马尔可夫逻辑网络

2019年度南京大学“专创融合”特色示范课程培育项目

高 阳

http://cs.nju.edu.cn/rl, 2019.10.29

马尔科夫逻辑网络

从贝叶斯信念网到马尔科夫逻辑网

高　阳

http://cs.nju.edu.cn/gaoy, 2019.10.29

大纲

马尔科夫网

马尔科夫逻辑网

马尔科夫网络推理

近似推理

马尔科夫网

Markov链

俄罗斯数学家

考虑一个N(有限)状态$\{s_1,s_2,\ldots,s_n\}$的系统：

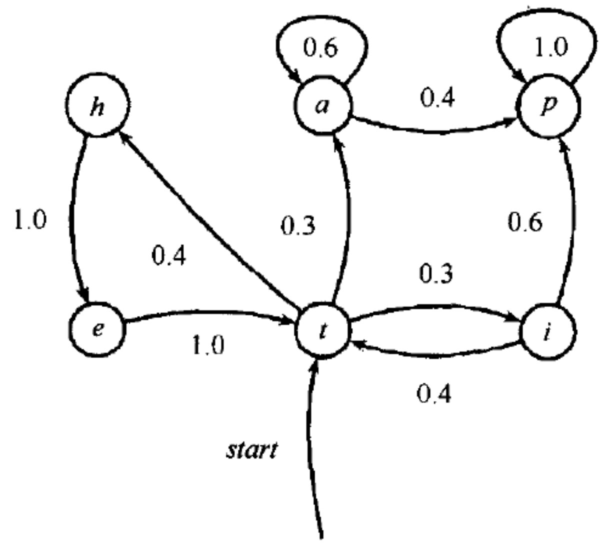
在离散的时间序列中$\{t_1,t_2,\ldots\}$，其在时间t的状态记为X_t,

系统在时间t处在状态$s_j(1\leq j \leq n)$的概率取决与其在时间1,2,…,t-1的状态：

$$\mathrm{P}(X_t = s_k|X_1,X_2,\ldots,X_{t-1})$$

一阶马尔科夫链

$$\mathrm{P}(X_t = s_k|X_1,X_2,\ldots,X_{t-1})=\mathrm{P}(X_t = s_k|X_{t-1})$$



Markov网

有向图 + 概率分布：Bayesian Network

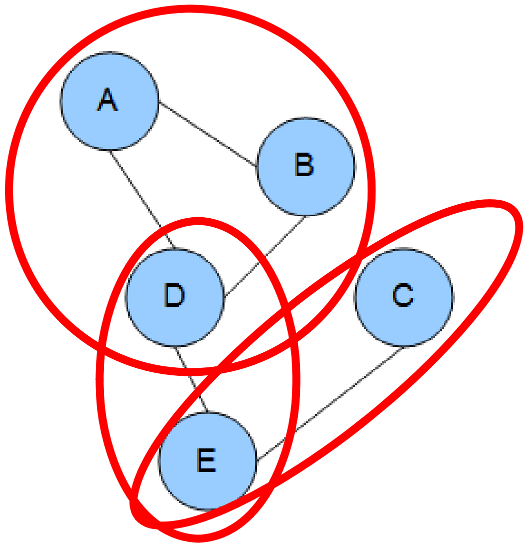
无向图 + 概率分布：Markov Network(Markov Random Field, MRF)， G=(V, E)

顶点：随机变量 V

边：变量间的依赖关系 E

团(Clique)：各边所在的最大全联通子图

势函数：为非负函数，表示团的一个状态



联合分布函数

- 吉布斯测度：已知一种分配方案x，求MLN刚好为该情况的联合分布概率

$$P(X=x)=\frac{1}{Z}\prod_k f_k(x_{\{k\}})$$

- 配分函数：所有分配方案的测度总和

$$Z=\sum_{x\in\mathcal{X}}\prod_k f_k(x_{\{k\}})$$

对数线性模型

- 对数线性模型：马尔可夫网络常表示为该模型，通过引入特征函数 ϕ_k

$$f_k = \exp\left(w_k^\top \phi_k(x_{\{k\}})\right)$$

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_k w_k^\top \phi_k(x_{\{k\}})\right)$$

- 配分函数：

$$Z = \sum_{x \in \mathcal{X}} \exp\left(\sum_k w_k^\top \phi_k(x_{\{k\}})\right)$$

将每个团的势函数表示为指数函数，指数项为对应团的加权特征量

Markov逻辑网的独立性

Pairwise Markov Property: 给定所有其他变量，任意两个不相邻变量条件独立。

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

Local Markov Property：一个变量如果给定所有邻居变量后与所有其他变量条件独立。

$$X_v \perp\!\!\!\perp X_{V \setminus \mathrm{cl}(v)} \mid X_{\mathrm{ne}(v)}$$

Global Markov Property：A、B两个子集间任何一条路径都经过子集S，则给定S后，A、B两个子集相互条件独立。

$$X_A \perp\!\!\!\perp X_B \mid X_S$$

一个无向图G=(V, E)，是马尔可夫网络的充分必要条件是：

当且仅当其满足以上三条独立性质。

构建马尔可夫网络

- 方法一：基于Pairwise Markov Property
  - ✓ 如果满足下面条件，则在X和Y之间加边；

$$P \not\models (X \perp Y \mid \mathcal{X} - \{X, Y\}).$$

- 方法二：基于Local Markov Property
  - ✓ 集合U是X的马尔可夫毯(Markov blanket，X的邻居节点集合)，给定U，X独立于余下的变量；

$$(X \perp \mathcal{X} - \{X\} - U \mid U) \in \mathcal{I}(P).$$

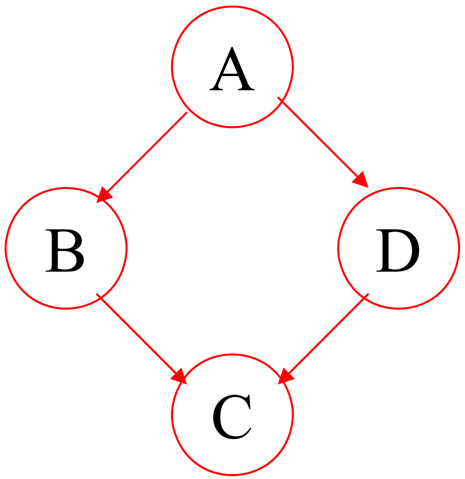
一个班上有四名同学，A、B、C、D，他们每次做作业都会与其他人讨论，但是由于某种不明的原因，A和C、B和D从不一起讨论。当老师上课讲错知识点时，每个学生有一定几率自己发现这个错误，当他发现错误时有可能将这个发现告诉自己讨论的对象。这时我们有了四个随机变量{A,B,C,D}，他们取值为1时表示发现了错误，为0时表示未发现，我们可以获得如下独立性质:

$$(A \perp C | \{B,D\})、(B \perp D | \{A,C\})$$

给定B,D后，A,C独立；给定A,C后，B,D独立。

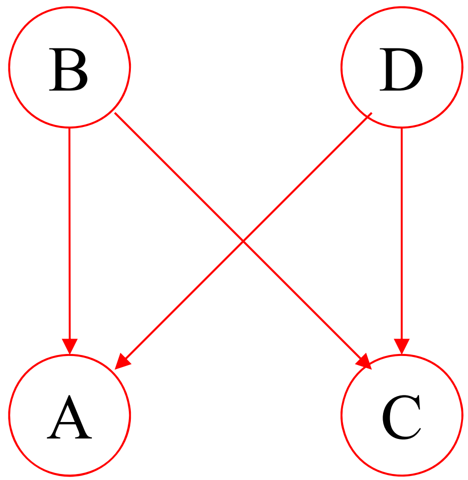
构造贝叶斯网络

- 第一次尝试(假设变量序存在因果)



