid \underbrace{\boldsymbol{X}^{(0:t)}}_{\text{state at 0..t}})$$

chain rule for probabilities

$$(\underbrace{\boldsymbol{X}^{(t+1)}}_{\text{next step}} \perp \underbrace{\boxed{\boldsymbol{X}^{(0:t-1)}}}_{\text{past}} \mid \underbrace{\boldsymbol{X}^{(t)}}_{\text{present}})$$

forgetting

$$P(\boldsymbol{X}^{(0:T)}) = P(\boldsymbol{X}^{(0)}) \prod_{t=0}^{T-1} P(\boldsymbol{X}^{(t+1)} \mid \underbrace{\boldsymbol{X}^{(t)}})$$

Is this true?

X = Location of robot

$L^{t+1} \perp L^{t-1} \mid L^{t}$?

probably not

velocity

enrich state by adding v^t and other variables

(adding dependencies back in time - semi. Markov)

Daphne Koller

Time Invariance

next time

current time

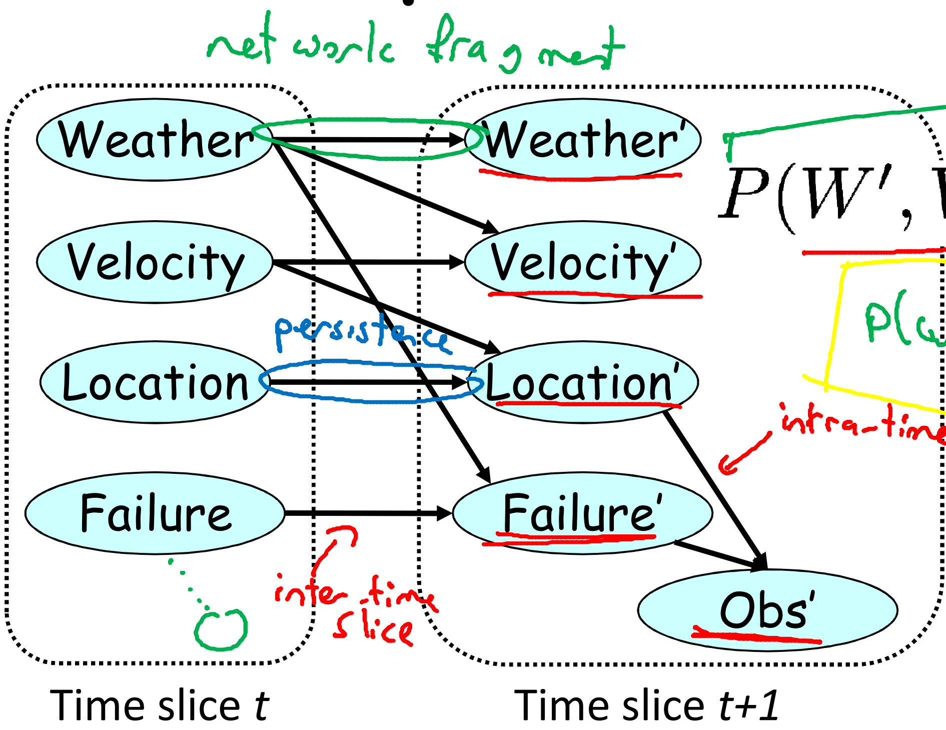
- Template probability model $P(X' \mid X)$
- For all t:

$$P(\boldsymbol{X}^{(t+1)} \mid \boldsymbol{X}^{(t)}) = P(\boldsymbol{X}' \mid \boldsymbol{X})$$

traffic   time of day, day of week, football

enrich model by including

Template Transition Model



cond dist t+1 given t

$$P(\underbrace{W', V', L', F', O'}) \mid \underbrace{W, V, L, F}) =$$

intra-time-slice

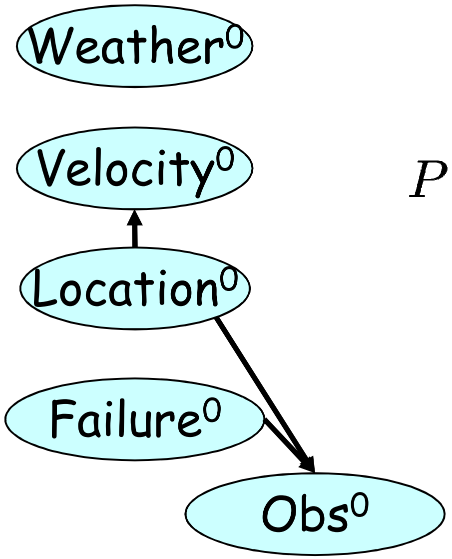
$$P(W' \mid W) \cdot P(V' \mid W, V) \cdot P(L' \mid L, V) \cdot \underbrace{P(F' \mid F, W)}_{t \text{ to } t+1} \cdot \underbrace{P(O' \mid L', F')}_{t+1}$$

