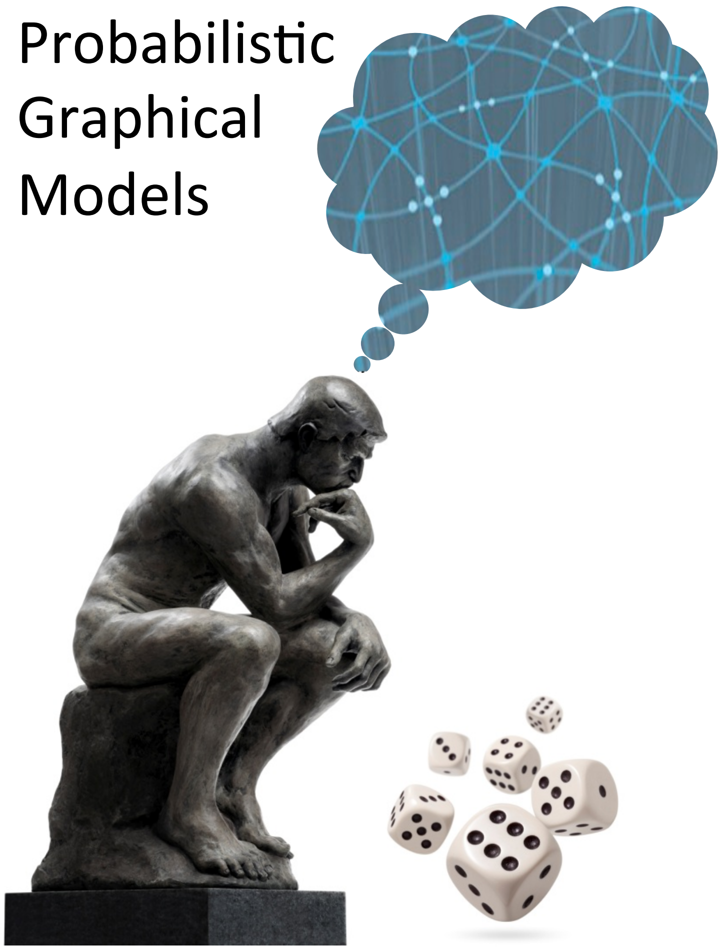
Probabilistic
Graphical
Models

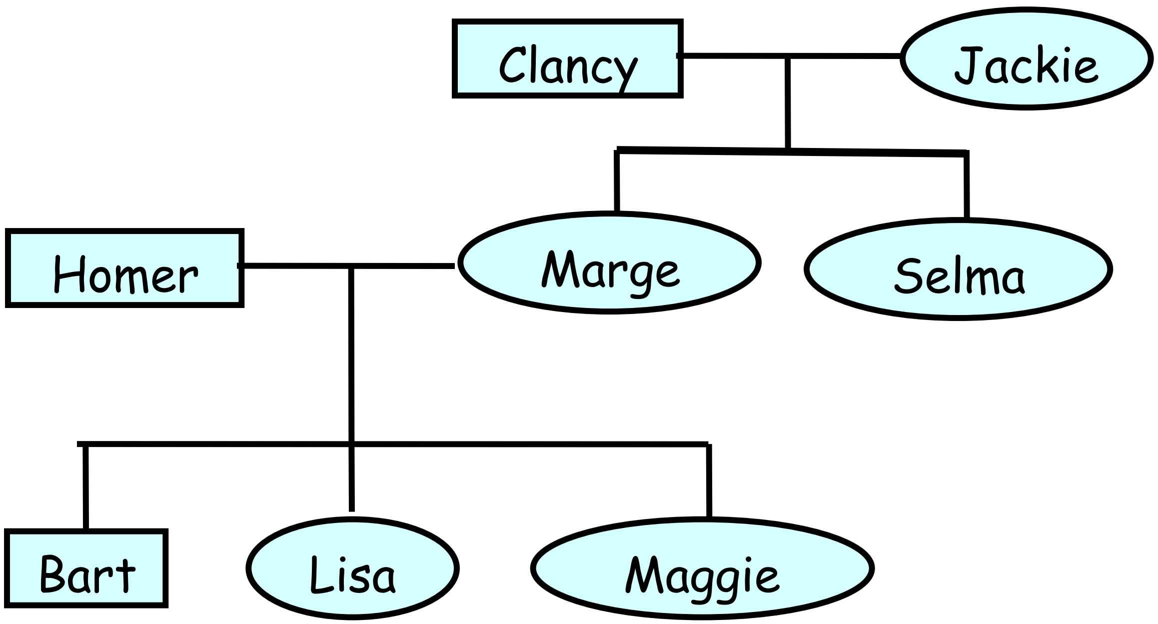


Representation

Template Models

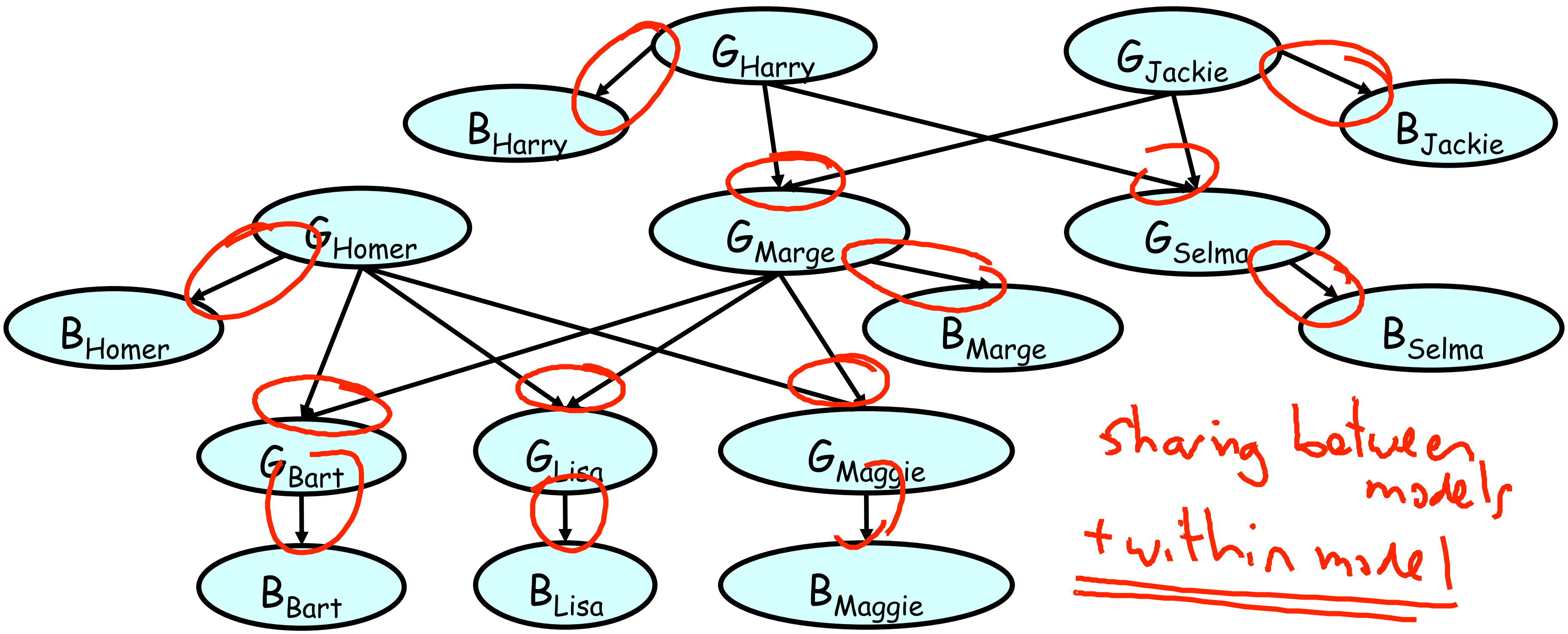
Overview

Genetic Inheritance



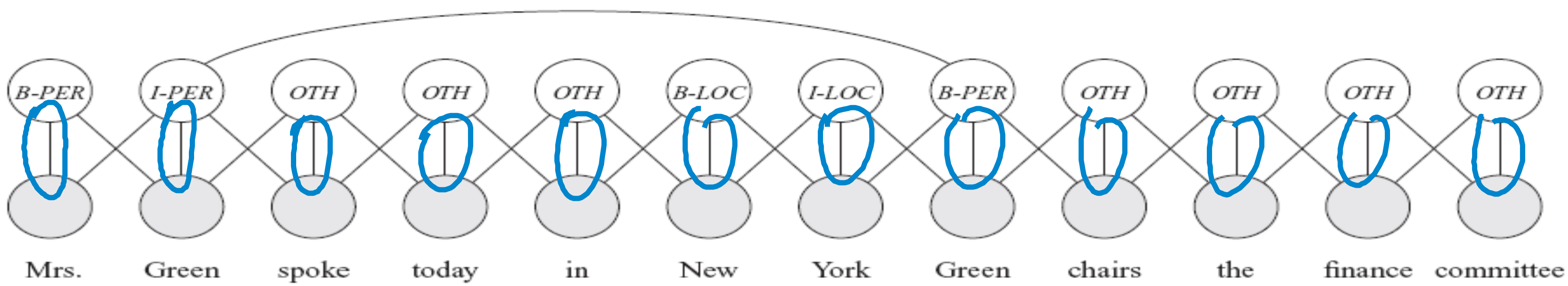
pedigree
trait

Genetic Inheritance



sharing between models
+ within model

NLP Sequence Models



KEY

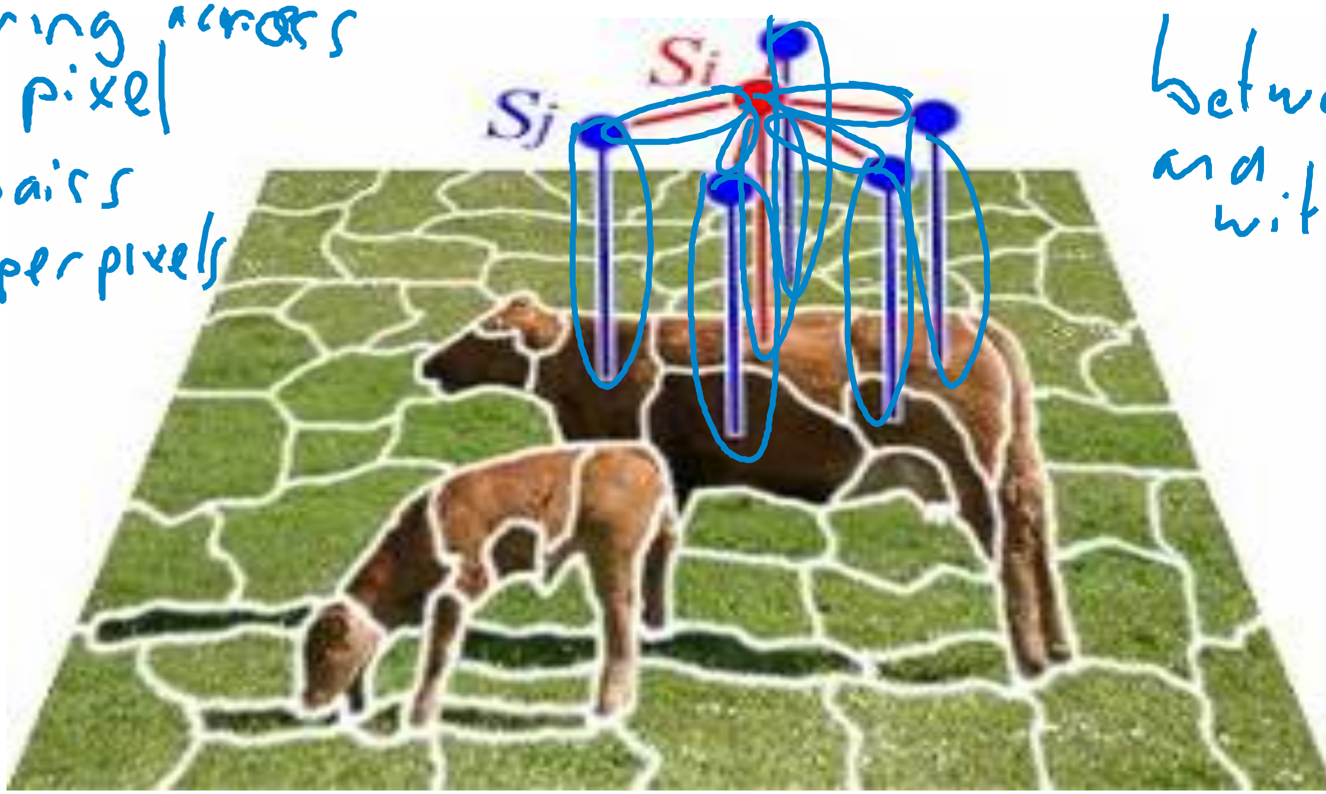
B-PER	Begin person name	*I-LOC*	Within location name
I-PER	Within person name	*OTH*	Not an entitiy
B-LOC	Begin location name		