- 不满足$(B \perp D \mid \{A,C\})$

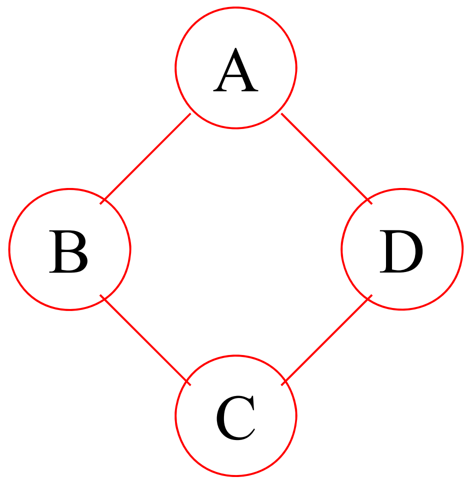
- 第二次尝试(假设B，D是A，C的因)



- 不满足$(B \perp D \mid \{A,C\})$

构造马尔科夫网络

- 将可能的节点之间连上无向边



- 满足$(A \perp C | \{B,D\})$、$(B \perp D | \{A,C\})$

马尔科夫逻辑网

一阶谓词演算：语义

表达式的真值依赖于常元、变元、谓词、函词到论域中的映射；在论域中关系的真假决定了相应表达式的真假。

闭谓词(ground predicate)为不包含变元的谓词。

friends(X, Y)

例如：

friends(george, susie)

friends(george, kate)

一阶谓词演算：语义

一个可能世界(A Possible World)：为每个可能的闭谓词指定真值。

可满足：一个一阶公式是可满足的，当且仅当该公式至少在一个世界中为真。

一阶逻辑的推理：判断一个知识库中是否包含公式F，即F是否在所有满足知识库的世界中为真。

friends(X, Y)

例如：

friends(george, susie)

friends(george, kate)

在一阶逻辑知识库中：每个可能世界必须满足知识库中的所有公式，否则该世界不可能存在（发生概率为0）

MLN起源

概率逻辑学习(Probabilistic Logic Learning, PLL)：

关系逻辑表示、概率推理机制(不确定性处理)、机器学习和数据挖掘，以获得关系数据中的似然模型

Markov逻辑网(Markov Logic Networks, MLNs)：

Markov网+一阶逻辑，其本质是

公式附加权值的一阶逻辑知识库

2004年，Richardson, Domingos, Kok, Singla

Markov逻辑网

基本思想：将一阶逻辑的限制放松，即一个可能世界违反公式越多，其发生的概率越小，但未必为0。

公式权重：表示公式限制强度的大小。权值越大，满足该公式世界的发生概率与不满足该公式世界的发生概率之间的差越大。

$$P(\text{world}) \propto \exp(\sum \text{可满足公式的权重})$$

Markov逻辑网的定义

马尔科夫逻辑网L：是（F_i , w_i）对的集合，其中 F_i代表一阶逻辑规则，w_i是用一个实数表示权重；有限的常数集为$C = \{c_1,c_2,...c_n\}$。

马尔科夫逻辑网$M_{L,C}$ ：

- ✓ L中的任意闭原子（ground atom）都对应了$M_{L,C}$中的一个二值节点。若此闭原子为真，则对应的二值节点取值为1；若为假，则取值为0。
- ✓ L中的任意闭规则（ground formula）都对应着一个特征值。若此闭规则为真，则对应的特征值为1；若为假，则特征值为0。并且这个特征值F_i的权重为二元项中该规则对应的权重w_i。

一个例子

- ✓ 学生的智商(Intelligence)影响了他的SAT成绩；
- ✓ 而考题难度(Difficulty)和智商共同决定影响了他的平时成绩(Grade)；
- ✓ 平时成绩和SAT影响到他能不能拿到导师的推荐信(Letter)；
- ✓ 两个成绩和推荐信又影响到他能不能找到工作(Job) ；
- ✓ 有没有工作决定了他是不是快乐(Happy)。

给出上述例子的MLN

另一个例子

- ✓ 某单位有一个健身房，所有员工都有权利去健身；
- ✓ 有两个同事是好朋友，那么他们有很大的概率(w=3)一起去健身或者不一起去健身；
  权重
- ✓ 如果一个员工整天去健身，那么他/她有较高的概率(w=2)被炒鱿鱼。
  权重

基于一阶逻辑的知识表示

- ✓ 某单位有一个健身房，所有员工都有权利去健身；
- ✓ 有两个同事是好朋友，那么他们有很大的概率(w=3)一起去健身或者不一起去健身；

权重

$$\forall x \forall y (Friends(x, y) \Rightarrow (Plays(x) \Leftrightarrow Plays(y)))$$

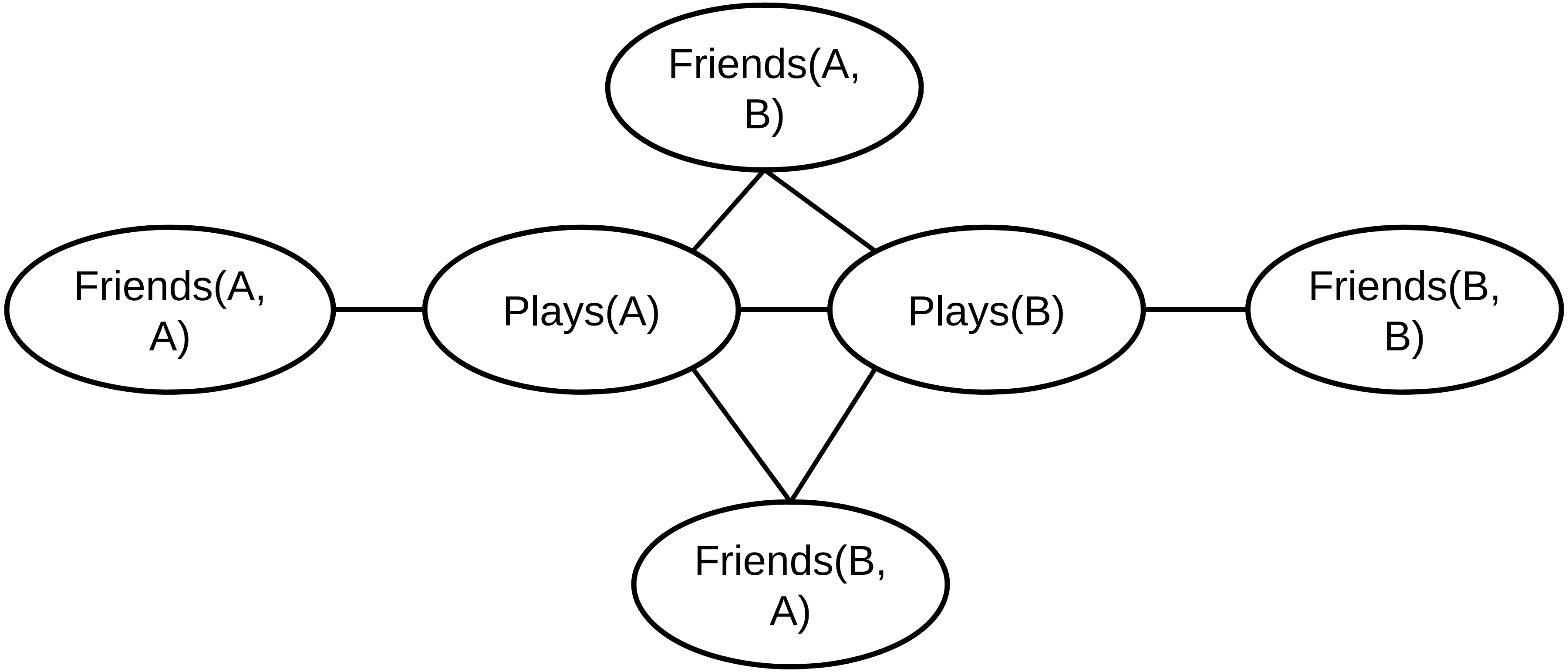
$$\left.\begin{array}{l} \neg Friends(x, y) \vee Plays(x) \vee \neg Plays(y), \\ \neg Friends(x, y) \vee \neg Plays(x) \vee Plays(y) \end{array}\right\}$$

