BITS Pilani

Pilani | Dubai | Goa | Hyderabad

PROBABILISTIC GRAPHICAL MODEL

SESSION # 6 : UNDIRECTED GRAPHICAL MODEL

SEETHA PARAMESWARAN

seetha.p@pilani.bits-pilani.ac.in

The instructor is gratefully acknowledging the authors who made their course materials freely available online.

Table of Contents

1 Undirected Graphical Models

Scenario 1

$P(A/C) = P(A)$

$P(A/C, B, D) = P(A/B, D)$

- Four people; Alice, Bob, Charlie, Diana; go out for dinner in different groups of two.
- Alice goes out with Bob, Bob goes out with Charlie, Charlie with Diana, and Diana with Alice.
- Bob doesn't go with Diana, and Alice doesn't go with Charlie.
- Let's think about the probability of them ordering food of the same cuisine.
- From our social experience, we know that people interacting with each other may influence each others choice of food.
- Alice can influence Bob's choice of cuisine. Bob can influence Charlie's choice of cuisine. But Alice and Charlie wont agree.
- How can we represent this in Bayesian Network?

Misconception Example

$P(A C) = P(A) P(C)$

Alice $\perp$ *Charlie* | *Bob, Diana*

Bob $\perp$ *Diana* | *Alice, Charlie*

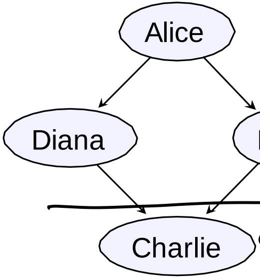
Scenario 1

sprinkler rain

Lawn is watered

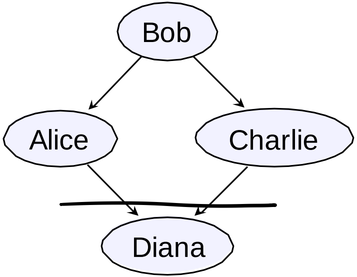
$$\boxed{Alice \perp Charlie \mid Bob,\ Diana} \qquad (1)$$

$$Bob \perp Diana \mid Alice,\ Charlie \qquad (2)$$



converging node

Satisfies Eq(1) but not Eqn(2).



Satisfies Eq(2) but not Eqn(1).

innovate achieve lead

What is the problem?

Scenario 1:

Bob and Diana are both serial nodes, so specifying them makes the paths from Alice to Charlie inactive

Thus $A \perp C \mid B, D$

innovate achieve lead

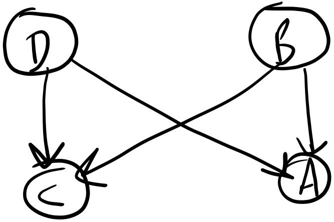
What is the problem?

Consider $B \perp D \mid A, C$

Here C is a converging node, so specifying it creates a path of influence between B and D

$\therefore B \not\perp D \mid A, C$

What's wrong here?



$A \perp C \mid B, D$

$B \not\perp D \mid A, C$

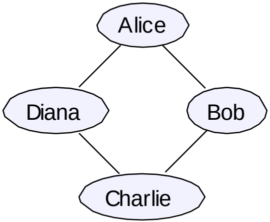
B, D marginally independent (i.e. $B \perp D \mid \phi$)

Scenario 1

- Directed models have a limitation that they cannot represent symmetric interactions.
- Undirected graphical model to encode influence flows in both directions.
- Example:

$$Alice \perp Charlie \mid Bob, \ Diana$$

$$Bob \perp Diana \mid Alice, \ Charlie$$



Markov Network

Definition

Markov network is an undirected graph, where

- the nodes represent the random variables and
- the dependencies or direct probabilistic interaction between these random variables are represented with undirected edges.

- No parent-child relationship.
- So we do not use CPD.
- Use factor to represent how likely it is for some states of a variable to agree with the states of other variables.

Parameterizing Markov Network

$\prod P(X_i / Pa(X_i))$ vs $\prod \phi_i(X_i)$

- Markov Networks are parameterized using factors.
- Factors help in symmetric parameterization of random variables.
- Factors capture the affinities between related variables.
- Factors do not represent the probability.
- Factors are not constrained to sum up to 1 or to be in the range [0,1].
- The parameterization of the Markov network defines the local interactions between directly related variables.
- The scope of a factor to be the set of random variables over which it is defined.

