

Sieci bayesowskie

Marek J. Drużdżel

Politechnika Białostocka

Wydział Informatyki

m.druzdzel@pb.edu.pl

<http://www.wi.pb.edu.pl/~druzdzel/>

Plan wykładu

- **Bayesowskie modele graficzne**
- **Wnioskowanie w sieciach bayesowskich**
- **Rozszerzona rodzina modeli graficznych**

Bayesowskie modele graficzne

Systemy i modele

Systemy - części rzeczywistego świata, które można badać w oderwaniu od reszty świata

Modele - (subiektywne) abstrakcje systemów, stosowane w nauce lub codziennym myśleniu

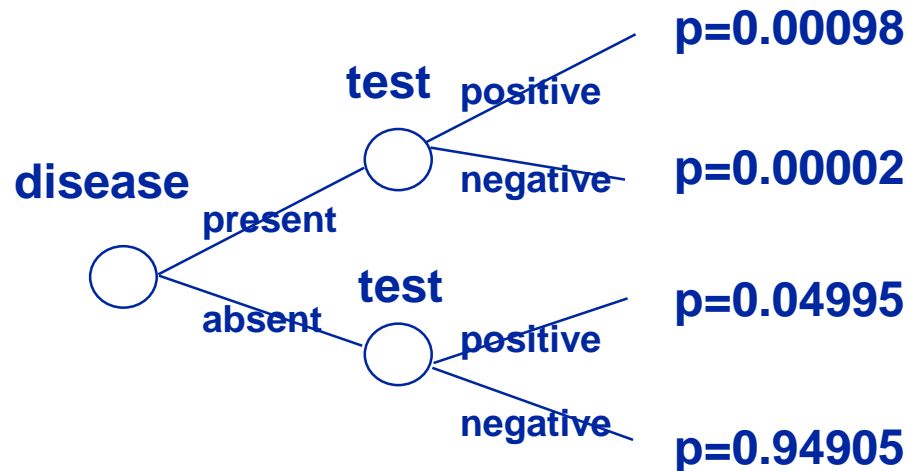
Sieci bayesowskie to modele 😊.

Reprezentacje wiedzy probabilistycznej

- Model probabilistyczny (bayesowski) koduje **łączny rozkład prawdopodobieństwa** dla swoich zmiennych.
- Znajomość łącznego rozkładu prawdopodobieństwa jest wystarczająca, aby wyznaczyć dowolne prawdopodobieństwo brzegowe i warunkowe dla zmiennych modelu (i wszystkiego innego, co mogłoby nas zainteresować!).

Drzewa prawdopodobieństwa

Najprostsza i całkiem naturalna graficzna reprezentacja rozkładu prawdopodobieństwa łącznego dla zmiennych dyskretnych



$$P(\text{disease present} \wedge \text{test positive}) = P(D \cap +) = 0.00098$$

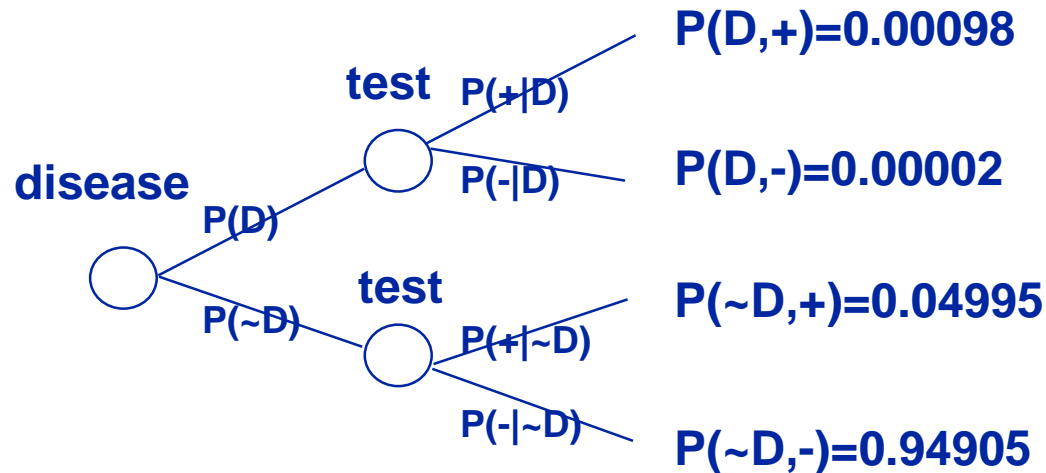
$$P(\text{disease present} \wedge \text{test negative}) = P(D \cap -) = 0.00002$$

$$P(\text{disease absent} \wedge \text{test positive}) = P(\sim D \cap +) = 0.04995$$

$$P(\text{disease absent} \wedge \text{test negative}) = P(\sim D \cap -) = 0.94905$$

Obliczenia w drzewach prawdopodobieństwa

Na podstawie rozkładu prawdopodobieństwa zakodowanego w drzewie możemy obliczyć dowolny rozkład prawdopodobieństwa brzegowego lub warunkowego.