需要同时满足！

- ✓ 如果一个员工整天去健身，那么他/她有较高的概率(w=2)被炒鱿鱼。

权重

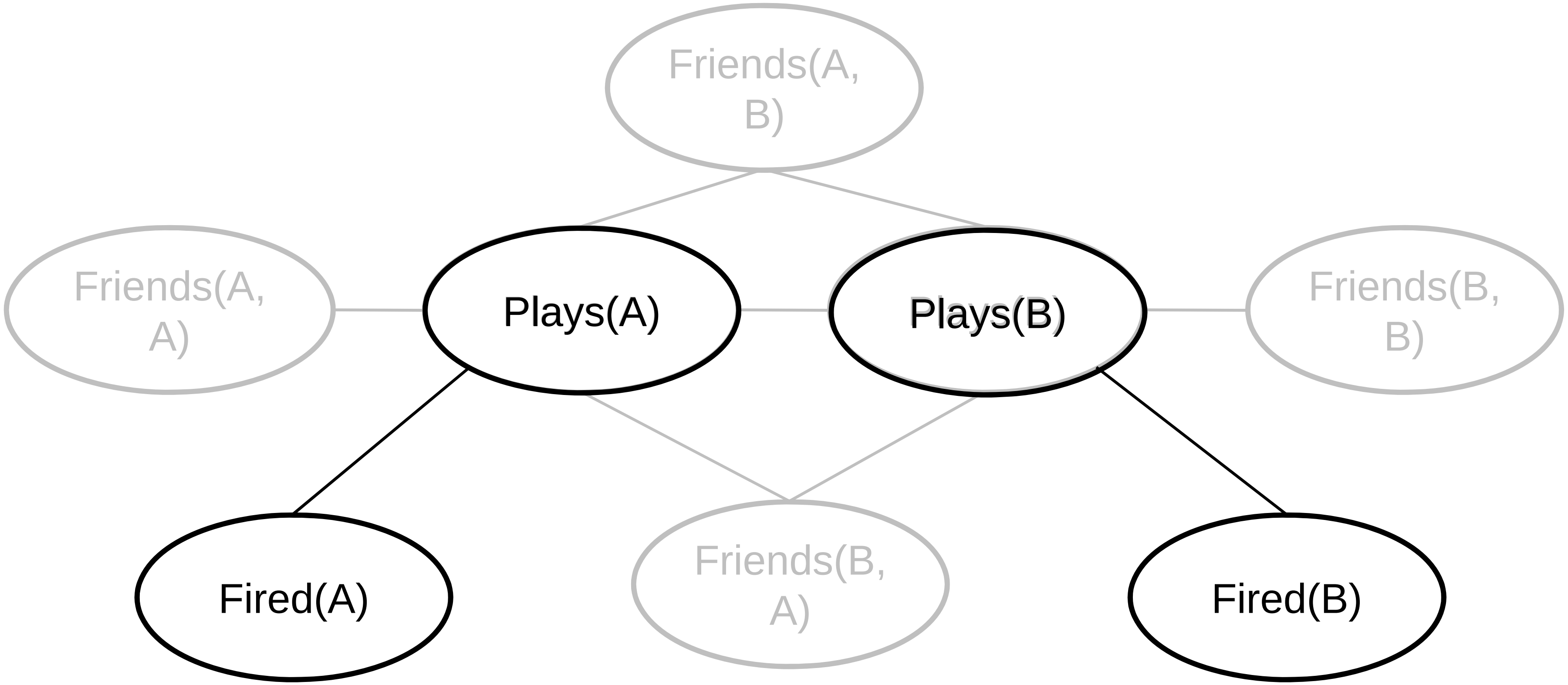
$$\forall x (Plays(x) \Rightarrow Fired(x)) \qquad \neg Plays(x) \vee Fired(x)$$



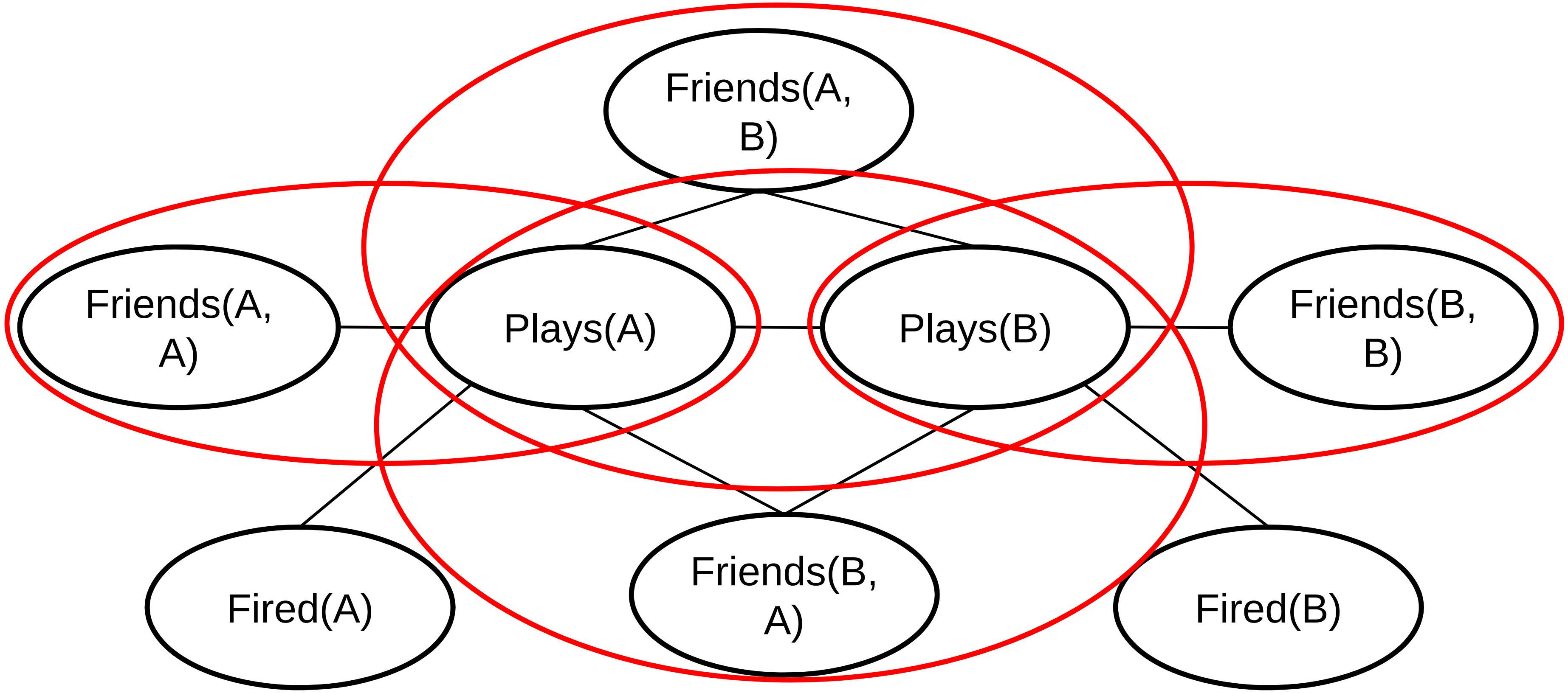
$$\neg Friends(x,y) \vee Plays(x) \vee \neg Plays(y), w = 3$$
$$\neg Friends(x,y) \vee \neg Plays(x) \vee Plays(y), w = 3$$

