PSDDs: Probabilistic SDDs

(and Conditional PSDDs)

Learning From Data and Knowledge

Learning Distributions over Combinatorial Objects

Learning Distributions over Structured Spaces

Learning From Data and Knowledge

probability space

A	B	C	D
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

knowledge

expressed in logic

structured space

A	B	C	D
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

Tractable Boolean Circuit (SDD)

data

A	B	C	D	
0	0	0	0	
0	0	0	1	
0	0	1	0	6.0%
0	0	1	1	54.0%
0	1	0	0	
0	1	0	1	
0	1	1	0	
0	1	1	1	10.0%
1	0	0	0	3.6%
1	0	0	1	
1	0	1	0	1.8%
1	0	1	1	0.6%
1	1	0	0	14.5%
1	1	0	1	
1	1	1	0	7.2%
1	1	1	1	2.4%

Tractable Probabilistic Circuit (PSDD)

- Supervised
- Unsupervised

- Impact of knowledge:
  - ✓ amount of data needed
  - ✓ robustness of ML systems
  - ✓ generality of ML systems

Basic Concepts

- Probability distributions
- SDD circuits
- Structured spaces

Probability Distributions

world	Earthquake (E)	Burglary (B)	[illegible]	Pr(.)
ω_1	true	true	[illegible]	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

notation : A, E

$A \wedge E$

- marginal probability

$$Pr(A) = Pr(w_1) + Pr(w_3) + Pr(w_5) + Pr(w_7)$$

$$Pr(\neg A) = Pr(w_2) + Pr(w_4) + Pr(w_6) + Pr(w_8)$$

$$Pr(A, E) = Pr(w_1) + Pr(w_3)$$

- conditional probability

$$Pr(E \mid A) = \frac{Pr(A, E)}{Pr(A)}$$

$$Pr(\alpha \mid \beta) = \frac{Pr(\alpha, \beta)}{Pr(\beta)}$$

Sentential Decision Diagrams (SDDs)

(**X**,**Y**)-Partition

$$f(\mathbf{X},\mathbf{Y}) = g_1(\mathbf{X})h_1(\mathbf{Y}) + \ldots + g_n(\mathbf{X})h_n(\mathbf{Y})$$

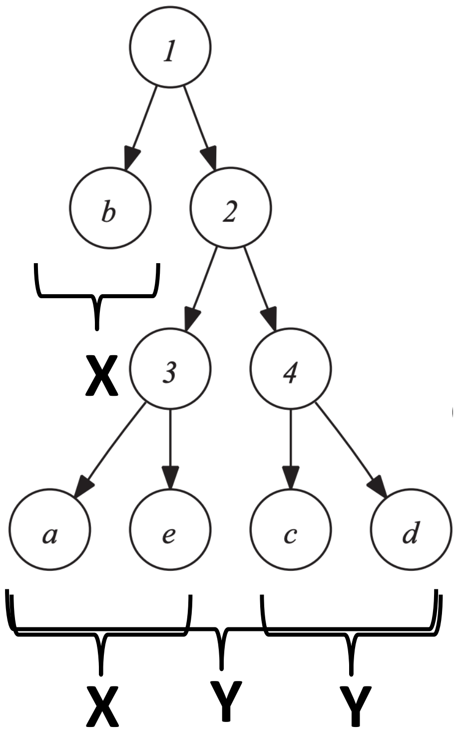
primes: g_i subs: h_i

$$g_i \neq 0 \text{ for all } i$$

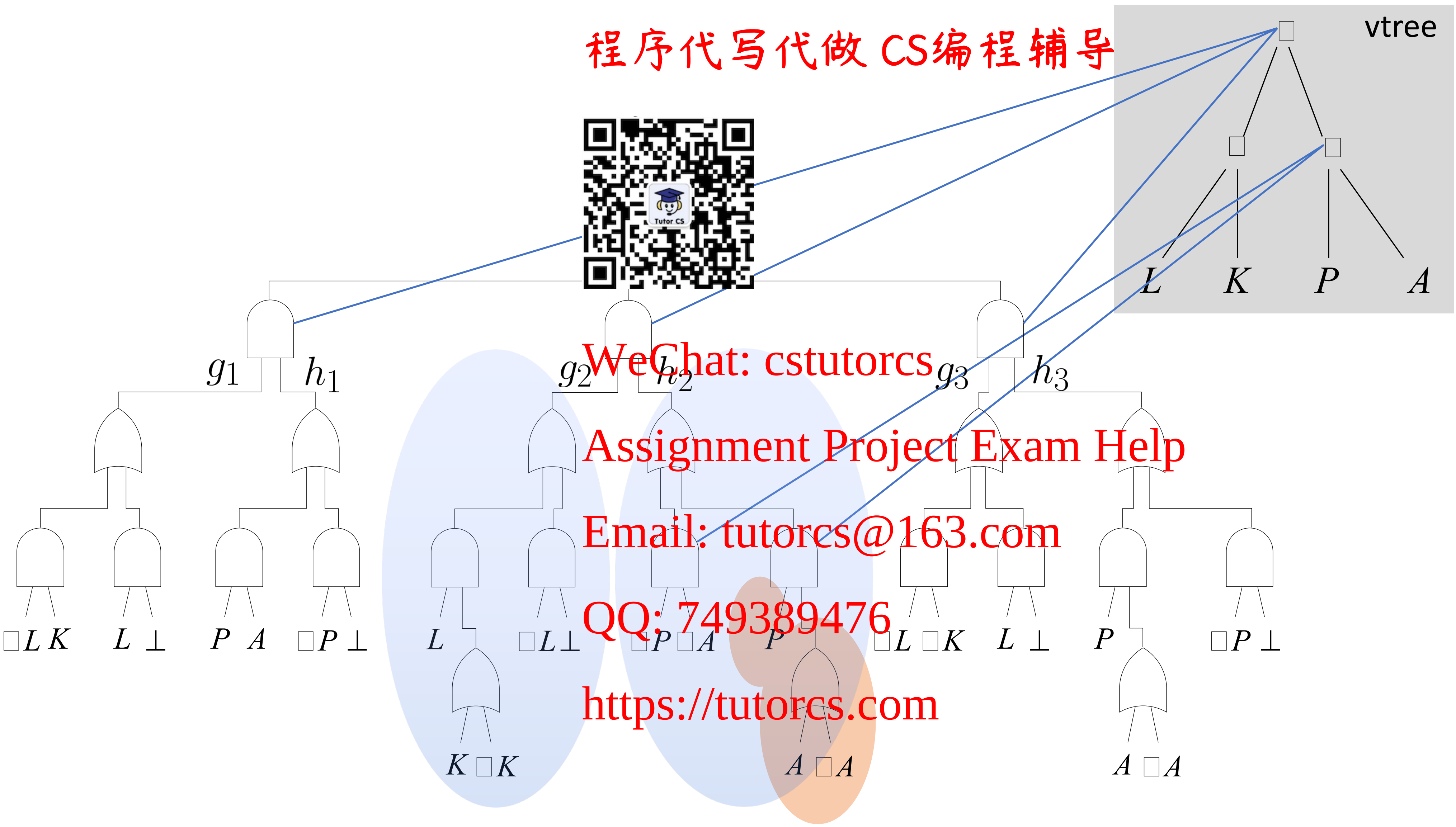
$$g_i g_j = 0 \text{ for } i \neq j$$

$$g_1 + \ldots + g_n = 1$$

vtree



compressed (**X**,**Y**)-Partition is unique


vtree
L K P A
g1 h1 g2 h2 g3 h3
□L K
L ⊥
P A
□P ⊥
L
□L⊥
□P □A
P
□L □K
L ⊥
P
□P ⊥
K □K
A □A
A □A

Sentential Decision Diagrams (SDDs)

(X,Y)-Partition

$$f(\mathbf{X}, \mathbf{Y}) = g_1(\mathbf{X})h_1(\mathbf{Y}) + \ldots + g_n(\mathbf{X})h_n(\mathbf{Y})$$

primes: g_i subs: h_i

$g_1 \cdots g_n$

$Pr(\alpha) = Pr(\alpha | g_1) Pr(g_1) +$

$\vdots$

$Pr(\alpha | g_n) Pr(g_n)$

$$g_i \neq 0 \text{ for all } i$$

$$g_i g_j = 0 \text{ for } i \neq j$$

$$g_1 + \ldots + g_n = 1$$

Structured Spaces

three examples

Structured Space

unstruct

L	K	P	
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

structured

L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

Structured Space: Domain Constraints

Background Knowledge

Logic (L)
Knowledge Representation (K)
Probability (P)
Artificial Intelligence (A)

Must take at least one of Probability or Logic.
Probability is a prerequisite for AI.
The prerequisites for KR is either AI or Logic.

$P \vee L \quad A \Rightarrow P \quad K \Rightarrow (A \vee L)$

Data

L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3

Structured Space

unstructured

L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

$$P \vee L$$

$$A \Rightarrow P$$

$$K \Rightarrow (P \vee L)$$

structured

L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

Structured Space

$$P \vee L$$
$$A \Rightarrow P$$
$$K \Rightarrow (P \vee L)$$

L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

Structured Space: Combinatorial Objects

rank	sushi
1	fatty tuna
2	sea urchin
3	salmon roe
4	shrimp
5	tuna
6	squid
7	tuna roll
8	see eel
9	egg
10	cucumber roll

rank	[illegible]
1	[illegible]
2	sea urchin
3	salmon roe
4	fatty tuna
5	tuna
6	squid
7	tuna roll
8	see eel
9	egg
10	cucumber roll

10 items:
3,628,800
rankings

20 items:
2,432,902,008,176,640,000
rankings

Structured Space: Rankings

rank	sushi
1	fatty tuna
2	sea urchin
3	salmon roe
4	shrimp
5	tuna
6	squid
7	tuna roll
8	see eel
9	egg
10	cucumber roll

rank	[illegible]
1	[illegible]
2	sea urchin
3	salmon roe
4	fatty tuna
5	tuna
6	squid
7	tuna roll
8	see eel
9	egg
10	cucumber roll

A_{ij} item i at position j (n items require n^2 Boolean variables)

An item may be assigned to more than one position

A position may contain more than one item

Structured Space: Rankings

A_{ij} : item i at positi

	pos 1	pos 2	po	
item 1	A_{11}	A_{12}	A_{13}	A_{14}
item 2	A_{21}	A_{22}	A_{23}	A_{24}
item 3	A_{31}	A_{32}	A_{33}	A_{34}
item 4	A_{41}	A_{42}	A_{43}	A_{44}

constraint: each item i assigned to a unique position (n constraints)

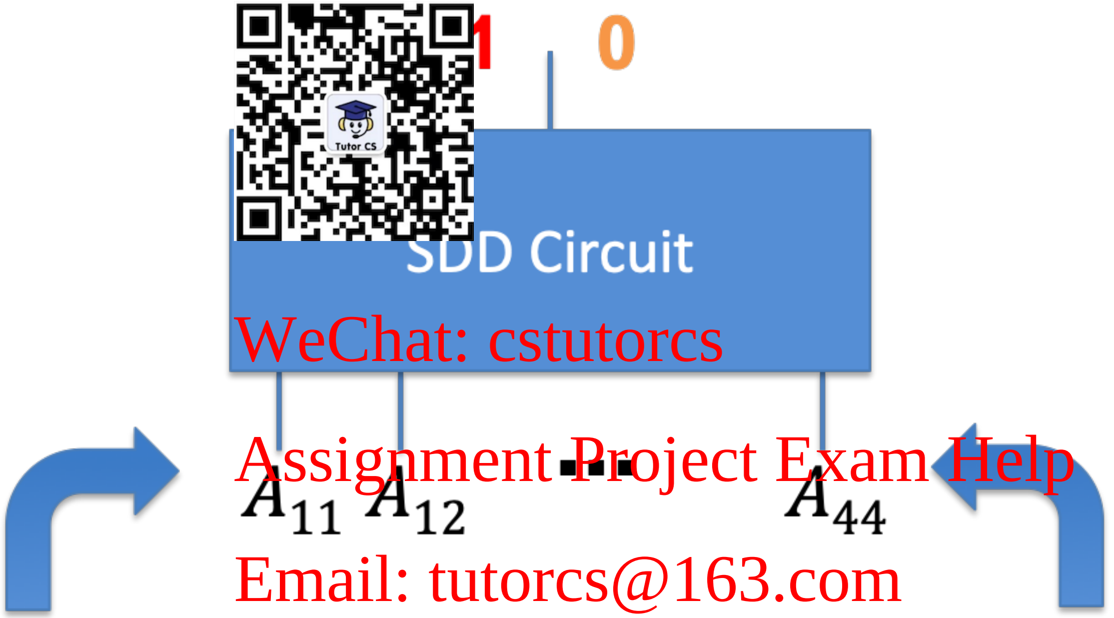
$$\bigvee_j A_{ij} \wedge \left(\bigwedge_{k \neq j} \neg A_{ik} \right)$$

constraint: each position j assigned a unique item (n constraints)

$$\bigvee_i A_{ij} \wedge \left(\bigwedge_{k \neq i} \neg A_{kj} \right)$$

total constraints	$2n$
unstructured space	2^{n^2}
structured space	$n!$

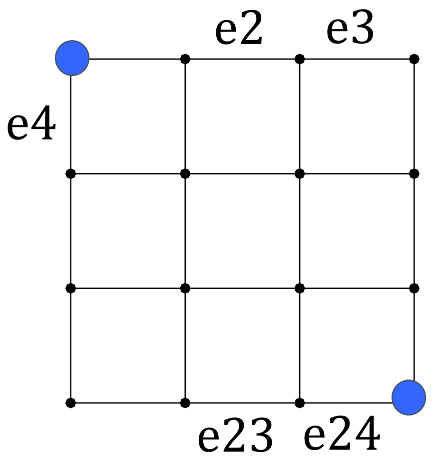
Structured Space: Rankings

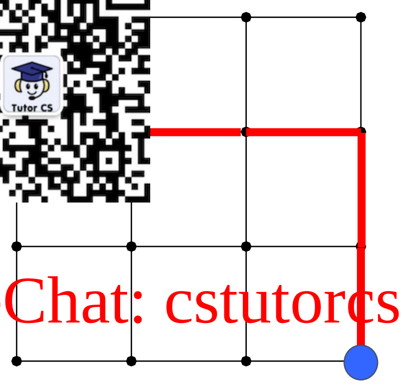


	pos 1	pos 2	pos 3	pos 4
item 1	A_{11}	A_{12}	A_{13}	A_{14}
item 2	A_{21}	A_{22}	A_{23}	A_{24}
item 3	A_{31}	A_{32}	A_{33}	A_{34}
item 4	A_{41}	A_{42}	A_{43}	A_{44}

	pos 1	pos 2	pos 3	pos 4
item 1	A_{11}	A_{12}	A_{13}	A_{14}
item 2	A_{21}	A_{22}	A_{23}	A_{24}
item 3	A_{31}	A_{32}	A_{33}	A_{34}
item 4	A_{41}	A_{42}	A_{43}	A_{44}