Factor

- A factor Φ is a function or a table that maps a set of random variables to a real value.

$$\Phi : Val(X_1, \ldots, X_n) \rightarrow \mathbb{R} \quad (3)$$

- The argument of the factor is called scope of the factor.

$$Scope : D = \{\underline{X_1, \ldots, X_n}\} \quad (4)$$

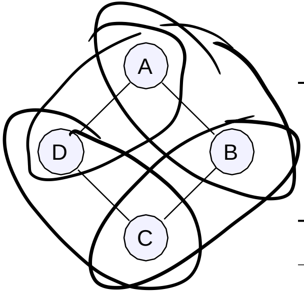
- Operations on a factor (Refer Session 3 for details)
  - Marginalize a factor φ whose scope is W with respect to a set of random variables X, sum out all the entries of X, to reduce its scope to $\{W - X\}$. .
  - Reduction of a factor φ whose scope is W to the context $X = x^i$ means removing all the entries from the factor where $X \neq x^i$. This reduces the scope to $\{W - X\}$.
  - Factor product refers to the product of factors φ_1 with a scope X and φ_2 with scope Y to produce a factor φ_3 with a scope $X \cup Y$.

Factor

4

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

A	B	$\varphi(A, B)$
a^0	b^0	90
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10



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

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

Queries using Factors

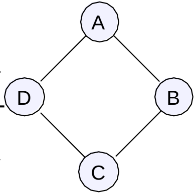
- Compute the probability corresponding to a^1, b^1, c^0, d^1.

$$\tilde{P}(a^1, b^1, c^0, d^1) = \varphi_1(a^1, b^1) \times \varphi_2(b^1, c^0) \times \varphi_3(c^0, d^1) \times \varphi_4(d^1, a^1)$$
$$= 10 * 1 * 100 * 100 = 700,000$$

Factor Product

D	A	$\phi_4(D, A)$
d^0	a^0	80
d^0	a^1	60
d^1	a^0	20
d^1	a^1	10

C	D	$\phi_3(C, D)$
c^0	d^0	10
c^0	d^1	1
c^1	d^0	100
c^1	d^1	90



A	B	$\phi_1(A, B)$
a^0	b^0	90
a^0	b^1	100
a^1	b^0	1
a^1	b^1	10

B	C	$\phi_2(B, C)$
b^0	c^0	10
b^0	c^1	80
b^1	c^0	70
b^1	c^1	30

A	B	C	D	$\tilde{P}(A, B, C, D) = \Phi(A, B, C, D)$	
a^0	b^0	c^0	d^0	$90 \cdot 10 \cdot 10 \cdot 80$	=720, 000
a^0	b^0	c^0	d^1	$90 \cdot 10 \cdot 1 \cdot 20$	=18, 000
a^0	b^0	c^1	d^0	$90 \cdot 80 \cdot 100 \cdot 80$	=57600, 000
a^0	b^0	c^1	d^1	$90 \cdot 80 \cdot 90 \cdot 20$	=12960, 000
a^0	b^1	c^0	d^0	$100 \cdot 70 \cdot 10 \cdot 80$	=5600, 000
a^0	b^1	c^0	d^1	$100 \cdot 70 \cdot 1 \cdot 20$	=140, 000
a^0	b^1	c^1	d^0	$100 \cdot 30 \cdot 100 \cdot 80$	=24000, 000
a^0	b^1	c^1	d^1	$100 \cdot 30 \cdot 90 \cdot 20$	=5400, 000
a^1	b^0	c^0	d^0	$1 \cdot 10 \cdot 10 \cdot 60$	=6, 000
a^1	b^0	c^0	d^1	$1 \cdot 10 \cdot 1 \cdot 10$	=100
a^1	b^0	c^1	d^0	$1 \cdot 80 \cdot 100 \cdot 60$	=480, 000
a^1	b^0	c^1	d^1	$1 \cdot 80 \cdot 90 \cdot 10$	=72, 000
a^1	b^1	c^0	d^0	$10 \cdot 70 \cdot 10 \cdot 60$	=420, 000
a^1	b^1	c^0	d^1	$10 \cdot 70 \cdot 1 \cdot 10$	=70, 000
a^1	b^1	c^1	d^0	$10 \cdot 30 \cdot 100 \cdot 60$	=1800, 000
a^1	b^1	c^1	d^1	$10 \cdot 30 \cdot 90 \cdot 10$	=270, 000

Factor Product

- Factor Product

$$\tilde{P}(A, B, C, D) = \varphi_1(A, B) \times \varphi_2(B, C) \times \varphi_3(C, D) \times \varphi_4(D, A) \qquad (5)$$

is un-normalized. It is not a probability distribution.