Named entity recognition

Image Segmentation

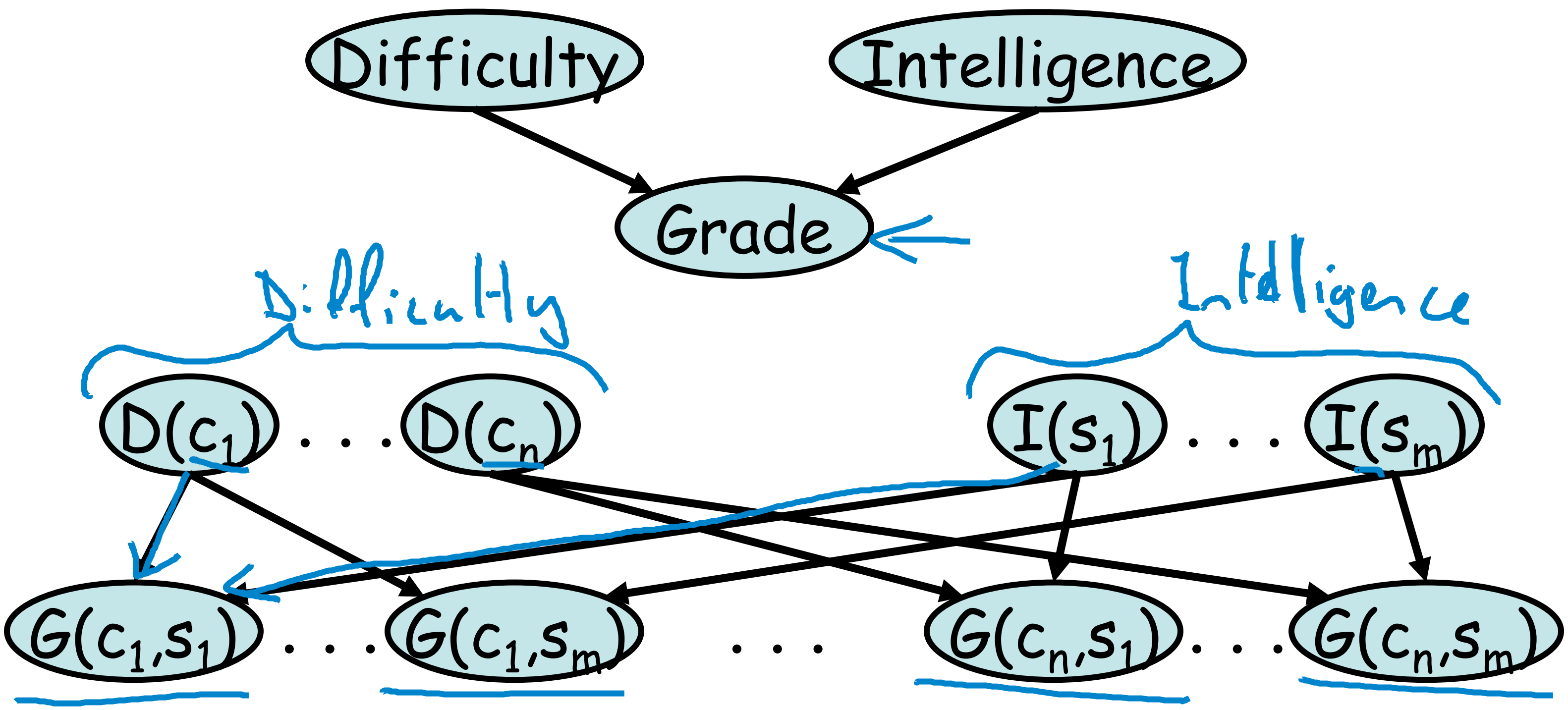
sharing across pixel

and pairs of superpixels



between and within

The University Example



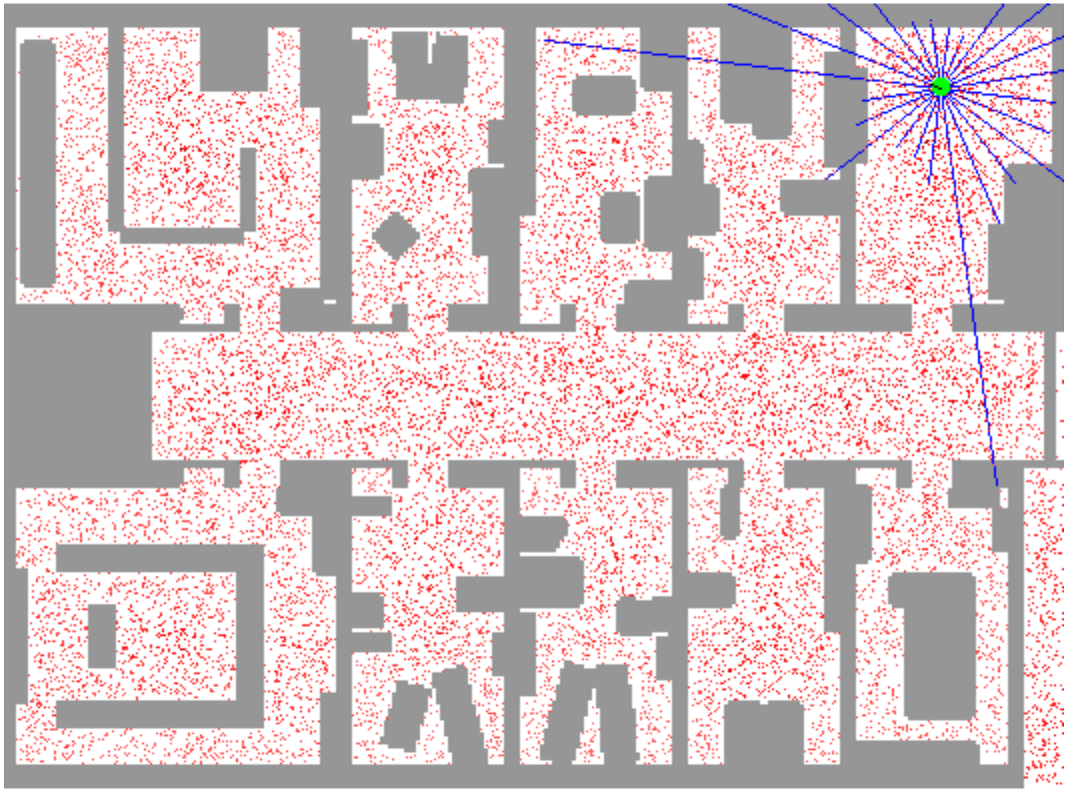
Daphne Koller

Robot Localization

time series

position at time t changes over time

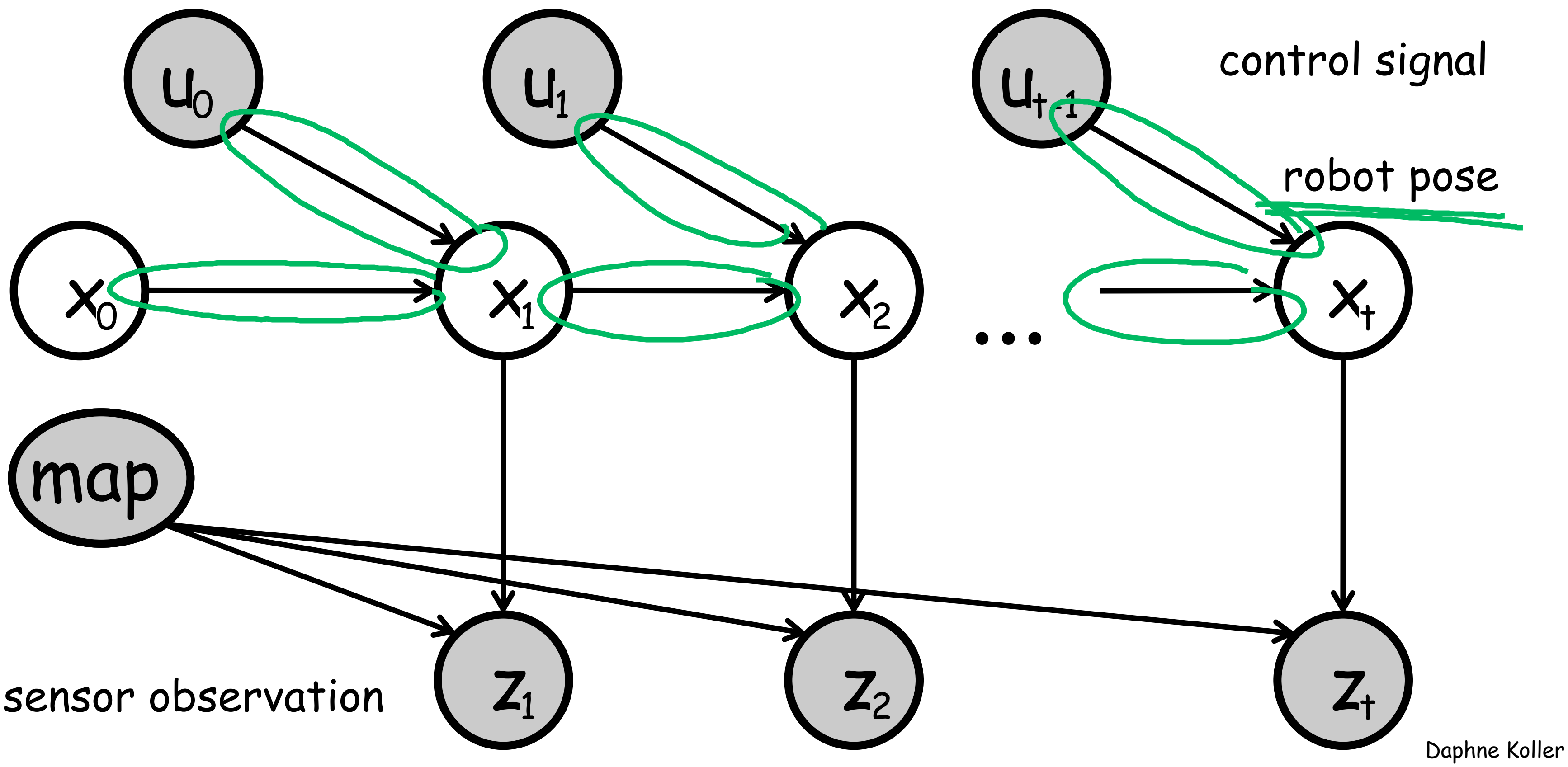
robot dynamics are fixed



Fox, Burgard, Thrun

Daphne Koller

Robot Localization



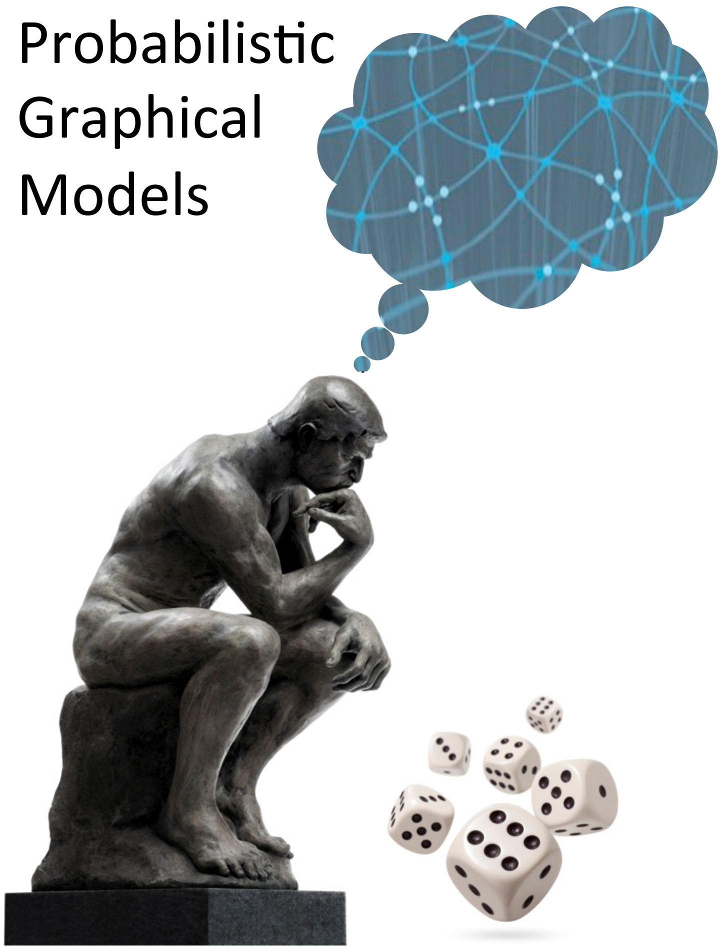
Template Variables

- 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

Probabilistic
Graphical
Models



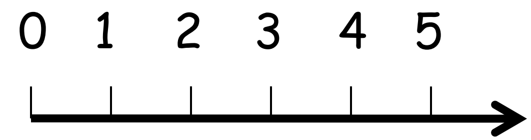
Representation

Template Models

Temporal Models

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

chain rule for probabilities

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

time flows forward