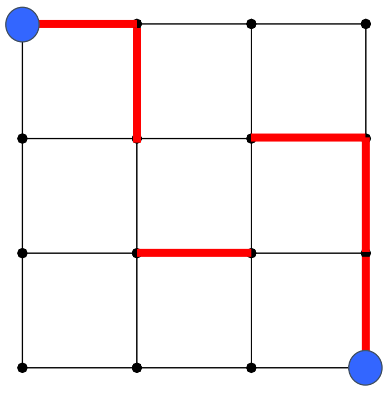
Structured Space: Routes





Good variable assignment
(represents route)

184



Bad variable assignment
(does not represent route)

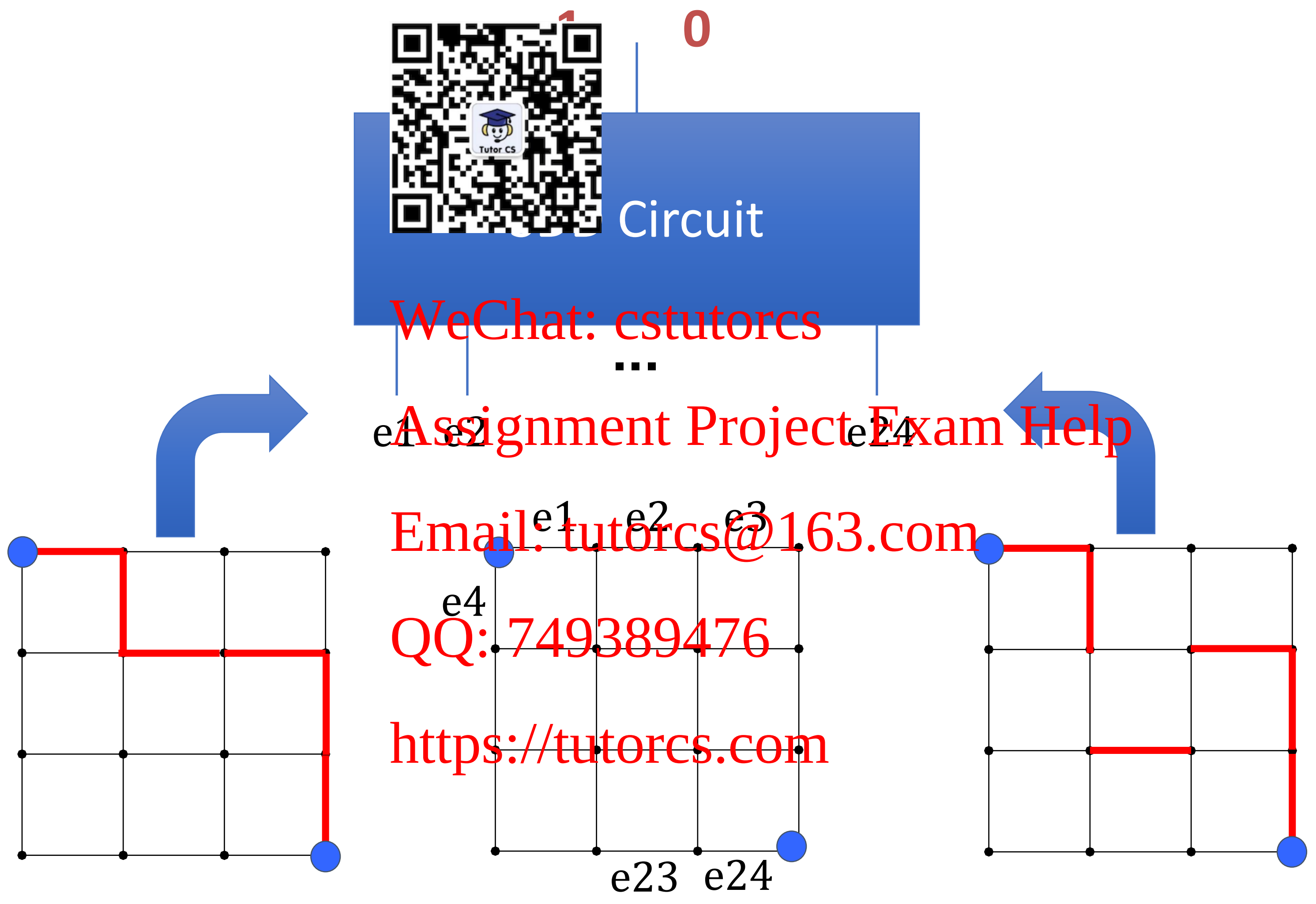
16,777,032

Space can be encoded using logical constraints

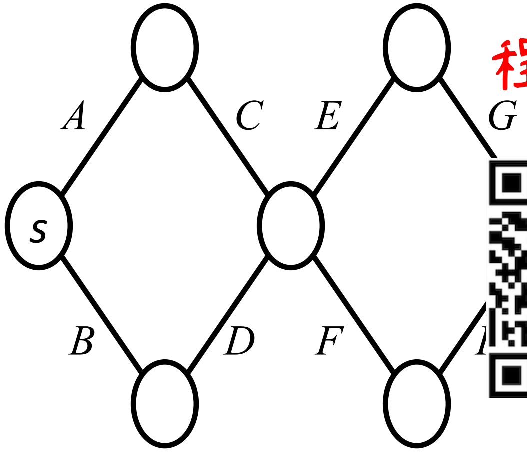
Unstructured probability space: 184+16,777,032 = 2^{24}

Constraints may include other desires (e.g., landmarks)

Structured Space: Routes



Map: nodes are cities, edges are streets

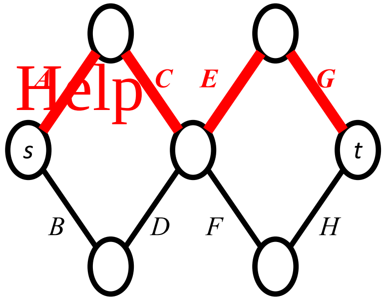


example routes from *s* to *t*:
A-C-E-G
B-D-E-G

A street is represented by a Boolean variable X
X=1 means street X on route
X=0 means street X is not on route

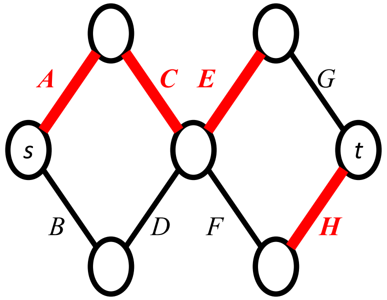
A route is a variable instantiation

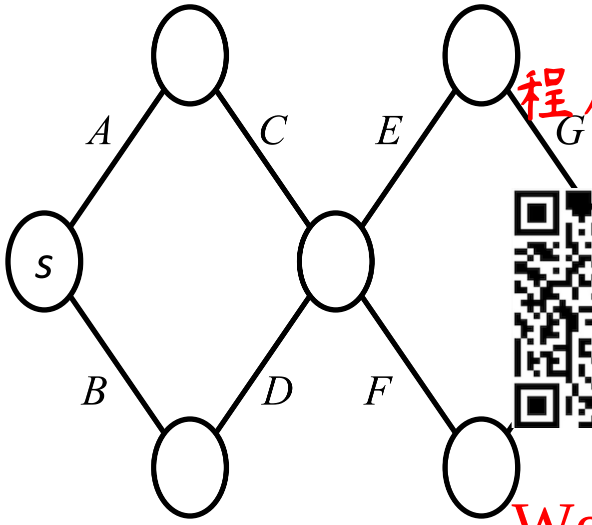
instantiation: A=1, B=0, C=1, D=0, E=1, F=0, G=1, H=0
represents route: A-C-E-G



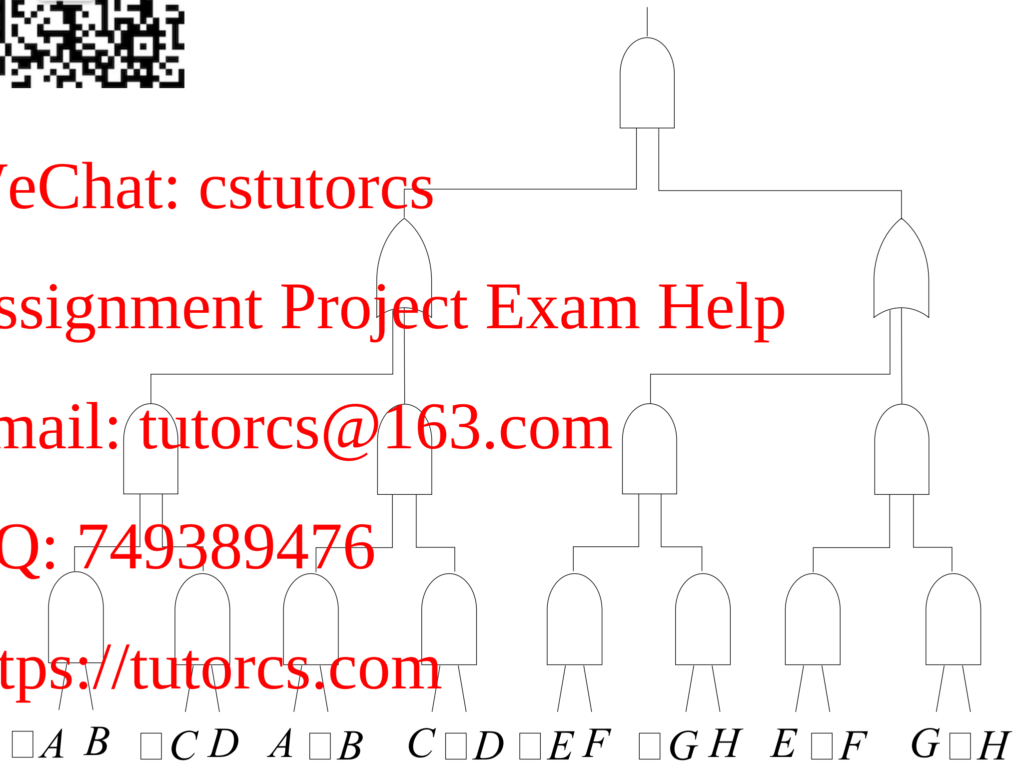
Not every variable instantiation is a route

instantiation: A=1, B=0, C=1, D=0, E=1, F=0, <u>G=0, H=1</u>
does not represent a route (disconnected)