- Normalize $\tilde{P}(A, B, C, D)$ using partition function Z. Z is called the partition function and is a function of the parameters.

$$Z = \sum_{A,B,C,D} \tilde{P}(A, B, C, D) \qquad (6)$$

Sum over all possible assignments to all variables

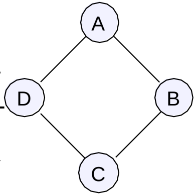
- Normalized factor product

$$\boxed{P(A, B, C, D) = \frac{1}{Z}\tilde{P}(A, B, C, D)} \qquad (7)$$

Normalized Factor Product

D	A	$\phi_4(D, A)$
d^0	a^0	80
d^0	a^1	60
d^1	a^0	20
d^1	a^1	10

A	B	$\phi_1(A, B)$
a^0	b^0	90
a^0	b^1	100
a^1	b^0	1
a^1	b^1	10



C	D	$\phi_3(C, D)$
c^0	d^0	10
c^0	d^1	1
c^1	d^0	100
c^1	d^1	90

B	C	$\phi_2(B, C)$
b^0	c^0	10
b^0	c^1	80
b^1	c^0	70
b^1	c^1	30

A	B	C	D	$\tilde{P}(A, B, C, D)$	$P(A, B, C, D)$
a^0	b^0	c^0	d^0	720,000	0.0055
a^0	b^0	c^0	d^1	18,000	0.0001
a^0	b^0	c^1	d^0	57600,000	0.4365
a^0	b^0	c^1	d^1	12960,000	0.0982
a^0	b^1	c^0	d^0	5600,000	0.0424
a^0	b^1	c^0	d^1	140,000	0.0011
a^0	b^1	c^1	d^0	24000,000	0.1819
a^0	b^1	c^1	d^1	5400,000	0.0409
a^1	b^0	c^0	d^0	6,000	0.0000
a^1	b^0	c^0	d^1	100	0.0000
a^1	b^0	c^1	d^0	480,000	0.0036
a^1	b^0	c^1	d^1	72,000	0.0005
a^1	b^1	c^0	d^0	420,000	0.0318
a^1	b^1	c^0	d^1	70,000	0.0005
a^1	b^1	c^1	d^0	1800,000	0.1364
a^1	b^1	c^1	d^1	270,000	0.0205
				109493,100	1.0

Queries using Factor Product

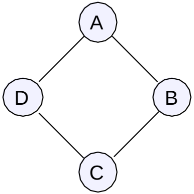
- Compute the probability of B.
  Marginalize wrt A,C,D

$$P(b^1) = 0.4555$$
$$P(b^0) = 0.5445$$

- Compute the probability of B agreeing with C given c^0.

$$P(b^1 | c^0) = 0.0759$$

Factors vs Probability Distribution



A	B	$\phi(A, B)$
a^0	b^0	90
a^0	b^1	100
a^1	b^0	1
a^1	b^1	10

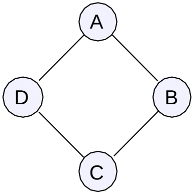
factor

Marginal Probability of A and B

prob dist

A	B	C	D	$P(A, B, C, D)$	$P_\varphi(A, B)$
a^0	b^0	c^0	d^0	0.0055	
a^0	b^0	c^0	d^1	0.0001	
a^0	b^0	c^1	d^0	0.4365	
a^0	b^0	c^1	d^1	0.0982	0.5403
a^0	b^1	c^0	d^0	0.0424	
a^0	b^1	c^0	d^1	0.0011	
a^0	b^1	c^1	d^0	0.1819	
a^0	b^1	c^1	d^1	0.0409	0.2663
a^1	b^0	c^0	d^0	0.0000	
a^1	b^0	c^0	d^1	0.0000	
a^1	b^0	c^1	d^0	0.0036	
a^1	b^0	c^1	d^1	0.0005	0.0042
a^1	b^1	c^0	d^0	0.0318	
a^1	b^1	c^0	d^1	0.0005	
a^1	b^1	c^1	d^0	0.1364	
a^1	b^1	c^1	d^1	0.0205	0.1892

Factors vs Probability Distribution



A	B	$\phi_1(A,B)$
a^0	b^0	$\boxed{90}$
a^0	b^1	$\boxed{100}$
a^1	b^0	1
a^1	b^1	10

A	B	$P_\varphi(A,B)$
a^0	b^0	$\boxed{0.5403}$
a^0	b^1	$\boxed{0.2663}$
a^1	b^0	0.0042
a^1	b^1	0.1892

There is no natural mapping between factors and probability distribution.