rapidly-acting

$$\cancel{P(W, V, L, F)}$$

Daphne Koller

Initial State Distribution



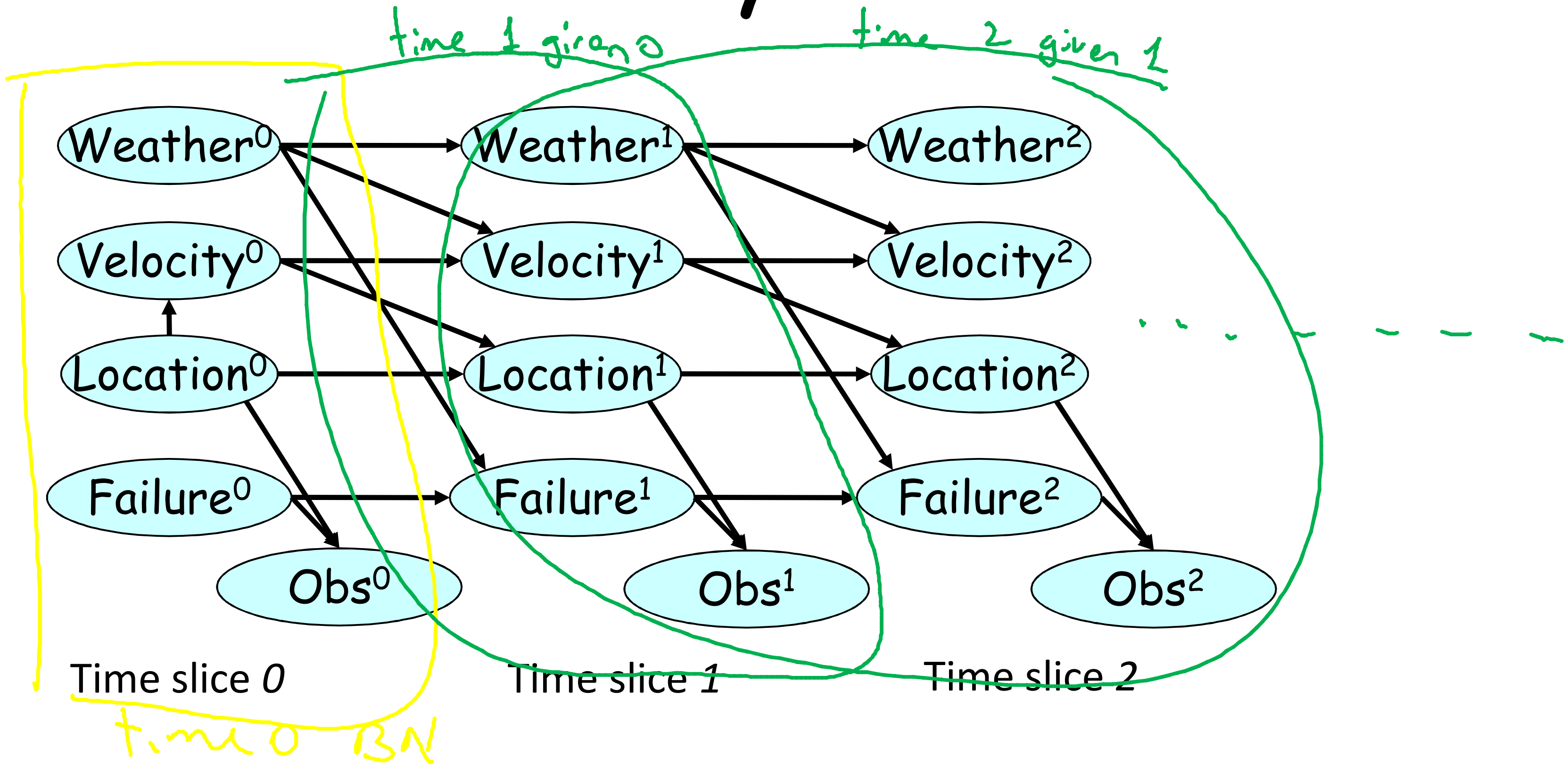
Time slice *0*

$$\boxed{P(W^{(0)}, V^{(0)}, L^{(0)}, F^{(0)}, O^{(0)})} =$$

$$P(W^{(0)})P(V^{(0)} \mid L^{(0)})P(L^{(0)})P(F^{(0)})P(O^{(0)} \mid F^{(0)}, L^{(0)})$$

chain rule

Ground Bayesian Network



2-time-slice Bayesian Network

- A transition model (2TBN) over template $X_1,\dots,X_n$ is specified as a BN fragment such that:
  - The nodes include $X_1',\dots,X_n'$ and a subset of $X_1,\dots,X_n$
  - Only the nodes $X_1',\dots,X_n'$ have parents and a CPD

the time-t vars that directly affect state at t+1

- The 2TBN defines a conditional distribution

$$P(\boldsymbol{X}' \mid \boldsymbol{X}) = \prod_{i=1}^{n} P(X_i' \mid \mathbf{Pa}_{X_i'})$$

chain rule

Daphne Koller

Dynamic Bayesian Network

- A dynamic Bayesian network (DBN) over $X_1,\ldots,X_n$ is defined by a
  - $\underbrace{\text{2TBN } BN_{\rightarrow}}$ over $X_1,\ldots,X_n$ dynamics
  - time 0 – a Bayesian network $BN^{(0)}$ over $X_1^{(0)},\ldots,X_n^{(0)}$

Ground Network

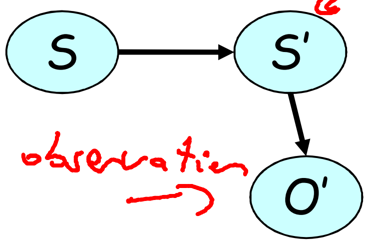
- For a trajectory over $\underbrace{0,\ldots,T}$ we define a ground (unrolled network) such that
  - time 0 – The dependency model for $X_1^{(0)}, \ldots, X_n^{(0)}$ is copied from $BN^{(0)}$
  - transition – The dependency model for $X_1^{(t)}, \ldots, X_n^{(t)}$ for all $t > 0$ is copied from $BN_{\rightarrow}$