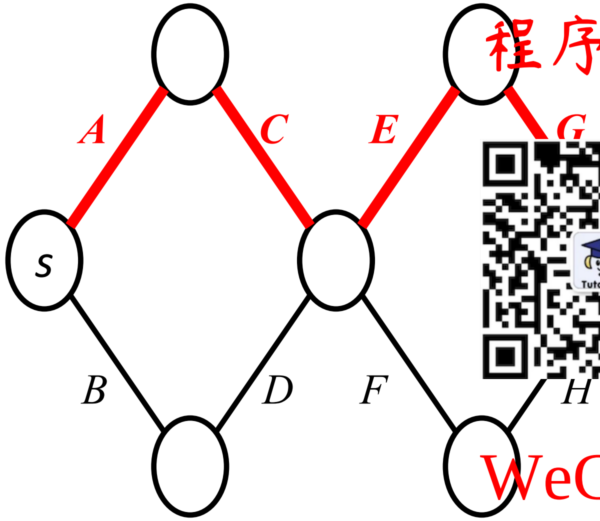




compile map into Boolean circuit

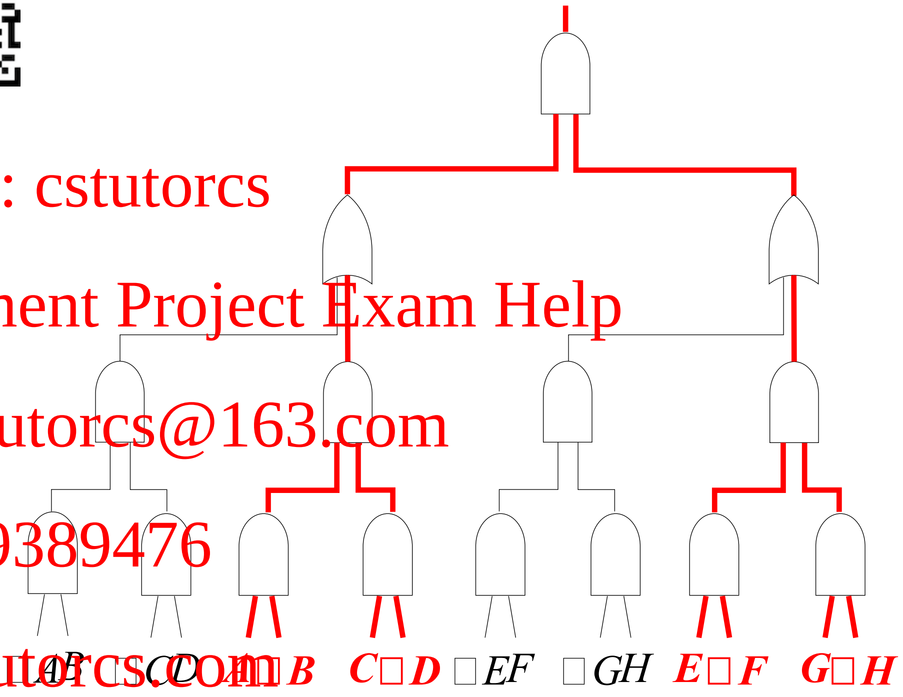


circuit representing routes on a map



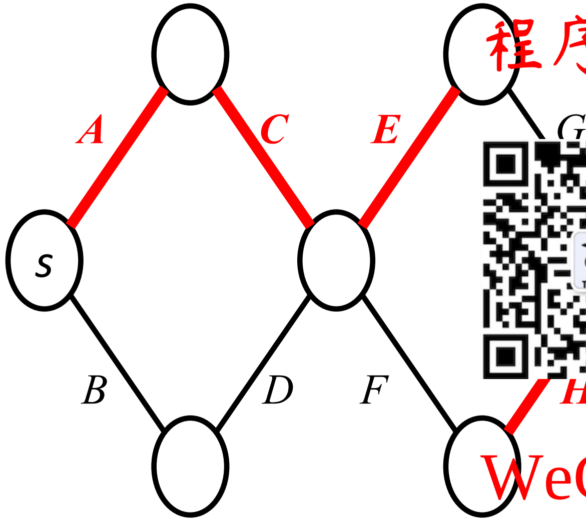
route

Evaluates to 1



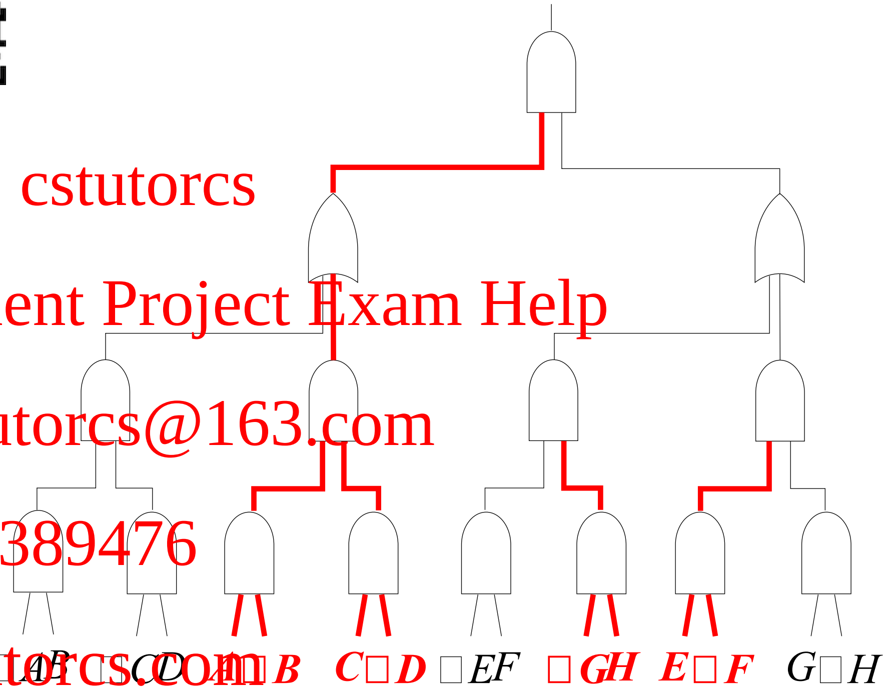
instantiation:
A=1, B=0, C=1, D=0, E=1, F=0, G=1, H=0

circuit representing routes on a map



not a route

Evaluates to 0

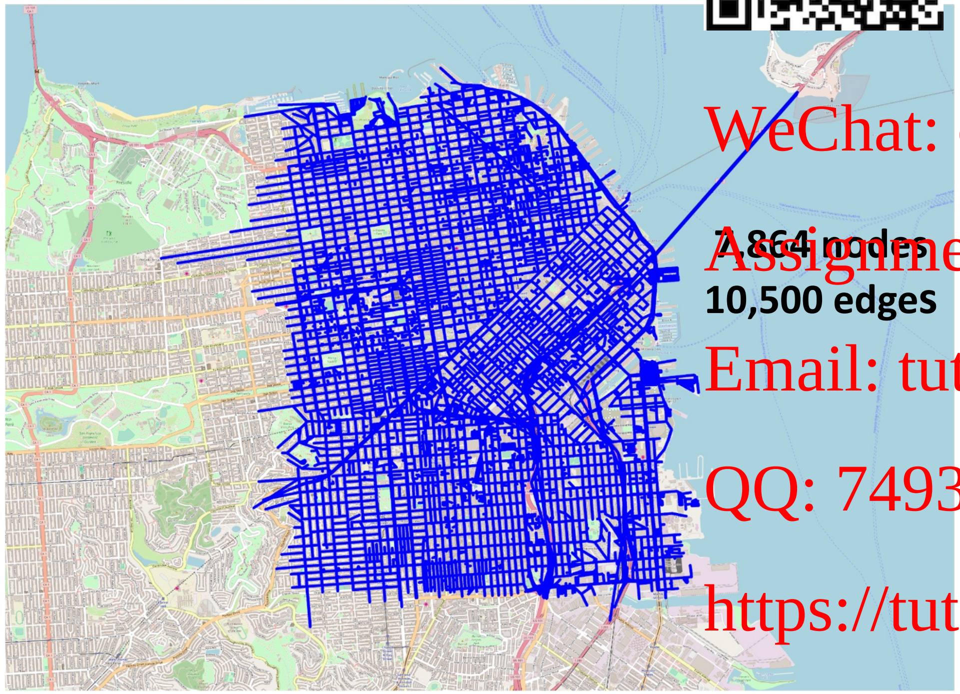


instantiation:
A=1, B=0, C=1, D=0, E=1, F=0, G=0, H=1

circuit representing routes on a map

Learning from Data (GPS) & Knowledge (Map)

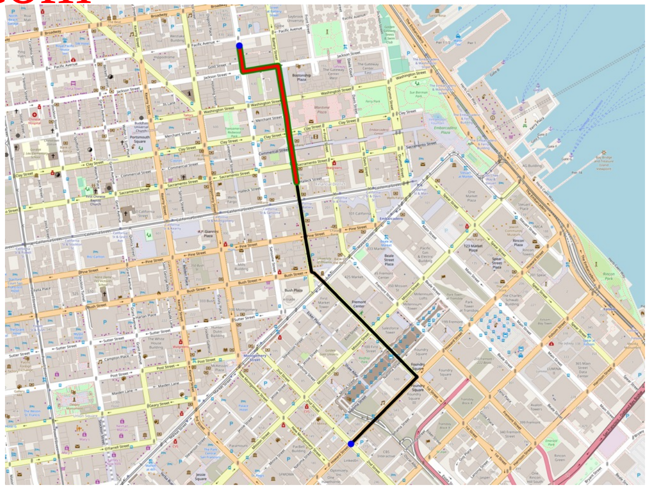
downtown SF



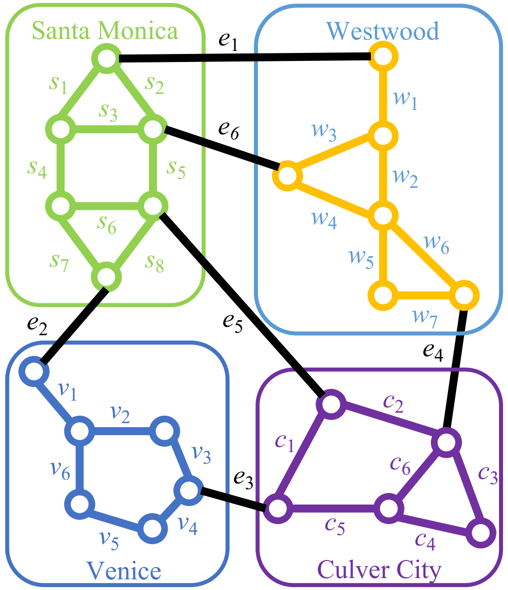
SDD had few million nodes

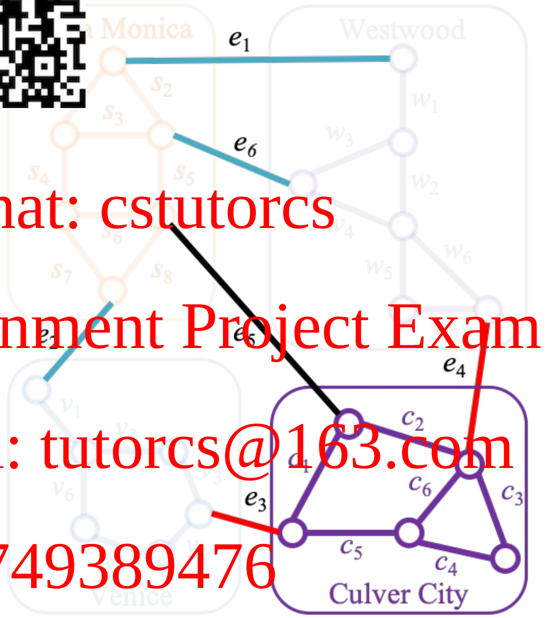
identifying drivers from routes

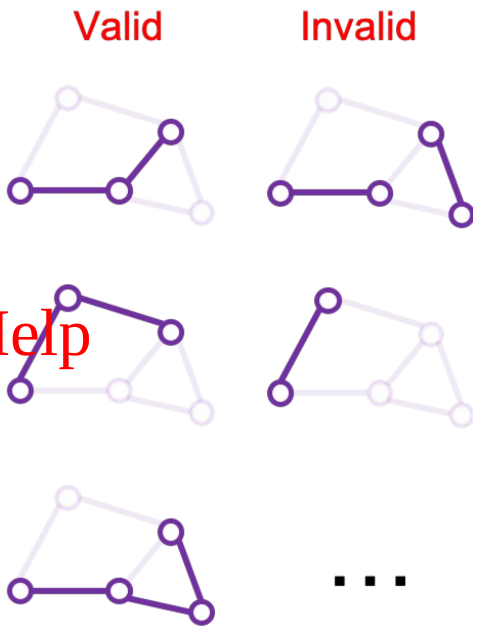
predicting route completion



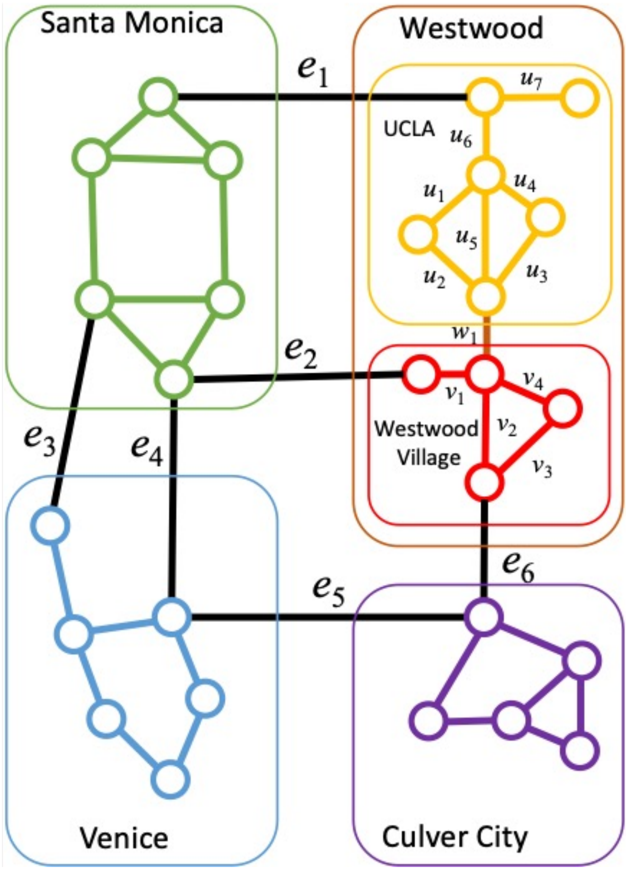
Conditional Structured Spaces



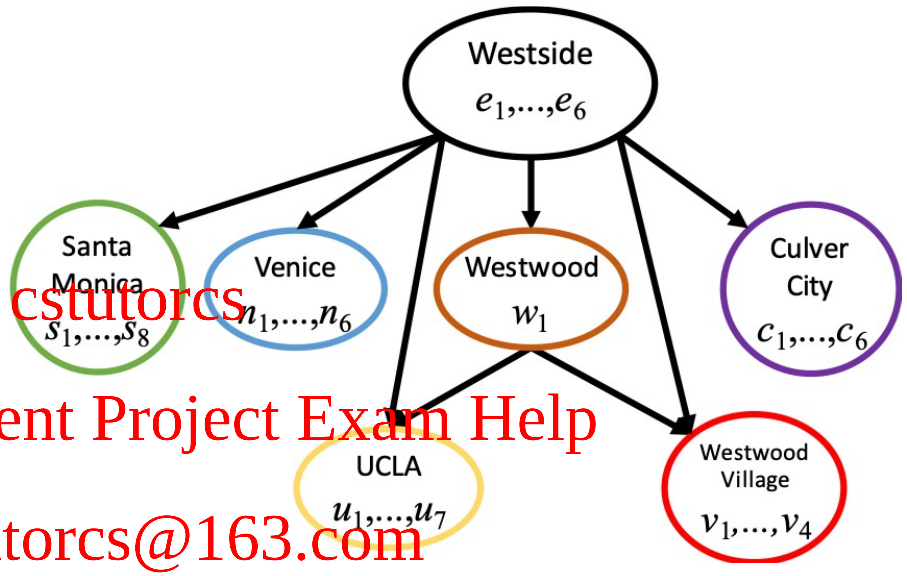




hierarchical map



Structured Bayesian Network



Structure of Conditional PSDDs compiled from knowledge (maps)
Parameters of Conditional PSDDs learned from GPS data
Conditional PSDDs multiplied into a monolithic PSDD used for inference

Probabilistic Sentential Decision Diagrams (PSDDs)

1. Syntax
2. Semantics
3. Inference
4. Learning

January 12, 2024 | 5PM CET | ONLINE

WORLD LOGIC DAY 2024

Vienna Center for Logic and Algorithms

TU WIEN Informatics

Logic is n… culus of truth and falsehoo… d foremost a calculus of events th… ue or false; probable or improbable; postulated, proven or learned; obvious or need to be explained.