state at t+1 state at 0..t

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

forgetting

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

Is this true?

X = Location of robot probably not

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

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 $P(\mathbf{X}' \mid \mathbf{X})$
- For all t:

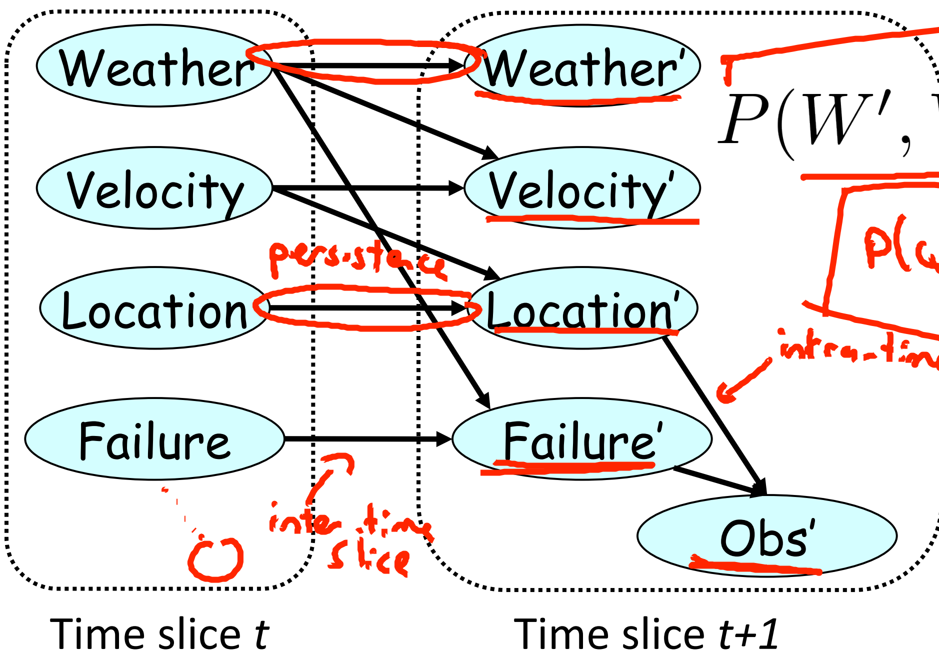
$$P(\boldsymbol{X}^{(t+1)} \mid \boldsymbol{X}^{(t)}) = P(\boldsymbol{X}' \mid \boldsymbol{X})$$

traffic time of day, day of week, football

enrich model by including

Template Transition Model

network fragment



cond dist t+1 given t

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

$P(W'|W) \cdot P(V'|W,V) \cdot P(L'|L,V)$

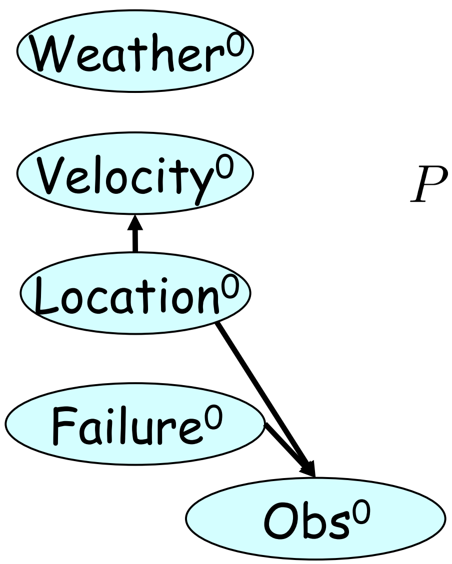
$\underbrace{P(F'|F,W)}_{t \text{ to } t+1} \cdot \underbrace{P(O'|L',F')}_{t+1}$