speech
robotics
vision
language

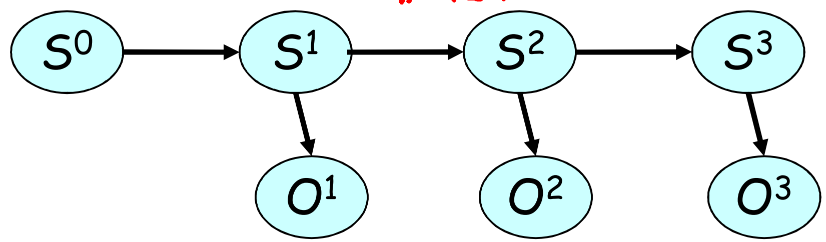
Hidden Markov Models

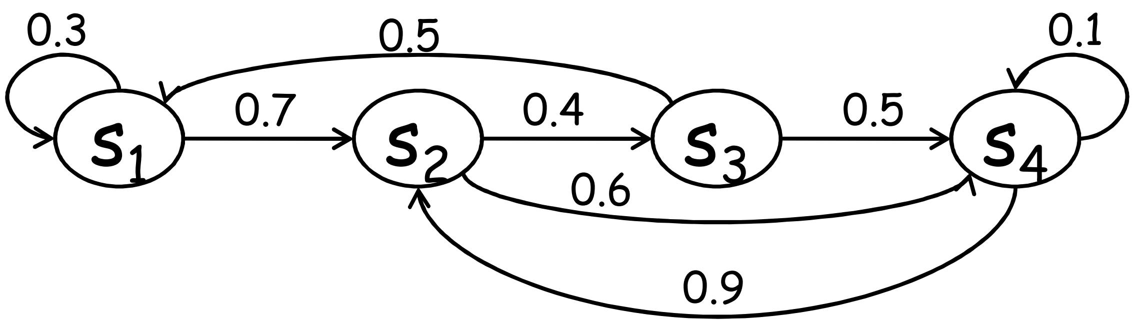
hidden state

observation

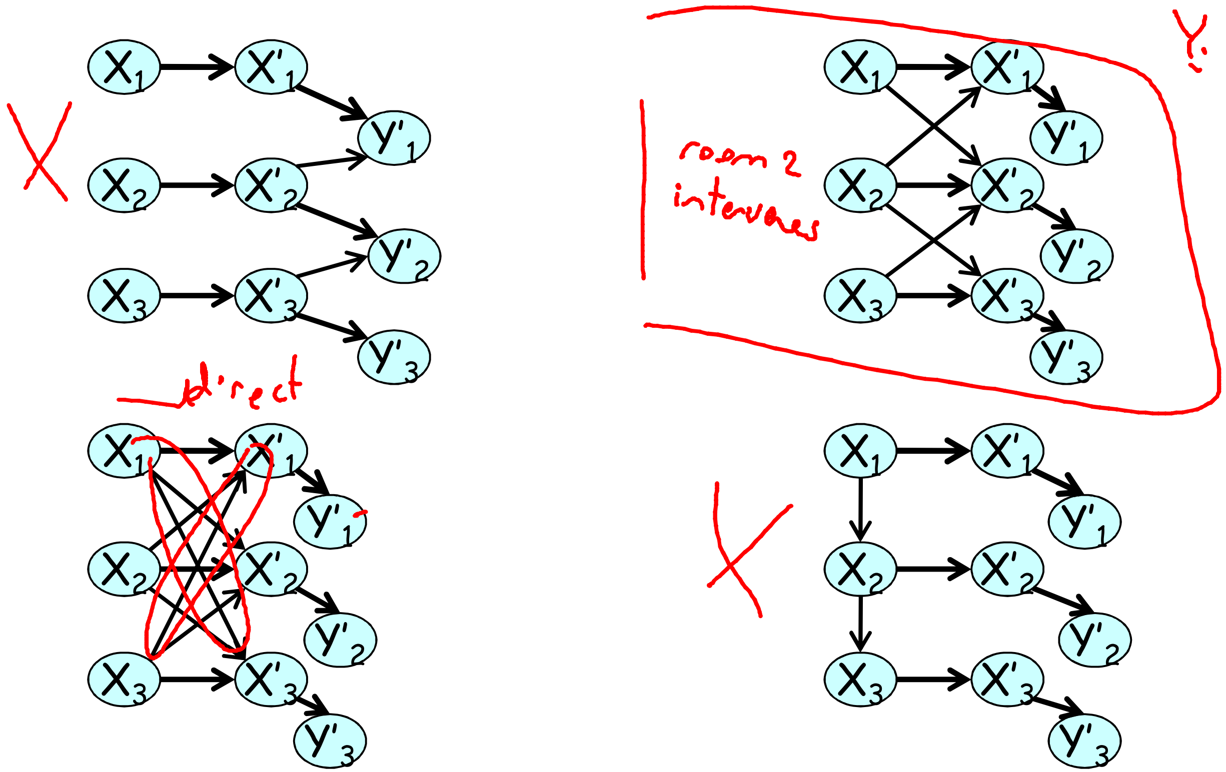


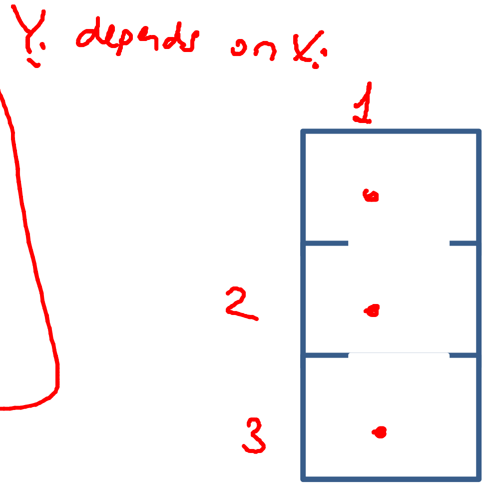
unrolled



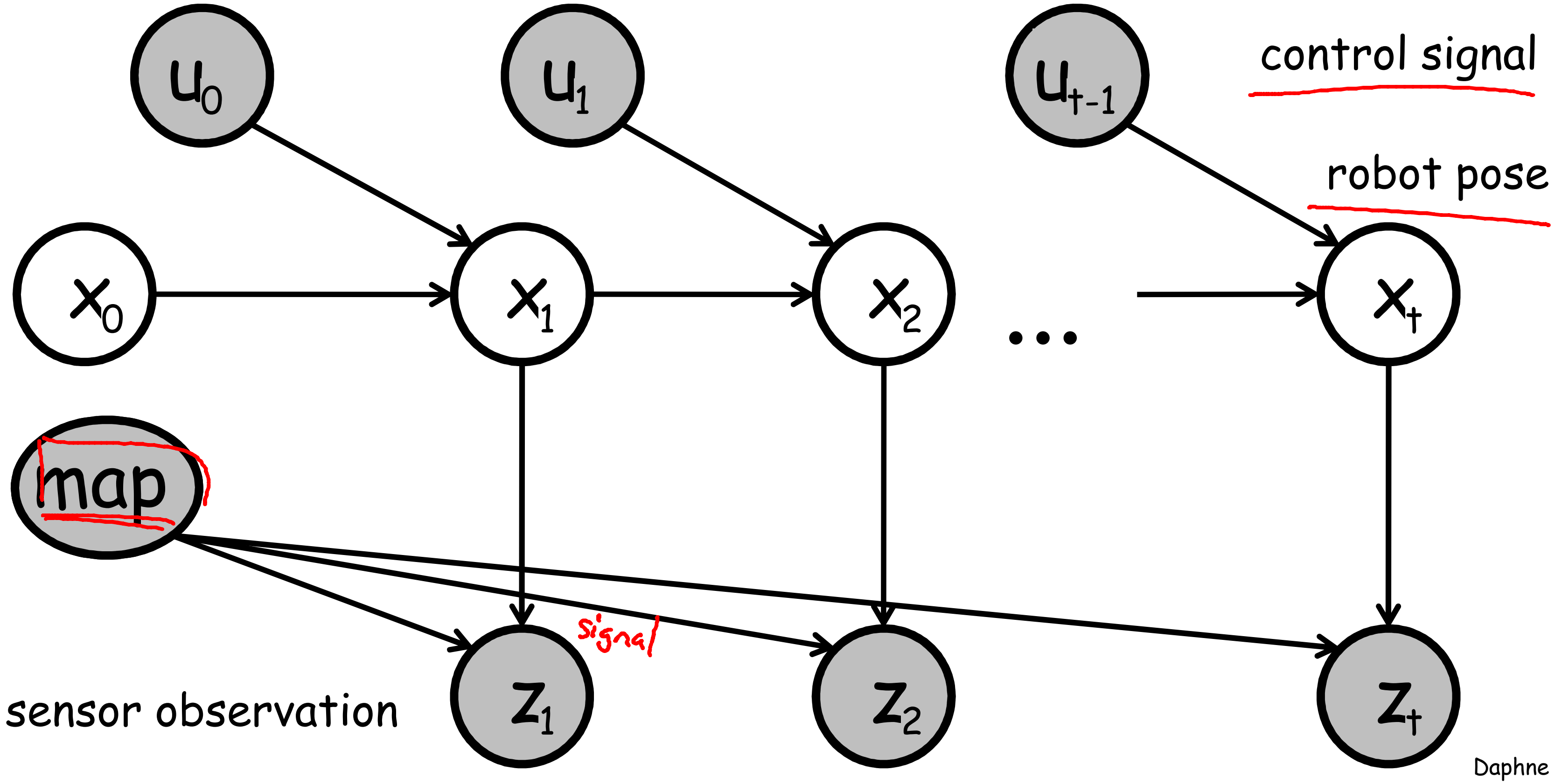


Consider a smoke detection tracking application, where we have 3 rooms connected in a row. Each room has a true smoke level (X) and a smoke level (Y) measured by a smoke detector situated in the middle of the room. Which of the following is the best DBN structure for this problem?

