电子科技大学
University of Electronic Science and Technology of China

Chap. 3-4
图模型的两种表示方法

黄峰 & 黄晨

Data Mining Lab,
Big Data Research Center, UESTC
Email：huangchen.uestc@gmail.com

目录

- **第一部分：贝叶斯网络**
  - ✓ 表示
  - ✓ 推断的例子
  - ✓ 。。。

- **第二部分：马尔科夫随机场**
  - ✓ MRF的初步认识
  - ✓ 越来越熟悉的MRF
  - ✓ 熟悉到陌生的BN和MRF

友情提示：为了和峰峰的风格保持一致，我继续用中文做ppt，并且同样加入新的风格，避免枯燥的理论讲解，迈出人类坚实的又一小步。

第二部分

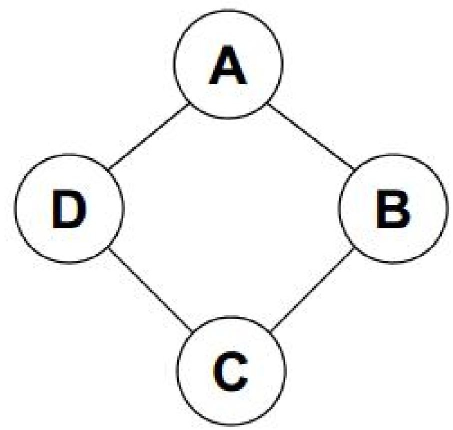
马尔科夫随机场

从入门到放弃

马尔科夫随机场的初步认识

举个栗子

- 四个学生ABCD，被安排两两一起讨论做毕设
- 但AC有矛盾，BD也有矛盾
- 四个人对毕设的要求可能存在偏差（0存在，1不存在）
- 假设一开始，AB意见一致的可能性比不一致的可能性更大

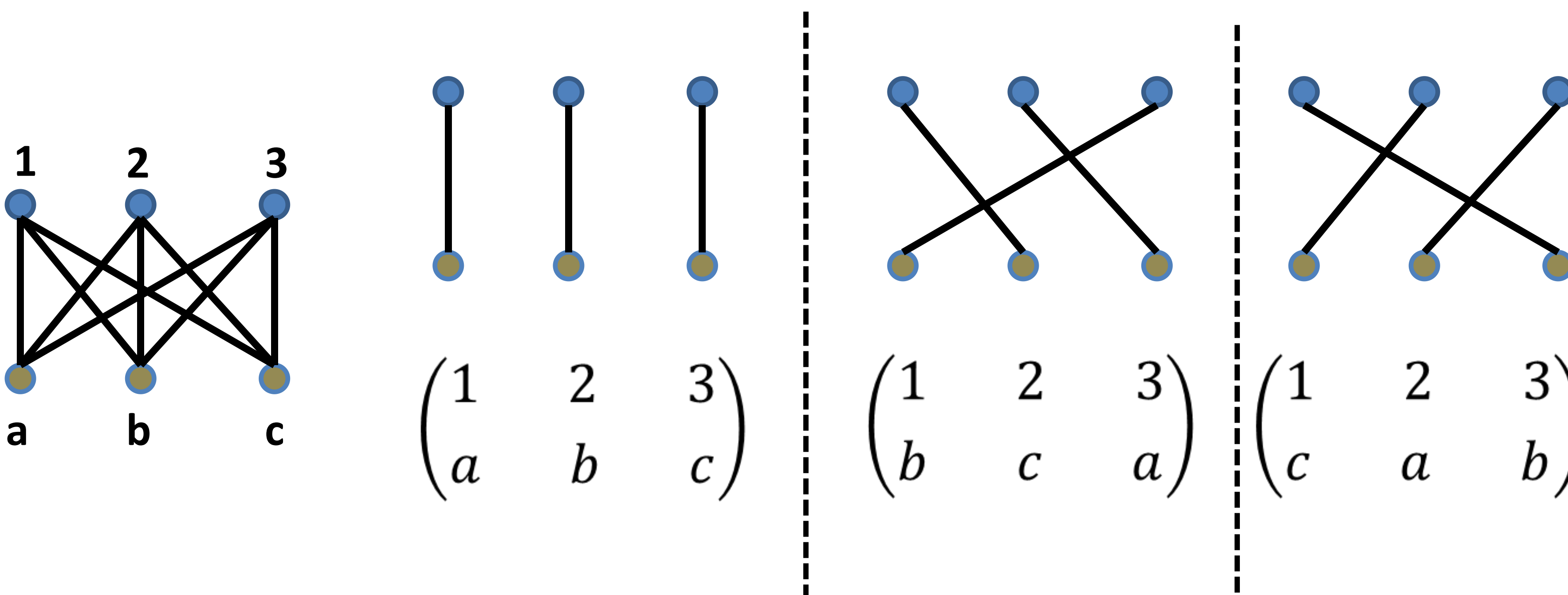


MRF

- 怎么刻画这个关系？ → 因子
  → 因子是什么

*图论中的因子

- **因子**：G的一个至少含有一条边的生成子图(点不变，边子集)
- **N-因子**：N度正则因子（每个节点度数都为N的因子）
- **因子分解**：G划分成为边不相交（无公共边）的因子的并
- **N-因子分解**：G划分成为边不相交（无公共边）的N-因子的并

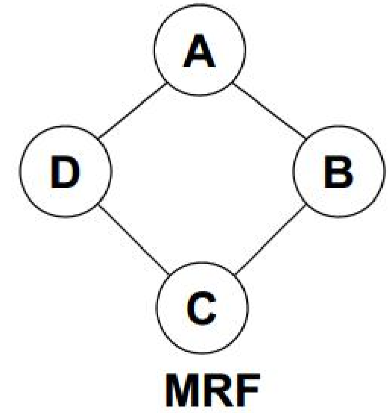


$$\begin{pmatrix} 1 & 2 & 3 \\ a & b & c \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ b & c & a \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ c & a & b \end{pmatrix}$$

图G的一种1-因子分解 ⟶ 映射？N-因子分解？因子分解？

关系刻画

- **因子**∅：定义为变量集X到实数域R上的一个映射
- 这个变量集叫做∅的**辖域**，记为Scope[∅]
- 刻画变量之间的局部交互影响



$\phi_1(A,B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

(a)

$\phi_2(B,C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

(b)

$\phi_3(C,D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

(c)

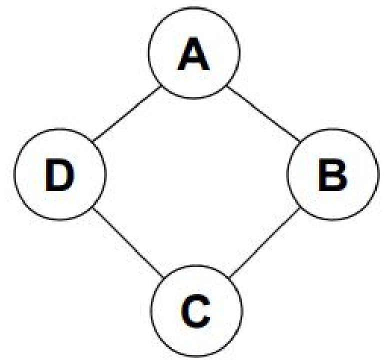
$\phi_4(D,A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

