

# Графические модели: ориентированное и неориентированное представление

Александр Адуенко

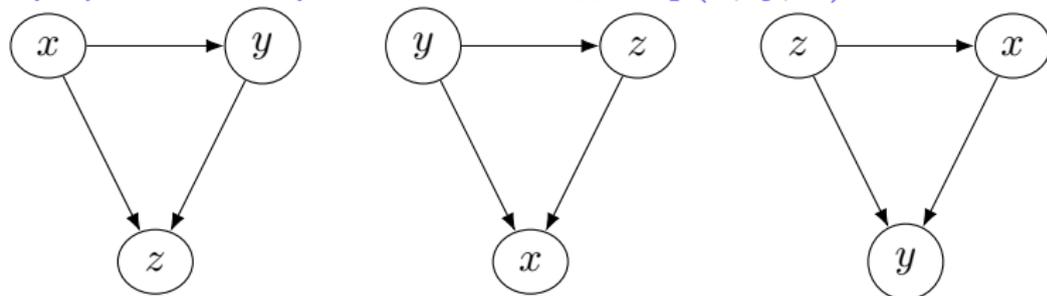
7е марта 2023

- EM-алгоритм. Использование EM-алгоритма для отбора признаков в байесовской линейной регрессии.
- Вариационный EM-алгоритм и его использование для вывода в смеси моделей линейной регрессии.
- Ориентированные графические модели и их представление plate notation. Критерий условной независимости D-separation.

**Идея:** Представим совместное распределение переменных в виде графа.

**Пример:**  $p(x, y, z) = p(x)p(y|x)p(z|x, y)$ .

Графическая вероятностная модель  $p(x, y, z)$



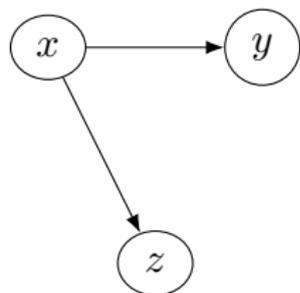
**Вопрос 1:** Чему соответствуют представления на средней и правой картинке?

$$p(x_1, \dots, x_K) = p(x_1)p(x_2|x_1) \cdot \dots \cdot p(x_K|x_1, \dots, x_{K-1}).$$

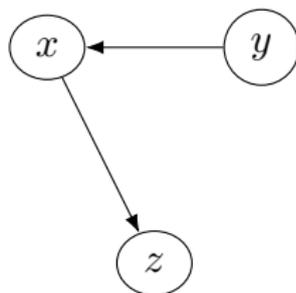
**Вопрос 2:** Для каких распределений выполнено разложение выше?

**Вопрос 3:** Какое представление получается для  $p(x_1, \dots, x_K)$  при таком разложении?

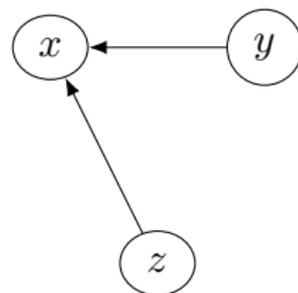
Идея: Представим совместное распределение переменных в виде графа.



Граф 1



Граф 2



Граф 3

Вопрос: Одинаковые ли совместные распределения соответствуют графическим представлениям выше?

Граф 1:  $p(x, y, z) = p(x)p(y|x)p(z|x)$ ;

Граф 2:  $p(x, y, z) = p(y)p(x|y)p(z|x)$ ;

Граф 3:  $p(x, y, z) = p(z)p(y)p(x|z, y)$ .

# Смесь моделей линейной регрессии

## Вероятностная модель генерации данных

- Веса моделей в смеси  $\pi$  получены из априорного распределения  $p(\pi|\mu)$ ;
- Векторы параметров моделей  $\mathbf{w}_k$  получены из нормального распределения  $p(\mathbf{w}_k|\mathbf{A}_k) = \mathcal{N}(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k^{-1})$ ,  $k = 1, \dots, K$ ;
- Для каждого объекта  $\mathbf{x}_i$  выбрана модель  $f_{k_i}$ , которой он описывается, причем  $p(k_i = k) = \pi_k$ ;
- Для каждого объекта  $\mathbf{x}_i$  целевая переменная  $y_i$  определена в соответствии с моделью  $f_{k_i}$ :  $y_i \sim \mathcal{N}(y_i|\mathbf{w}_{k_i}^\top \mathbf{x}_i, \sigma_{k_i}^2)$ .