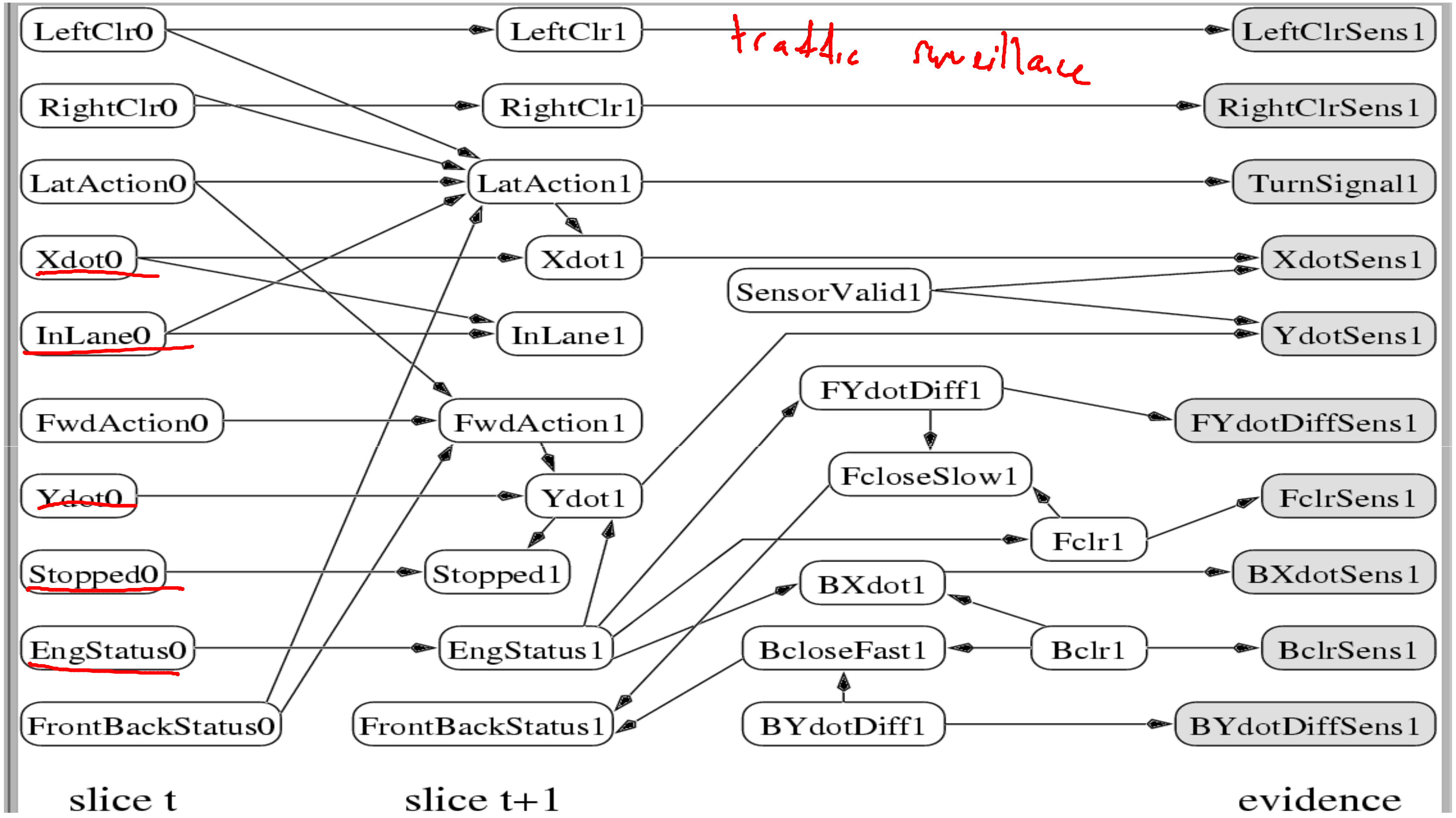




Robot Localization

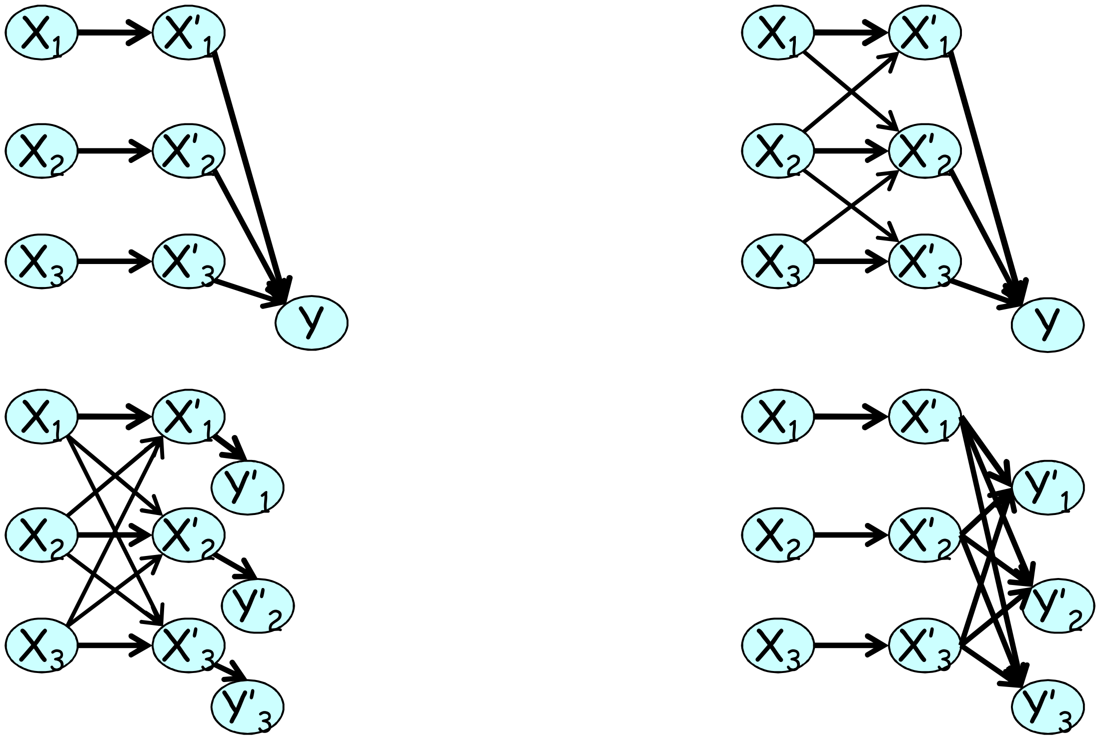


Daphne Koller



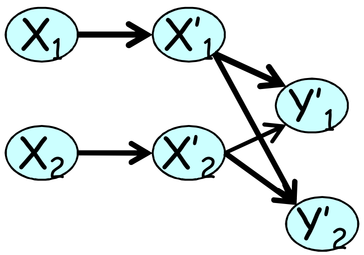
Tim Huang, Dieter Koller, Jitendra Malik, Gary Ogasawara, Bobby Rao, Stuart Russell, J. Weber

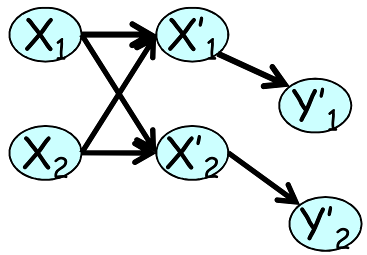
Consider an application, where we have 3 sources, each making sound (X) simultaneously (e.g., a TV, a blender, and a ringing cellphone). We want to separate the observed aggregate signal (Y) into its unobserved constituent signals. Which of the following is the best DBN structure for this problem?

