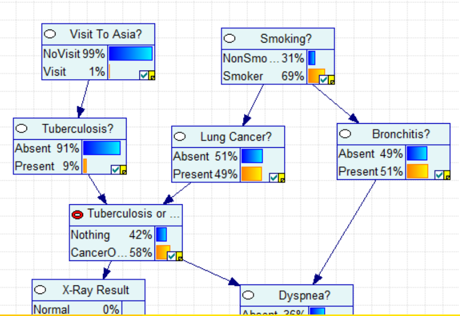
Bayesian Networks

Definition (Bayesian Network)

Bayesian network is a pair (G, P), where

- $G = (V, E)$ is a DAG (directed acyclic graph with set of vertexes V and set of edges E) and
- P is a list of conditional probability distributions such that for every vertex X there is a probability table $P(X|pa(X))$ where $pa(X)$ denotes the set of parents of a vertex in G.

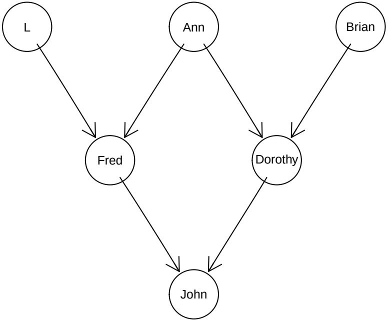
The graph represents conditional independencies of the join probability distribution $\Pi_{X \in V} P(X|pa(X))$.



Příklad: Studfarm

- Existuje genetická nemoc, která je recesivní, tj. projeví se jen pokud jsou oba rodiče nosiči.
- Známe rodokmen, objevilo se nemocné dítě – vytvořte model, který určí pravděpodobnost genetické chyby u všech členů rodokmenu.
- Rodokmen chcete modelovat do prarodičů nemocného.

	Prior
aA	0.01
AA	0.99



Nachlazení nebo Angína

- Nachlazení nebo angína
- Teplota, Bolest v krku, Skrvny v krku
- single nebo multiple faults

Trobleshooting, Klasifikace dokumentů

Naive Bayes Klasifikátor

- Klasifikujeme do třídy C; $sp(C) = \{1, \ldots, K\}$
- Pro každý atribut $A_1, \ldots, A_n$ známe podmíněnou pravděpodobnost $P(A_i|C = k)$ pro každou třídu $k \in sp(C)$
  - Bernoulli (0-1): $P(a_1, \ldots, a_n|C = k) = \Pi_{i=1}^{n} P(a_i|C = k)$
  - Gaussovský $p(x = v|k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$
  - Multinomial: $P(x_1, \ldots, x_n|C = k) = \frac{(\Sigma_{i=1}^{n} x_i)!}{\Pi_{i=1}^{n} x_i!} \Pi_{i=1}^{n} P(A_i = 1|C = k)^{x_i}$
- Predikce:
  - $\hat{g} = argmax_{k=1,\ldots,K} P(C = k) \Pi_{i=1}^{n} P(A = a_i|C = k)$

Příklad

- Snědli jsme 1000 bonbónů a zapsali, co jsme pozorovali:

	W=red		W=green	
	H=1	H=0	H=1	H=0
F=cherry	273	93	104	90
F=lime	79	100	94	167

- Vytvořte naive bayes model s veličinou F jakožto cíl (nahoře).

F=cherry	273 + 93 + 104 + 90	=560	0.56
F=lime	79 + 100 + 94 + 167	=440	0.44

- $P(W|F)$

	F=cherry	F=lime
W=red	$\frac{273+93}{560}$	$\frac{79+100}{440}$
W=green	$\frac{104+90}{560}$	$\frac{94+167}{440}$

- $P(H|F)$

	F=cherry	F=lime
H=1	$\frac{273+104}{560}$	$\frac{79+94}{440}$
H=0	$\frac{93+90}{560}$	$\frac{100+167}{440}$

Tree augmented Naive Bayes

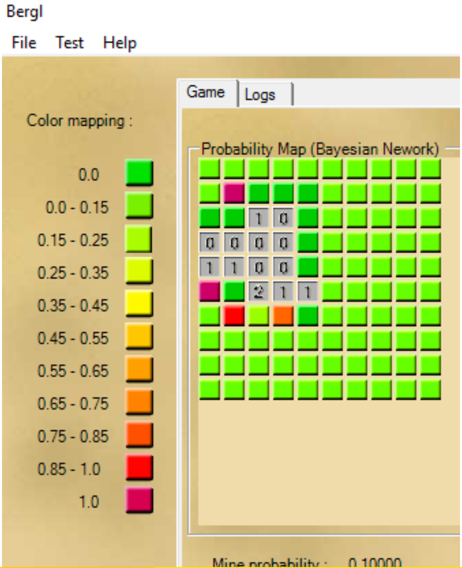
- Dovolíme stromovou strukturu na $\{A_i\}_i$ místo podmíněné nezávislosti dáno C.
- Pr, Bt, Ut, Sc
- potřeba reprezentovat Ho