Constants:
Alice (A) and Bob (B)

$$\neg Plays(x) \vee Fired(x), w = 2$$

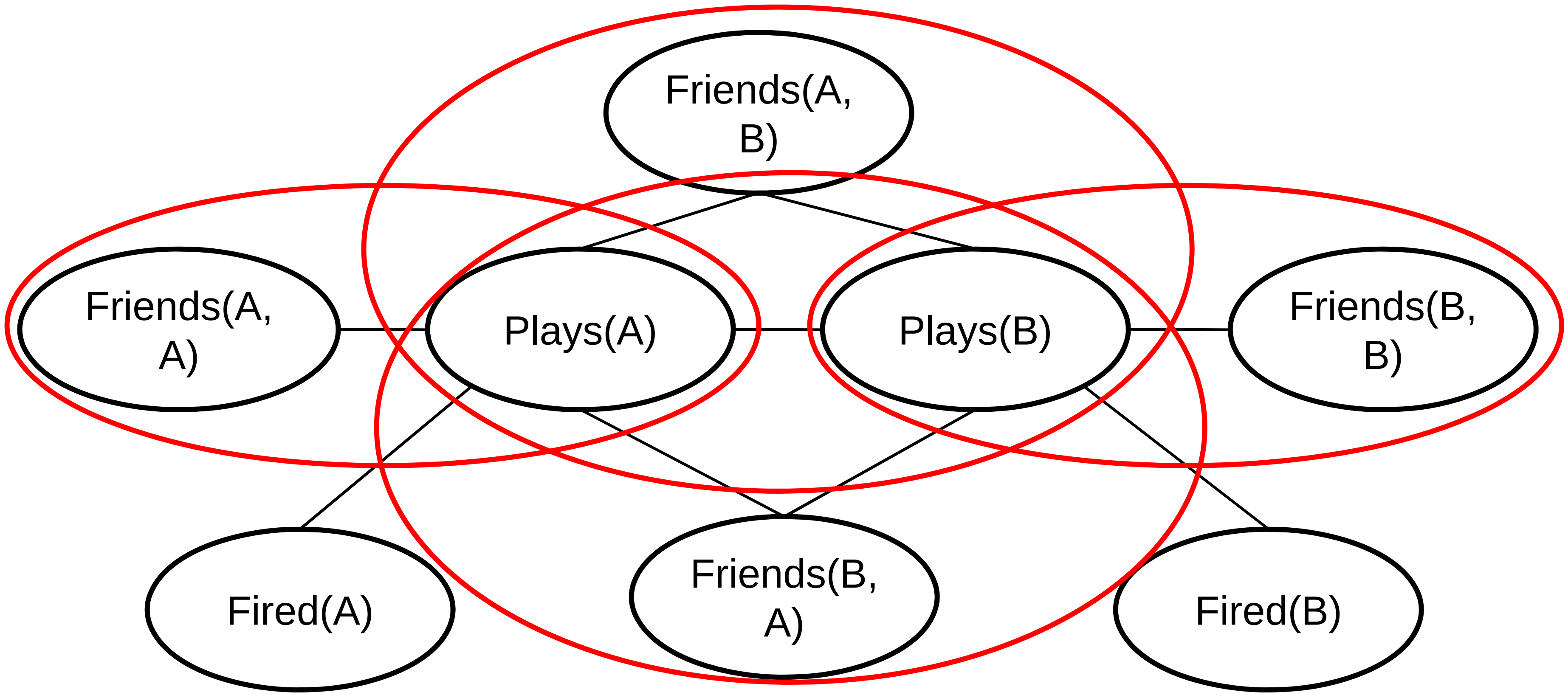


那么请问Alice和Bob是朋友，他们都整天去健身，而且他们都没有被炒鱿鱼的可能性？



$$\neg Friends(x,y) \vee Plays(x) \vee \neg Plays(y), w = 3$$

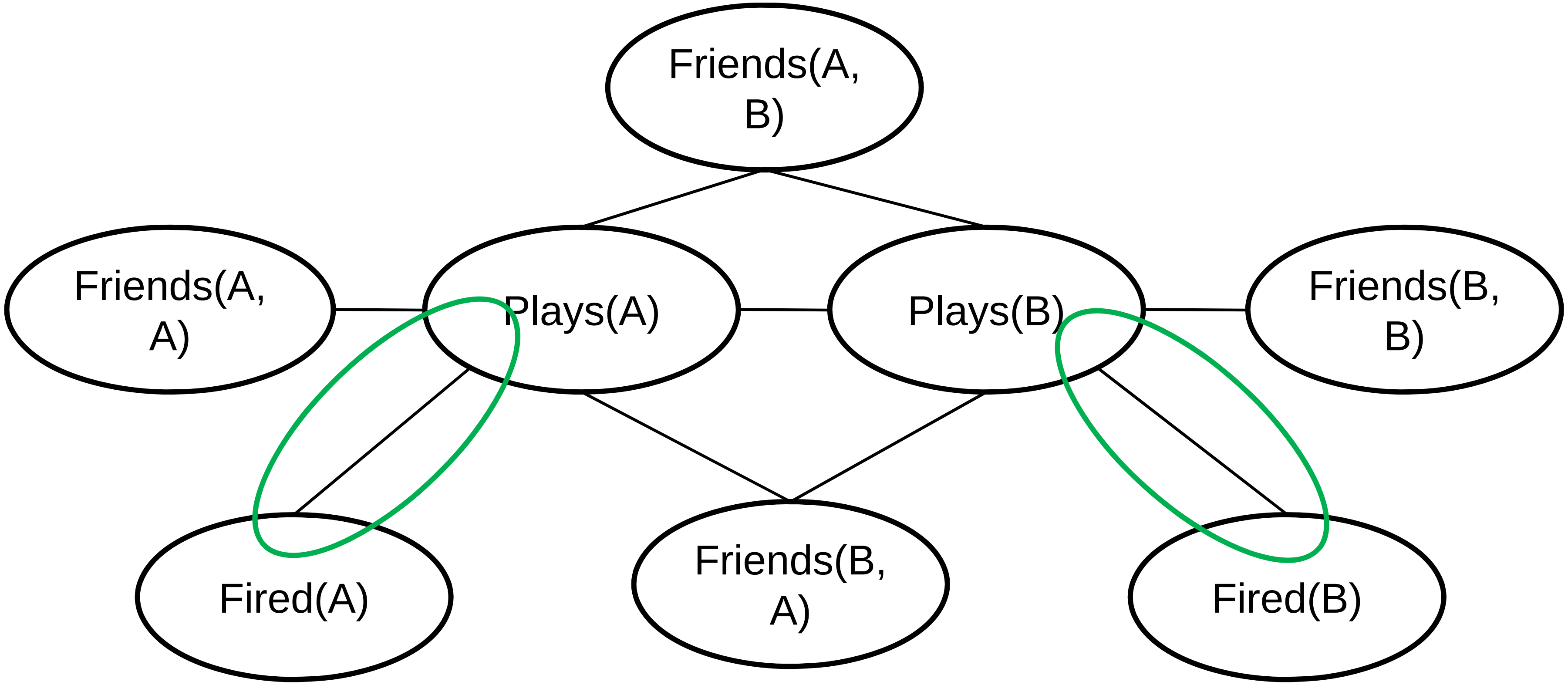
$$\neg Friends(x,y) \vee \neg Plays(x) \vee Plays(y), w = 3$$



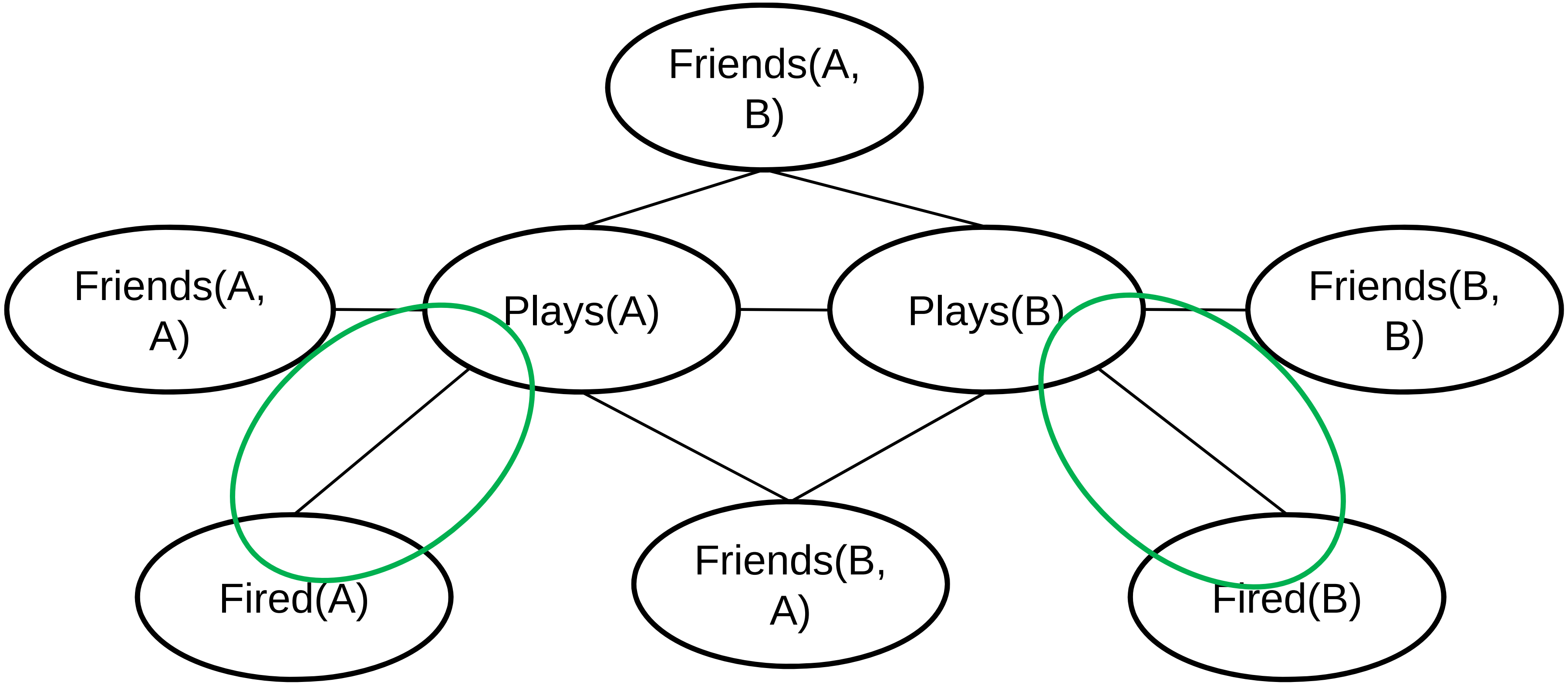
$$\neg Friends(x, y) \vee Plays(x) \vee \neg Plays(y), w = 3$$