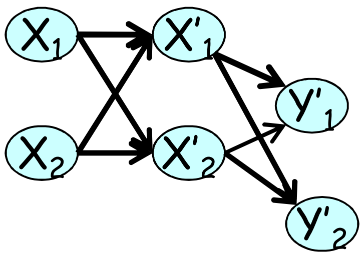


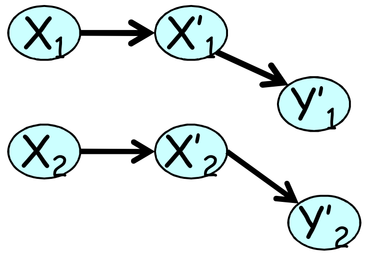
Daphne Koller

Consider a situation where we have two people in a conversation, each wearing a sensitive lapel microphone. We want to separate each person's speech (X) using the signal (Y) obtained from their microphone. Which of the following is the best DBN structure for this problem?







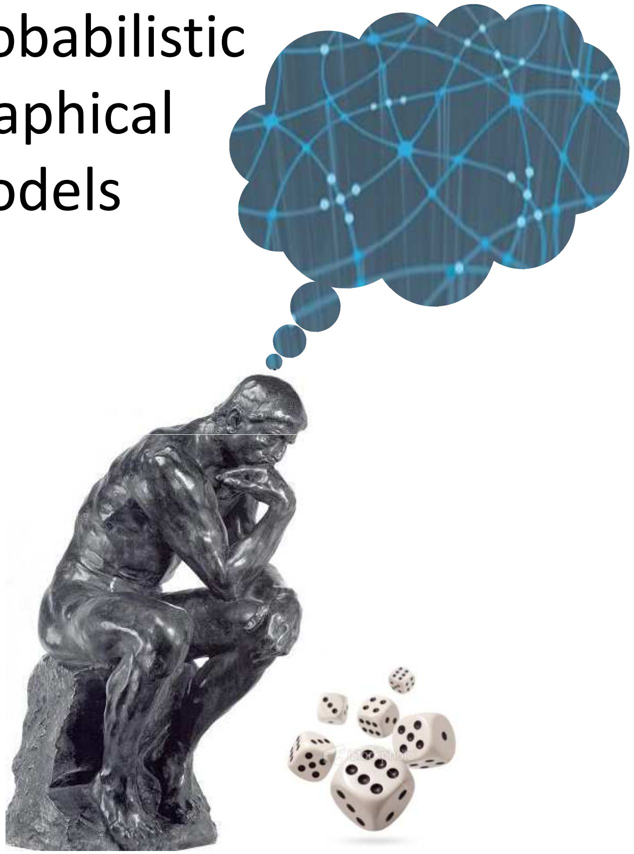


Daphne Koller

Summary

- DBNS are a compact representation for encoding structured distributions over arbitrarily long temporal trajectories
- They make assumptions that may require appropriate model (re)design:
  - Markov assumption ← independence
  - Time invariance   consistency over time