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

rapidly-acting

Daphne Koller

Initial State Distribution



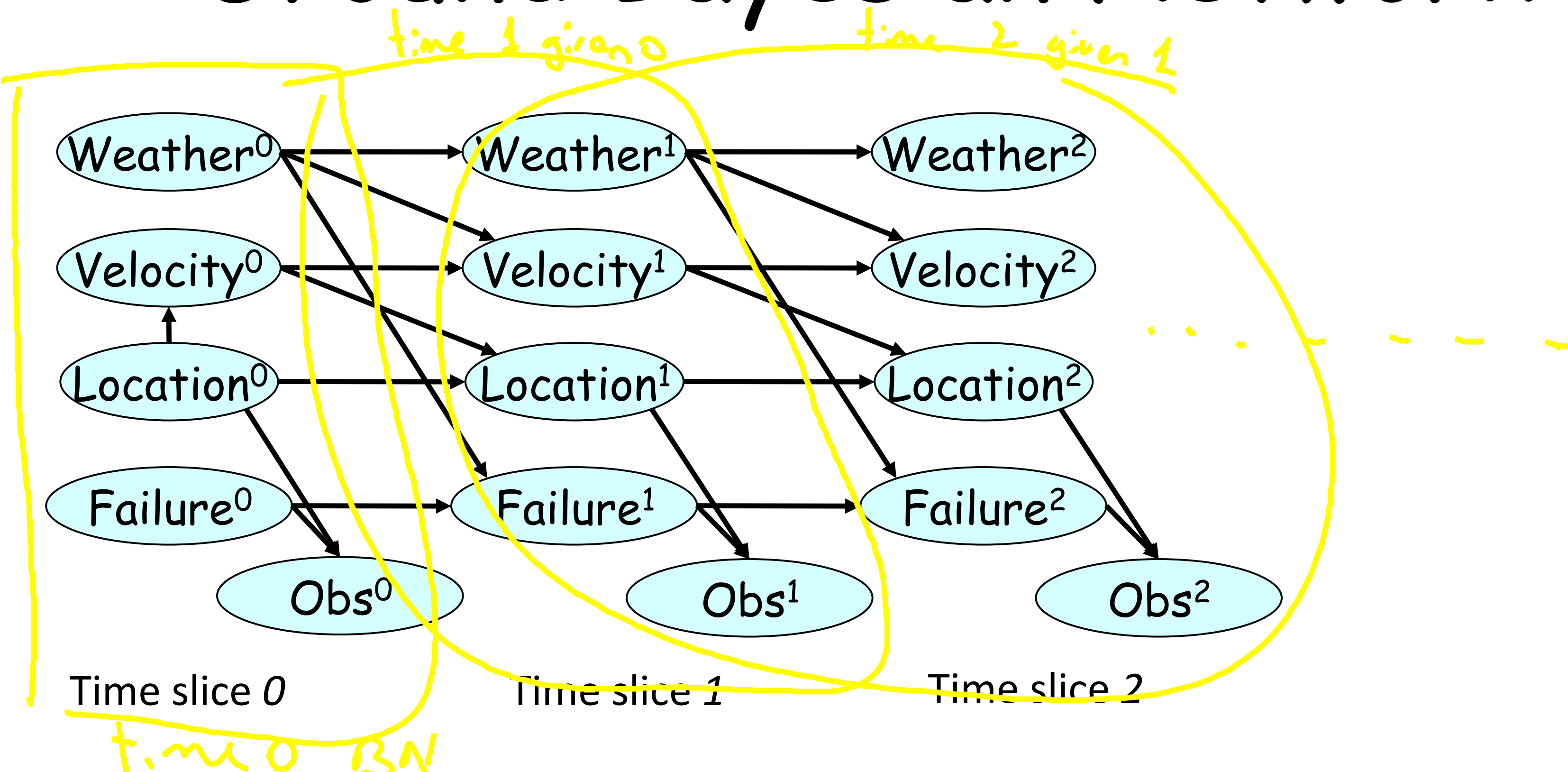
Time slice *0*

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

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

chain rule

Ground Bayesian Network



Daphne Koller

2-time-slice Bayesian Network

template

- A transition model (2TBN) over $X_1,\ldots,X_n$ is specified as a BN fragment such that:
  - The nodes include $X_1',\ldots,X_n'$ and a subset of $X_1,\ldots,X_n$

    the time t vars that directly affect state at t+1
  - Only the nodes $X_1',\ldots,X_n'$ have parents and a CPD
- The 2TBN defines a conditional distribution

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

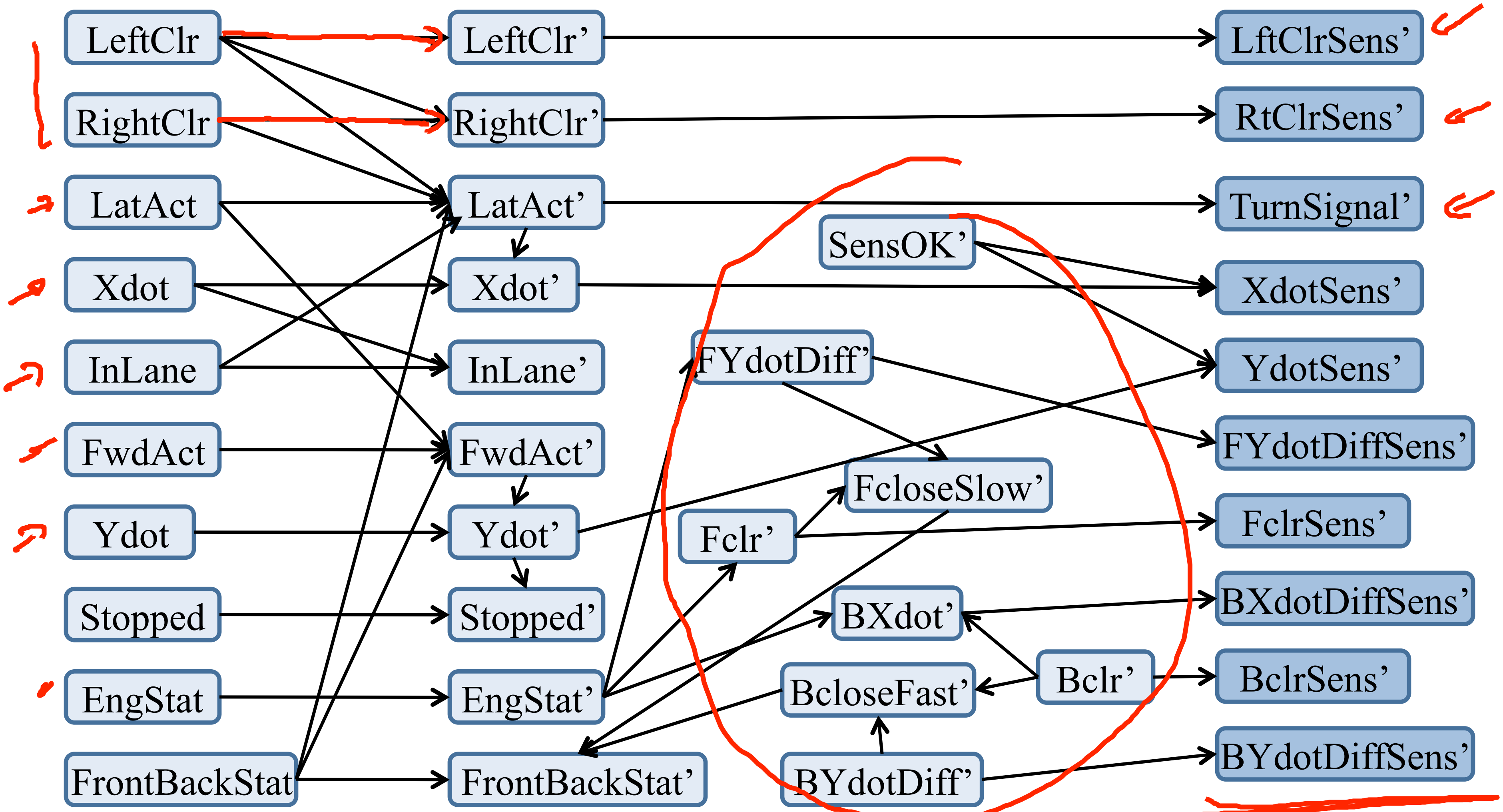
chain rule

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 0,…,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 $BN^{(0)}$
  - trans.t. – The dependency model for $X_1^{(t)}, \ldots, X_n^{(t)}$ for all $t > 0$ is copied from $BN_{\rightarrow}$