$$\neg Friends(x, y) \vee \neg Plays(x) \vee Plays(y), w = 3$$

Friends(x,y)	Plays(x)	Plays(y)	w
T	T	T	3
T	F	T	0
F	T	T	3
F	F	T	3
T	T	F	0
T	F	F	3
F	T	F	3
F	F	F	3



$$\neg Plays(x) \vee Fired(x), w = 2$$



$$\neg Plays(x) \vee Fired(x), w = 2$$

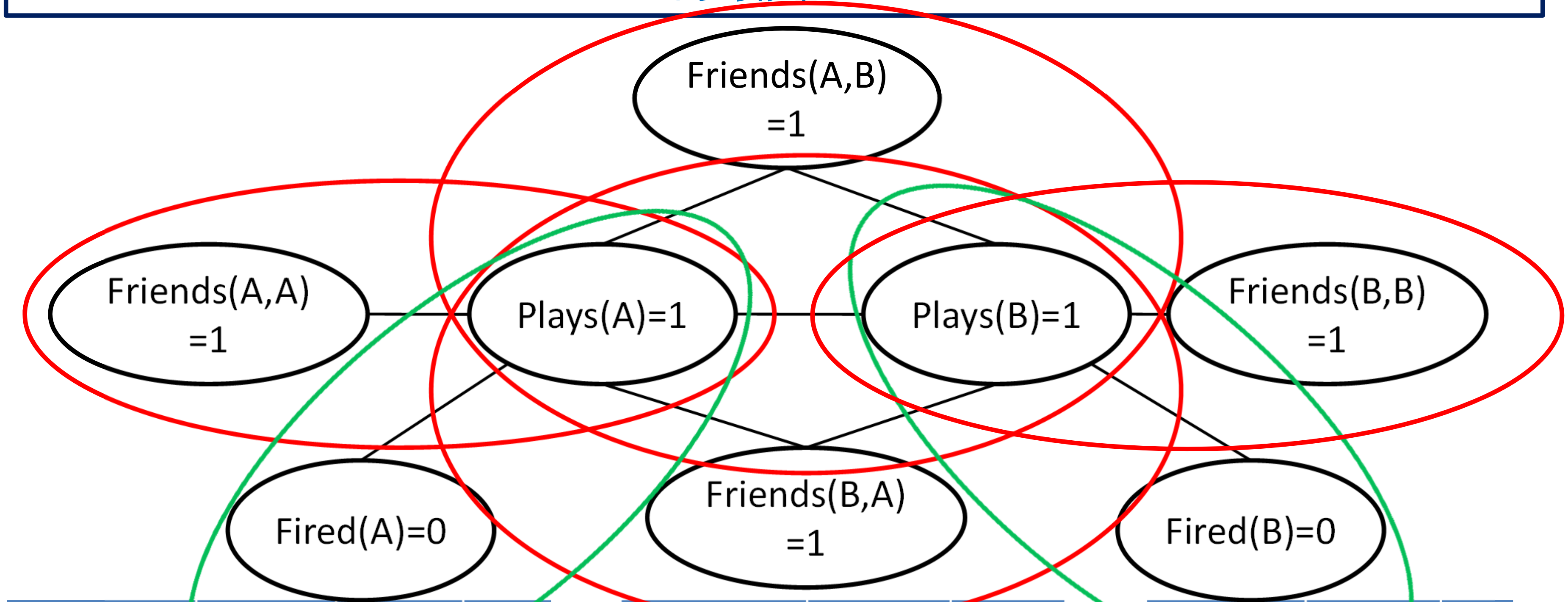
Plays(x)	Fired(x)	w
T	T	2
T	F	0
F	T	2
F	F	2

问题1：那么请问Alice和Bob是朋友，他们都整天去健身，而且他们都没有被炒鱿鱼的可能性？

$$P(X = x(Alice,Bob)) = \frac{1}{Z} exp\left(\sum_{i=1}^{3} \omega_i n_i (x(Alice,Bob)) \right)$$

- $n_i(x)$是F_i在X中所有取真值的公式的数量
- 而x{i}是F_i中为真的原子

问题1：那么请问Alice和Bob是朋友，他们都整天去健身，而且他们都没有被炒鱿鱼的可能性？



Friends(A,B)	Plays(A)	Plays(B)	ω
True	True	True	3

Friends(x,x)	Plays(x)	ω
True	True	3

Plays(x)	Fired(x)	ω
True	False	0

$$P(X = x(Alice, Bob)) = \frac{1}{Z}\exp\left(\sum_{i=1}^{3}\omega_i n_i(x(Alice, Bob))\right)$$

$$= \frac{1}{Z}\exp(3*4+3*4+2*0) = \frac{1}{Z}\exp(24)$$

MLN特例：一阶逻辑

- 只有一条规则

  $\forall x\ R(x) \Rightarrow S(x)$、权重为w，常数集C = {A}

- 只有4个事件

  $\{\neg R(A),\neg S(A)\},\{\neg R(A),S(A)\},\{R(A),\neg S(A)\},\{R(A),S(A)\}$

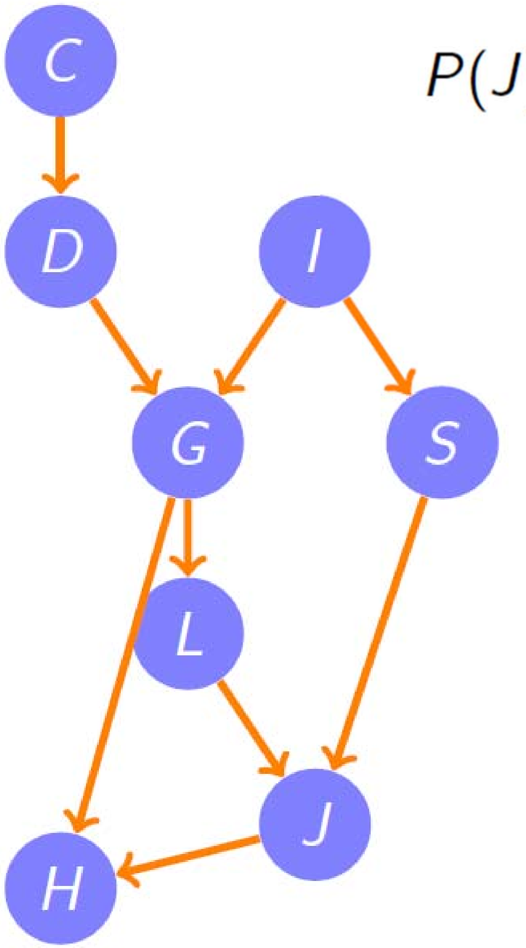
$P(\{R(A),\neg S(A)\})=1/(3e^{w}+1)$ 其他均为 $e^{w}/(3e^{w}+1)$