(d)

全局刻画

- 为了定义一个全局的模型，仿照DAG中概率乘积的方式，对局部交互项做组合

Assignment				Unnormalized	Normalized
a^0	b^0	c^0	d^0	300,000	0.04
a^0	b^0	c^0	d^1	300,000	0.04
a^0	b^0	c^1	d^0	300,000	0.04
a^0	b^0	c^1	d^1	30	$4.1 \cdot 10^{-6}$
a^0	b^1	c^0	d^0	500	$6.9 \cdot 10^{-5}$
a^0	b^1	c^0	d^1	500	$6.9 \cdot 10^{-5}$
a^0	b^1	c^1	d^0	5,000,000	0.69
a^0	b^1	c^1	d^1	500	$6.9 \cdot 10^{-5}$
a^1	b^0	c^0	d^0	100	$1.4 \cdot 10^{-5}$
a^1	b^0	c^0	d^1	1,000,000	0.14
a^1	b^0	c^1	d^0	100	$1.4 \cdot 10^{-5}$
a^1	b^0	c^1	d^1	100	$1.4 \cdot 10^{-5}$
a^1	b^1	c^0	d^0	10	$1.4 \cdot 10^{-6}$
a^1	b^1	c^0	d^1	100,000	0.014
a^1	b^1	c^1	d^0	100,000	0.014
a^1	b^1	c^1	d^1	100,000	0.014



MRF

$$P(a,b,c,d) = \frac{\emptyset_1(a,b)\emptyset_2(b,c)\,\emptyset_3(c,d)\emptyset_4(d,a)}{\sum_{a,b,c,d}\emptyset_1(a,b)\emptyset_2(b,c)\emptyset_3(c,d)\emptyset_4(d,a)}$$

因子积的意义

- 令XYZ是三个不相交的变量集
- $\phi_1(X,Y)$和$\phi_2(Y,Z)$是两个因子，那么两者的乘积可以定义为一个新的因子：

$$\varphi(X,Y,Z) = \phi_1(X,Y)\phi_2(Y,Z)$$

- 在更大的辖域上进行刻画

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

回顾BN

- 当计算$P(X,Y)$时，总是可以写成概率乘积的形式$P(X)P(Y|X)$或者$P(Y)P(X|Y)$，更一般来说：

$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i | Parents(X_i))$$

- 现在令

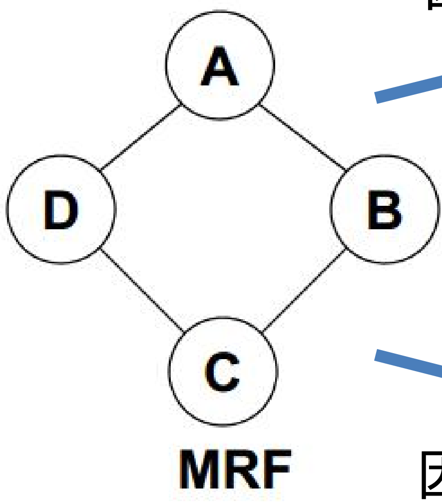
$$\emptyset_{X_i}\big(X_i, Parents(X_i)\big) = P(X_i | Parents(X_i))$$

- 则有

$$P(X_1, X_2, \dots, X_n) = \prod_i \emptyset_{X_i}\big(X_i, Parents(X_i)\big)$$

- 所以，BN把因子定义为变量的条件概率。

因子积的效果



MRF

因子乘积之前

$\phi_1(A, B)$

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

因子乘积之后

a^0	b^0	0.13
a^0	b^1	0.69
a^1	b^0	0.14
a^1	b^1	0.04

确实引入了变量之间的相互影响

*因子表示的好处

➢ 若$P_\phi(X_1,\dots,X_n)$满足以下

- 则称$P_\phi(X_1,\dots,X_n)$为在因子集$D=\{D_1,D_2,\dots,D_m\}$上的Gibbs分布，其中Z称为Partition Function

$$P_\emptyset(X_1,\dots,X_n)=\frac{1}{Z}\emptyset_1(D_1)\emptyset_2(D_2)\dots\emptyset_m(D_m)$$

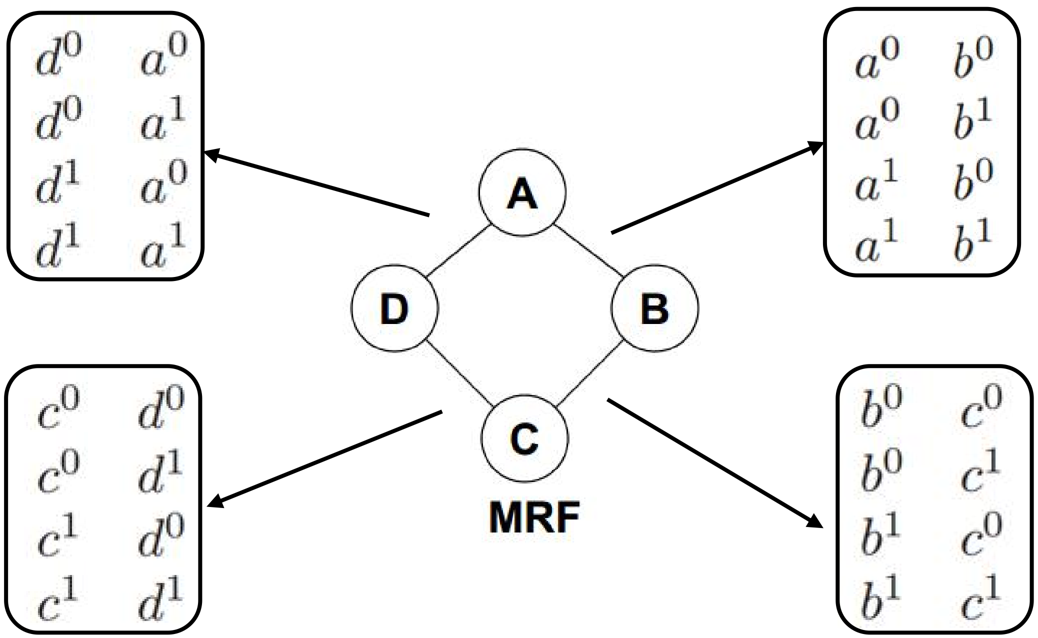
$$Z=\sum_{D_1,\dots,D_n}\emptyset_1(D_1)\emptyset_2(D_2)\dots\emptyset_m(D_m)$$