## Совместное правдоподобие модели

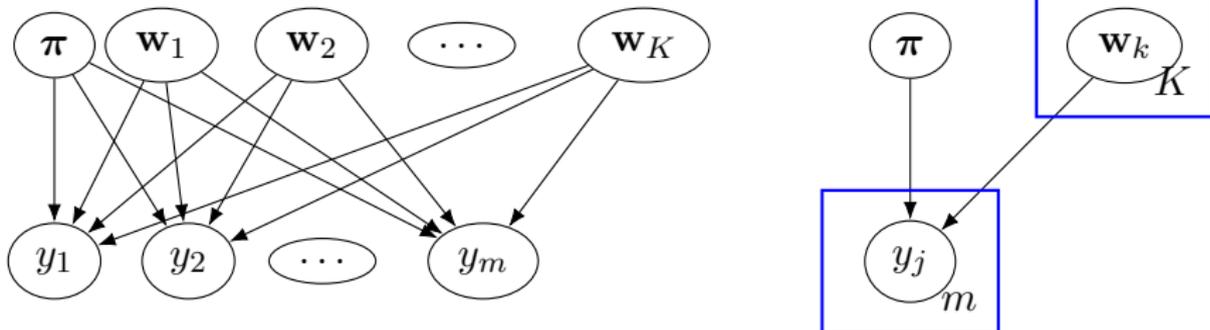
$$p(\mathbf{y}, \mathbf{w}_1, \dots, \mathbf{w}_K, \pi | \mathbf{X}, \mathbf{A}_1, \dots, \mathbf{A}_K, \sigma_1^2, \dots, \sigma_K^2, \mu) = p(\pi|\mu) \prod_{k=1}^K \mathcal{N}(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k^{-1}) \prod_{i=1}^m \left( \sum_{l=1}^K \pi_l \mathcal{N}(y_i|\mathbf{w}_l^\top \mathbf{x}_i, \sigma_l^2) \right).$$

Введем матрицу скрытых переменных  $\mathbf{Z} = \|z_{ik}\|$ , где  $z_{ik} = 1 \iff k_i = k$ .

$$p(\mathbf{y}, \mathbf{w}_1, \dots, \mathbf{w}_K, \pi, \mathbf{Z} | \mathbf{X}, \mathbf{A}_1, \dots, \mathbf{A}_K, \sigma_1^2, \dots, \sigma_K^2, \mu) = p(\pi|\mu) \prod_{k=1}^K \mathcal{N}(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k^{-1}) \prod_{i=1}^m \prod_{l=1}^K \left( \pi_l \mathcal{N}(y_i|\mathbf{w}_l^\top \mathbf{x}_i, \sigma_l^2) \right)^{z_{il}}.$$

# Представление смеси моделей в виде графа

Граф 3:  $p(x, y, z) = p(z)p(y)p(x|z, y)$ .



**Вопрос 1:** Как в представлении учитывается то, что смесь составлена из моделей линейной регрессии?

**Вопрос 2:** Как учесть, что  $p(\pi) = \text{Dir}(\mu)$ ?

**Вопрос 3:** Как указать наличие наблюдаемого признакового описания  $x_1, \dots, x_m$  и гиперпараметров модели  $\mathbf{A}, \sigma^2$ ?

**Вопрос 4:** Зачем нам графическое представление вероятностных моделей?

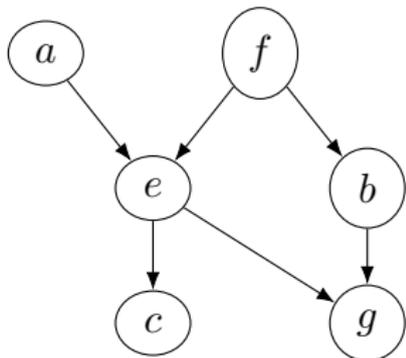
# Критерий условной независимости D-separation

Рассмотрим две группы переменных  $A$ ,  $B$  и проверим их условную независимость при условии группы переменных  $C$ .