Marginal Probability of A and B

A	B	C	D	$P(A,B,C,D)$	$P_\varphi(A,B)$
a^0	b^0	c^0	d^0	0.0055	
a^0	b^0	c^0	d^1	0.0001	
a^0	b^0	c^1	d^0	0.4365	
a^0	b^0	c^1	d^1	0.0982	0.5403
a^0	b^1	c^0	d^0	0.0424	
a^0	b^1	c^0	d^1	0.0011	
a^0	b^1	c^1	d^0	0.1819	
a^0	b^1	c^1	d^1	0.0409	0.2663
a^1	b^0	c^0	d^0	0.0000	
a^1	b^0	c^0	d^1	0.0000	
a^1	b^0	c^1	d^0	0.0036	
a^1	b^0	c^1	d^1	0.0005	0.0042
a^1	b^1	c^0	d^0	0.0318	
a^1	b^1	c^0	d^1	0.0005	
a^1	b^1	c^1	d^0	0.1364	
a^1	b^1	c^1	d^1	0.0205	0.1892

Factorization and Independencies

$\psi_5(B, D, A)\ \psi_6(B, D, C)$

- $P \models (B \perp D \mid A, C)$ should have a decomposition

$G(A, C, D)$

$F(A, B, C)$

$$P = \frac{1}{Z}[\varphi_1(A, B) \times \varphi_2(B, C)] \times \varphi_3(C, D) \times \varphi_4(D, A)$$

B and D are separated given $\boxed{\text{A and C.}}$

$P = F(B)\,G(D)$

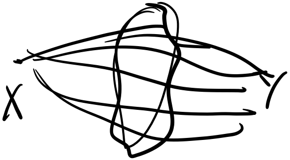
- $P \models (A \perp C \mid B, D)$ should have a decomposition

$F(A)$ $G(C)$

$$P = \frac{1}{Z}[\varphi_4(D, A) \times \varphi_1(A, B)] \times \varphi_2(B, C) \times \varphi_3(C, D)$$

A and C are separated given B and D.

Factorization and Independencies



$$P \models (X \perp Y \mid Z) \quad \text{if and only if} \quad P = \boxed{\varphi_1(X, Z)\varphi_2(Y, Z)} \tag{8}$$

contains

- Independence properties of the distribution P correspond directly to separation properties in the graph over which P factorizes.

Factors can be misleading

a^0	b^0	c^0	d^0	0.04
a^0	b^0	c^0	d^1	0.04
a^0	b^0	c^1	d^0	0.04
a^0	b^0	c^1	d^1	4.1×10^{-6}
a^0	b^1	c^0	d^0	6.9×10^{-5}
a^0	b^1	c^0	d^1	6.9×10^{-5}
a^0	b^1	c^1	d^0	0.69
a^0	b^1	c^1	d^1	6.9×10^{-5}
a^1	b^0	c^0	d^0	1.4×10^{-5}
a^1	b^0	c^0	d^1	0.14
a^1	b^0	c^1	d^0	1.4×10^{-5}
a^1	b^0	c^1	d^1	1.4×10^{-5}
a^1	b^1	c^0	d^0	1.4×10^{-5}
a^1	b^1	c^0	d^1	0.014
a^1	b^1	c^1	d^0	0.014
a^1	b^1	c^1	d^1	0.014

Joint Distribution for the Misconception Example

Factors for Misconception Example

$\phi_1(A, B)$

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

$\phi_2(B, C)$

b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

$\phi_3(C, D)$

c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_4(D, A)$

d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

The factor $\phi_1(A, B)$ suggests that A and B are mostly <u>in agreement</u>.