Huang, Koller, Malik, Ogasawara, Rao, Russell, Weber, AAAI 94

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
  - Time invariance

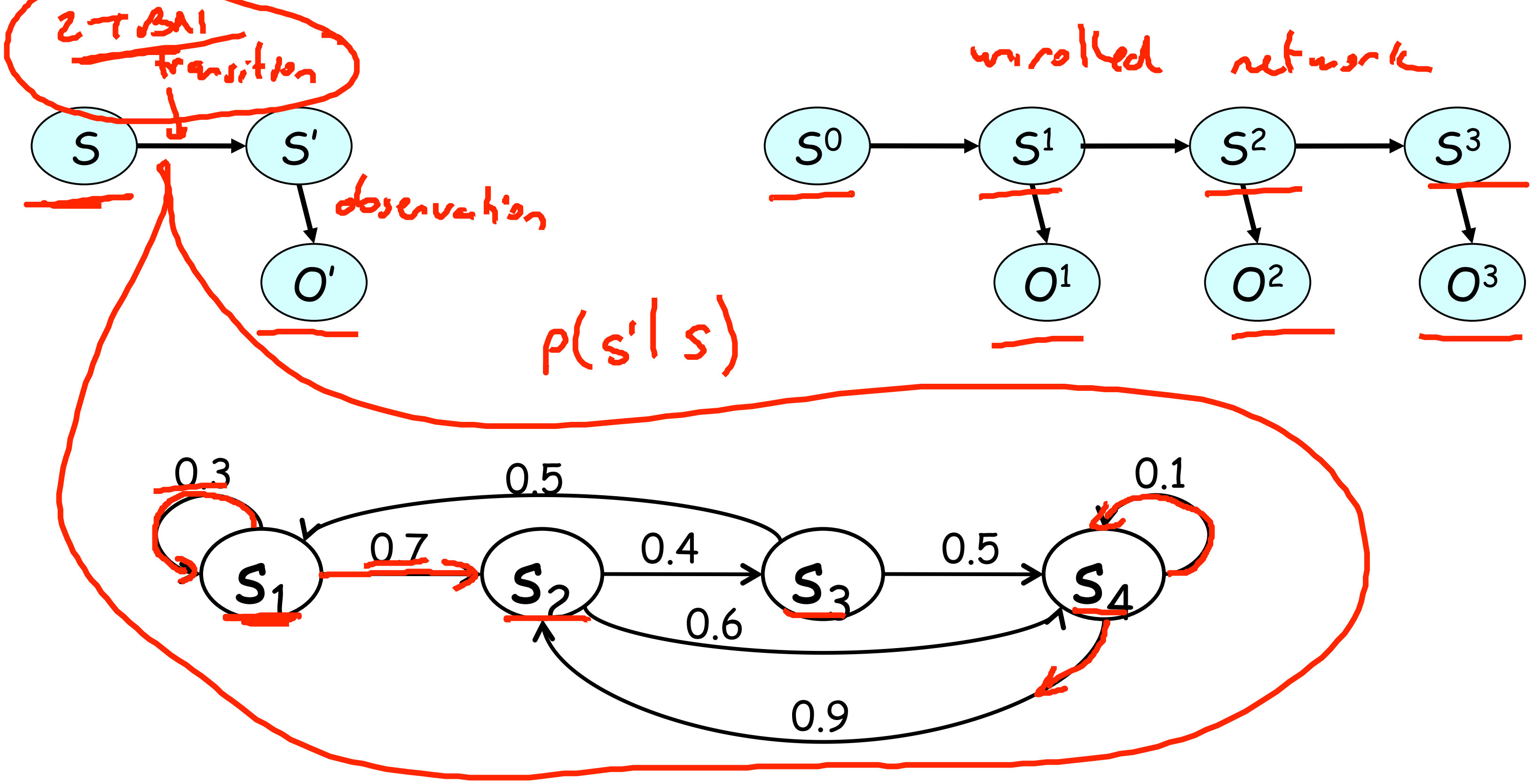
Probabilistic
Graphical
Models

Representation

Template Models

Hidden Markov Models

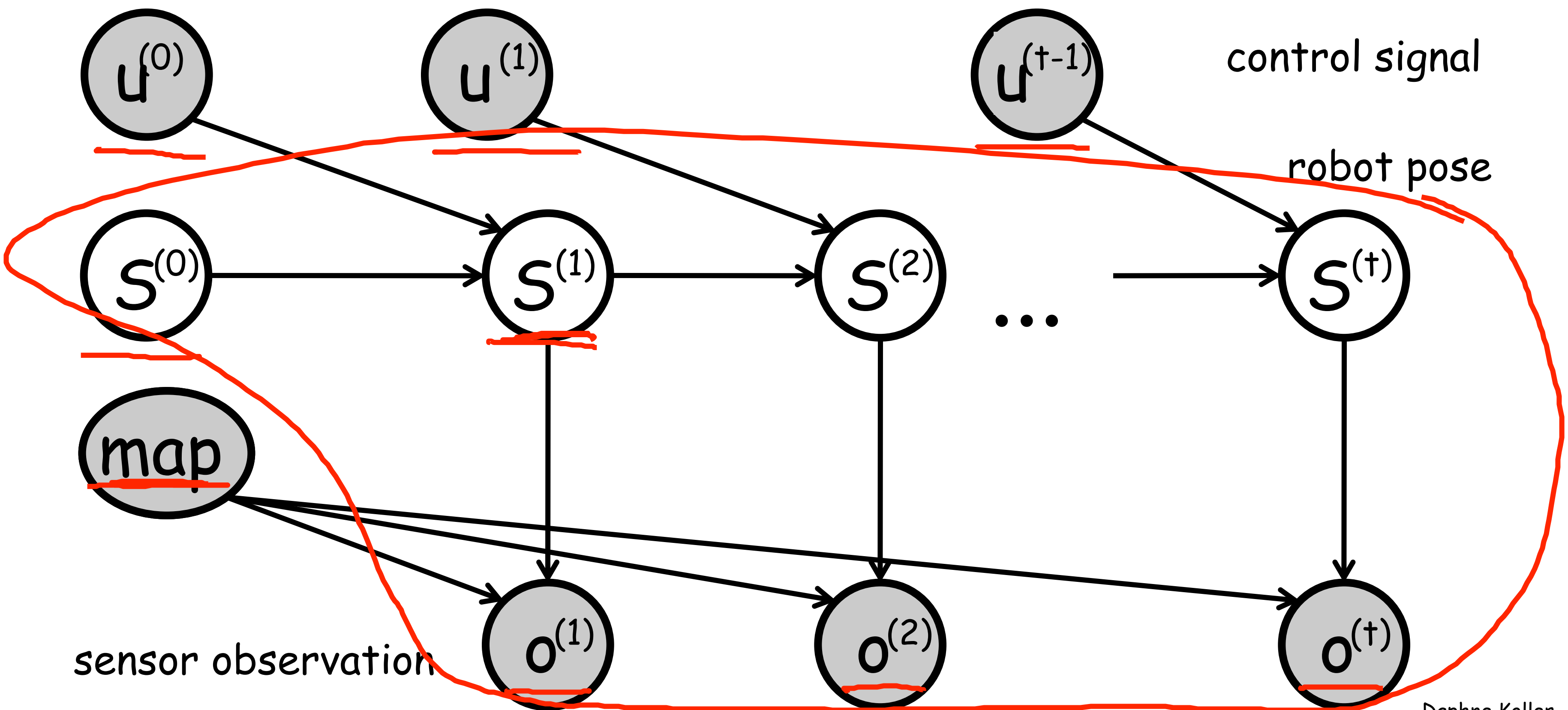
Hidden Markov Models



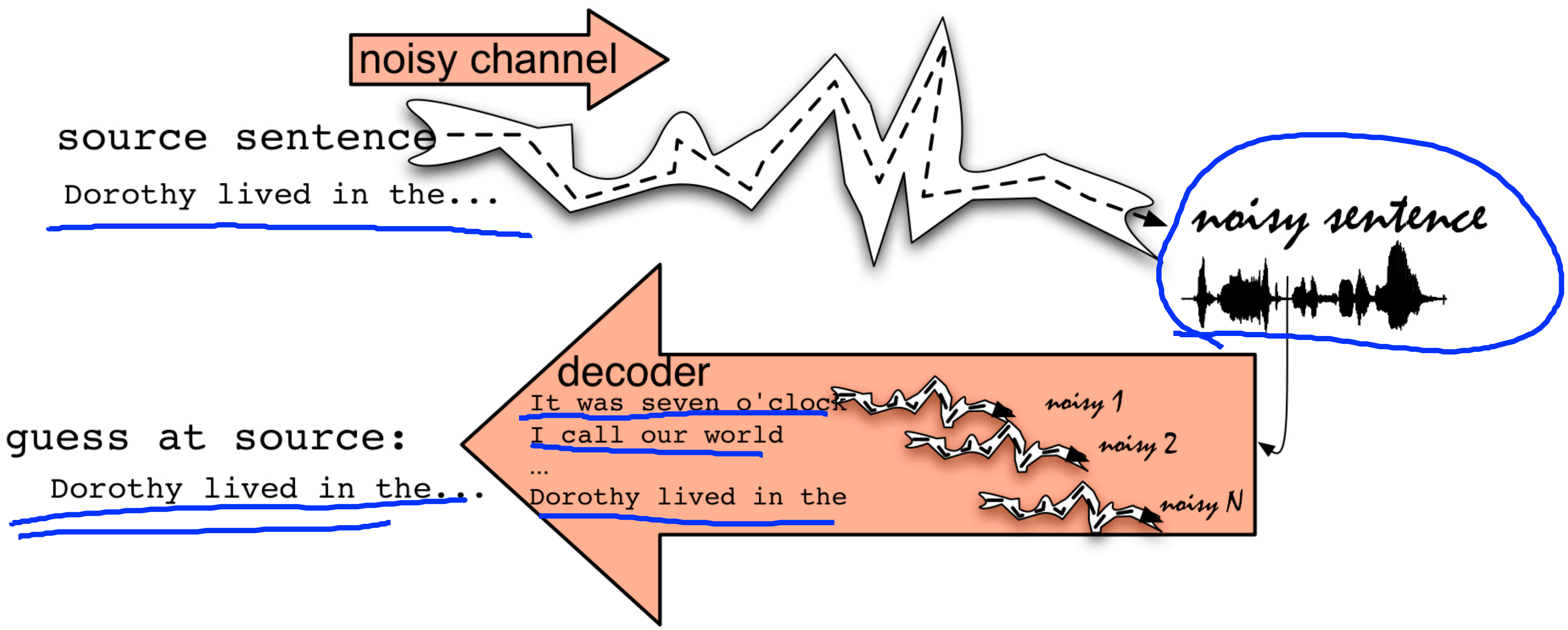
Numerous Applications

- Robot localization
- Speech recognition
- Biological sequence analysis
- Text annotation