马尔科夫网推理

MLN推理

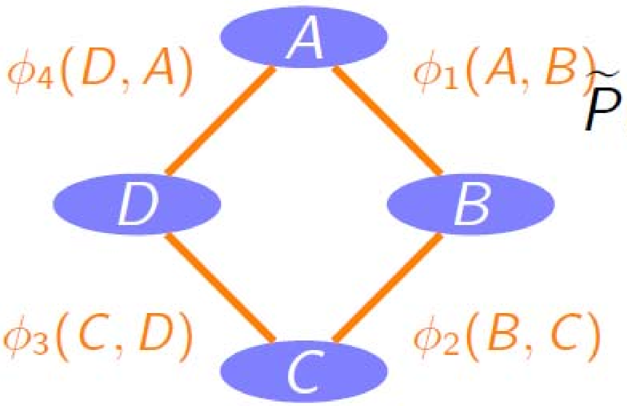
- 条件概率查询(Conditional Probability Query)
  - ✓ 证据：$\vec{E} = \vec{e}$
  - ✓ 查询：变量子集$\vec{Y}$
  - ✓ 计算：$P(\vec{Y}|\vec{E} = \vec{e})$
- 最大化后验(Maximum a Posterior, MAP)
  - ✓ 证据：$\vec{E} = \vec{e}$
  - ✓ 查询：所有其他变量$\vec{Y}(\vec{Y} = \{X_1, X_2, \dots, Xn\} - \vec{E})$
  - ✓ 计算：$\mathrm{MAP}(\vec{Y}|\vec{E} = \vec{e}) = \operatorname{argmax}_{\vec{y}} P(\vec{Y} = \vec{y}|\vec{E} = \vec{e})$

BN中的和积(Sum-Product)



$$P(J) = \sum_{C,D,I,G,S,L,H} \phi_C(C)\phi_D(C,D)\phi_I(I)\phi_G(G,I,D)$$
$$\phi_S(S,I)\phi_L(L,G)\phi_J(J,L,S)\phi_H(H,G,J)$$

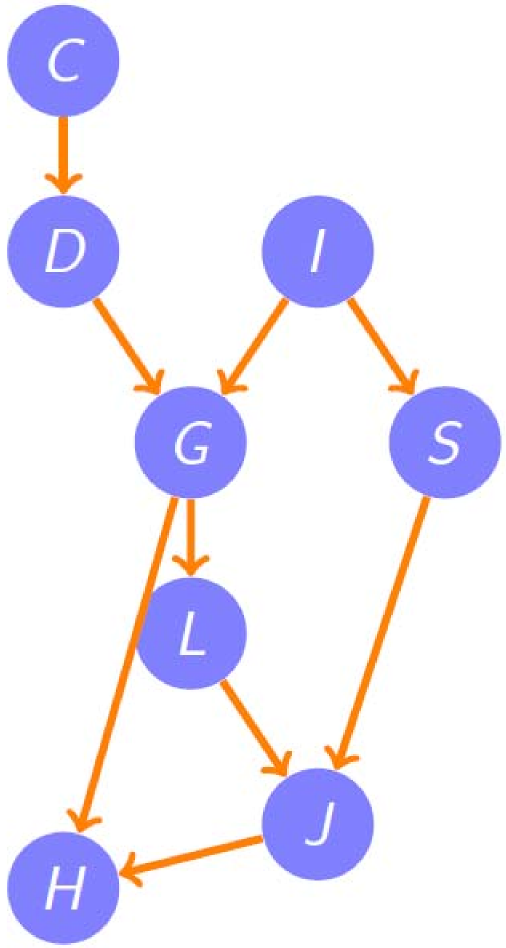
MN中的和积(Sum-Product)



$$\widetilde{P}(D) = \sum_{A,B,C} \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(A, D)$$

$$P(D) = \frac{1}{Z}\widetilde{P}(D)$$

证据：Reduced Factors



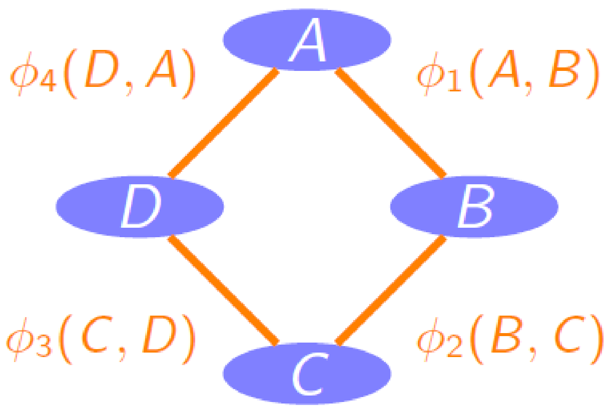
$$P(J) = \sum_{C,D,I,G,S,L,H} \phi_C(C)\phi_D(C,D)\phi_I(I)\phi_G(G,I,D)$$
$$\phi_S(S,I)\phi_L(L,G)\phi_J(J,L,S)\phi_H(H,G,J)$$

$$P(J, I=i, H=h) =$$
$$\sum_{C,D,I,G,S,L,H} \phi_C(C)\phi_D(C,D)\boxed{\phi_I(i)}\boxed{\phi_G(G,i,D)}$$
$$\boxed{\phi_S(S,i)}\phi_L(L,G)\phi_J(J,L,S)\boxed{\phi_H(h,G,J)}$$

证据：Reduced Factors

$$P(\vec{Y}|\vec{E}=\vec{e}) = \frac{\boxed{P(\vec{Y},\vec{E}=\vec{e})}}{P(\vec{E}=\vec{e})}$$

$$\text{Let } \vec{W} = \{X_1,\ldots,X_n\} - \vec{Y} - \vec{E}$$



$$\begin{aligned}
\boxed{P(\vec{Y},\vec{E}=\vec{e})} &= \sum_{\vec{W}} P(\vec{Y},\vec{W},\vec{E}=\vec{e}) \\
&= \sum_{\vec{W}} \frac{1}{Z} \prod_k \boxed{\phi_k(\vec{D}_k,\vec{E}=\vec{e})} \\
&= \sum_{\vec{W}} \boxed{\frac{1}{Z}} \prod_k \boxed{\phi'_k(\vec{D}'_k)} \\
&\propto \sum_{\vec{W}} \prod_k \phi'_k(\vec{D}'_k)
\end{aligned}$$

证据：Reduced Factors

$$P(\vec{Y}|\vec{E}=\vec{e}) = \frac{\boxed{P(\vec{Y},\vec{E}=\vec{e})}}{\boxed{P(\vec{E}=\vec{e})}}$$

Let $\vec{W} = \{X_1,\ldots,X_n\} - \vec{Y} - \vec{E}$

$$\boxed{P(\vec{Y},\vec{E}=\vec{e})} = \sum_{\vec{W}} \frac{1}{Z} \prod_k \phi'_k(\vec{D}'_k)$$

$$\boxed{P(\vec{E}=\vec{e})} = \sum_{\vec{Y}} \sum_{\vec{W}} \frac{1}{Z} \prod_k \phi'_k(\vec{D}'_k)$$

Compute $\sum_{\vec{W}} \prod_k \phi'_k(\vec{D}'_k)$ and renormalize

变量消除算法

算法

- 精确算法
  - ✓ 变量消除(Variable elimination)
- 近似算法
  - ✓ 信念传播(Belief propagation)
  - ✓ 随机采样(Random sampling)
    - ✓ Markov chain
    - ✓ Gibbs sampling