Ponožky

- dva páry ponožek
- různého vzoru a barvy
- výrazně sepravé
- do modelu přidáme podmínku, že každého typu jsou právě dvě ponožky

Příklad: Miny

- Navrhněte model, který spočte pravděpodobnost miny ve hře Minesweeper.
- Navrhněte aproximaci, která nebude mít uzel s více než 10 rodiči.

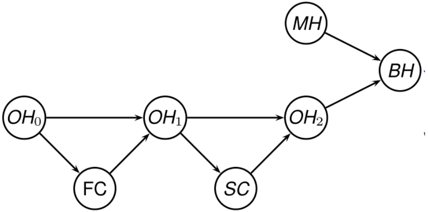


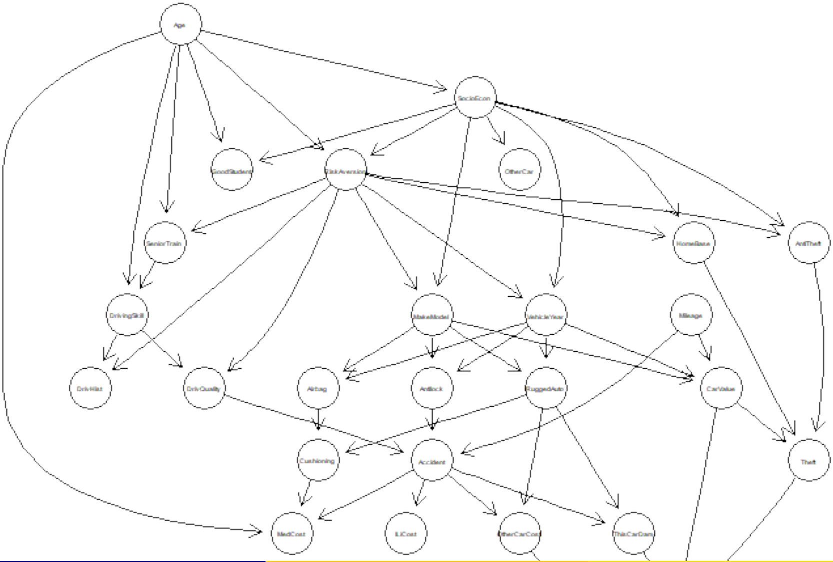
Parent divorsing - půjčka

- 11 atributů: příjem, typ práce, auta v rodině, finanční situace, počet dětí, počet adres v posledních 5 letech, ...
- Při 5-ti hodnotových atributech $5^{11} \approx 5M$ konfigurací.
- Pro snadný odhad pravděpodobnosti potřebuji 'aspoň 5' příkladů v nejméně časté kombinaci.

Poker

- Vytvořte zjednodušený model hry Poker s veličinou odhadující, jestli je lepší moje ruka nebo soupeřova.




Age
SocioEcon
GoodStudent
RiskAversion
OtherCar
SeniorTrain
HomeBase
AntiTheft
DrivingSkill
MakeModel
VehicleYear
Mileage
DrivHist
DrivQuality
Airbag
Antilock
RuggedAuto
CarValue
Cushioning
Accident
Theft
MedCost
ILiCost

Čtení v bayesovské síti

- Typický dotaz: známe evidenci e o veličinách $dom(e)$; ptáme se na jednorozměrné podmíněné pravděpodobnosti všech ostatních veličin X: $P(X|e)$.

Struktura grafu nám umožnuje:

- Optimalizovat výpočet.
- Určit některé veličiny jako 'zbytečné', které je možno vypustit z modelu.
- Můžeme zkoumat nezávislost dvojice veličin dáno 'evidence'.

Výpočet marginálních podmíněných pravděpodobností v bayesovské síti

Úmluva: Zajímáme se pouze o bayesovské sítě, jejichž graf je spojitý. Jinak uvažujeme každou komponentu zvlášť.