Beyond tru…
Logic as a …

Adnan Darwiche

UCLA

14 JAN…

WORLD LOGIC …

Date: January 12, 2024

Time: 8am PST | 11am EST | 1pm GMT-3 | **5pm CET (Vienna)**

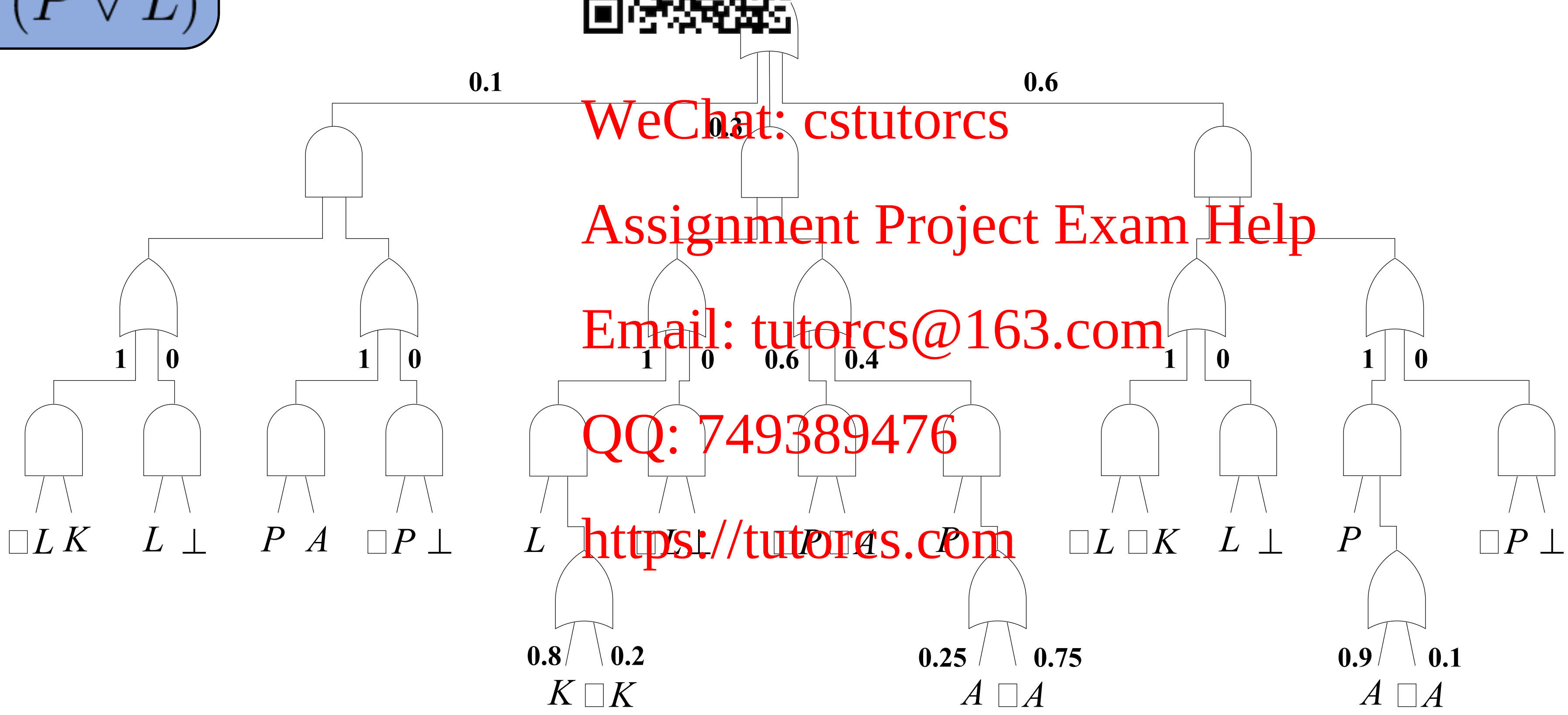
This event will take place **online via Zoom** with a **livestream on the VCLA youtube channel** @viennacenterforlogicandalg5604.

Probabilistic Sentential Decision Diagrams (PSDDs)

1. Syntax
2. Semantics
3. Inference
4. Learning

PSDD: Syntax

$$P \vee L$$
$$A \Rightarrow P$$
$$K \Rightarrow (P \vee L)$$



L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

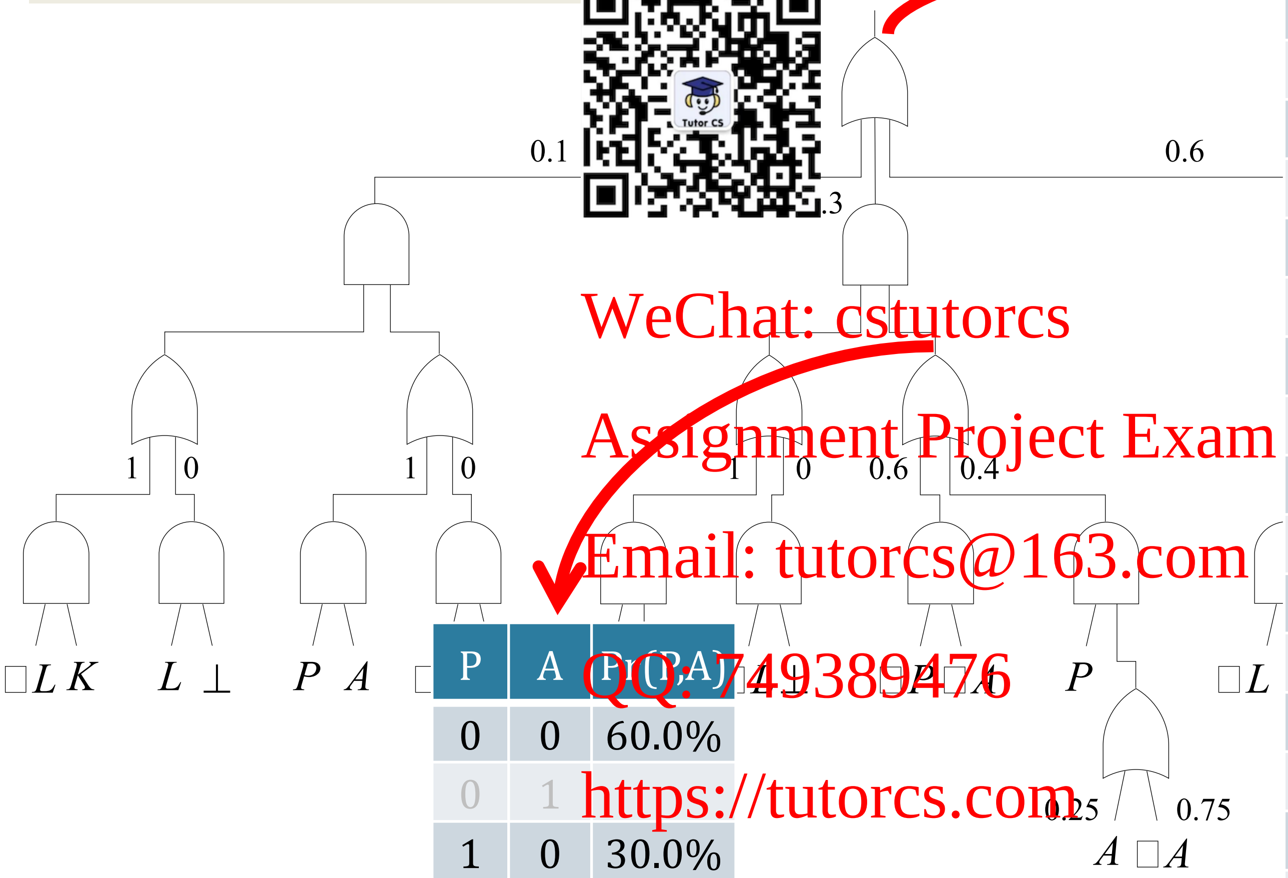
PSDD: Semantics

L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

Input: L, K, P, A

$\Pr(L,K,P,A) = 0.3 \times 1.0 \times 0.8 \times 0.4 \times 0.25 = 0.024$

induces a normalized distribution over satisfying assignments (structured probability space)



P	A	Pr(P,A)
0	0	60.0%
0	1	
1	0	30.0%
1	1	10.0%

L	K	P	A	P(L,K,P,A)
0	0	0	0	
0	0	0	1	
0	0	1	0	6.0%
0	0	1	1	54.0%
0	1	0	0	
0	1	0	1	
0	1	1	0	
0	1	1	1	10.0%
1	0	0	0	3.6%
1	0	0	1	
1	0	1	0	1.8%
1	0	1	1	0.6%
1	1	0	0	14.5%
1	1	0	1	
1	1	1	0	7.2%
1	1	1	1	2.4%

Interpreting PSDD Parameters

$Pr(p_1) = 0.1$
$Pr(p_2) = 0.3$
$Pr(p_3) = 0.6$

$Pr(q_1 \mid p_2) = 0.6$
$Pr(q_2 \mid p_2) = 0.4$

Computing Marginal Probabilities

Probability Distributions

world	E Earthquake	B Burglary	[illegible]	Pr(.)
ω_1	true	true	[illegible]	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

- marginal probability

$$Pr(A) = Pr(w_1) + Pr(w_3) + Pr(w_5) + Pr(w_7)$$

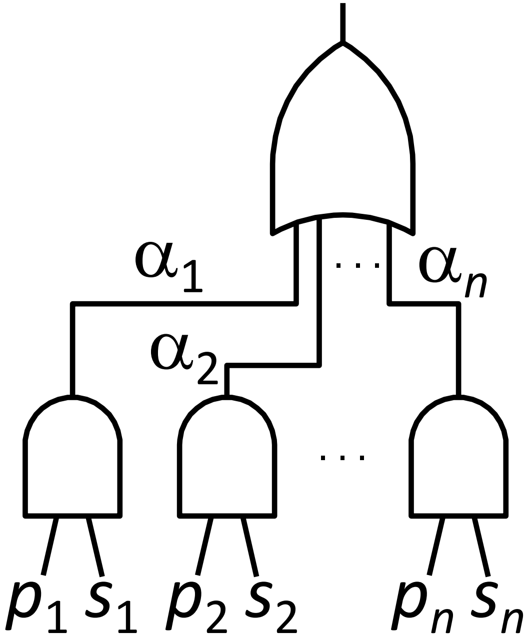
$$Pr(\neg A) = Pr(w_2) + Pr(w_4) + Pr(w_6) + Pr(w_8)$$