Marginal Distribution for A, B in Misconception Example

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

Here we see that A and B are mostly in <u>disagreement</u> unlike in $\phi_1(A, B)$

This is because of the influence of the other Factors on the distribution

Influence of other factors

$\phi_3(C, D)$ asserts that Charles and Debbie disagree

$\phi_2(B, C)$ asserts that Bob and Charles agree.

$\phi_4(D, A)$ asserts that Debbie and Alice agree

The implication of the above is that Alice and Bob disagree.

$B - C \not- D - A$

Gibbs Distribution

Definition

A distribution P_Φ is called a Gibbs distribution parameterized by a set of factors $\Phi = \{\varphi_1(D_1), \ldots, \varphi_k(D_k)\}$ if it can be expressed as product of the factors.

$$P_\Phi(X_i, \ldots, X_n) = \frac{1}{Z_\Phi}[\varphi_1(D_1) \times \ldots \times \varphi_k(D_k)]$$

$$\tilde{P}(X_i, \ldots, X_n) = \prod_{i=1}^{k} \varphi_i(D_i) \quad (9)$$

$$\boxed{Z_\Phi = \sum_{X_i,\ldots,X_n} \tilde{P}(X_i, \ldots, X_n)} \quad (10)$$

$$P_\Phi(X_i, \ldots, X_n) = \frac{1}{Z_\Phi} \tilde{P}(X_i, \ldots, X_n) \quad (11)$$

Gibbs Distribution

$\prod P(X_i / Pa(X_i))$

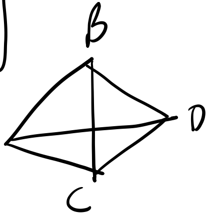
Definition

A distribution P_Φ with $\Phi = \{\varphi_1(D_1), \ldots, \varphi_k(D_k)\}$ factorizes over a Markov Network H if each D_k is a complete subgraph of H.

- The factors that parameterize a Markov network are often called clique potentials.
- Reduce the number of factors in the parameterization by allowing factors only for maximal cliques.
- Let $C_1, \ldots, C_k$ be the cliques in H.
- Parameterize P using a set of factors $\varphi_1(C_1), \ldots, \varphi_l(C_l)$.

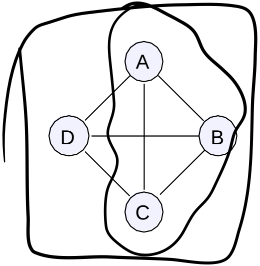
$\varphi_1(A, B, C, D)$

$\varphi_1(A, B)\, \varphi_2(B, D)\, \varphi_3(B, A, C)$



Gibbs Distribution Example

$P_1 \neq \phi_1(A, B)\, \phi_2(B, C)\, \phi_3(C, D) \ldots$



- Cliques (Option 1): $\{A, B\}, \{B, C\}, \{C, D\}, \{D, A\}, \{D, B\}, \{A, C\}$
- Cliques (Option 2): $\{A, B, D\}, \{B, C, D\}$
- Cliques (Option 3): $\{A, B, C\}, \{A, C, D\}$

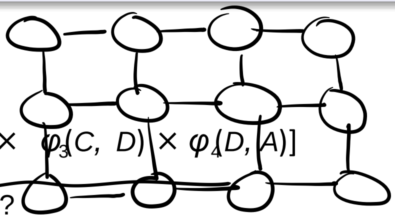
$\varphi(A, B) \leftarrow$ too many parameters not needed

Pairwise Markov Network

Definition

Pairwise Markov Network is an undirected graph whose nodes $X_i, \ldots, X_n$ and edges $X_i - X_j$ are associated with a factor $\varphi_{ij}(X_i, X_j)$.

- A subclass of Markov networks.
- Eg:



$$P(A, B, C, D) = \frac{1}{Z}\left[\varphi_1(A, B) \times \varphi_2(B, C) \times \varphi_3(C, D) \times \varphi_4(D, A)\right]$$

- How many parameters for n RV with d values each?

$$\text{Number of parameters in Pairwise Markov Network} = O(n^2 d^2) \quad (13)$$

$P(X_1, X_2, \ldots X_n) = \varphi_1(X_1, X_2) \times \varphi_2(X_2, X_3) \times \varphi_3(X_3, X_4) \ldots \varphi_{n-1}(X_{n-1}, X_n)$

Induced Markov Network

Definition

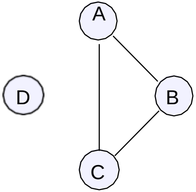
For a set of factors φ_i, with a scope D_i, the Induced Markov Network H_Φ, has an edge between a pair of variables X_i and X_j whenever there exists a factor $\varphi_m \in \Phi$ such that $X_i, X_j \in D_m$.