**D-separation.** Группы переменных  $A$  и  $B$  условно независимы, если все неориентированные пути из  $A$  в  $B$  заблокированы  $C$ .

Путь из вершины  $a$  в вершину  $b$  называется заблокированным набором вершин  $C$ , если выполнено хотя бы одно из условий

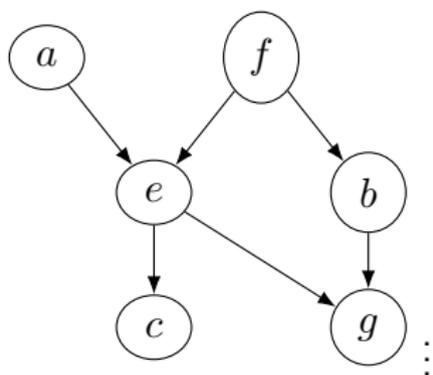
- Стрелки на пути встречаются перед-хвост или хвост-хвост в вершине  $c \in C$ ;
- Стрелки на пути встречаются перед-перед в вершине  $x$ , такой, что  $x \notin C$  и все ее ориентированные потомки  $y \notin C$ .



Пути из  $a$  в  $b$ :  $a \rightarrow e \rightarrow g \leftarrow b$ ;  $a \rightarrow e \leftarrow f \rightarrow b$ .

Вопрос: Зависимы ли переменные  $a$  и  $b$ ?

Условия условной независимости:

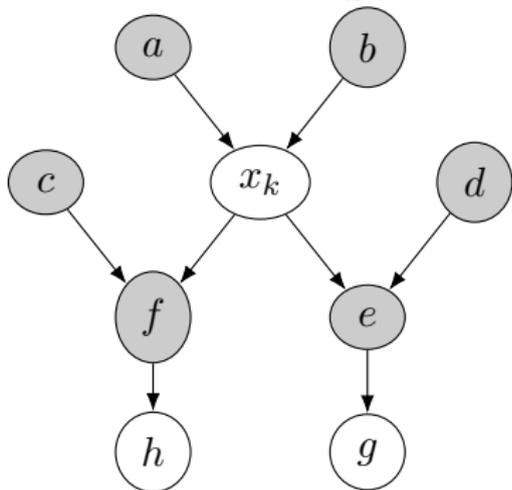


- $a \perp b | \emptyset$ ;
- $a \perp f | \emptyset$ ;
- $a \not\perp f | e$ ;
- $a \not\perp c | \emptyset$ ;
- $a \perp c | e$ ;



**Утверждение:** Множество распределений, представимых ориентированным графом заданного вида, совпадает с множеством распределений, удовлетворяющим всем условиям условной независимости, им порожденным.

$$p(x_1, \dots, x_K) = \prod_k p(x_k | Pa_k).$$



$$p(x_k | \mathbf{x}_{j \neq k}) = \frac{\prod_l p(x_l | Pa_l)}{\int \prod_l p(x_l | Pa_l) dx_k}.$$

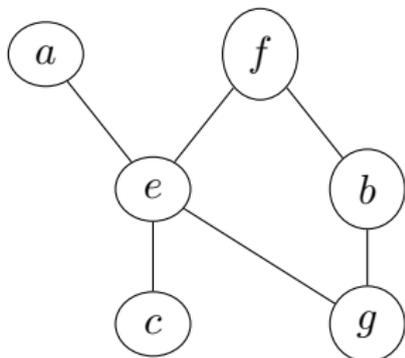
$$p(x_1, \dots, x_K) = p(a)p(b)p(x_k|a, b)p(c)p(d)p(e|d, x_k)p(f|c, x_k)p(g|e)p(h|f).$$

$$p(x_k | \mathbf{x}_{j \neq i}) = \frac{p(x_k|a, b)p(e|d, x_k)p(f|c, x_k)}{\int p(x_k|a, b)p(e|d, x_k)p(f|c, x_k) dx_k}.$$

Вопрос: От каких переменных действительно зависит  $x_k | \mathbf{x}_{j \neq k}$ ?

# Неориентированные графические модели

Неориентированные графические модели  $\equiv$  марковские случайные поля.



**Идея:** Определить критерий условной независимости в терминах разделимости, не  $d$ -разделимости.

**Вопрос:** Независимы ли  $a$  и  $b$

- Безусловно?
- При условии  $e$ ?
- При условии  $f$ ?
- При условии  $f, g$ ?