$$Pr(A, E) = Pr(w_1) + Pr(w_3)$$

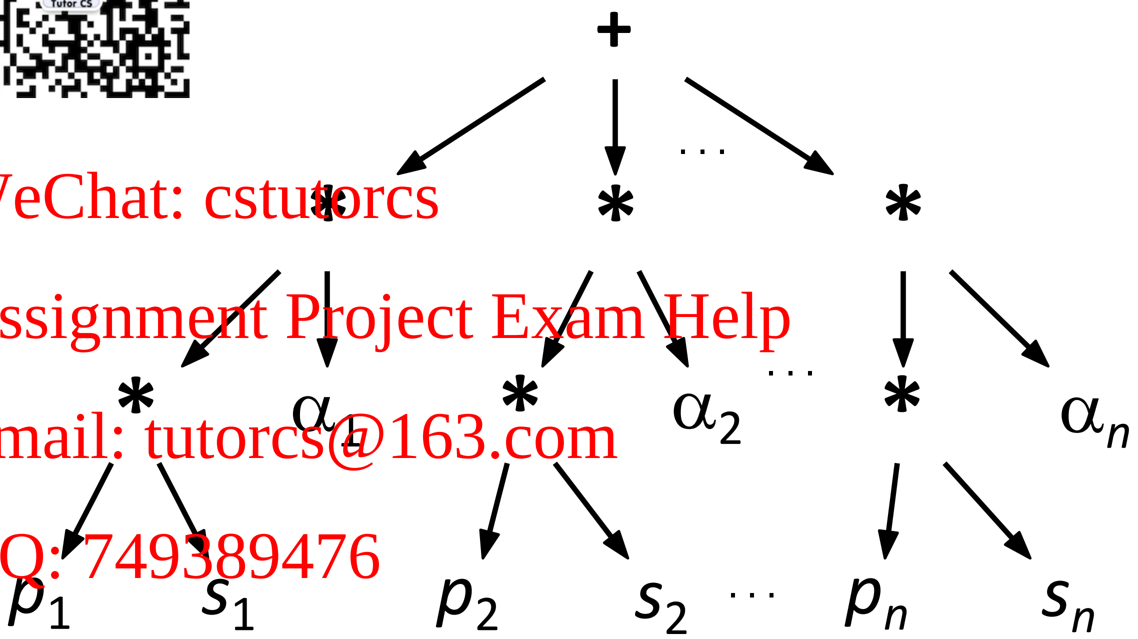
- conditional probability

$$Pr(\alpha \mid \beta) = \frac{Pr(\alpha, \beta)}{Pr(\beta)}$$

PSDDs are ACs



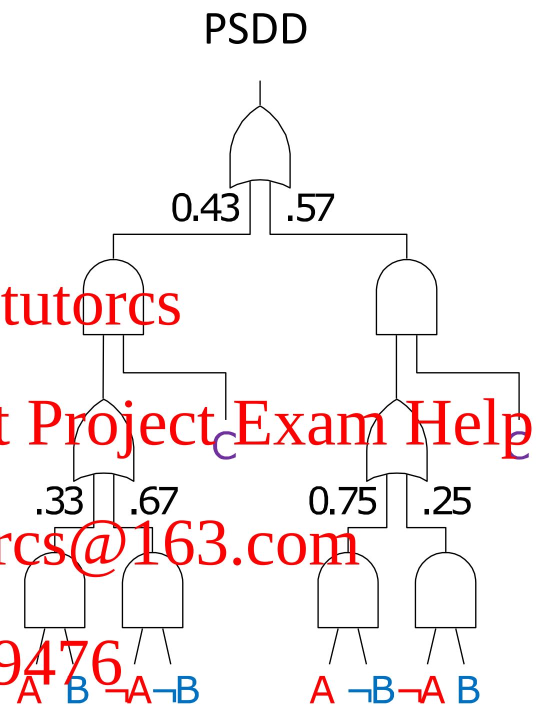
PSDD



AC

Marginal Probabilities

A	B	C	
T	T	T	1/7
T	T	F	0
T	F	T	3/7
T	F	F	0
F	T	T	1/7
F	T	F	0
F	F	T	2/7
F	F	F	0



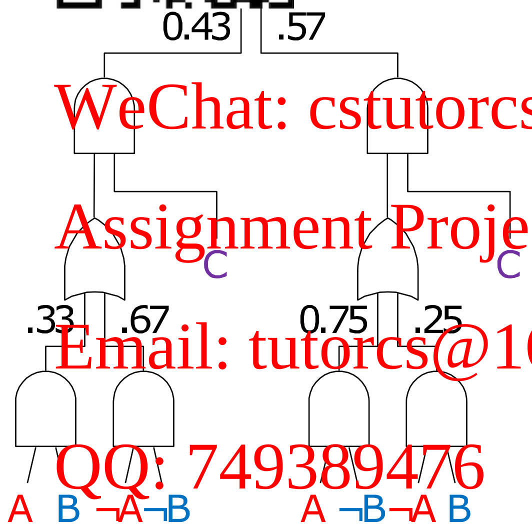
values of leaves (literals) given evidence e

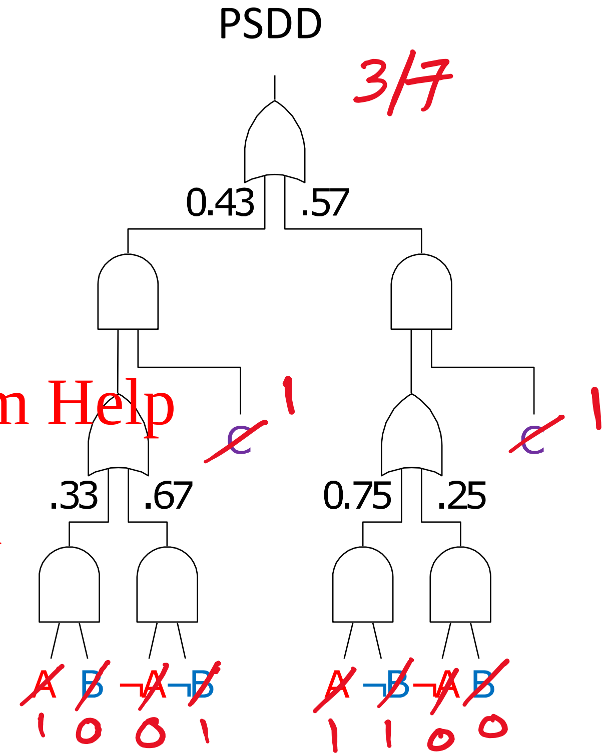
Value of literal ℓ :

$$\begin{cases} 0 & \text{if } \ell \text{ inconsistent with evidence } e \\ 1 & \text{otherwise} \end{cases}$$

Marginal Probabilities

A	B	C	
T	T	T	1/7
T	T	F	0
T	F	T	3/7
T	F	F	0
F	T	T	1/7
F	T	F	0
F	F	T	2/7
F	F	F	0





evidence

$A = T,\ B = F,\ C = T$

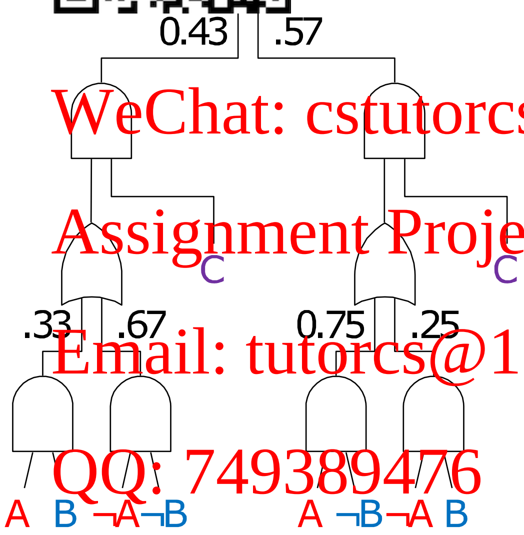
$A \leftarrow 1$ $\quad B \leftarrow 0$ $\quad C \leftarrow 1$

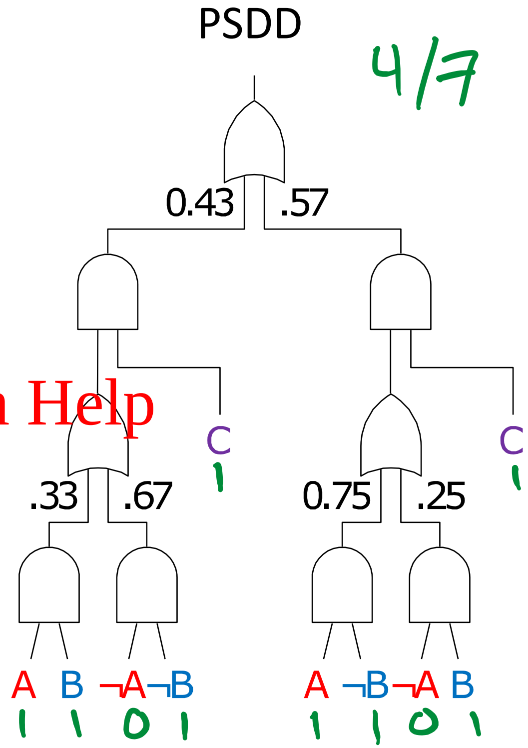
$\neg A \leftarrow 0$ $\quad \neg B \leftarrow 1$ $\quad \neg C \leftarrow 0$

Marginal Probabilities

PSDDs are tractable for computing marginals (feed-forward)

A	B	C	
T	T	T	1/7
T	T	F	0
T	F	T	3/7
T	F	F	0
F	T	T	1/7
F	T	F	0
F	F	T	2/7
F	F	F	0





evidence

A = T

$A \leftarrow 1$ $\quad$ $B \leftarrow 1$ $\quad$ $C \leftarrow 1$

$\neg A \leftarrow 0$ $\quad$ $\neg B \leftarrow 1$ $\quad$ $\neg C \leftarrow 1$

Learning PSDD Parameters

supervised vs unsupervised learning

Learning with Background Knowledge

Background Knowledge

Logic (L)
Knowledge Representation (K)
Probability (P)
Artificial Intelligence (A)

Must take at least one of Probability or Logic.
Probability is a prerequisite for AI.
The prerequisites for KR is either AI or Logic.

$P \vee L \quad A \Rightarrow P \quad K \Rightarrow (A \vee L)$

Data

L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3

complete data

Maximum Likelihood Parameters

data

L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3

- examples

  e_1
  e_2
  $\vdots$
  e_n

- m[illegible] $Pr_\theta(\cdot)$ (PSDD with parameters θ)
- Likelihood of parameters θ:

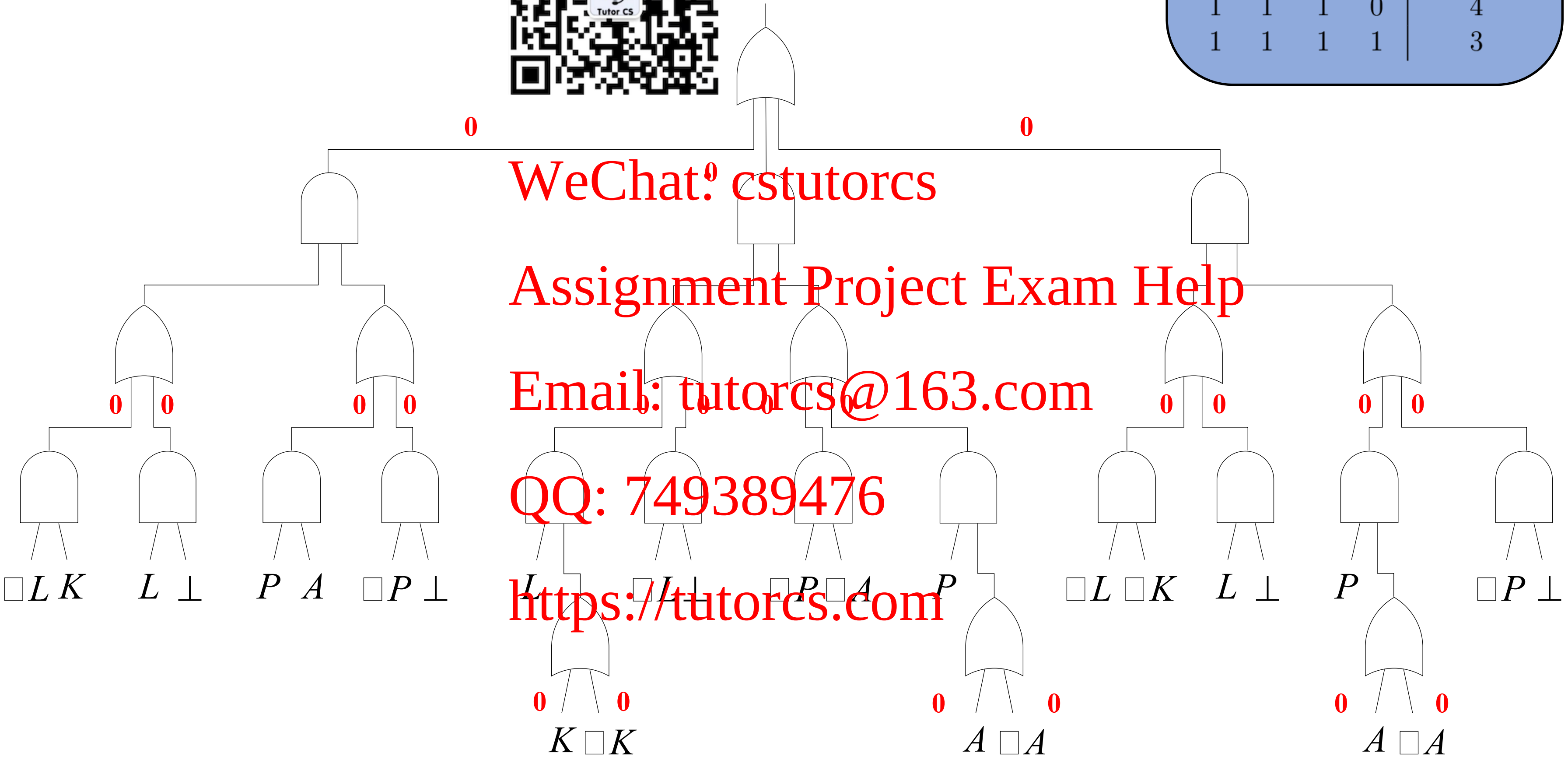
  $$L(Data \mid \theta) = Pr_\theta(e_1) \times \cdots \times Pr_\theta(e_n)$$

- Learning: find the maximum likelihood parameters
- Difficulty & properties depend on type of data (complete, incomplete)

Learning Parameters

data

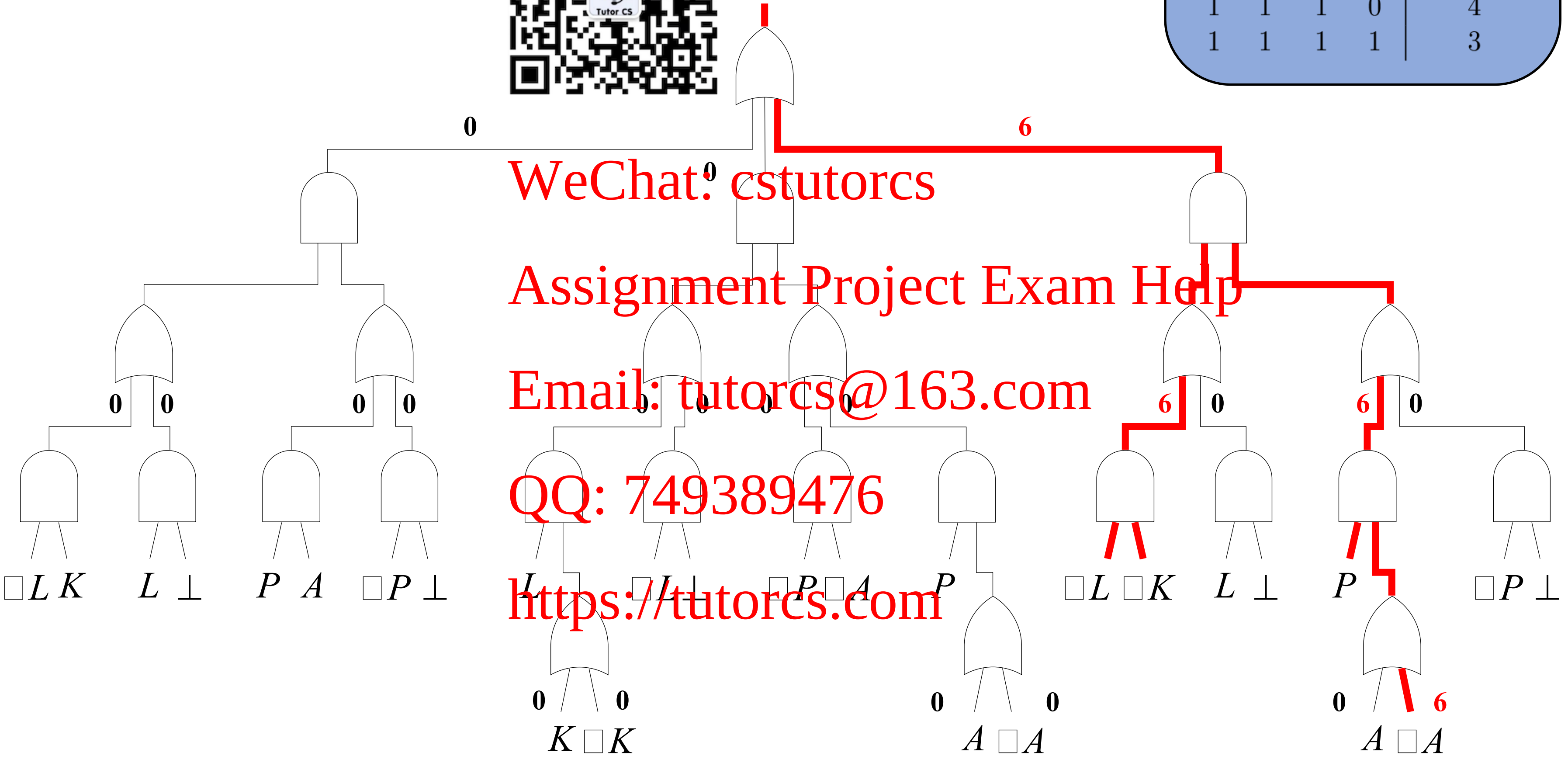
L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3