变量消除算法

- 基本思想
  - ✓ 为得到精确解，必须计算和积
  - ✓ 通常一个变量不会出现在所有的因子中

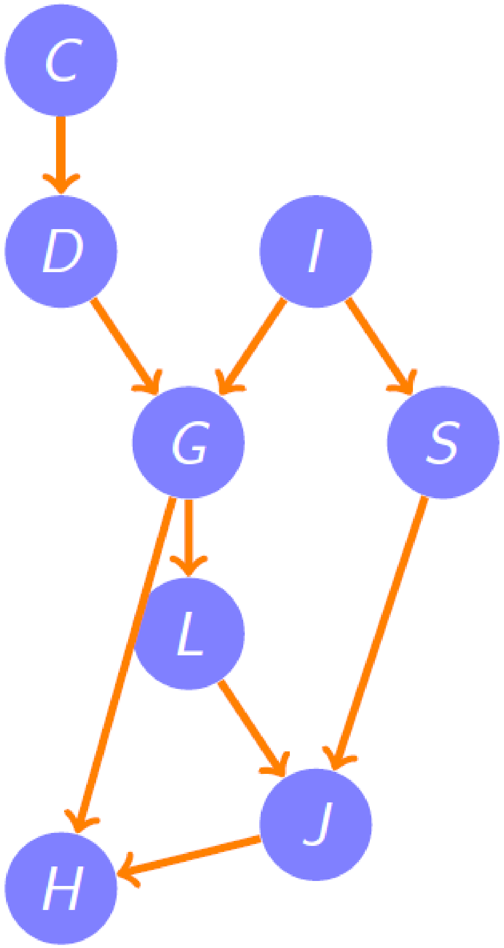
变量消除算法

$$
\begin{aligned}
P(E) &\propto \sum_D \sum_C \sum_B \sum_A \widetilde{P}(A, B, C, D, E) \\
&= \sum_D \sum_C \sum_B \sum_A \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, E) \\
&= \sum_D \sum_C \sum_B \phi_2(B, C)\phi_3(C, D)\phi_4(D, E) \boxed{\sum_A \phi_1(A, B)} \\
&= \sum_D \sum_C \sum_B \phi_2(B, C)\phi_3(C, D)\phi_4(D, E) \boxed{\tau_1(B)}
\end{aligned}
$$

变量消除算法

$$
\begin{aligned}
P(E) &\propto \sum_D \sum_C \sum_B \phi_2(B,C)\phi_3(C,D)\phi_4(D,E)\tau_1(B) \\
&= \sum_D \sum_C \phi_3(C,D)\phi_4(D,E) \boxed{\sum_B \phi_2(B,C)\tau_1(B)} \\
&= \sum_D \sum_C \phi_3(C,D)\phi_4(D,E) \boxed{\tau_2(C)} \\
&= \sum_D \phi_4(D,E) \boxed{\sum_C \phi_3(C,D)\tau_2(C)} \\
&= \sum_D \phi_4(D,E) \boxed{\tau_3(D)}
\end{aligned}
$$

图模型中的变量消除



$$P(J)$$

$$= \sum_{C,D,I,G,S,L,H} \phi_J(J,L,S)\phi_L(L,G)\phi_S(S,I)$$

$$\phi_G(G,I,D)\phi_H(H,G,J)\phi_I(I)\phi_C(C)\phi_D(C,D)$$

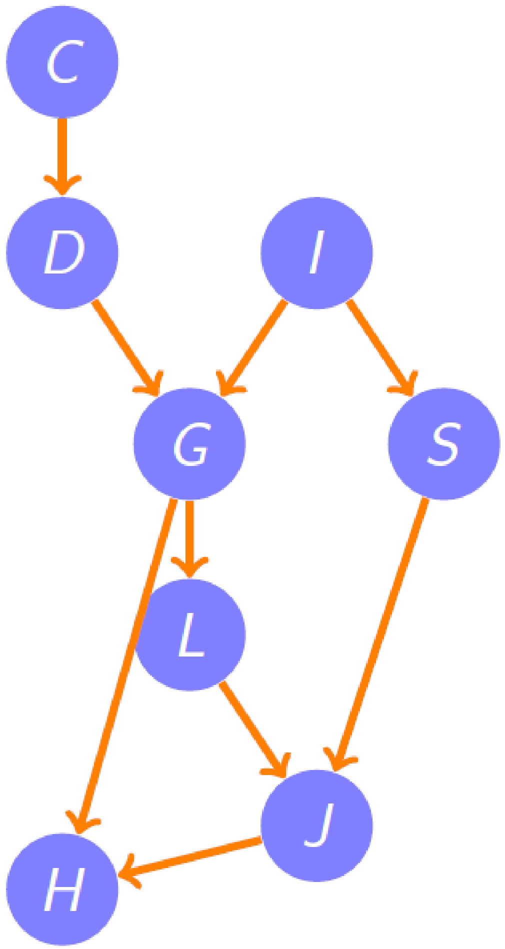
$$= \sum_{D,I,G,S,L,H} \phi_J(J,L,S)\phi_L(L,G)\phi_S(S,I)$$

$$\phi_G(G,I,D)\phi_H(H,G,J)\phi_I(I)\boxed{\sum_C \phi_C(C)\phi_D(C,D)}$$

$$= \sum_{D,I,G,S,L,H} \phi_J(J,L,S)\phi_L(L,G)\phi_S(S,I)$$

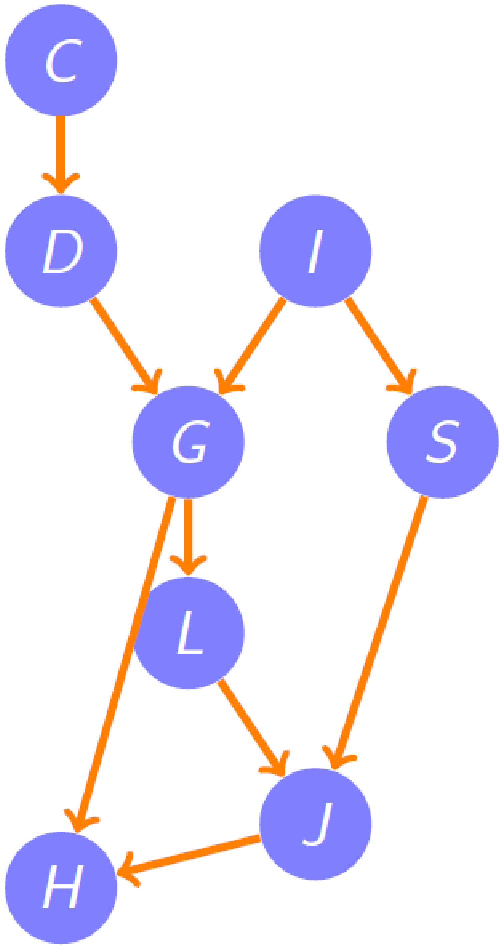
$$\phi_G(G,I,D)\phi_H(H,G,J)\phi_I(I)\boxed{\tau_1(D)}$$

图模型中的变量消除



$$
\begin{aligned}
&P(J) \\
&= \sum_{I,G,S,L,H} \phi_J(J,L,S)\phi_L(L,G)\phi_S(S,I) \\
&\phi_H(H,G,J)\phi_I(I)\boxed{\tau_2(G,I)} \\
&= \sum_{G,S,L,H} \phi_J(J,L,S)\phi_L(L,G) \\
&\phi_H(H,G,J)\boxed{\sum_I \phi_S(S,I)\phi_I(I)\tau_2(G,I)} \\
&= \sum_{G,S,L,H} \phi_J(J,L,S)\phi_L(L,G) \\
&\phi_H(H,G,J)\boxed{\tau_3(G,S)}
\end{aligned}
$$