**Вопрос:** Что есть марковское «одеяло» для неориентированных графических моделей?

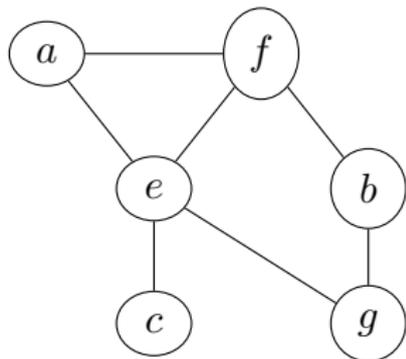
**Замечание:** Если  $x_i$  и  $x_j$  не соединены ребром, то они независимы при условии всех остальных переменных:

$$p(x_i, x_j | \mathbf{x} \setminus \{i, j\}) = p(x_i | \mathbf{x} \setminus \{i, j\}) p(x_j | \mathbf{x} \setminus \{i, j\}).$$

# Неориентированные графические модели (НГМ)

**Замечание:** Если  $x_i$  и  $x_j$  не соединены ребром, то они независимы при условии всех остальных переменных:

$$p(x_i, x_j | \mathbf{x} \setminus \{i, j\}) = p(x_i | \mathbf{x} \setminus \{i, j\}) p(x_j | \mathbf{x} \setminus \{i, j\}).$$



$$p(a, b, c, e, f, g) = \frac{1}{Z} \psi_{afe}(a, f, e) \psi_{ec}(e, c) \psi_{eg}(e, g) \psi_{bg}(b, g) \psi_{bf}(b, f).$$

$$Z = \int \prod_i \psi_{C_i}(\mathbf{x}_{C_i}).$$

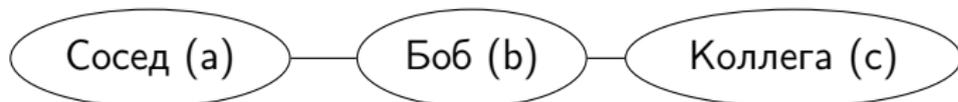
**Вопрос:** Какими свойствами должны обладать  $\psi_{C_i}(\mathbf{x}_{C_i})$ , чтобы задать корректное распределение?

**Теорема (Hammersley-Clifford).** Предположим, что все потенциалы строго положительны  $\psi_C(\mathbf{x}_C) > 0 \forall \mathbf{x}_C$ . Тогда факторизация по максимальным кликам графа задает то же множество распределений, что и набор условий условной независимости в терминах графовой разделимости.

## Пример построения НГМ

Пусть моделируется заболеваемость простудой для трех человек: Боба, его соседа и коллеги.

Метка 0 соответствует тому, что человек здоров, а 1 – болезни.



$$p(a, b, c) = \frac{1}{Z} \psi_{ab}(a, b) \psi_{bc}(b, c), \quad a, b, c \in \{0, 1\}.$$

$$\psi_{a,b}(a, b) = \begin{cases} 20, & a = 0, b = 0, \\ 2, & a = 1, b = 1, \\ 1, & a = 0, b = 1, \\ 1, & a = 1, b = 0. \end{cases}, \quad \psi_{b,c}(b, c) = \begin{cases} 5, & b = 0, c = 0, \\ 0.7, & b = 1, c = 1, \\ 5, & b = 0, c = 1, \\ 0.1, & b = 1, c = 0. \end{cases}$$

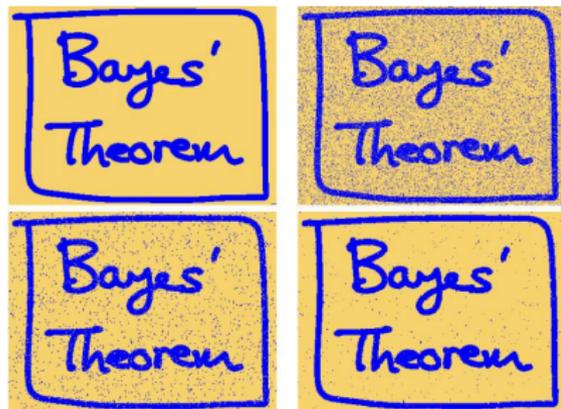
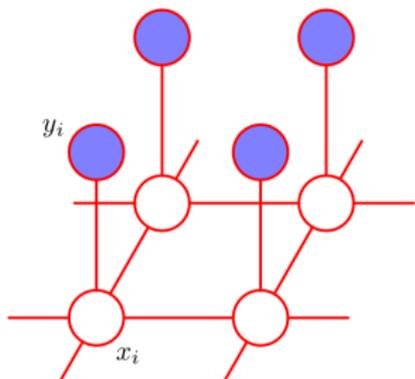
# Задание потенциала через энергию