Probabilistic Graphical Models



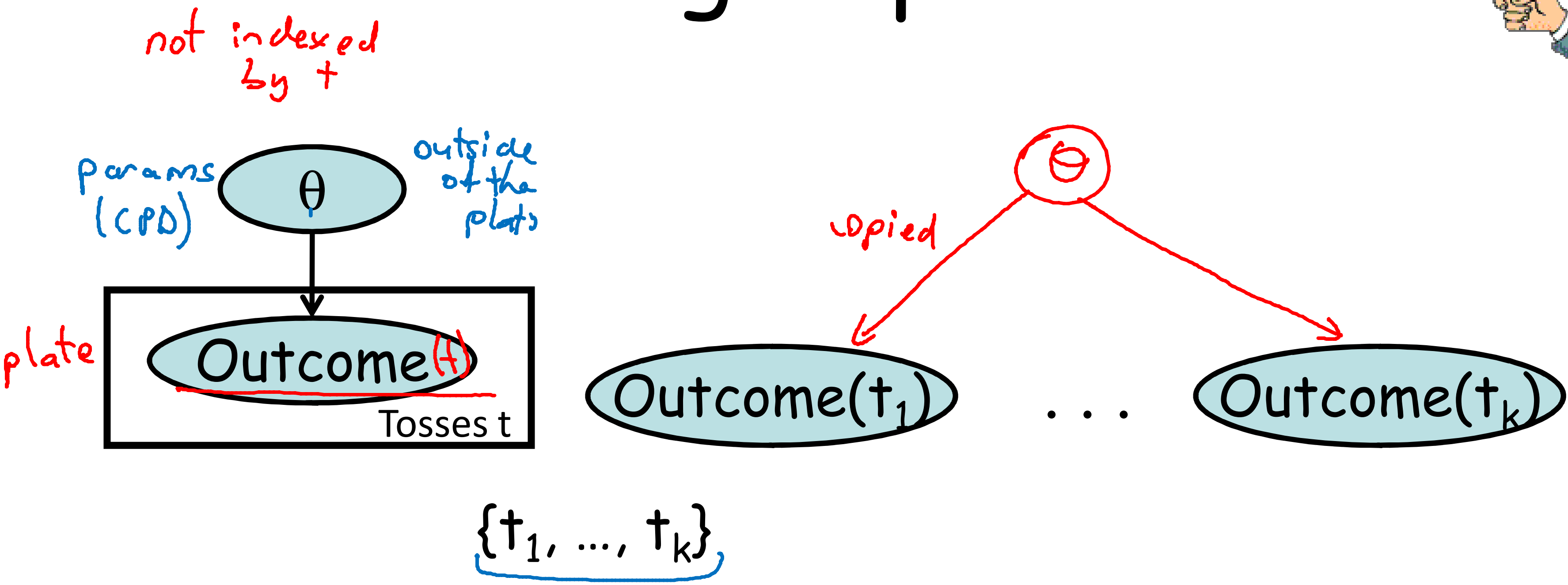
Representation

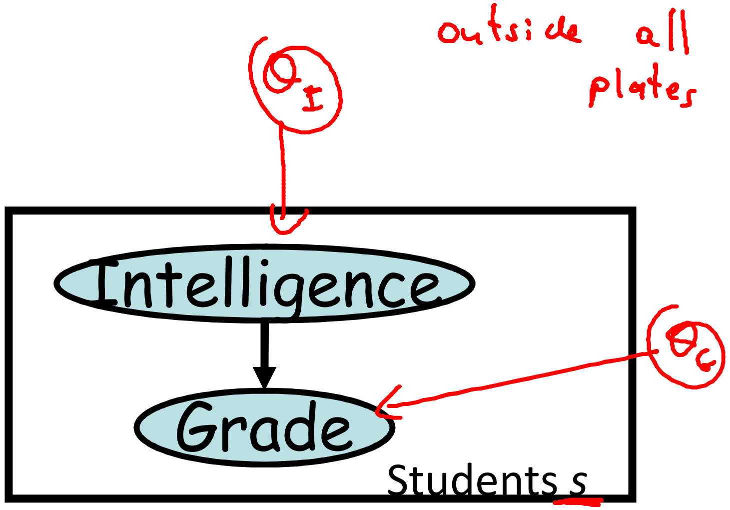
Template Models

Plate Models

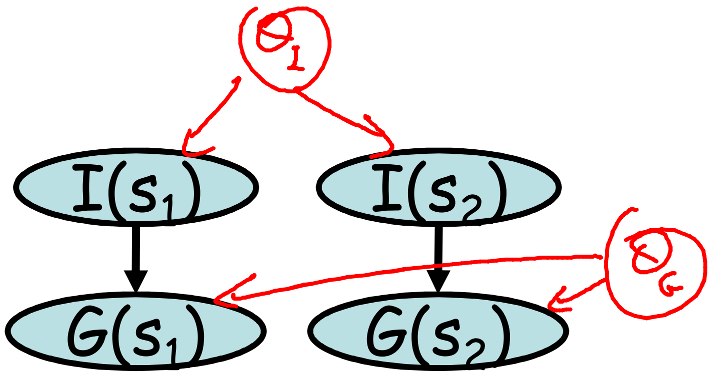
objects of same type

Modeling Repetition





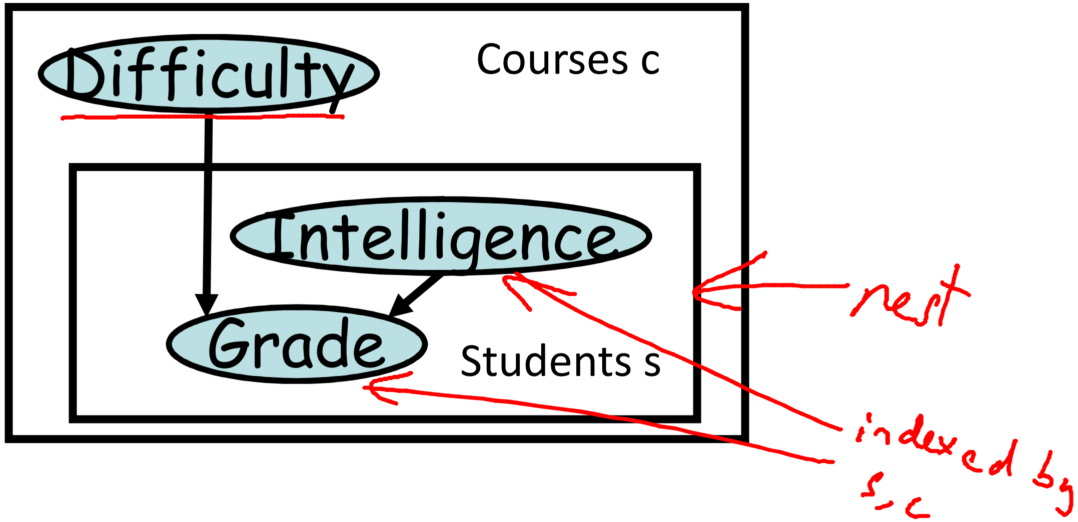
θ_I
outside all plates
Intelligence
Grade
θ_G
Students s