很生硬的引入：
无向图模型上通过因子分解的方式正好对应了一个Gibbs分布

因子表示的好处

- 回想BN，因子分解提供了联合概率的紧凑表示
- 那么，下图联合概率需要2^4-1个参数，因子分解表示之后呢？

- 现在考虑n个节点，全连通MRF的情况：



$$2^n - 1 \ \to 3C_n^2$$

MRF的初步认识

不管你承不承认，反正现在我已经给你介绍了无向图模型的大致轮廓了。
那到目前为止，你认识到了什么？

- MRF通过定义**因子**来刻画变量之间的相互关系，这种因子的表示方变量联合概率提供了紧凑表示，简化计算
- MRF与**Gibbs**有联系

给个缓冲

人生若只如初见，何事秋风悲画扇

事物的结果并不像人们最初想象的那样美好，在发展的过程中往往会变化的超出人们最初的理解。

越来越熟悉的MRF

填空题

➢ **根据你的了解，请在空白处填上适当的语句。**

- MRF通过在 __________ 上，定义因子函数 _________ 来刻画变量之间的相互关系，这种因子的表示方变量联合概率提供了紧凑表示，简化计算。
- MRF与**Gibbs**的联系是 _____________

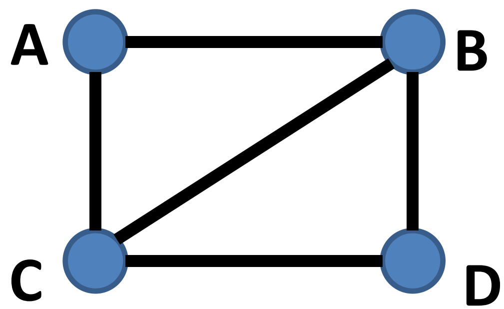
➢ **听完这一部分之后，你需要知道：**

- 因子的辖域如何选择？为什么这么选择？
- 因子定义成什么？为什么这么定义？
- MRF的具体定义是？
- MRF与Gibbs分布的关系是？

因子辖域选择

➢团（Clique）！

- 全连通子图
- 最大团（Max Clique）



最大团= ABC, BCD

➡ 由图结构导出的独立性不会被破坏

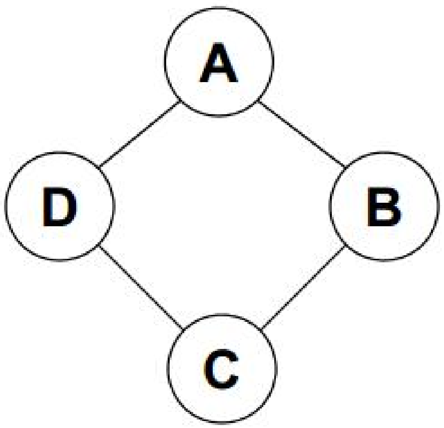
➡ MRF中的独立性是什么？

无向图中的独立性

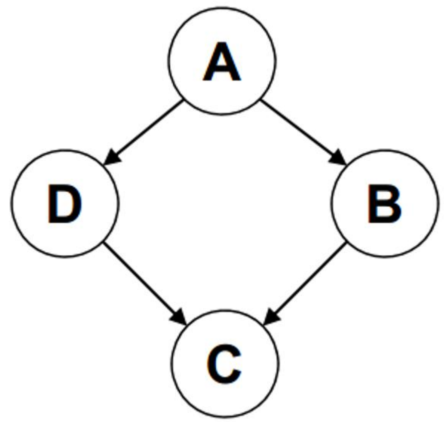
数据挖掘实验室 Data Mining Lab

➢ 马尔科夫毯（Markov Blanket）

- $\mathrm{MB}_{\mathrm{MRF}}(\boldsymbol{X})$={ X的所有邻居 }
- $\mathrm{MB}_{\mathrm{DAG}}(\boldsymbol{X})$={ X的父节点, X的子节点， X的子节点的父节点 }



MRF



BN1

$$MB(D) = \{A, C\}_{UGM} = \{A, C, B\}_{DAG}$$

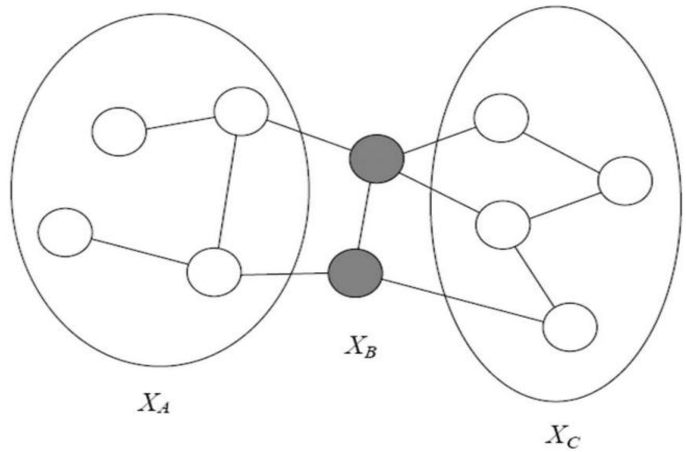
➢ 局部马尔科夫独立性

$$\mathrm{I}_l(H): \{X_i \perp V - \{X_i\} - MarkovBlanket(X_i)\}$$

无向图中的独立性

➢全局马尔科夫独立性

- 对于点集A,B,C，B **分离(separate)** A and C, 如果每一条从A到C的路都经过B



$$Sep(A;C|B)$$

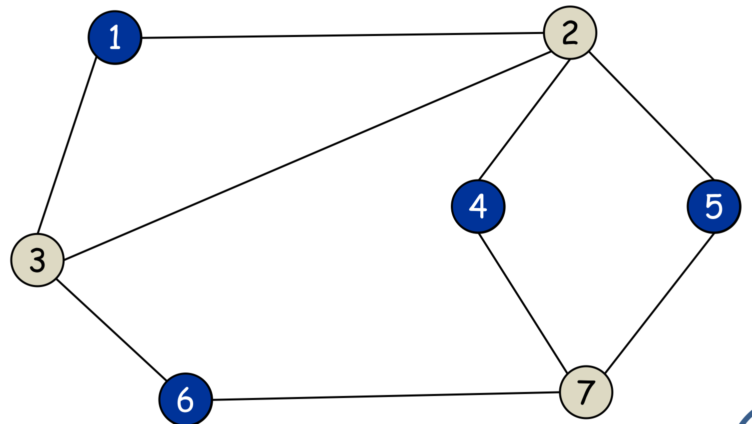
- *可靠性+完备性

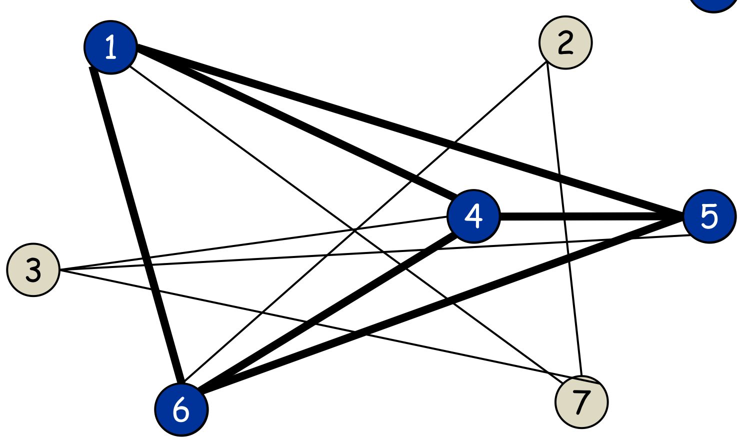
图上因子分解的Gibbs分布满足与图相关的独立性
图中的非独立性，在部分分布上可能也非独立