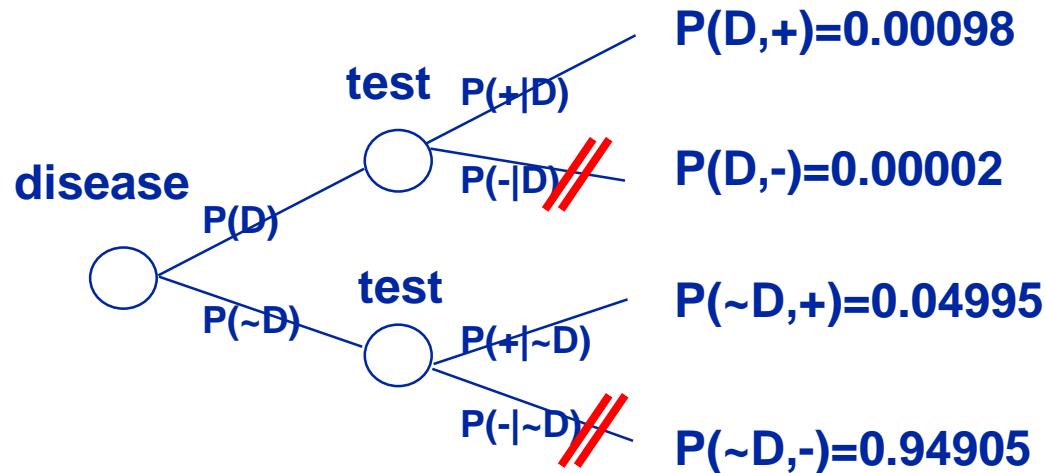


Jakie jest prawdopodobieństwo występowania tej choroby?

$$P(D) = 0.00098 + 0.00002 = 0.001$$

Obliczenia w drzewach prawdopodobieństwa

Na podstawie rozkładu prawdopodobieństwa zakodowanego w drzewie możemy obliczyć dowolny rozkład prawdopodobieństwa brzegowego lub warunkowego.

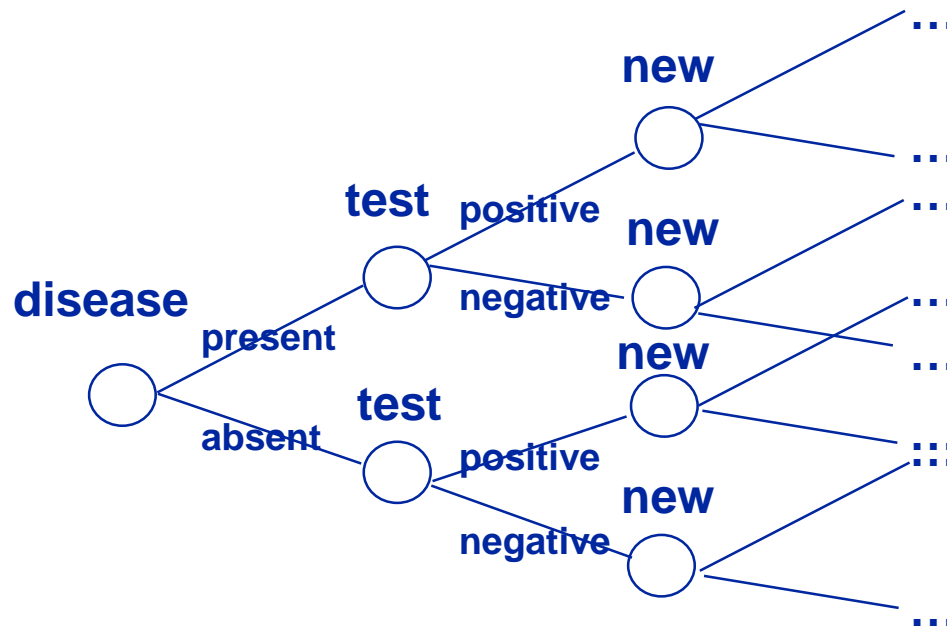


Jakie jest prawdopodobieństwo obecności choroby przy dodatnim wyniku testu? Obserwacja dodatniego wyniku testu czyni niektóre gałęzie drzewa niemożliwymi. Musimy po prostu znormalizować pozostałe (tj. te, które są zgodne z obserwacjami) gałęzie!

$$P(D|+) = 0.00098 / (0.00098 + 0.04995) \approx 0.01924$$

Co jest nie tak z drzewami prawdopodobieństwa?

Drzewa rosną wykładniczo wraz z liczbą zmiennych



Dla n zmiennych binarnych mamy 2^n gałęzi.

Dla $n=10$, całkowita liczba gałęzi wyniesie 1.024

Dla $n=11$, liczba gałęzi drzewa wyniesie 2.048

...