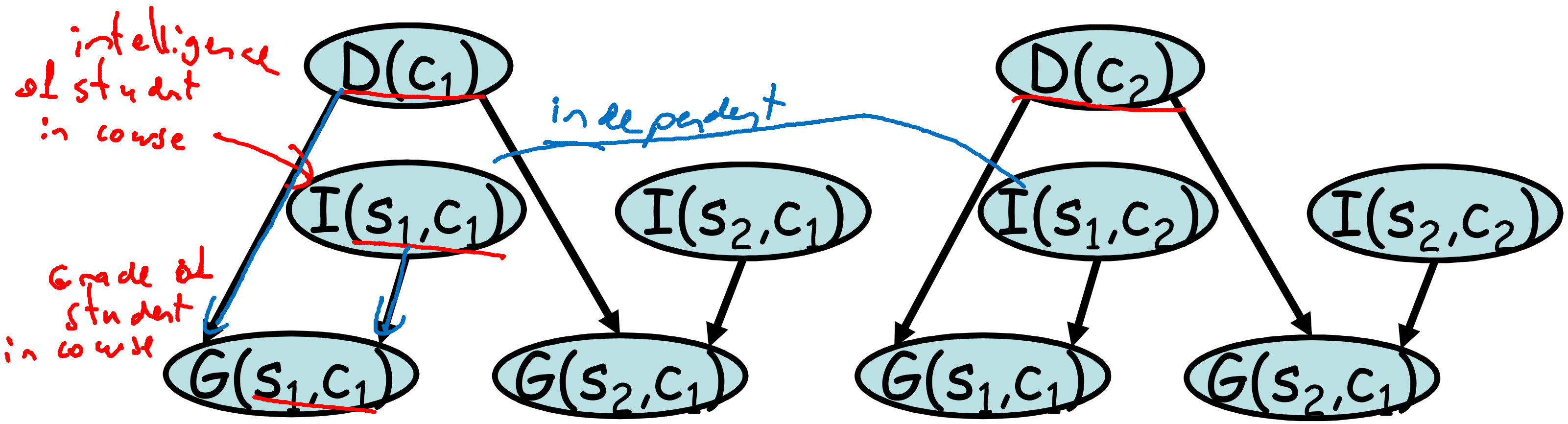


θ_I
$I(s_1)$
$I(s_2)$
θ_G
$G(s_1)$
$G(s_2)$

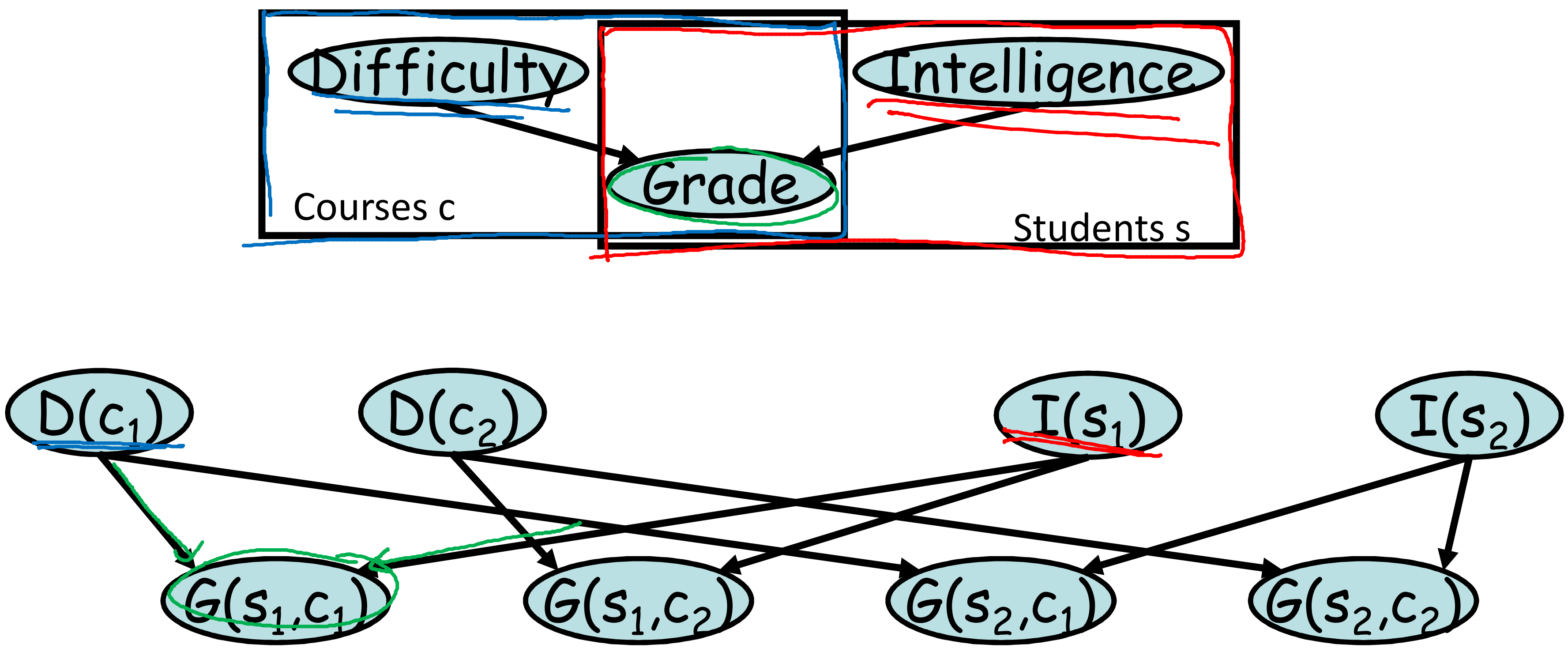
Nested Plates

courses c
students s



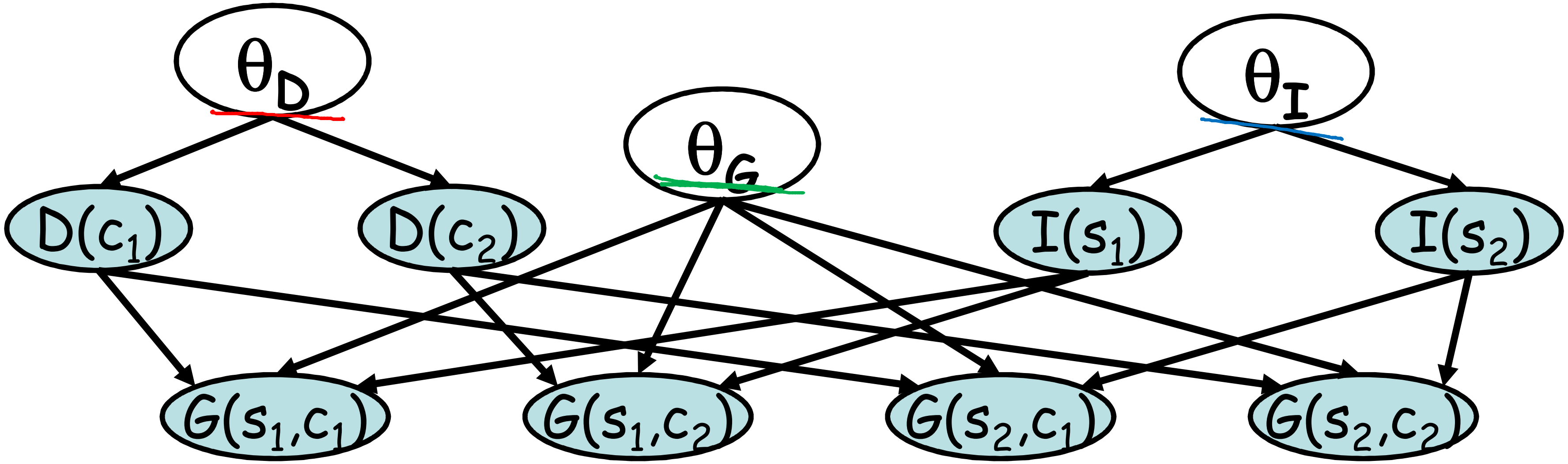


Overlapping Plates

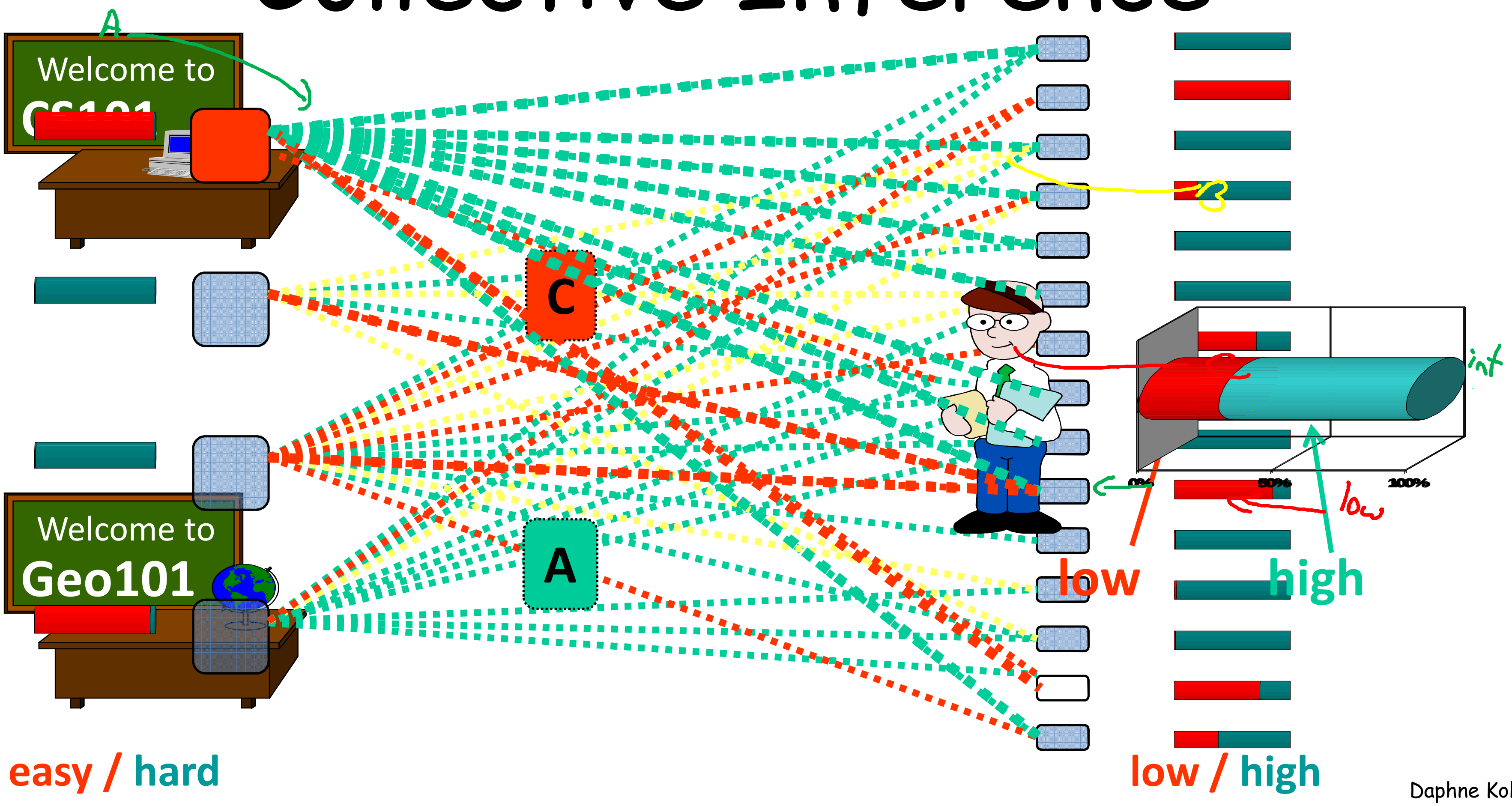


Daphne Koller

Explicit Parameter Sharing



Collective Inference



Daphne Koller

Plate Dependency Model

indexed

- For a template variable $A(U_1,\ldots,U_k)$:
  - Template parents $B_1(U_1),\ldots,B_m(U_m)$

    $\subseteq$

    $U_i \subseteq \{U_1,\ldots,U_k\}$

    $I(s) \rightarrow G(s,c) \leftarrow D(c)$

    $G(s,c) \rightarrow Honors(s)$

    unbounded # of parents

    aggregator CPD

    template

  - CPD $P(A \mid B_1,\ldots, B_m)$

Ground Network

Let $A(U_1,\ldots,U_k)$ with parents $B_1(\mathbf{U}_1),\ldots,B_m(\mathbf{U}_m)$

- for any instantiation $u_1,\ldots,u_k$ to $U_1,\ldots,U_k$ we would have:

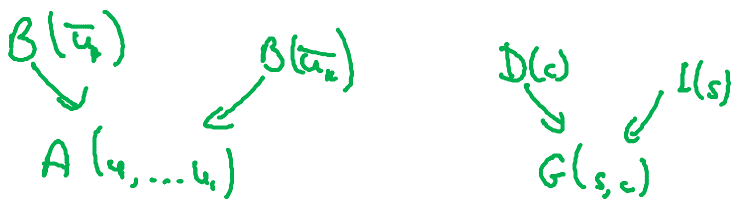