因子辖域选择

➢图中的独立性有什么用？

- 直观的，BN根据独立性进行因子分解
- MRF中的因子分解是根据团，那MRF的独立性就没用了?(生硬)
- **独立集和团：** **Independent Set $\leq_p$ Clique**





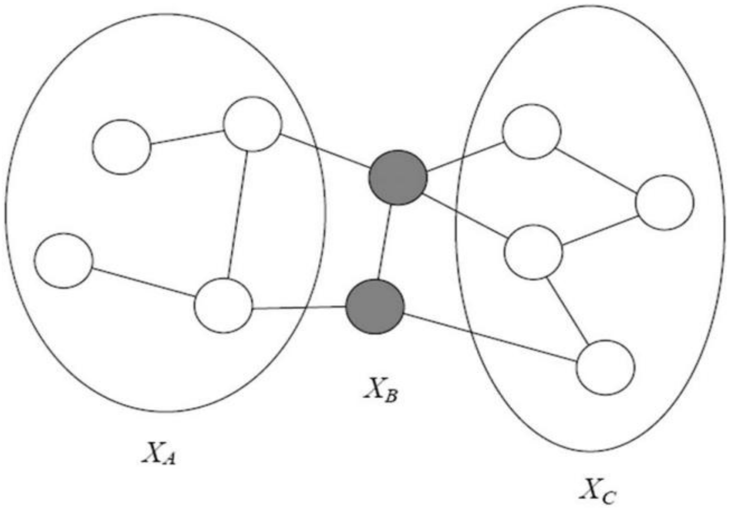
*因子为什么定义在团上

数据挖掘实验室
Data Mining Lab

- MRF根据独立性因子分解，但表现形式是团
- 图中，在给定点集B的情况下，点集A条件独立于点集C

图中任意的团在A∪B ，B ∪ C中

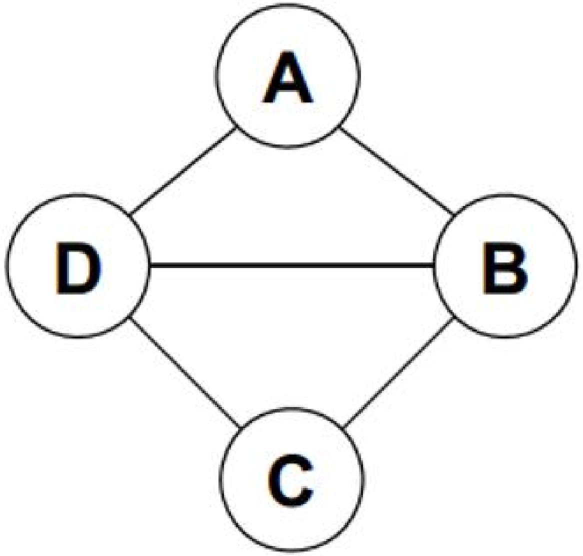
$$P(X_1,\dots,X_n) \propto \prod_{i\in X_A \cup X_B} \phi_i(D_i) \prod_{i\in X_C \cup X_B} \phi_i(D_i)$$



团的大小选择

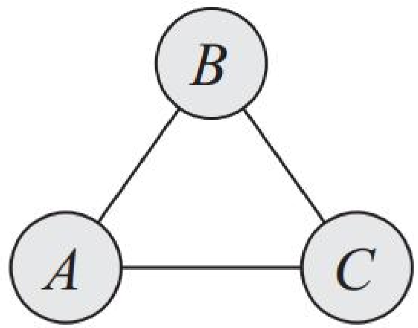
➢因子定义在团上

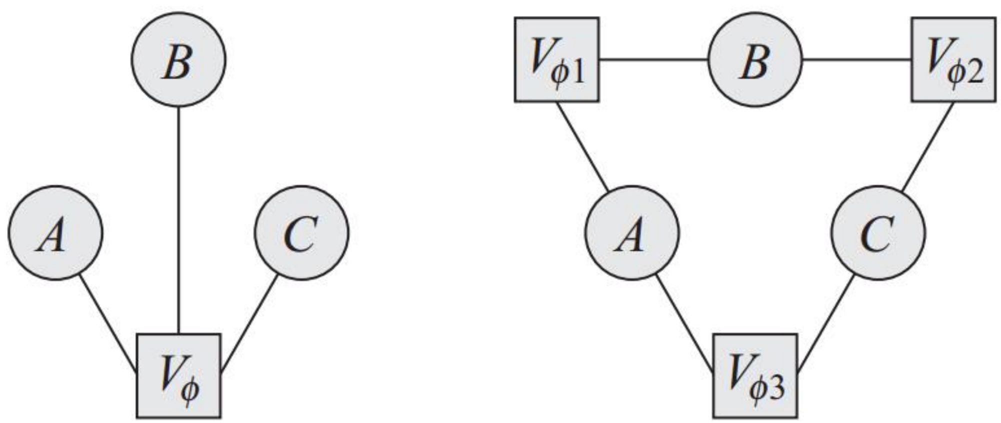
- 应该如何选团的大小呢？
- 团的大小对MRF有什么影响？



因子图

- 给定一个无向图G，可以将它表示两种节点组合而成的新图：变量节点和因子节点。
- 在因子图上，可以直观的看出团的辖域

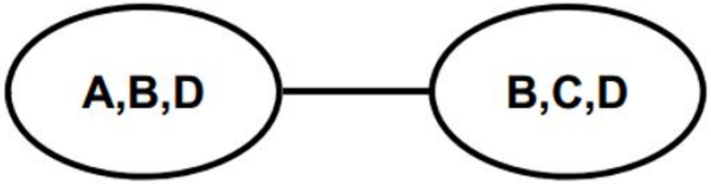




(a) (b)