Robot Localization



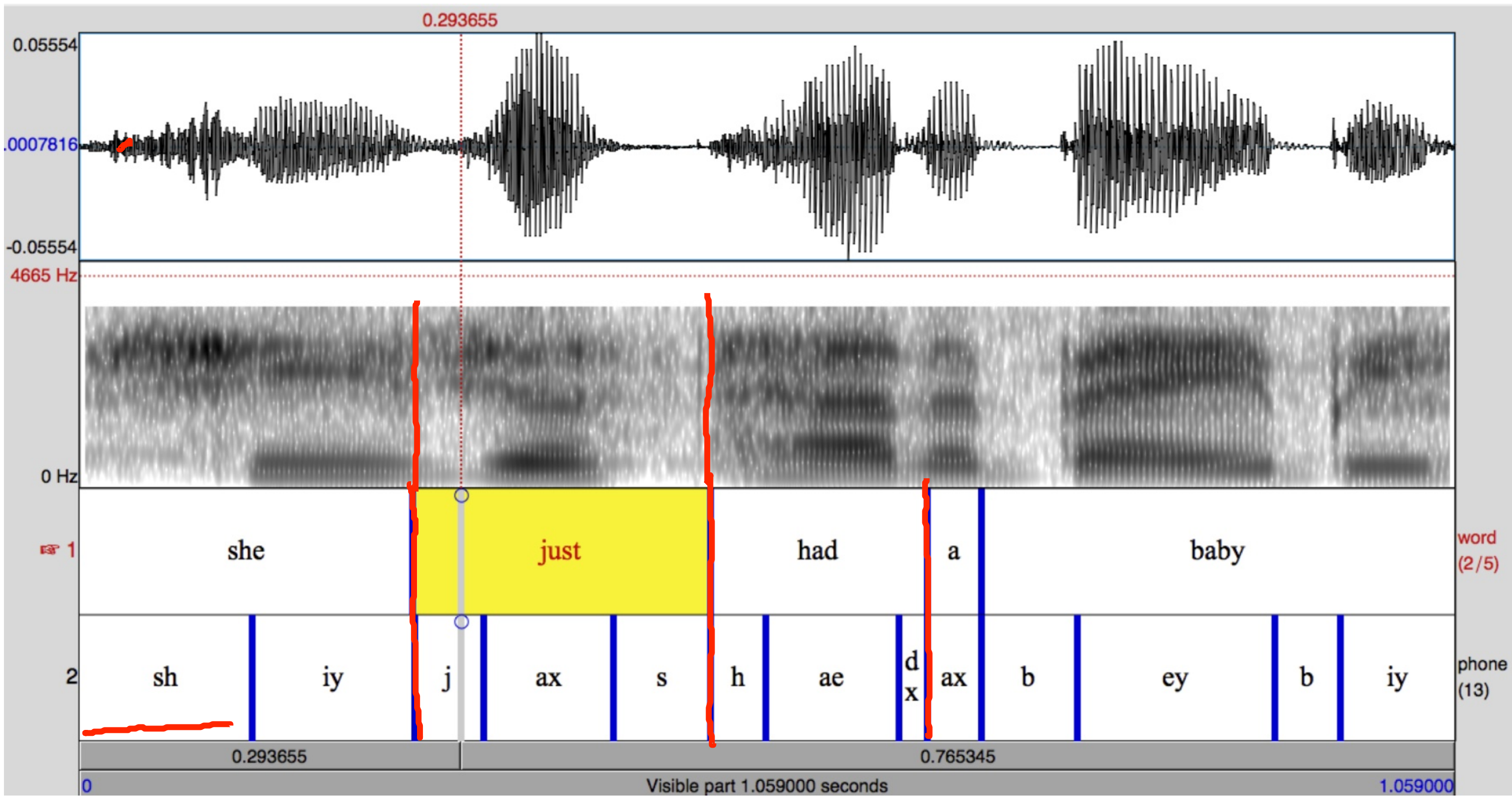
Speech Recognition



Dan Jurafsky, Stanford

Daphne Koller

Segmentation of Acoustic Signal



phones

Dan Jurafsky, Stanford

Daphne Koller

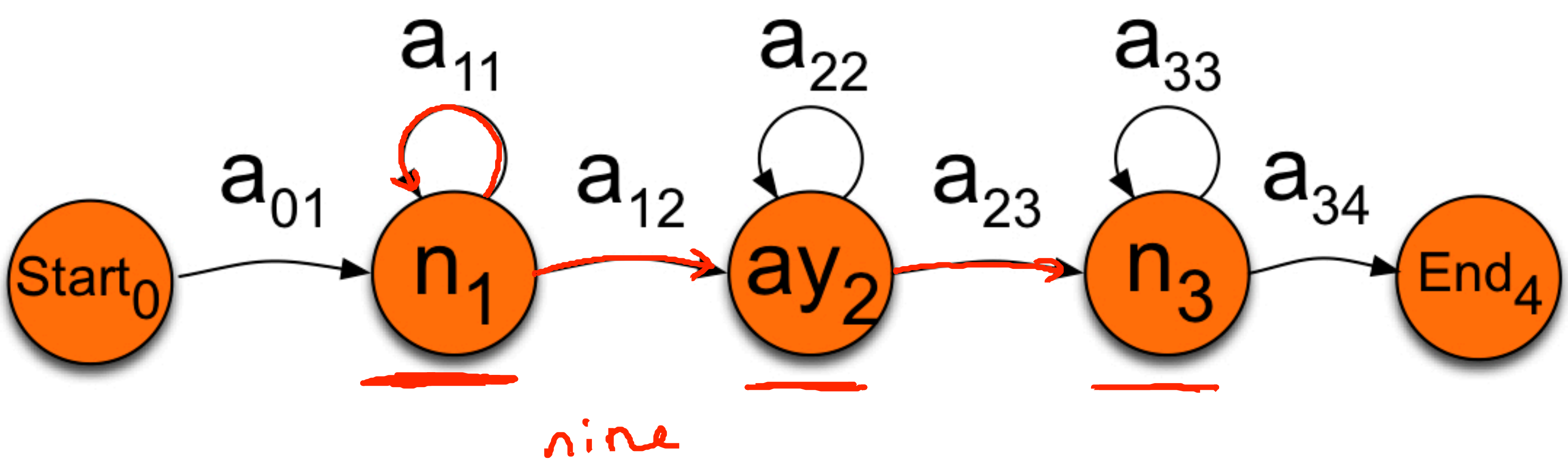
Phonetic Alphabet

AA	odd	AA D
AE	at	AE T
AH	hut	HH AH T
AO	ought	AO T
AW	cow	K AW
AY	hide	HH AY D
B	be	B IY
CH	cheese	CH IY Z
D	dee	D IY
DH	thee	DH IY
EH	Ed	EH D
ER	hurt	HH ER T
EY	ate	EY T
F	fee	F IY
G	green	G R IY N
HH	he	HH IY
IH	it	IH T
IY	eat	IY T
JH	gee	JH IY
K	key	K IY
L	lee	L IY
M	me	M IY
N	knee	N IY
NG	ping	P IH NG
OW	oat	OW T
OY	toy	T OY
P	pee	P IY
R	read	R IY D
S	sea	S IY
SH	she	SH IY
T	tea	T IY
TH	theta	TH EY T AH
UH	hood	HH UH D
UW	two	T UW
V	vee	V IY
W	we	W IY
Y	yield	Y IY L D
Z	zee	Z IY
ZH	seizure	S IY ZH ER

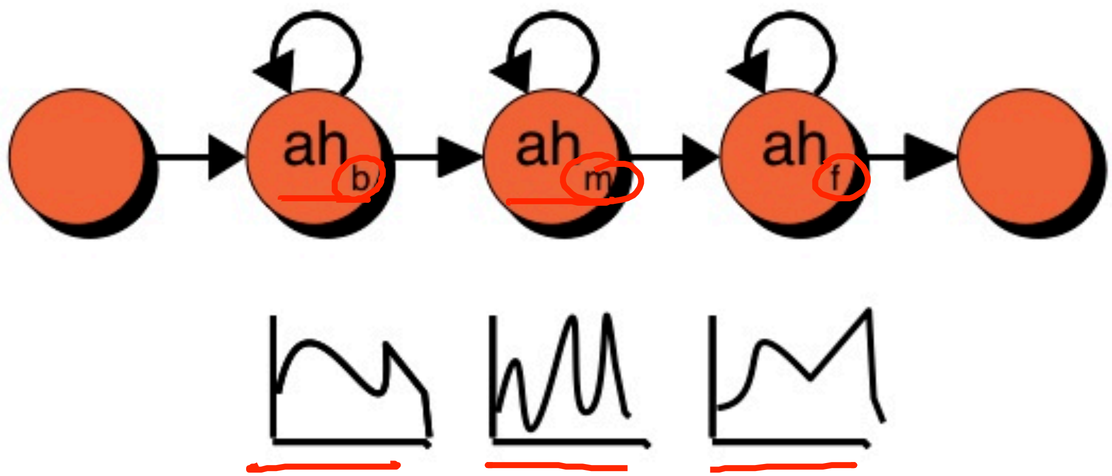
http://www.speech.cs.cmu.edu/cgi-bin/cmudict