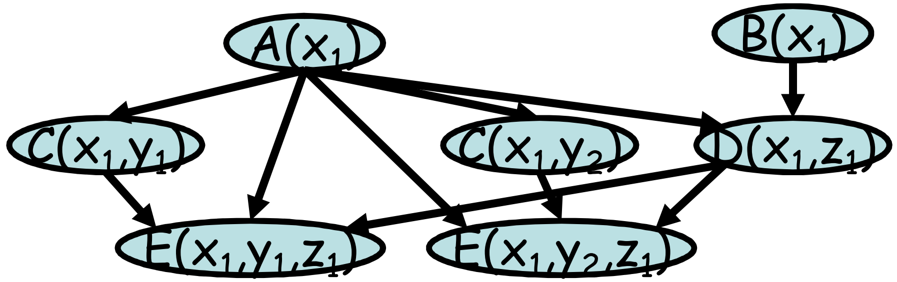


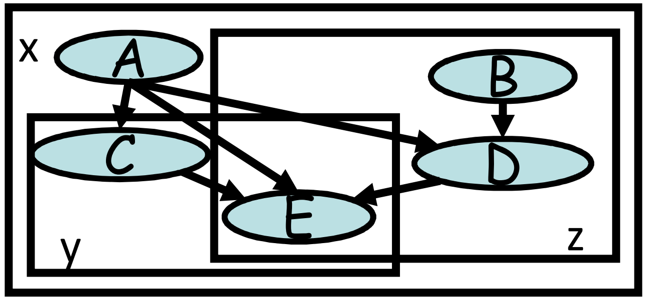
Plate Dependency Model

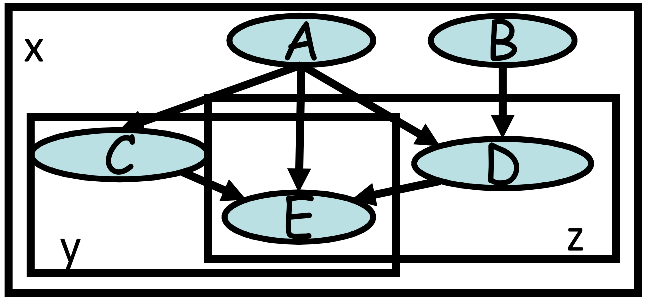
Let $A(U_1,\ldots,U_k)$ with parents $B_1(\mathbf{U}_1),\ldots,B_m(\mathbf{U}_m)$

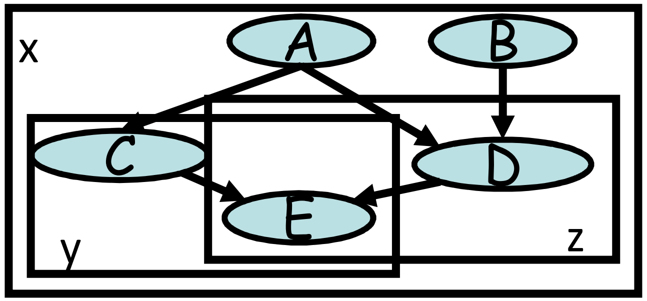
- For each i, we must have $\mathbf{U}_i \subseteq U_1,\ldots,U_k$
  - No indices in parent that are not in child

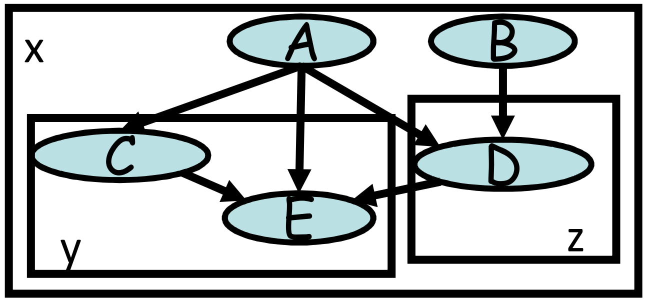
Which of the following plate models could have induced the ground network shown on the right?











Summary

$G(m) \rightarrow G(c) \leftarrow G(p)$ $\qquad X^{t-1} \rightarrow X^t$

- Template for an infinite set of BNs, each induced by a different set of domain objects
- Parameters and structure are reused within a BN and across different BNs
- Models encode correlations across multiple objects, allowing collective inference
- Multiple "languages", each with different tradeoffs in expressive power