Algoritmus eliminace proměnných

- **INIT** Do seznamu Φ_1 dáme všechny tabulky $P(e, A_i|pa(A_i))$, v každé tabulce odstraním "řádky"nekonzistentní s evidencí, tj. s nulovou pravděpodobností. Tj. předem evidencí vynásobíme a marginalizujeme přes proměnné s evidencí.
- **ELIM** Postupně budeme eliminovat (následujícím algoritmem) všechny proměnné bez evidence, které nás nezajímají (dostaneme $P(A, e)$).
- **NORM** Nakonec eliminujeme i zbývající proměnné bez evidence, čímž spočteme normalizační konstatnu $\alpha = P(e)$; touto konstantou vydělíme tabulku z předchozího kroku a dostaneme podmíněnou pravděpodobnost $P(A|e)$.

Eliminace proměnné X v kroku i znamená:

1. Vyber z Φ_i všechny tabulky, které mají v doméně X, dej je do Φ_X.
2. Spočti $\phi = \sum_X \prod_{T \in \Phi_X} T$
3. Nové Φ_{i+1} se rovná: $\Phi_i \setminus \Phi_X \cup \{\phi\}$

Pozn: Pokud v Φ_{last} nakonec zbyde více tabulek, musíme je vynásobit.
Pozn2: Ne-eliminované A může být buď veličina, nebo množina veličin.

Charakteristika algoritmu Eliminace proměnných

- Snadný na pochopení a implementaci.
- Otázka: v jakém pořadí eliminovat? Špatné pořadí vede ke zbytečně velkým tabulkám ϕ.
- Pokud nás zajímají **všechny jednorozměrné** marginály, tak bychom nemuseli počítat vše pro každou zvlášť, dost výpočtů se opakuje.

Proto většina software používá jiné algoritmy, my se podíváme, co používá Hugin, ostatní mají různé modifikace.

Miny: Špatné a lepší pořadí eliminace.

špatně Nejdřív nepozorované rodiče (např. v 'minách').

právně Nejdřív **barren** tj. uzly bez dětí a bez evidence.

obecně Nejdřív **simpliciální uzly**.

d-separace

Definition (d-separace)

Dvě veličiny $A, B \in V$ bayesovské sítě $G = (V, E)$ jsou **d-separované** $A \perp\!\!\!\perp_d B|\mathcal{C}$ množinou $\mathcal{C} \subseteq V \setminus \{A, B\}$ právě když pro každou (neorientovanou) cestu z A do B platí aspoň jedno z následujících:

- cesta obsahuje uzel $Blocking \in \mathcal{C}$ a hrany se v $Blocking$ **nesetkávají 'head-to-head'**,
- cesta obsahuje uzel $Blocking$ kde se hrany **setkávají** 'head-to-head' a **ani on ani nikdo z jeho následníků není** v $\mathcal{C}$, $\{Blocking, succ(Blocking)\} \cap \mathcal{C} = \emptyset$.

Theorem (d-separace)

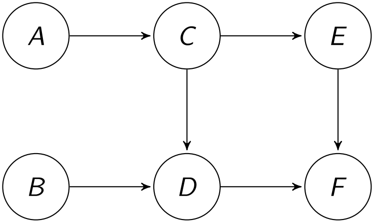
Pokud jsou A, B d-separované dáno $\mathcal{C}$ ($A \perp\!\!\!\perp_d B|\mathcal{C}$) v BN B, pak jsou i podmíněně nezávislé ($A \perp\!\!\!\perp B|\mathcal{C}$).

Příklad d-separace

Definition (d-separace)

Dvě veličiny $A, B \in V$ bayesovské sítě $G = (V, E)$ jsou **d-separované** $A \perp\!\!\!\perp_d B | \mathcal{C}$ množinou $\mathcal{C} \subseteq V \setminus \{A, B\}$ právě když pro každou (neorientovanou) cestu z A do B platí aspoň jedno z následujících:

- cesta obsahuje uzel *Blocking* $\in \mathcal{C}$ a hrany se v *Blocking* **nesetkávají 'head-to-head'**,
- cesta obsahuje uzel *Blocking* kde se hrany **setkávají** 'head-to-head' a ani on ani nikdo z jeho následníků není v $\mathcal{C}$, $\{Blocking, succ(Blocking)\} \cap \mathcal{C} = \emptyset$.



Platí následující?

- $E \perp\!\!\!\perp_d B$ ano
- $E \perp\!\!\!\perp_d D$ ne
- $E \perp\!\!\!\perp_d D | A$ ne
- $E \perp\!\!\!\perp_d D | C$ ano
- $E \perp\!\!\!\perp_d D | \{C, F\}$ ne
- $E \perp\!\!\!\perp_d B | F$ ne