Learning Parameters

data

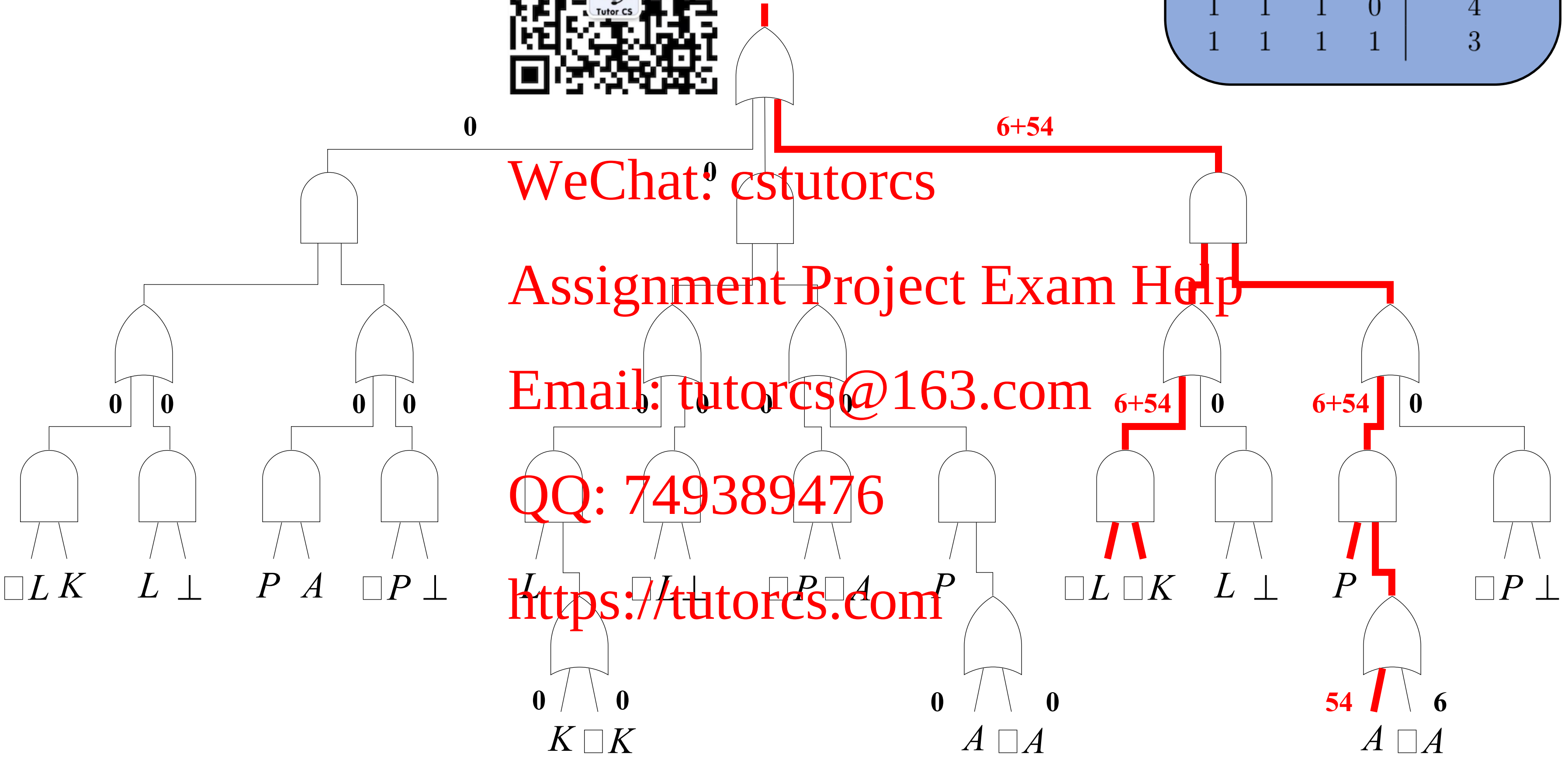
L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3



Learning Parameters

data

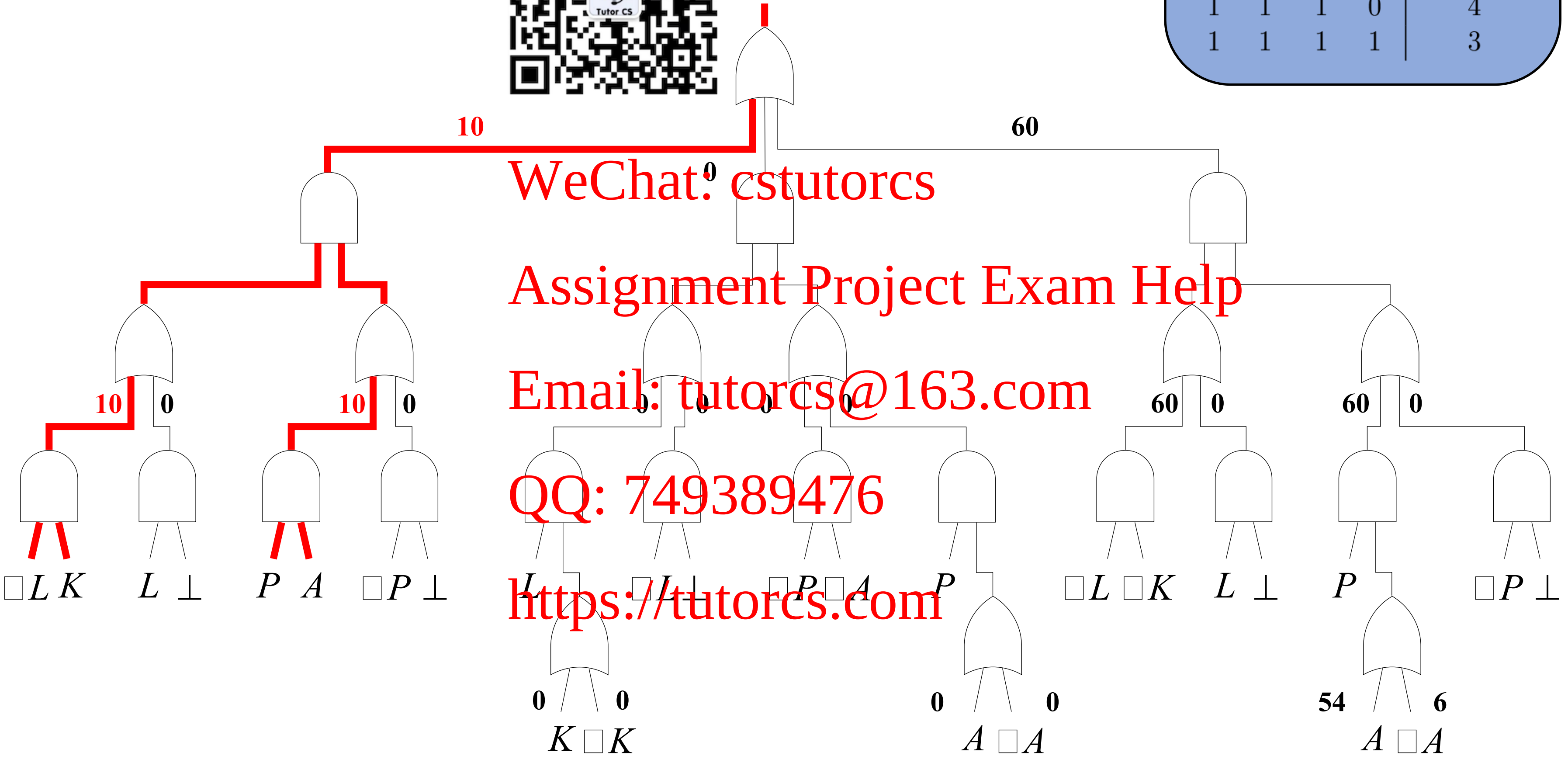
L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3



Learning Parameters

data

L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3



Types of Datasets

classical
complete dataset

id	X	Y	Z
1	x_1	y_2	z_1
2	x_2	y_1	z_2
3	x_2	y_1	z_2
4	x_1	y_1	z_1
5	x_1	y_2	z_2

closed-form
(maximum-likelihood
estimates are unique)

classical
incomplete dataset

id	X	Y	Z
1	x_1	y_2	?
2	x_2	y_1	?
3	?	?	z_2
4	?	y_1	z_1
5	x_1	y_2	z_2

EM algorithm
(on PSDDs)

non-classical
incomplete dataset

id	X	Y	Z
1	$X \equiv Z$		
2	x_2 and (y_2 or z_2)		
3	$x_2 \Rightarrow y_1$		
4	$X \oplus Y \oplus Z \equiv 1$		
5	x_1 and y_2 and z_2		

Missed in the
ML literature

Types of Datasets

classical **complete**
(e.g., total ranki

id	1^{st} sushi	2^{nd} sushi	3^{rd} sushi	□
1	fatty tuna	sea urchin	salmon roe	□
2	fatty tuna	tuna	shrimp	□
3	tuna	tuna roll	sea eel	□
4	fatty tuna	salmon roe	tuna	□
5	egg	squid	shrimp	□

classical **incomplete** dataset
(e.g., top-*k* rankings)

id	1^{st} sushi	2^{nd} sushi	3^{rd} sushi	□
1	fatty tuna	sea urchin	?	□
2	fatty tuna	?	?	□
3	tuna	tuna roll	?	□
4	fatty tuna	salmon roe	?	□
5	egg	?	?	□

Types of Datasets

classical **complete**
(e.g., total ranki

id	1^{st} sushi	2^{nd} sushi	3^{rd} sushi	□
1	fatty tuna	sea urchin	salmon roe	□
2	fatty tuna	tuna	shrimp	□
3	tuna	tuna roll	sea eel	□
4	fatty tuna	salmon roe	tuna	□
5	egg	squid	shrimp	□

non-classical **incomplete** dataset
(e.g., **partial** rankings)

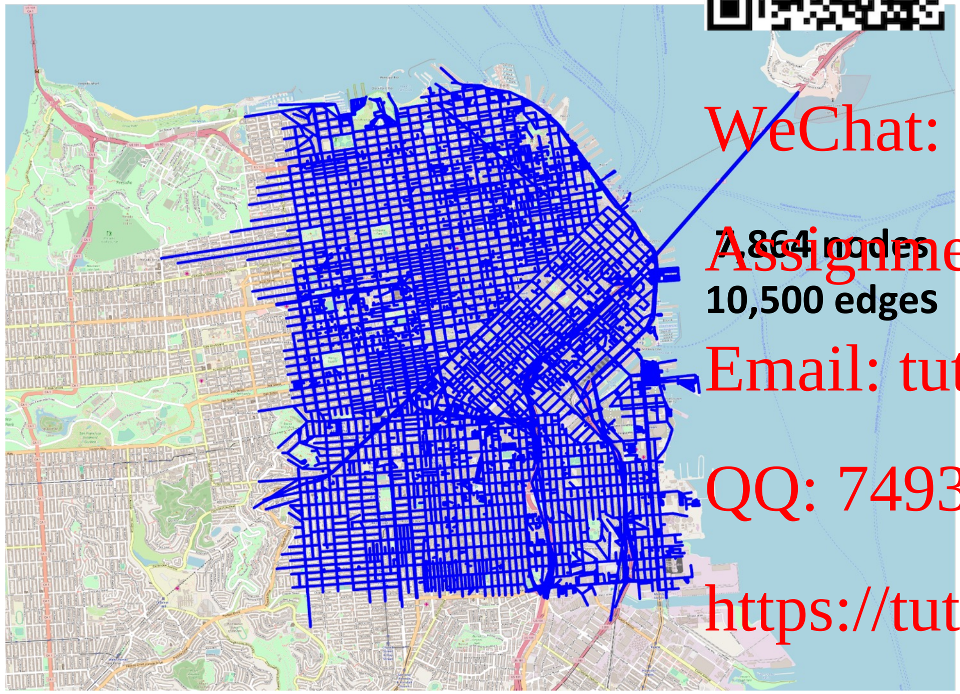
id	1^{st} sushi	2^{nd} sushi	3^{rd} sushi	□
1	(fatty tuna > sea urchin) and (tuna > sea eel)			□
2	(fatty tuna is 1^{st}) and (salmon roe > egg)			□
3	tuna > squid			□
4	egg is last			□
5	egg > squid > shrimp			□