**Идея:**  $\psi_{C_i}(\mathbf{x}_{C_i}) = \exp(-E_{C_i}(\mathbf{x}_{C_i}))$ .

Тогда  $\log p(\mathbf{x}) \propto -E(\mathbf{x}) = -\sum_i E_{C_i}(\mathbf{x}_{C_i})$ .

**Пример:** Пусть имеется бинарное изображение  $\mathbf{y}$ ,  $y_i \in \{-1, 1\}$ , которое зашумлено. Требуется восстановить исходное изображение  $\mathbf{x}$ .

$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{(i,j) \in \epsilon} x_i x_j - \eta \sum_i x_i y_i.$$

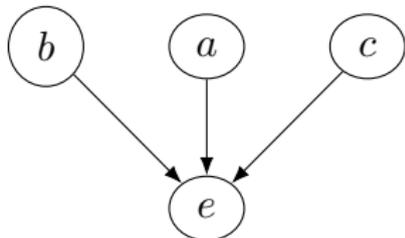


Графическая модель  $p(\mathbf{x}, \mathbf{y})$   
[Bishop, 2006]

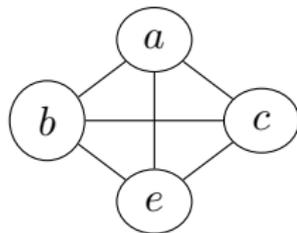
Иллюстрация шумоподавления [Bishop,  
2006]

# Связь между ориентированными и неориентированными моделями

**Вопрос:** Можно ли по ориентированной модели построить соответствующую ей неориентированную и наоборот?



$$p(a, b, c, e) = p(a)p(b)p(c)p(e|a, b, c).$$



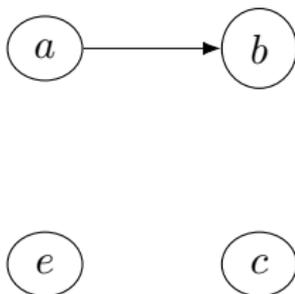
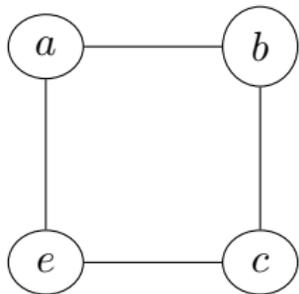
$$p(a, b, c, e) = \psi(a, b, c, e).$$

**Конверсия из ориентированного в неориентированный граф:**

- Соединяем ребрами попарно всех родителей каждой вершины («moralization»);
- Удаляем ориентацию на ребрах и получаем неориентированный граф;
- Для каждой клики задаем потенциал  $\psi_{C_i}(\mathbf{x}_{C_i}) = 1$ ;
- Для каждого условного распределения, умножаем на него потенциал той клики, к которой относятся все переменные из него.

# Из неориентированной модели к ориентированной

**Вопрос:** Верно ли, что ориентированная модель позволяет всегда «более аккуратно» задать зависимости, не потеряв условных независимостей между переменными?