The CMU Pronouncing Dictionary

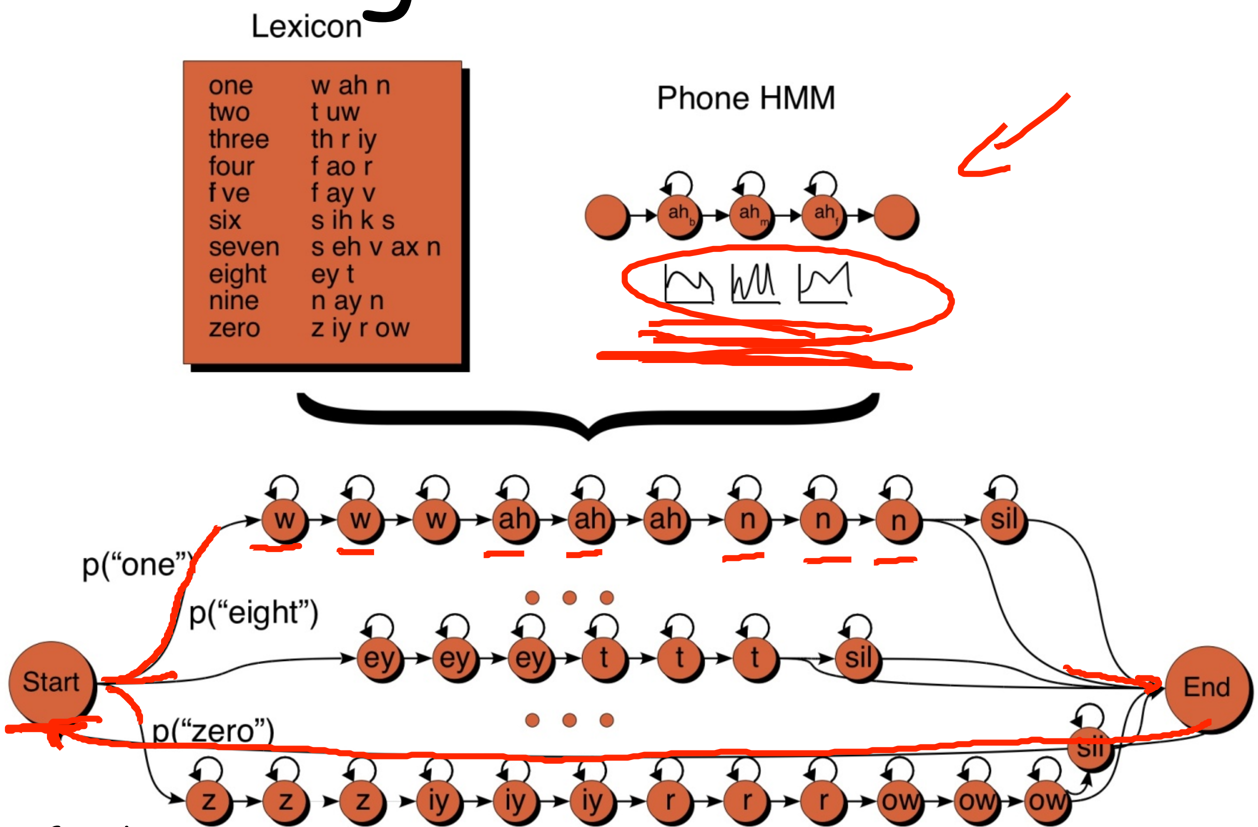
Word HMM



Phone HMM



Recognition HMM



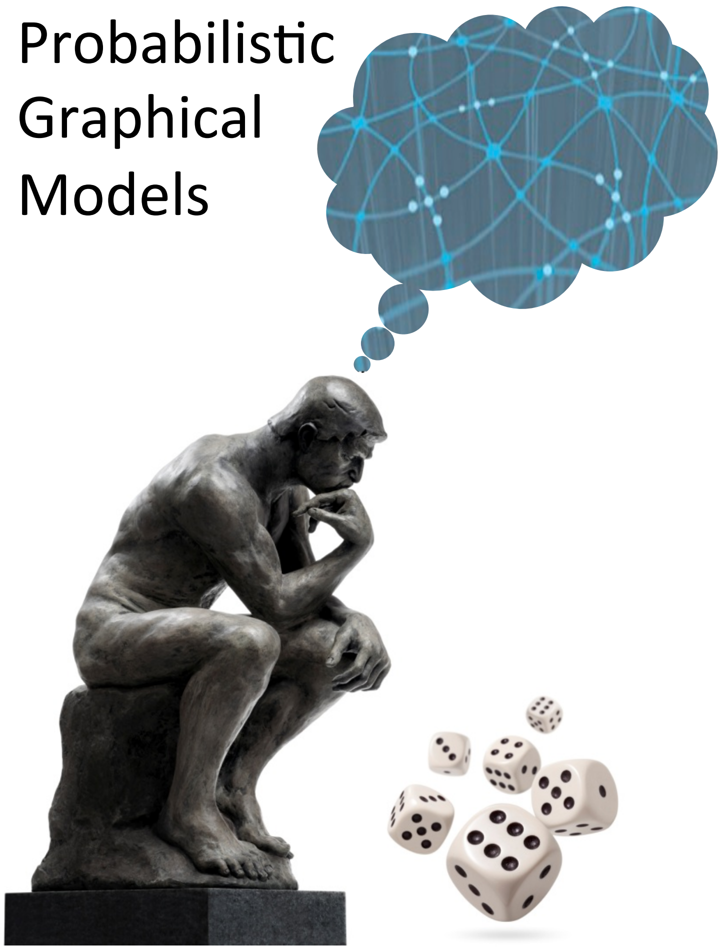
Dan Jurafsky, Stanford

Daphne Koller

Summary

- HMMs can be viewed as a subclass of DBNs
- HMMs seem unstructured at the level of random variables
- HMM structure typically manifests in sparsity and repeated elements within the transition matrix
- HMMs are used in a wide variety of applications for modeling sequences

Daphne Koller

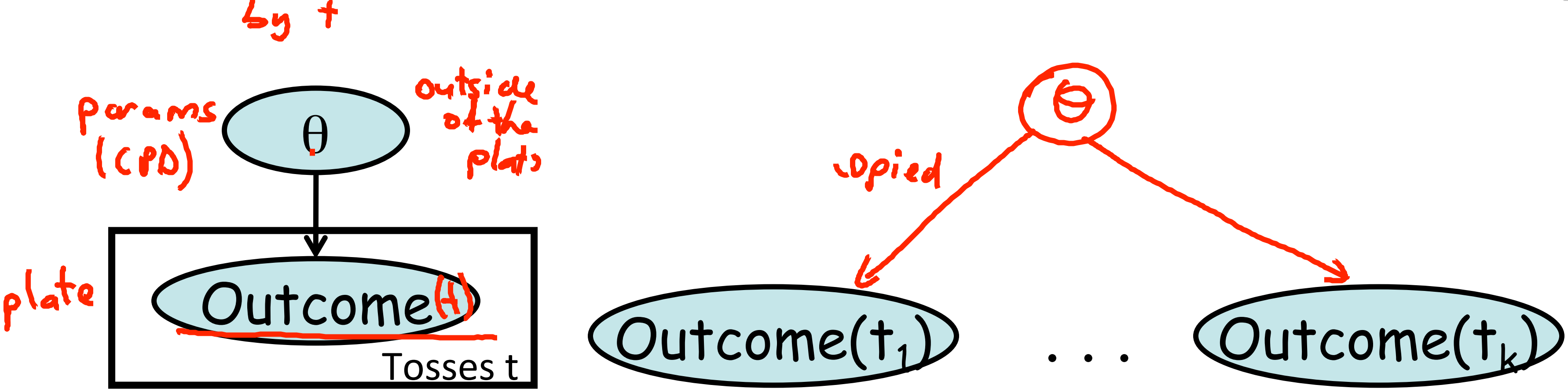


Representation

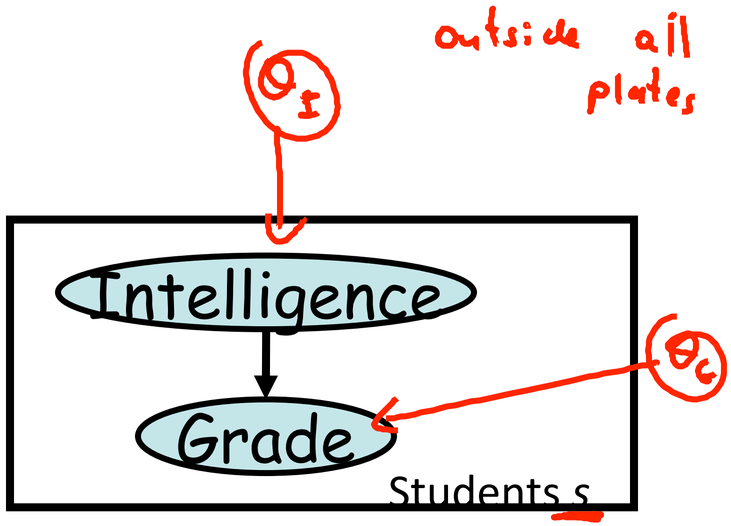
Template Models

Plate Models

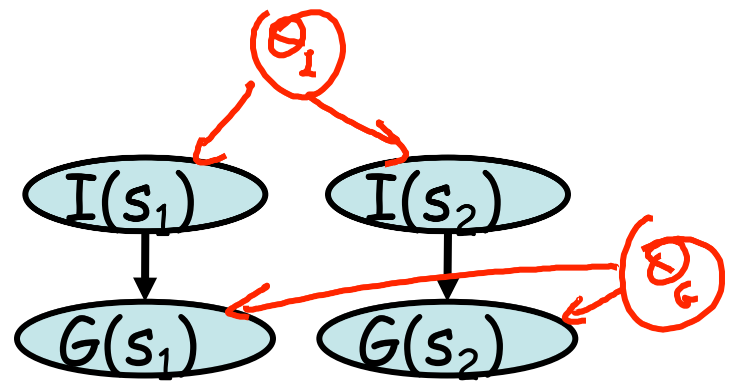
Modeling Repetition



$\underbrace{\{t_1, \ldots, t_k\}}$



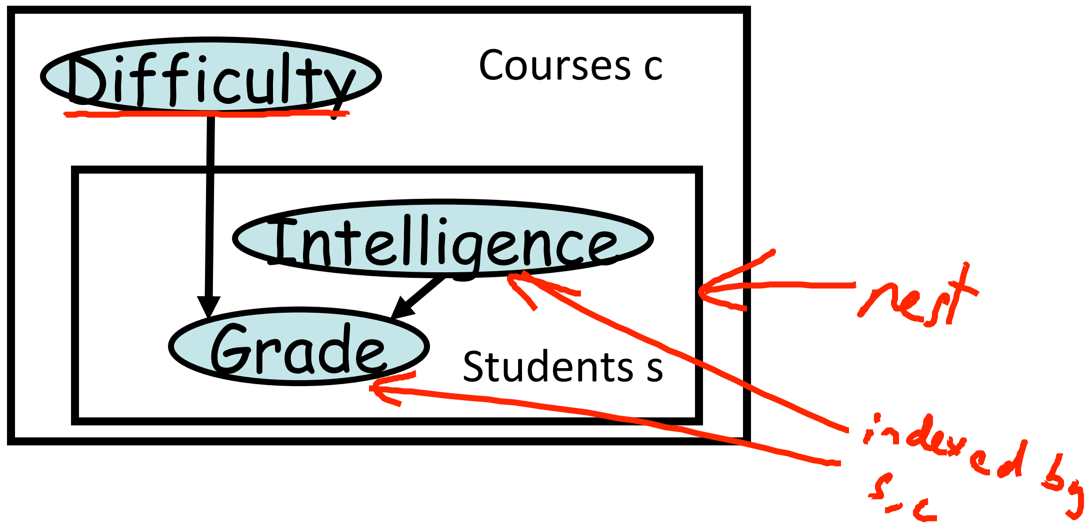
outside all plates
θ_I
Intelligence
Grade
θ_G
Students s