- X and Y will have an undirected edge
  - if they appear together in some factor φ
  - if there exists a factor $\varphi(X, Y)$.

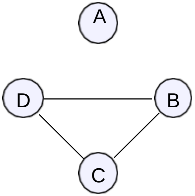
Induced Markov Network

Consider 4 RVs A,B,C,and D. The factor and its induced Markov Network is given below.

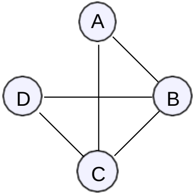
$\varphi_1(A, B, C)$



$\varphi_2(B, C, D)$



$\Phi = \varphi_1(A, B, C) \times \varphi_2(B, C, D)$



P factorizes H

Definition

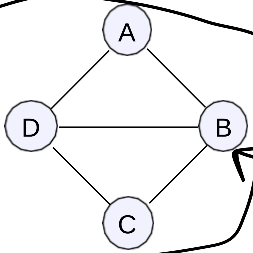
Gibbs distribution P factorizes a Markov Network H if there exists $\Phi = \{\varphi_1(D_1), \ldots \varphi_k(D_k)\}$ such that

- $P = P_\Phi$, normalized product of factors φ_i
- H is the induced graph for Φ.

$\frac{1}{Z}\tilde{P}$

P factorizes H

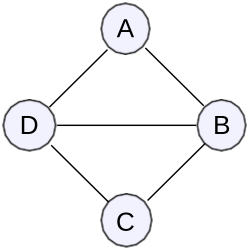
- From an induced Markov network H, we cannot read the factorization P_Φ from the graph, as there can be multiple possible factorizations.



- $\varphi_1(A, B, D)$, $\varphi_2(B, C, D)$
- $\varphi_5(A, B)$, $\varphi_6(B, C)$, $\varphi_7(C, D)$, $\varphi_8(D, A)$, $\varphi_9(A, B$

$\varphi_{10}(B, D)$

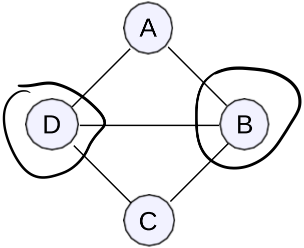
Flow of Influence



- $\varphi_1(A, B, D)$, $\varphi_2(B, C, D)$
- $\varphi_5(A, B)$, $\varphi_6(B, C)$, $\varphi_7(C, D)$, $\varphi_8(D, A)$, $\varphi_9(B, D)$
- When can B influence D?
- When can A influence C?

D-sep H-sep

Flow of Influence



- $\varphi_1(A, B, D)$, $\varphi_2(B, C, D)$
- $\varphi_5(A, B)$, $\varphi_6(B, C)$, $\varphi_7(C, D)$, $\varphi_8(D, A)$, $\varphi_9(B, D)$
- When can B influence D?
  - Direct influence
  - $\varphi_1(A, B, D)$
  - $\varphi_9(B, D)$
- When can A influence C?
  - Indirect influence
  - Through B or D
  - $\varphi_1(B, C, D)$
  - $\varphi_1(A, B, D)\varphi_2(B, C, D)$
  - $\varphi_5(A, B)$, $\varphi_6(B, C)$
  - $\varphi_7(C, D)$, $\varphi_8(D, A)$

A, D, C, A, ~~B~~, ~~B~~, C

A, B, D, C, A, B, C

Flow of Influence

- Parameterization of the distributions are different.
- The trails in the graph through which influence can flow are the same.
- Active trails depend only on the graph structure.

References

1. Probabilistic Graphical Models: Principles and Techniques by Daphne Koller and Nir Friedman. MIT Press. 2009
2. Artificial Intelligence: A Modern Approach (3rd Edition) by Stuart Russell, Peter Norvig
3. Mastering Probabilistic Graphical Models using Python by Ankur Ankan, Abhinash Panda. Packt Publishing 2015.
4. Learning in Graphical Models by Michael I. Jordan. MIT Press. 1999

Thank You !!!