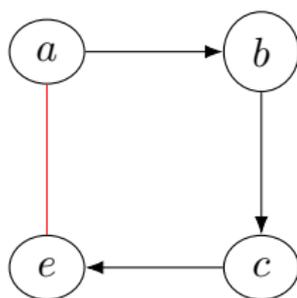
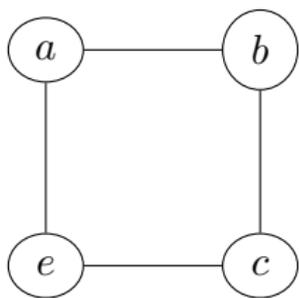


$$p(a, b, c, e) = \psi_1(a, b)\psi_2(b, c)\psi_3(c, e)\psi_4(e, a).$$

Свойства зависимости и независимости переменных:

- $a \not\perp c | \emptyset$ ;
- $a \perp c | b, e$ ;
- $b \perp e | a, c$ ;
- $\vdots$

## Из неориентированной модели к ориентированной 2

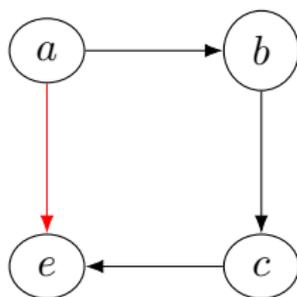
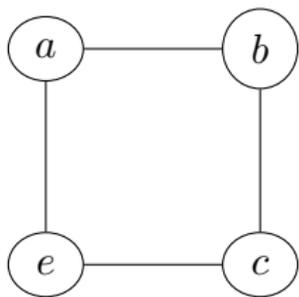


$$p(a, b, c, e) = \psi_1(a, b)\psi_2(b, c)\psi_3(c, e)\psi_4(e, a).$$

Свойства зависимости и независимости переменных:

- $a \not\perp c | \emptyset$ ;
- $a \perp c | b, e$ ;
- $b \perp e | a, c$ ;
- $\vdots$

**Вопрос:** Могли ли мы ориентировать ребро  $(b, c)$  иначе и выполнить условие  $a \perp c | b, e$ ?



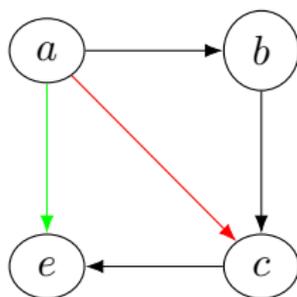
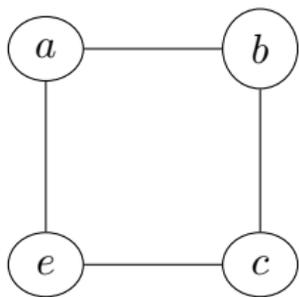
$$p(a, b, c, e) = \psi_1(a, b)\psi_2(b, c)\psi_3(c, e)\psi_4(e, a).$$

Свойства зависимости и независимости переменных:

- $a \not\perp c | \emptyset$ ;
- $a \perp c | b, e$ ;
- $b \perp e | a, c$ ;
- $\vdots$

**Вопрос:** Соответствует ли граф справа правдоподобию? Нет ли дополнительной условной независимости?

# Из неориентированной модели к ориентированной 4



**Слева:**  $p(a, b, c, e) = \psi_1(a, b)\psi_2(b, c)\psi_3(c, e)\psi_4(e, a)$ .

Свойства зависимости и независимости переменных:

$a \perp c | b, e, b \perp e | a, c, \dots$

**Справа:**  $p(a, b, c, e) = p(a)p(b|a)p(c|b, a)p(e|a, c)$ .

**Замечание 1:** Без ребра из a в c, была дополнительная условная независимость  $a \perp c | b$ .

**Вопрос 1:** Почему нельзя было бы провести вместо  $a \rightarrow c$  ребра  $c \rightarrow a, e \rightarrow b$ ?

**Вопрос 2:** Помогло ли бы ребро  $b \rightarrow e$  убрать  $a \perp c | b$ ?

**Вопрос:** Какую условную независимость мы потеряли?

- 1 Bishop, Christopher M. "Pattern recognition and machine learning". Springer, New York (2006). Pp. 381-394.
- 2 Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016. Pp. 549-576.
- 3 Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.
- 4 Bacchus, Fahiem, and Adam J. Grove. "Graphical models for preference and utility." arXiv preprint arXiv:1302.4928 (2013).
- 5 Mor, Bhavya, Sunita Garhwal, and Ajay Kumar. "A systematic review of hidden markov models and their applications." Archives of computational methods in engineering 28.3 (2021): 1429-1448.
- 6 Pearl, Judea. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan kaufmann, 1988.
- 7 Pearl, Judea. "Graphical models for probabilistic and causal reasoning." Quantified representation of uncertainty and imprecision (1998): 367-389.