(represents constraints on possible *total rankings*)

Learning from Data (GPS) & Knowledge (Map)

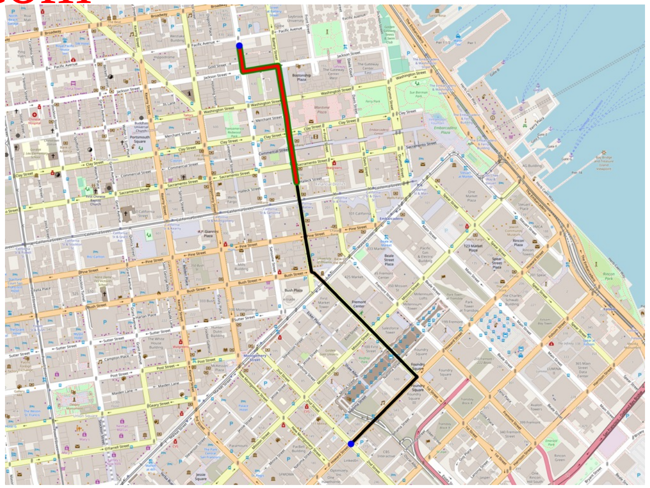
identifying drivers from routes

downtown SF



7,864 nodes
10,500 edges

predicting route completion



PSDD had few million nodes

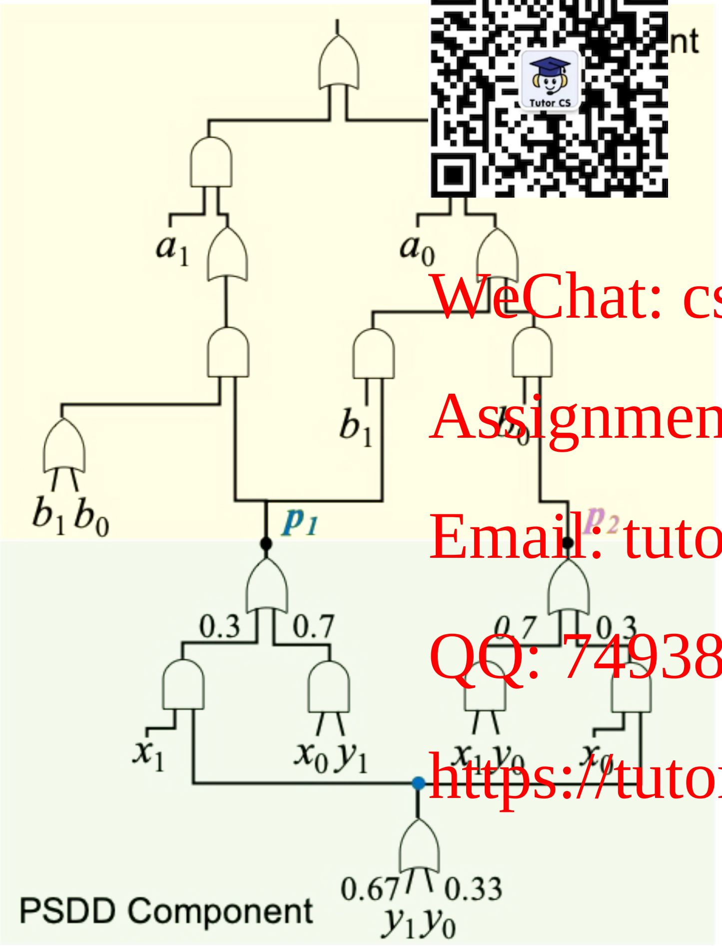
Conditional PSDDs

(conditional vtrees)

Conditional Distributions

[illegible]	[illegible]	[illegible]	Y	$Pr(X, Y \mid A, B)$
[illegible]	[illegible]	[illegible]	y_0	0.1
[illegible]	[illegible]	[illegible]	y_1	0.2
a_0	b_0	x_1	y_0	0.7
a_0	b_0	x_1	y_1	**0.0**
a_0	b_1	x_0	y_0	**0.0**
a_0	b_1	x_0	y_1	0.7
a_0	b_1	x_1	y_0	0.1
a_0	b_1	x_1	y_1	0.2
a_1	b_0	x_0	y_0	**0.0**
a_1	b_0	x_0	y_1	0.7
a_1	b_0	x_1	y_0	0.1
a_1	b_0	x_1	y_1	0.2
a_1	b_1	x_0	y_0	**0.0**
a_1	b_1	x_0	y_1	0.7
a_1	b_1	x_1	y_0	0.1
a_1	b_1	x_1	y_1	0.2

Conditional PSDDs

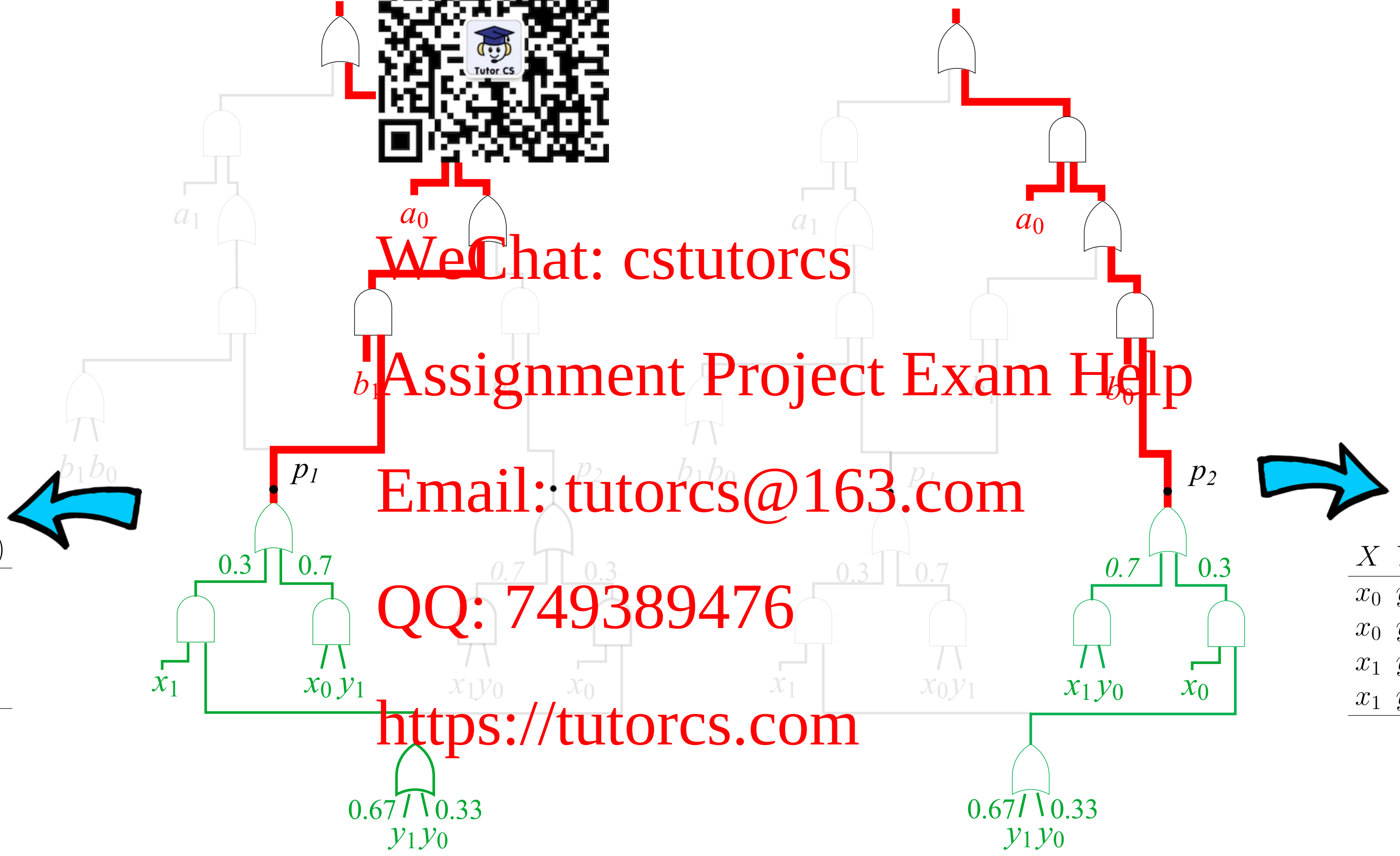


A	B	X	Y	$Pr(X, Y \mid A, B)$
a_0	b_0	x_0	y_0	0.1
a_0	b_0	x_0	y_1	0.2
a_0	b_0	x_1	y_0	0.7
a_0	b_0	x_1	y_1	**0.0**
a_0	b_1	x_0	y_0	**0.0**
a_0	b_1	x_0	y_1	0.7
a_0	b_1	x_1	y_0	0.1
a_0	b_1	x_1	y_1	0.2
a_1	b_0	x_0	y_0	**0.0**
a_1	b_0	x_0	y_1	0.7
a_1	b_0	x_1	y_0	0.1
a_1	b_0	x_1	y_1	0.2
a_1	b_1	x_0	y_0	**0.0**
a_1	b_1	x_0	y_1	0.7
a_1	b_1	x_1	y_0	0.1
a_1	b_1	x_1	y_1	0.2