团的大小选择

- 用最大团

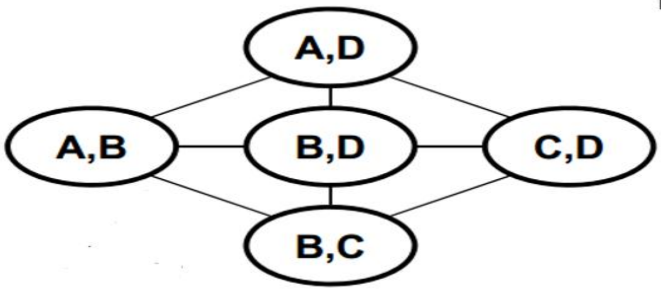


$\psi_c(\mathbf{x}_{124})$ $\psi_c(\mathbf{x}_{234})$

$$P'(x_1,x_2,x_3,x_4)=\frac{1}{Z}\psi_c(\mathbf{x}_{124})\times\psi_c(\mathbf{x}_{234})$$

$$Z=\sum_{x_1,x_2,x_3,x_4}\psi_c(\mathbf{x}_{124})\times\psi_c(\mathbf{x}_{234})$$

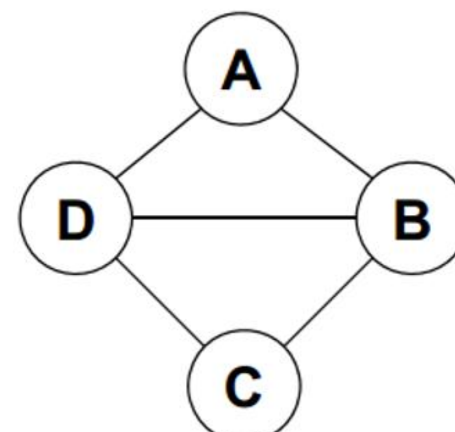
- 用比较小的团



$$\frac{1}{Z}\psi_{12}(\mathbf{x}_{12})\psi_{14}(\mathbf{x}_{14})\psi_{23}(\mathbf{x}_{23})\psi_{24}(\mathbf{x}_{24})\psi_{34}(\mathbf{x}_{34})$$

- 都试试

$$\frac{1}{Z}\psi_c(\mathbf{x}_{124})\times\psi_c(\mathbf{x}_{234})$$
$$\times\psi_{12}(\mathbf{x}_{12})\psi_{14}(\mathbf{x}_{14})\psi_{23}(\mathbf{x}_{23})\psi_{24}(\mathbf{x}_{24})\psi_{34}(\mathbf{x}_{34})$$
$$\times\psi_1(x_1)\psi_2(x_2)\psi_3(x_3)\psi_4(x_4)$$



团的大小

数据挖掘实验室
Data Mining Lab

➢最大团

- 忽略了更多的局部结构信息
- 表征空间的值更大（二维表能表示的信息，三维表一定能表示）
- 用更少的项来分解

➢用比较小的团

- Partition function更容易计算
- 便于做 Variable Elimination

➢用所有团

- …

这三种方式刻画的分布等价吗？

因子的定义

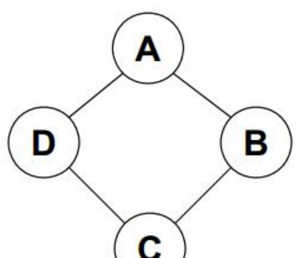
➢ **势函数**

- 正值函数φ_c
- 用来刻画团中各个节点之间的演化关系
- 一般采用指数形式

$$P(X)=\frac{1}{Z}\exp\left(-\sum \varnothing_c(X_C)\right)=\frac{1}{Z}\exp(-H(X))$$

- 其中∅'叫做能量函数(Energy Function)，能量越高，出现的可能性越低。 H(X)叫自由能量(Free Energy)

$$\varphi_\mathrm{c}=\exp(-\varnothing'(\mathrm{c}))$$



MRF

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

为什么用指数形式

- 保证正分布
- *把$\varphi_c = \exp(-f(c))$视为指数簇函数：

$$P(X|\eta) = H(X)\exp\left(T(X)^T\eta - A(\eta)\right)$$

- η叫natural parameter
- A(η)叫log-normalizer，是唯一一个不含有X的项，如果对原式子做积分等于1的话，可以反解出A就等于其他项在x上的积分，所以A也可以看做是一个partition function。
- T(X)叫Sufficient Statistics（充分统计量），它的意思是给定了T，我们不能再从X的概率分布中获得更多关于η的更多信息。

为什么用指数形式

数据挖掘实验室
Data Mining Lab

- 保证正分布
- *把$\varphi_c = \exp(-f(c))$视为指数簇函数，享受便利（MRF的求解一般比BN更难，为什么？）

$$P(X|\eta) = H(X)\exp\left(T(X)^T\eta - A(\eta)\right)$$

***指数簇的便利性**

- 矩生成
- MLE
- 共轭
- 变分推断

$$\frac{\partial^d A(\eta)}{\partial \eta^d} = E_{P(X|\eta)}[T(X)^d]$$

$$A'(\eta) = \frac{1}{n}\sum_{i=1}^{n} T(X_i)$$

$$\lambda = E_{Q(Z|\phi)}[\eta(Z,X)]$$

MRF

➢ **根据你的了解，请在空白处填上适当的语句。**

- MRF通过在 __**团c**__ 上，定义__**势函数ψ_C**__ 来刻画变量之间的相互关系，这种因子的表示方变量联合概率提供了紧凑表示，简化计算。形如：