图模型中的变量消除



$$P(J)$$

$$= \sum_{G,S,L,H} \phi_J(J,L,S)\phi_L(L,G)$$

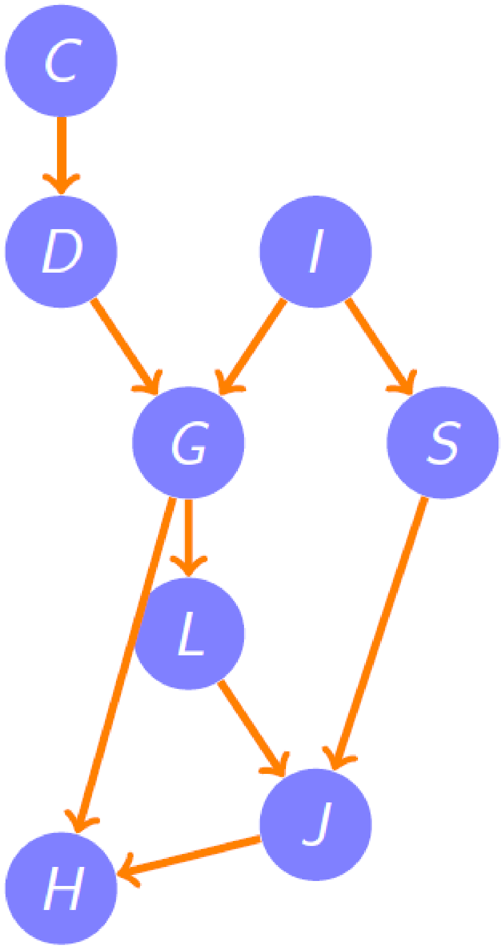
$$\phi_H(H,G,J)\tau_3(G,S)$$

$$= \sum_{G,S,L} \phi_J(J,L,S)\phi_L(L,G)$$

$$\tau_3(G,S) \boxed{\sum_H \phi_H(H,G,J)}$$

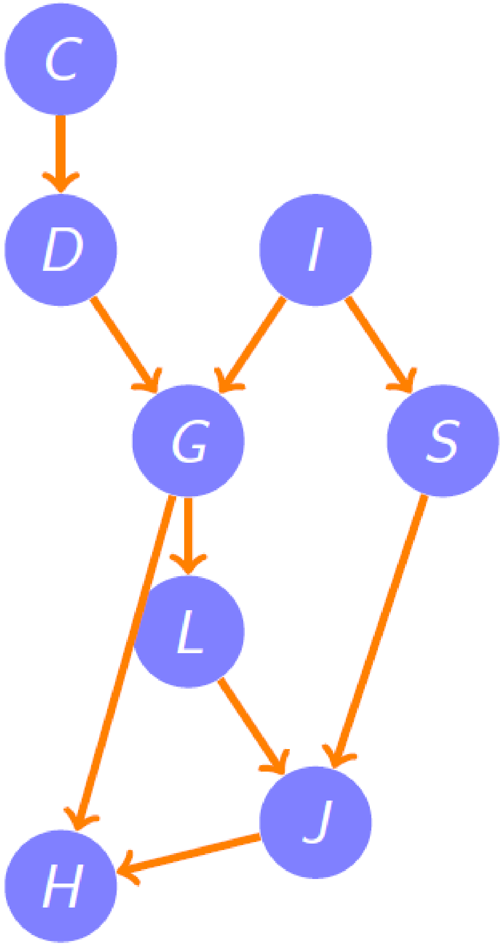
$$= \sum_{G,S,L} \phi_J(J,L,S)\phi_L(L,G)\tau_3(G,S) \boxed{\tau_4(G,J)}$$

图模型中的变量消除



$$
\begin{aligned}
&P(J) \\
&= \sum_{G,S,L} \phi_J(J,L,S)\phi_L(L,G)\tau_3(G,S)\tau_4(G,J) \\
&= \sum_{S,L} \phi_J(J,L,S) \boxed{\sum_{G} \phi_L(L,G)\tau_3(G,S)\tau_4(G,J)} \\
&= \sum_{S,L} \phi_J(J,L,S) \boxed{\tau_5(J,L,S)}
\end{aligned}
$$

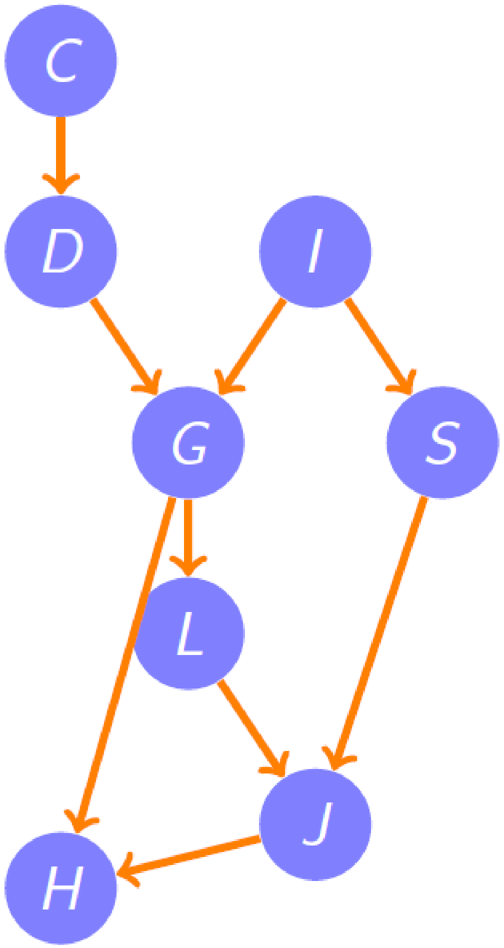
带证据的变量消除



$$
\begin{aligned}
& P(J, I = i, H = h) \\
&= \sum_{C,D,I,G,S,L,H} \phi_C(C)\phi_D(C,D)\boxed{\phi_I(i)}\boxed{\phi_G(G,i,D)} \\
& \boxed{\phi_S(S,i)}\phi_L(L,G)\phi_J(J,L,S)\boxed{\phi_H(h,G,J)} \\
&= \sum_{C,D,G,S,L} \phi_C(C)\phi_D(C,D)\boxed{\phi'_I()}\boxed{\phi'_G(G,D)} \\
& \boxed{\phi'_S(S)}\phi_L(L,G)\phi_J(J,L,S)\boxed{\phi'_H(G,J)}
\end{aligned}
$$

重复继续……

带证据的变量消除



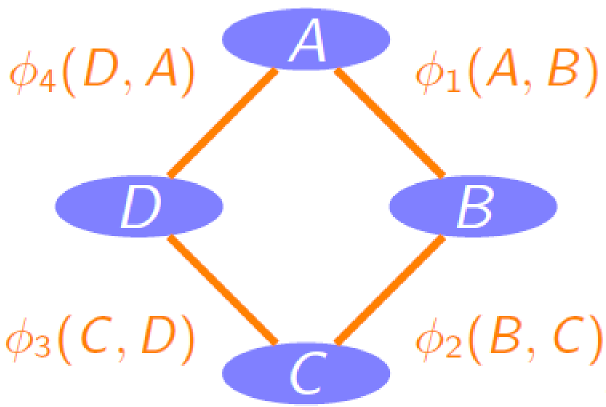
如何计算 $P(J|I=i,H=h)$?

$$P(J|I=i,H=h)=\frac{P(J,I=i,H=h)}{P(I=i,H=h)}$$

归一化

$$P(I=i,H=h)=\sum_{J}P(J,I=i,H=h)$$

MN中的变量消除



$$\sum_{A,B,C} \phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(A, D)$$

$$= \sum_{B,C} \phi_2(B, C)\phi_3(C, D) \boxed{\sum_A \phi_1(A, B)\phi_4(A, D)}$$

$$= \sum_{B,C} \phi_2(B, C)\phi_3(C, D) \boxed{\tau_1(B, D)}$$

$$\ldots$$

$$= \tau_3(D) = \widetilde{P}(D) \propto P(D)$$

归一化……

变量消除算法

基本思想

- ✓ 根据给定的证据，得出需要约简的因子集Φ
- ✓ 对每一个非查询的变量X，将其从Φ消除
- ✓ 连乘所有余下的因子
- ✓ 正则化……

$$\text{Let } \Phi' = \{\phi_i \in \Phi : X \in scope[\phi_i]\}$$

$$\psi = \prod_{\phi_i \in \Phi'} \phi_i$$

$$\tau = \sum_X \psi$$

$$\Phi := (\Phi - \Phi') \cup \{\tau\}$$

变量消除算法

- 小结
  - ✓ 对BN和MN均有效
    - ✓ BN中计算的是条件概率分布
    - ✓ MN中计算的是势函数
  - ✓ 任何一种消除次序都可以导致正确的结果
  - ✓ 计算复杂性NP-hard……

思考和讨论

1. 变量消除算法的计算复杂度分析？
2. 如何设计有效的变量消除次序？
3. MLN与一阶谓词逻辑的关系？
4. MLN与BN的异同和联系？

谢 谢 ！