Conditional PSDDs

$a_0\ b_1$

X	Y	$Pr(X, Y \mid a_0, b_1)$
x_0	y_0	**0.0**
x_0	y_1	0.7
x_1	y_0	0.1
x_1	y_1	0.2

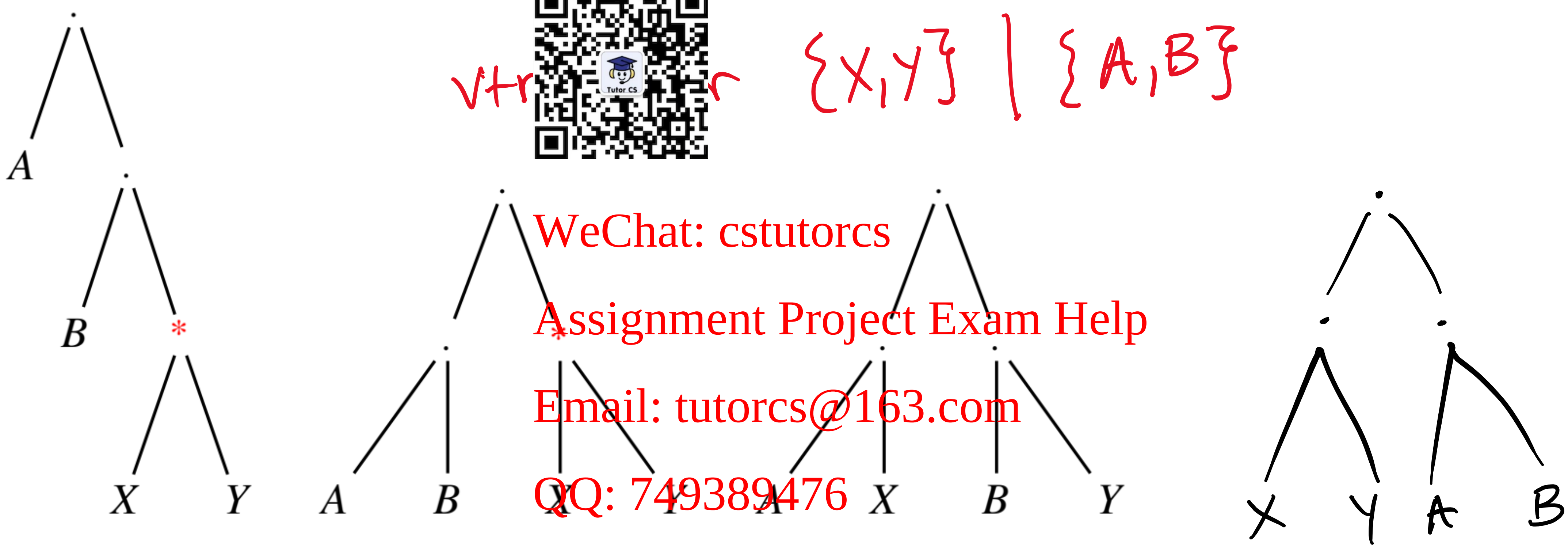


$a_0\ b_0$

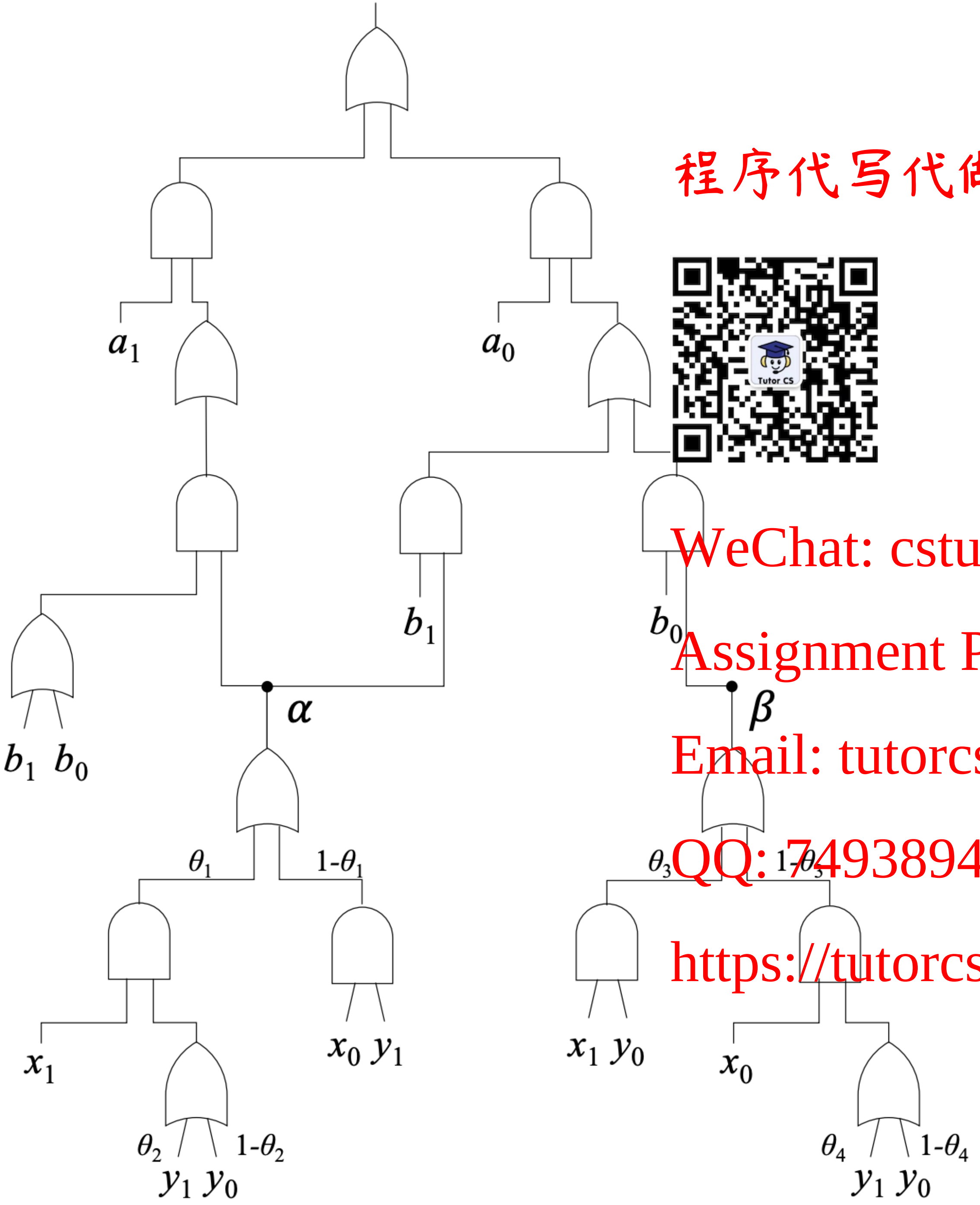
X	Y	$Pr(X, Y \mid a_0, b_0)$
x_0	y_0	0.1
x_0	y_1	0.2
x_1	y_0	0.7
x_1	y_1	**0.0**

Conditional Vtrees

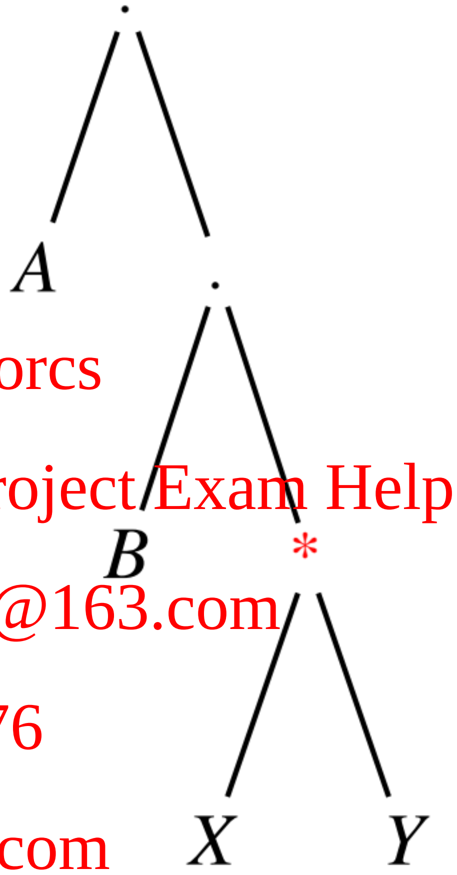
vtree for $\{X, Y\} \mid \{A, B\}$



Definition 1 (Conditional Vtrees) *Let v be a vtree for variables $\mathbf{X} \cup \mathbf{P}$ which has a node u that contains precisely the variables $\mathbf{X}$. If node u can be reached from node v by only following right children, then v is said to be a vtree for $\mathbf{X} \mid \mathbf{P}$ and u is said to be its $\mathbf{X}$-node.*



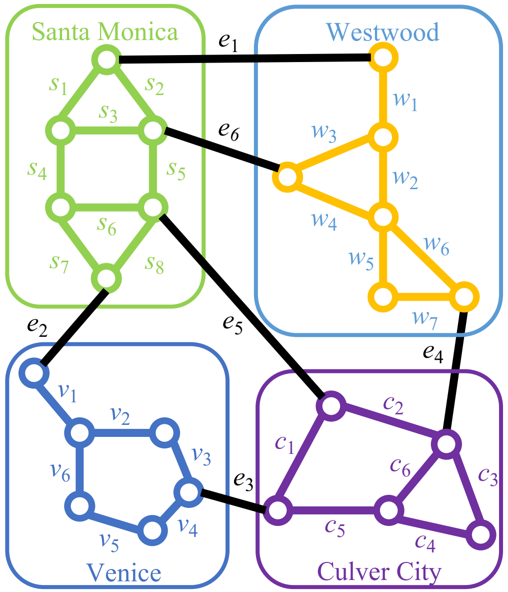
a_1
a_0
b_1
b_0
b_1 b_0
α
β
θ_1
1-θ_1
θ_3
1-θ_3
x_1
x_0 y_1
x_1 y_0
x_0
θ_2
1-θ_2
y_1 y_0
θ_4
1-θ_4
y_1 y_0

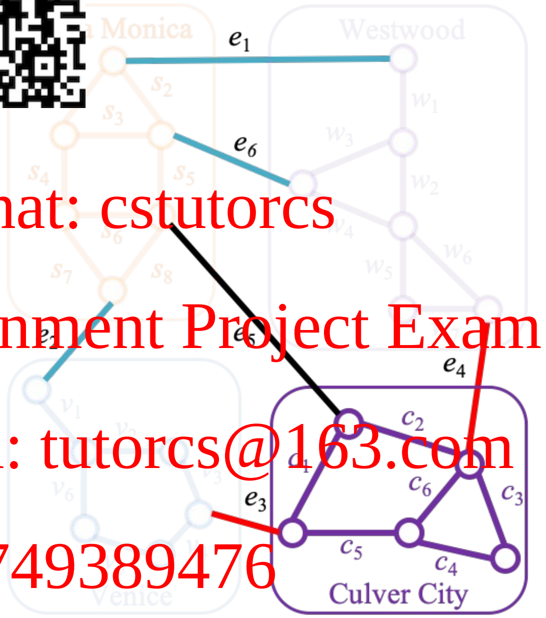


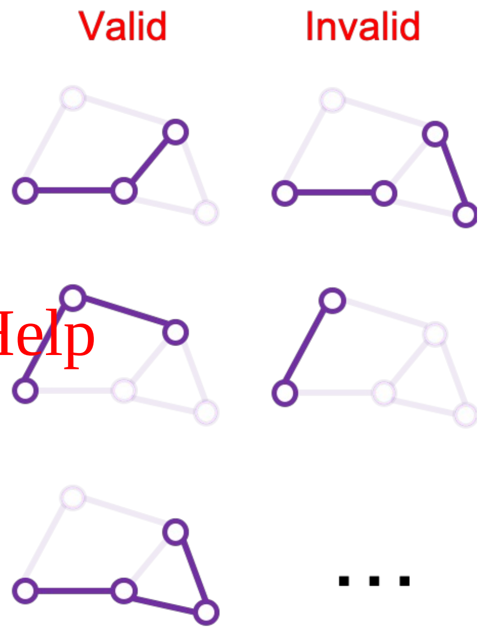
A
B
*
X
Y

$\Pr(X,Y|A,B)$

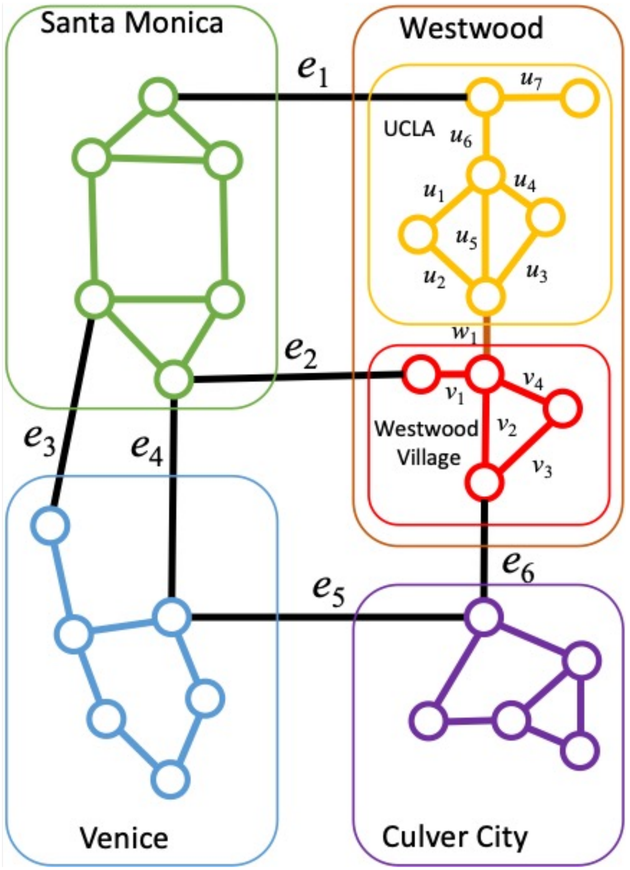
Conditional Structured Spaces



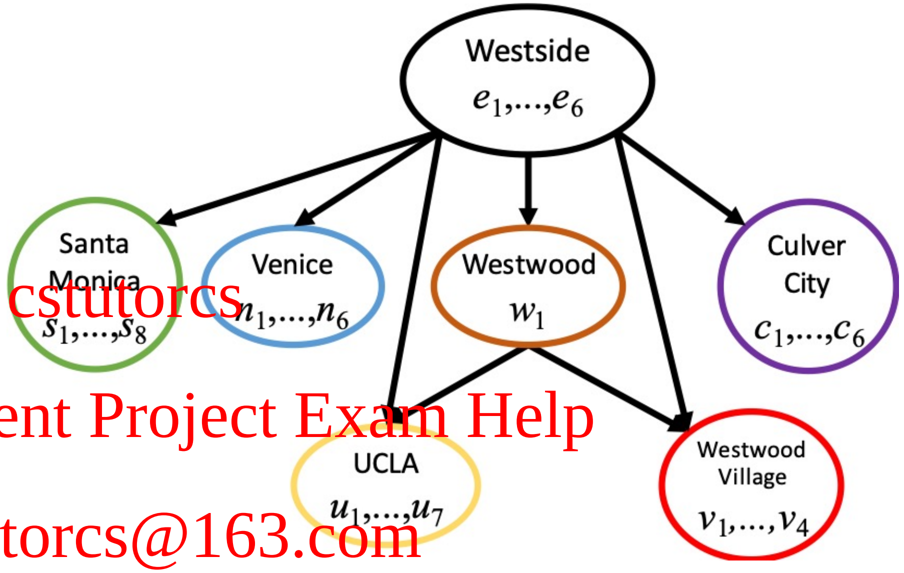




hierarchical map



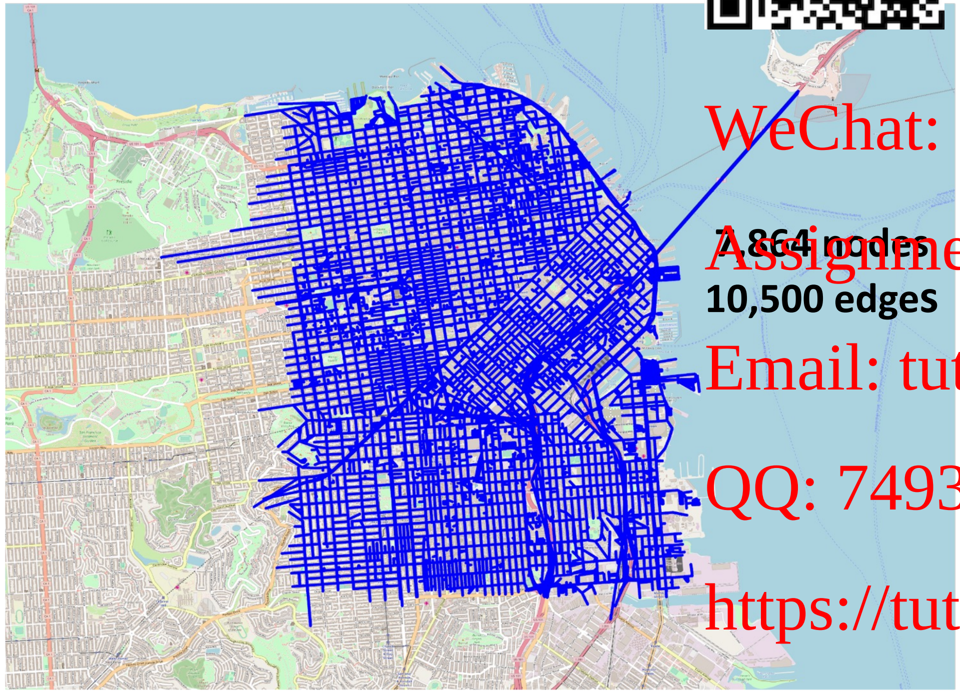
Structured Bayesian Network



Structure of Conditional PSDDs compiled from knowledge (maps)
Parameters of Conditional PSDDs learned from GPS data
Conditional PSDDs multiplied into a monolithic PSDD used for inference

Route Distributions: Maps + GPS data

downtown SF



7,864 nodes
10,500 edges

PSDD had few million nodes

identifying drivers from routes

predicting route completion