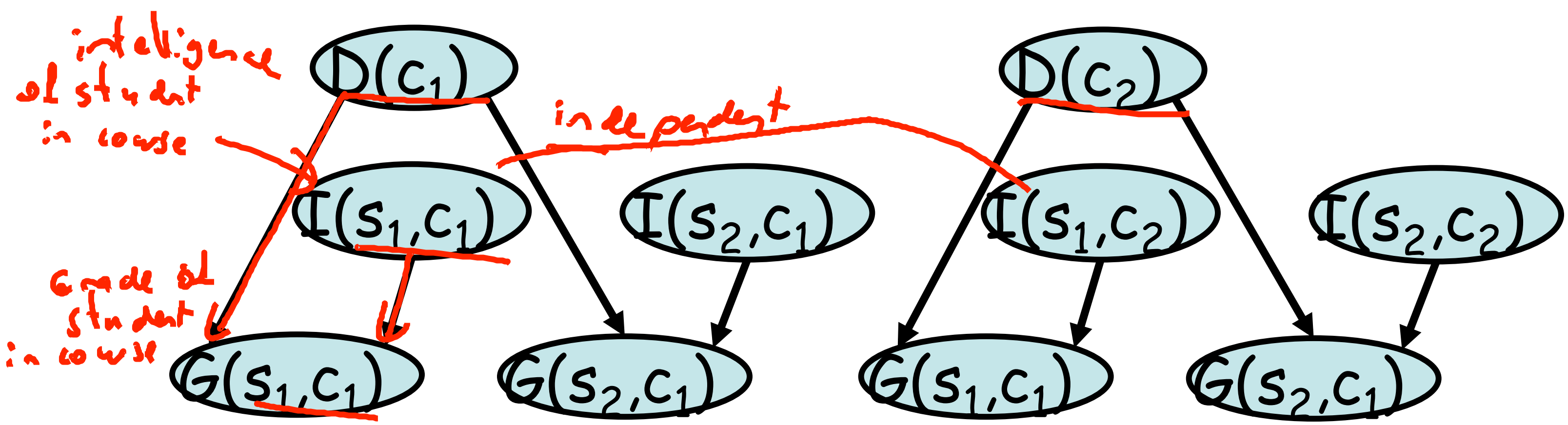


θ_I
$I(s_1)$
$I(s_2)$
$G(s_1)$
$G(s_2)$
θ_G

Nested Plates

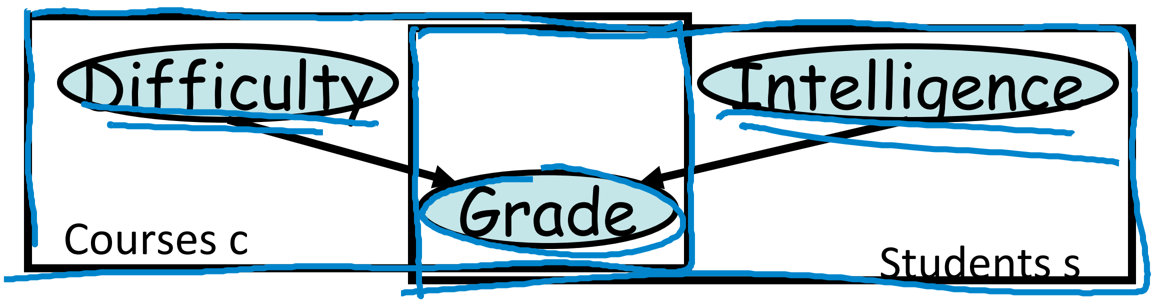
courses c
students s

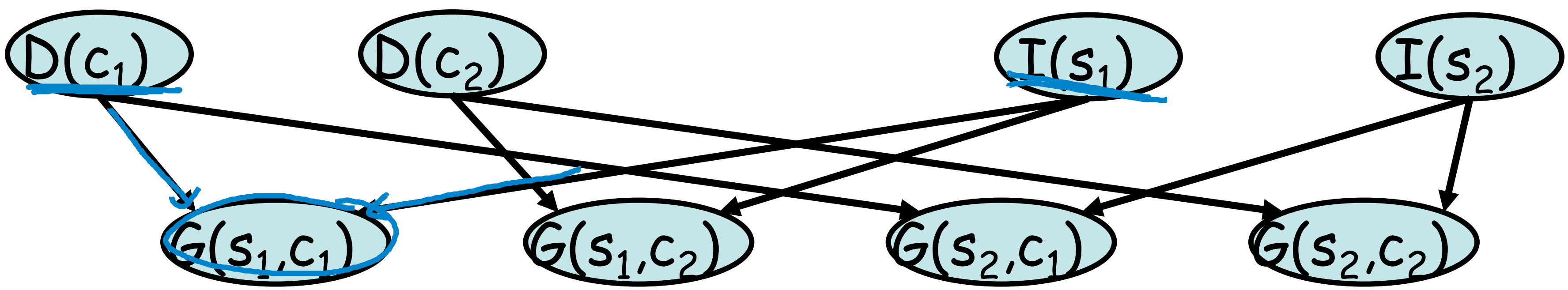




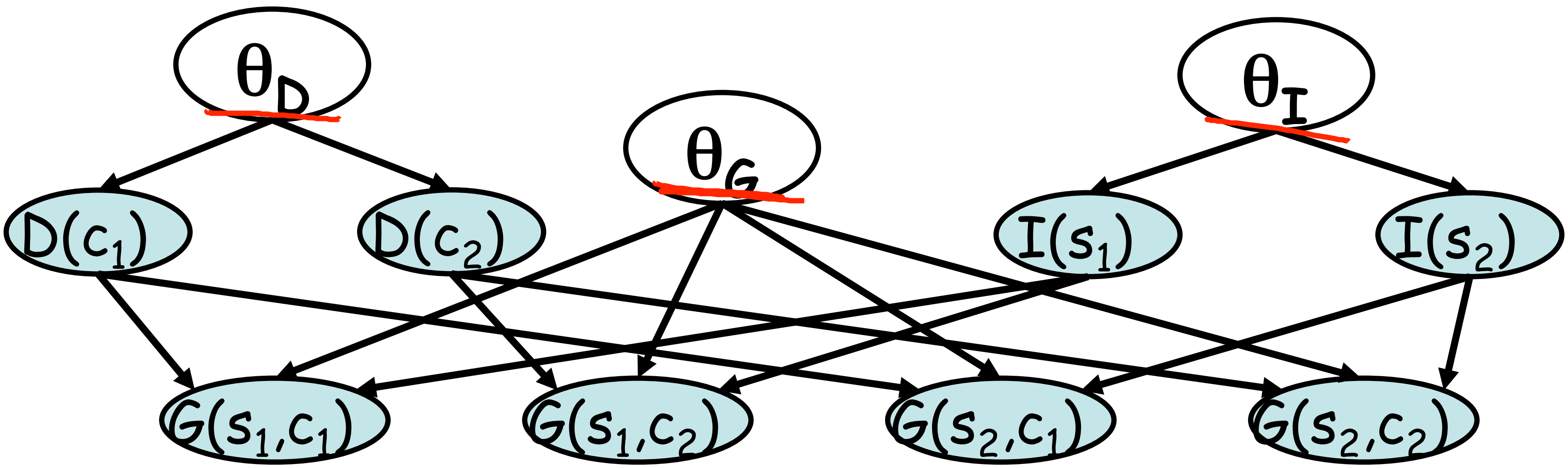
Daphne Koller

Overlapping Plates



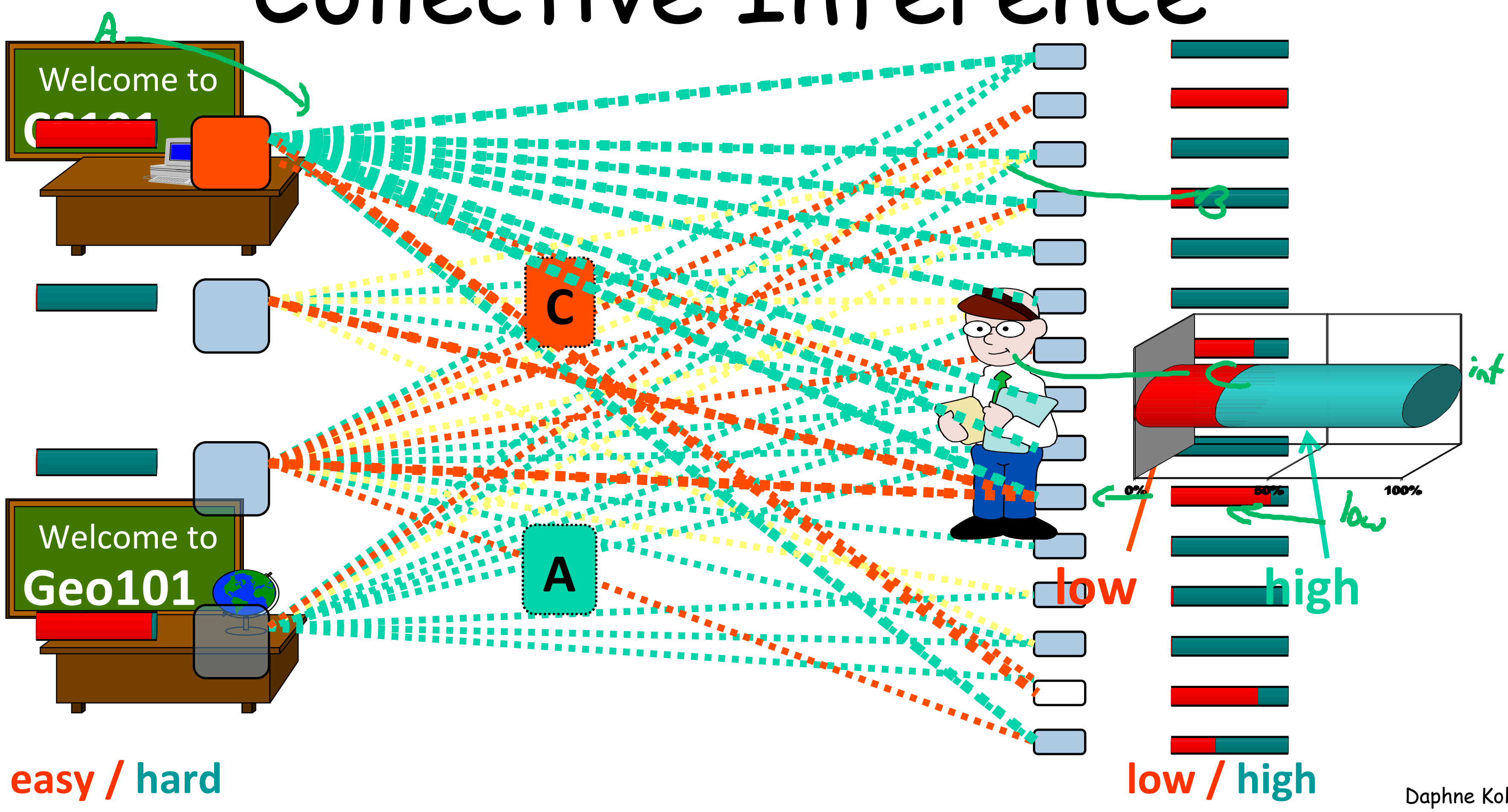


Explicit Parameter Sharing



Daphne Koller

Collective Inference



Daphne Koller

Plate Dependency Model

indexed

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

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

I(s) D(c)

G(s,c)

G(s,c)

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:

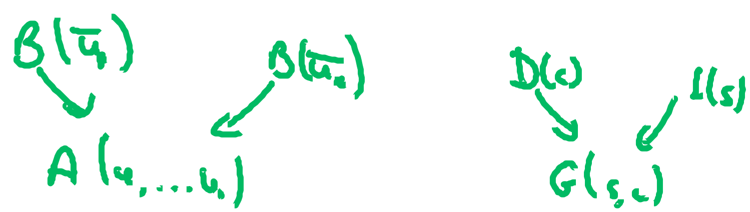


Plate Dependency Model

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

$G(m) \rightarrow G(c) \leftarrow G(p)$

Summary

$X^{t-1} \rightarrow X^{t}$

- 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