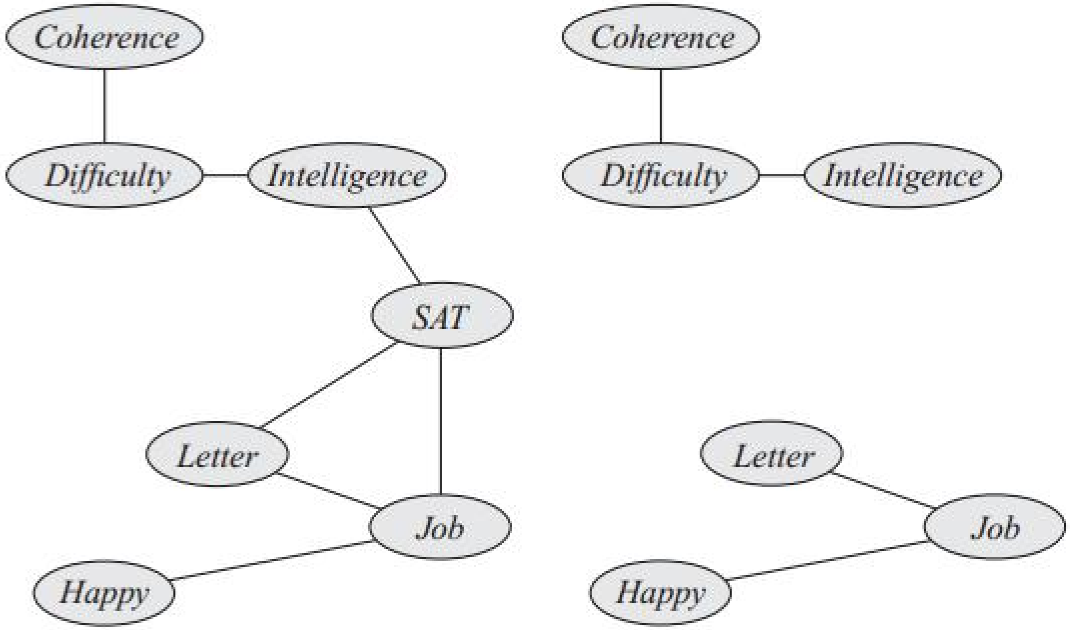
- 这个形式在统计上，称为对数线性模型（log-linear model）

$$P(X_1, \ldots, X_n) = \frac{1}{Z}\prod \psi_C(X_c)$$

$$Z = \sum \prod \psi_C(X_c)$$

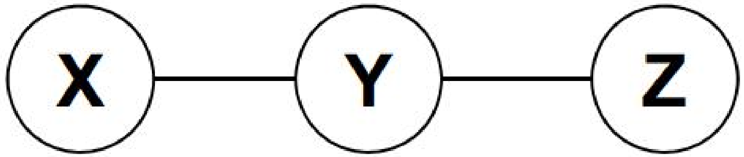
*约简的MRF

- 令H是节点集X上的MRF，若子集U⊆X已经被观察到，则H[U]是节点集W=X-U上的一个**约简的MRF**



团的势函数的解释

- 给定下图中三个变量



$$
\left.\begin{aligned}
P(X,Y,Z) &= P(Y)P(X|Y)P(Z|Y) \\
&= P(X,Y)P(Z|Y) \\
&= P(X|Y)P(Z,Y)
\end{aligned}\right\} \text{概率展开}
$$

$$
\propto \psi(x,y)\psi(y,z) \rightarrow \text{势函数乘积}
$$

在某些团上，势函数不能全部表示为边缘分布，或条件分布

*MRF与Gibbs

无向图

变量满足马尔科夫独立性

变量服从Gibbs分布

MRF

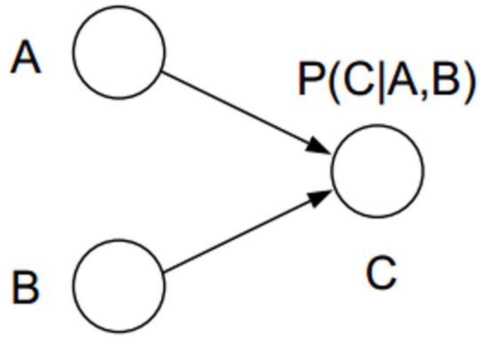
Hammersley-Clifford + 正分布

Gibbs Field

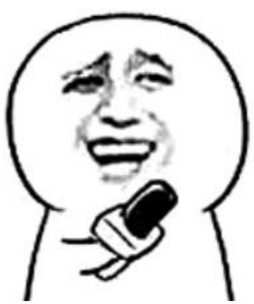
熟悉到陌生的BN与MRF

BN → MRF

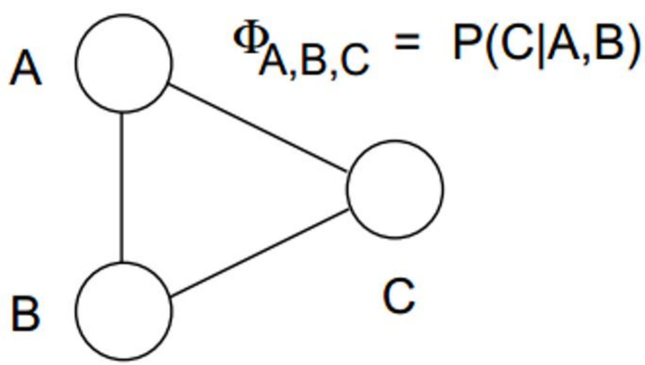
- 条件概率 → 因子势函数
- 考虑如图中A,B,C三个变量的条件概率



- 把上图的条件概率转化为因子势 → 不能，为什么？
- A,B,C都不在同一个团中！



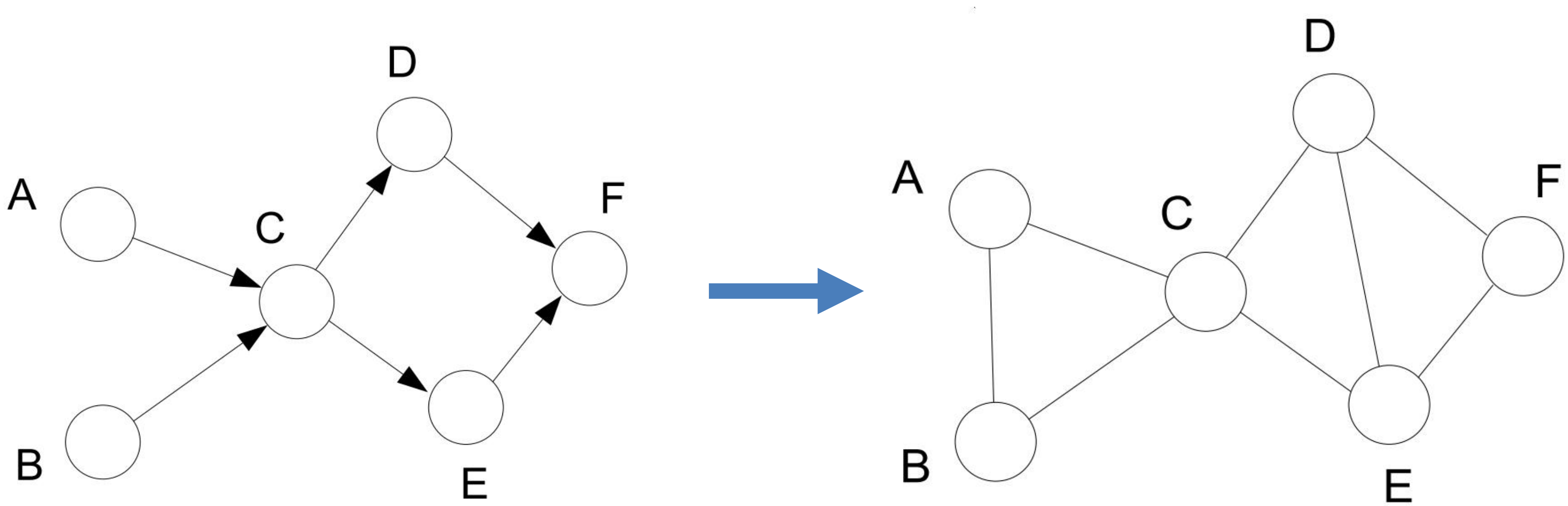
肿么办



BN → MRF

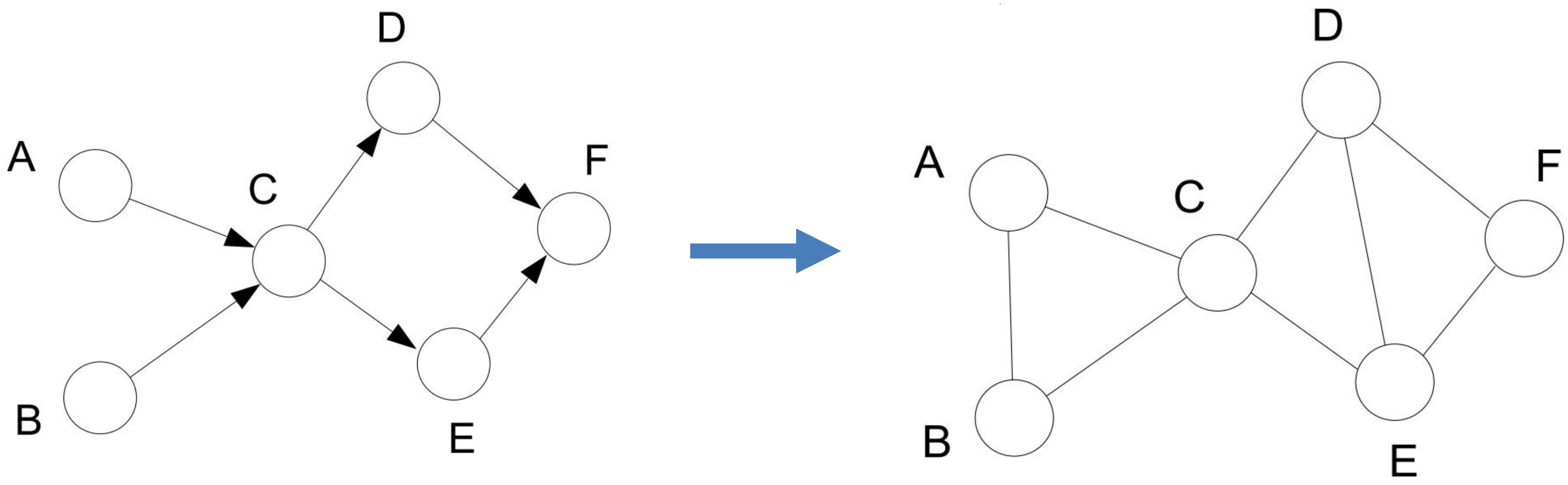
- Moral图：在一个团中的所有条件概率上定义因子势函数
- Moralization: “父母联姻” ，去除方向

$$
\begin{aligned}
&\mathrm{P}(A,B,C,D,E,F)\\
&= \underbrace{P(A)P(B)P(C|A,B)}_{\emptyset(A,B,C)}\underbrace{P(D|C)P(E|C)}_{\emptyset(C,D,E)}\underbrace{P(F|D,E)}_{\emptyset(D,E,F)}
\end{aligned}
$$



BN → MRF

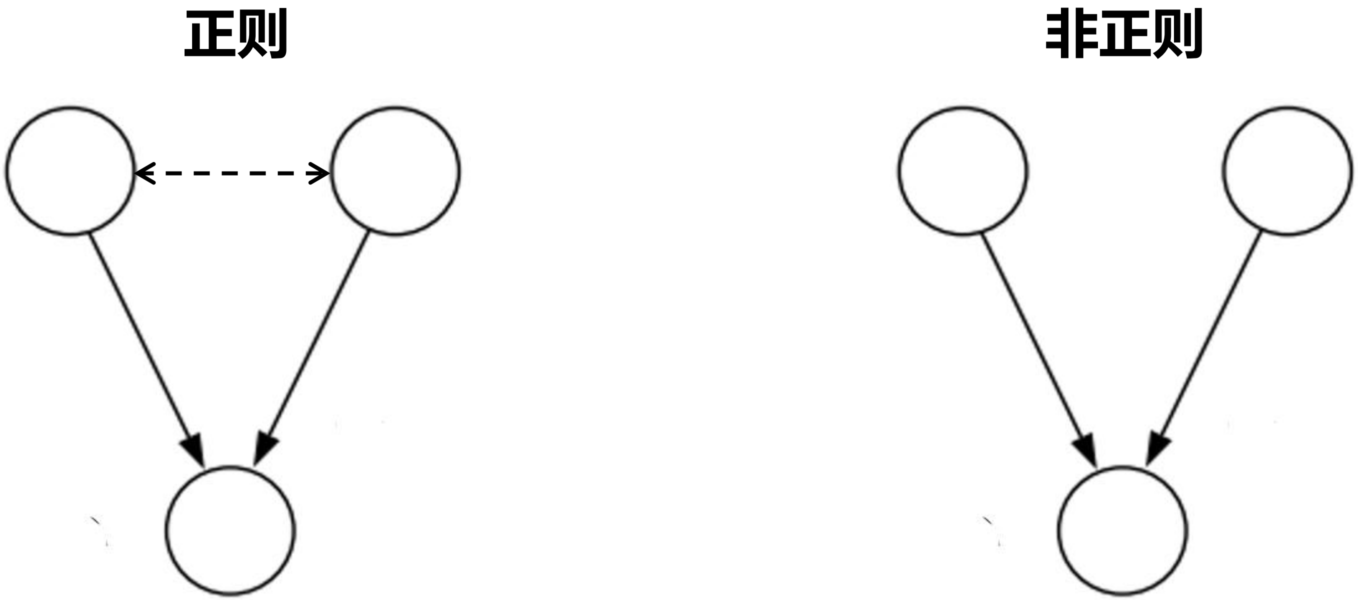
- 转化损失
  - 转化成MRF之后，AB不在边缘独立



*有向正则图可以在假定不丢失独立性的情况下转化为一个MRF，实际上，独立性的丢失都在于V-structure上。

BN补充

➢ 正则图（V-Structure）

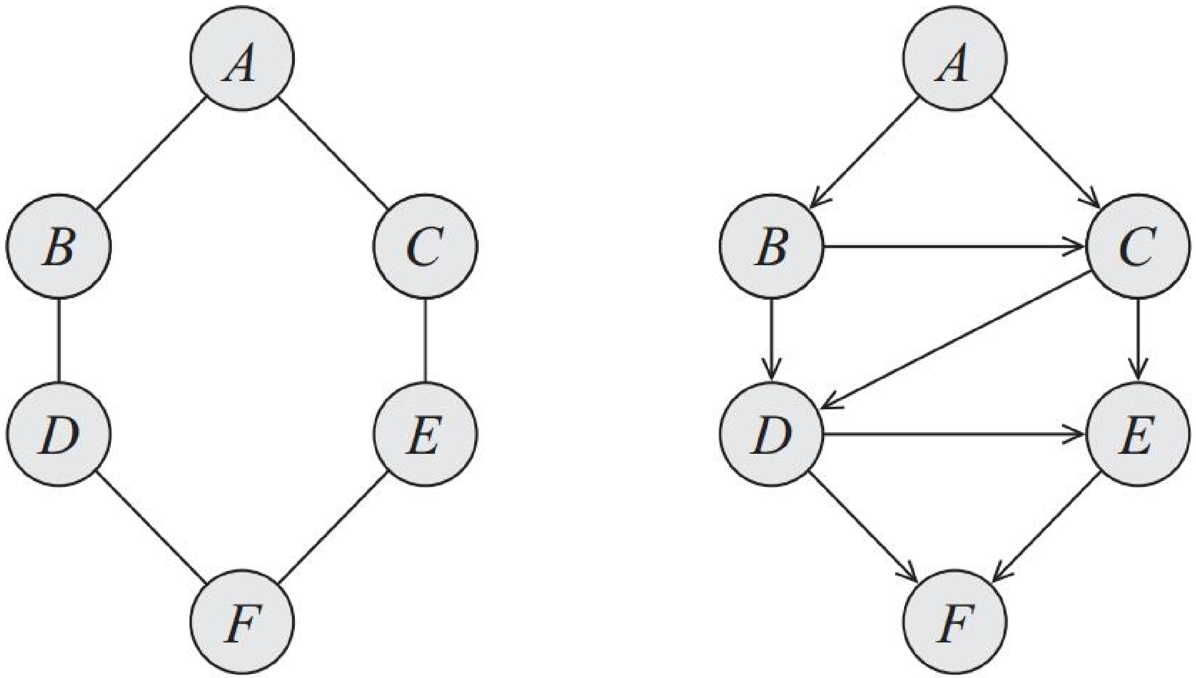


➢ 对于两个BN，如果满足

- 相同骨架 + 相同V-structure集 → I-等价
- 相同骨架 + 相同非正则V-structure集 ←→ I-等价

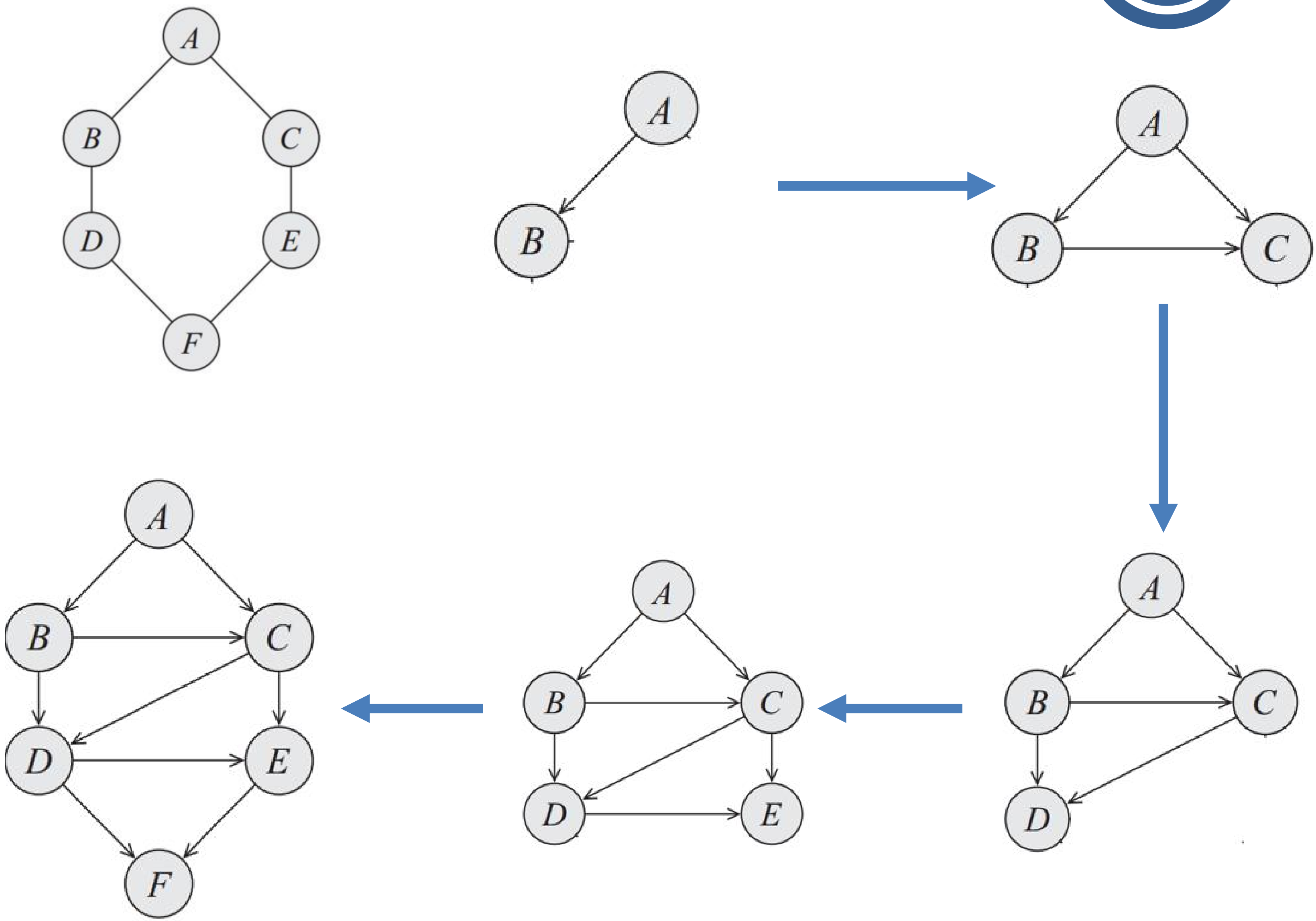
*MRF → BN

- 直观上就知道，这个很难
- 作为某个MRF的最小I-map的BN可能比原MRF复杂很多



- 需要添加很多边，构成三角形的，对应的MRF一定是弦图

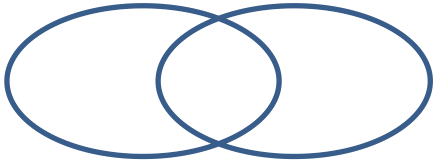
*MRF → BN



BN V.S. MRF

- **BN和MRF的区别？**

- 有向图与无向图
- 因果关系与对等联系(Affinity: 亲密性)
- 势函数不能直接表征概率
- 刻画独立性能力



- MRF更难求解 → 因为partition function
- 给定观察变量后，BN可能会引入新的依赖关系（V-Structure），而MRF只是对应约简的MRF
- 马尔科夫毯
- ……

Take-Home Message

MRF部分

- MRF的全局和局部独立性
- MRF通过因子乘积来传递变量间的相互影响
- MRF根据独立性因子分解，但表现形式是团
- 团不会破坏图结构导出的独立性
- 指数形式的势函数
- MRF与Gibbs
- MRF与BN的区别和相互转换
- Moralization和弦图

Thanks

By HC