Dla $n=20$, wyniesie 1.048.576 (co jest dużą liczbą 😊)

Genialny pomysł (mający mniej niż 50 lat)

**Wykorzystaj niezależność
zmiennych w rozkładzie
prawdopodobieństwa łącznego,
aby zredukować liczbę parametrów
w jego reprezentacji!**

Pionierska praca nad niezależnością probabilistyczną
dwóch naukowców, A. Philipa Dawida and Judea'y Pearla



**Wszystkie genialne pomysły są oczywiste
(ale tylko wtedy, gdy już je mamy 😊)**



**Czy koncepcja koła
jest oczywista?**

**Dlaczego więc
żadna z cywilizacji
w Ameryce go nie
miała?**

Rozkład łącznego rozkładu prawdopodobieństwa na czynniki

Każdy łączny rozkład prawdopodobieństwa można rozłożyć na czynniki, czyli zapisać go jako iloczyn warunkowych rozkładów prawdopodobieństwa każdej ze zmiennych modelu.

$$f(X_1, X_2, \dots, X_n) = f(X_1 | X_2, X_3, \dots, X_n) f(X_2 | X_3, \dots, X_n) \dots \\ f(X_{n-2} | X_{n-1}, X_n) f(X_{n-1} | X_n) f(X_n)$$

np. dla czterech zmiennych (a, b, c, d), mamy:

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

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

...

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

...

Czyli $n!$ różnych sposobów rozkładu na czynniki rozkładu łącznego prawdopodobieństwa n zmiennych.

Dla $n=4$ mamy $4!=24$ różne rozkłady na czynniki.

Rozkład łącznego rozkładu prawdopodobieństwa na czynniki

- Każdy rozkład na czynniki można uprościć, uwzględniając niezależność zmiennych.
- Te rozkłady, które stają się najprostsze, są lepsze od innych pod względem efektywności reprezentacji.

np. założmy, że wiemy o następujących niezależnościach:
 $B \perp D | C$, $D \perp A | C$, and $A \perp C$

Możemy wtedy uprościć

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

na

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

Sieci bayesowskie

- To leży u podstaw idei sieci bayesowskich.
- Rysujemy graf skierowany z łukiem od zmiennych warunkujących do zmiennych warunkowanych.

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

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

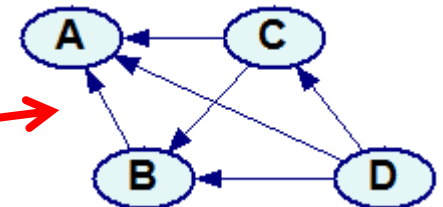
...

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

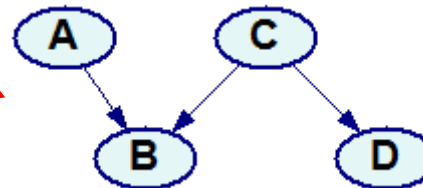
...

$$B \perp D | C, D \perp A | C, A \perp C$$

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



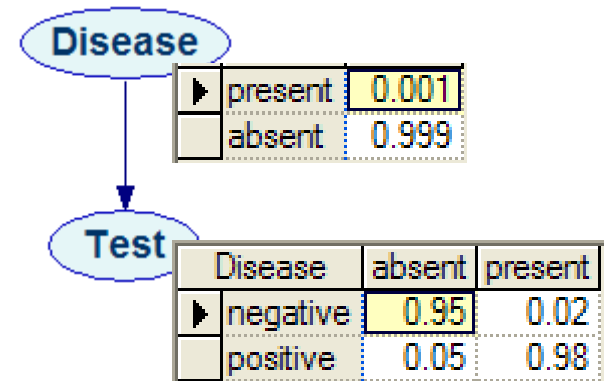
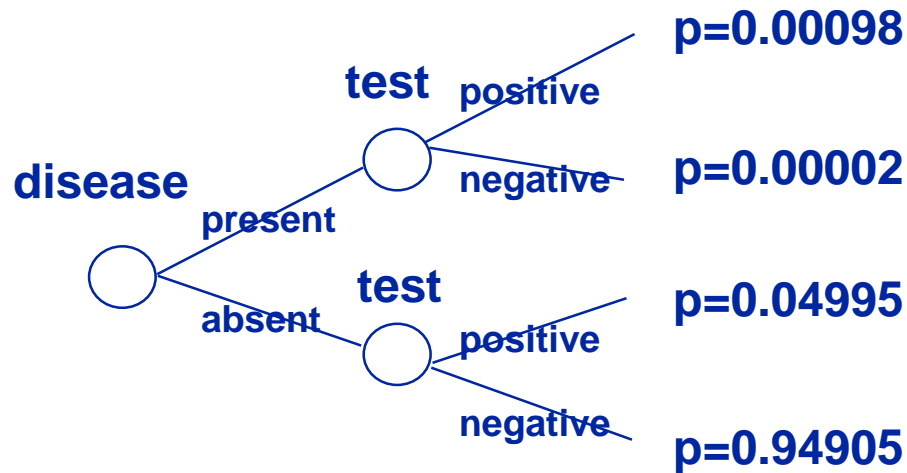
**Brak łuku jest
graficzną
reprezentacją
niezależności!**



Drzewa prawdopodobieństwa i sieci bayesowskie

drzewo prawdopodobieństwa

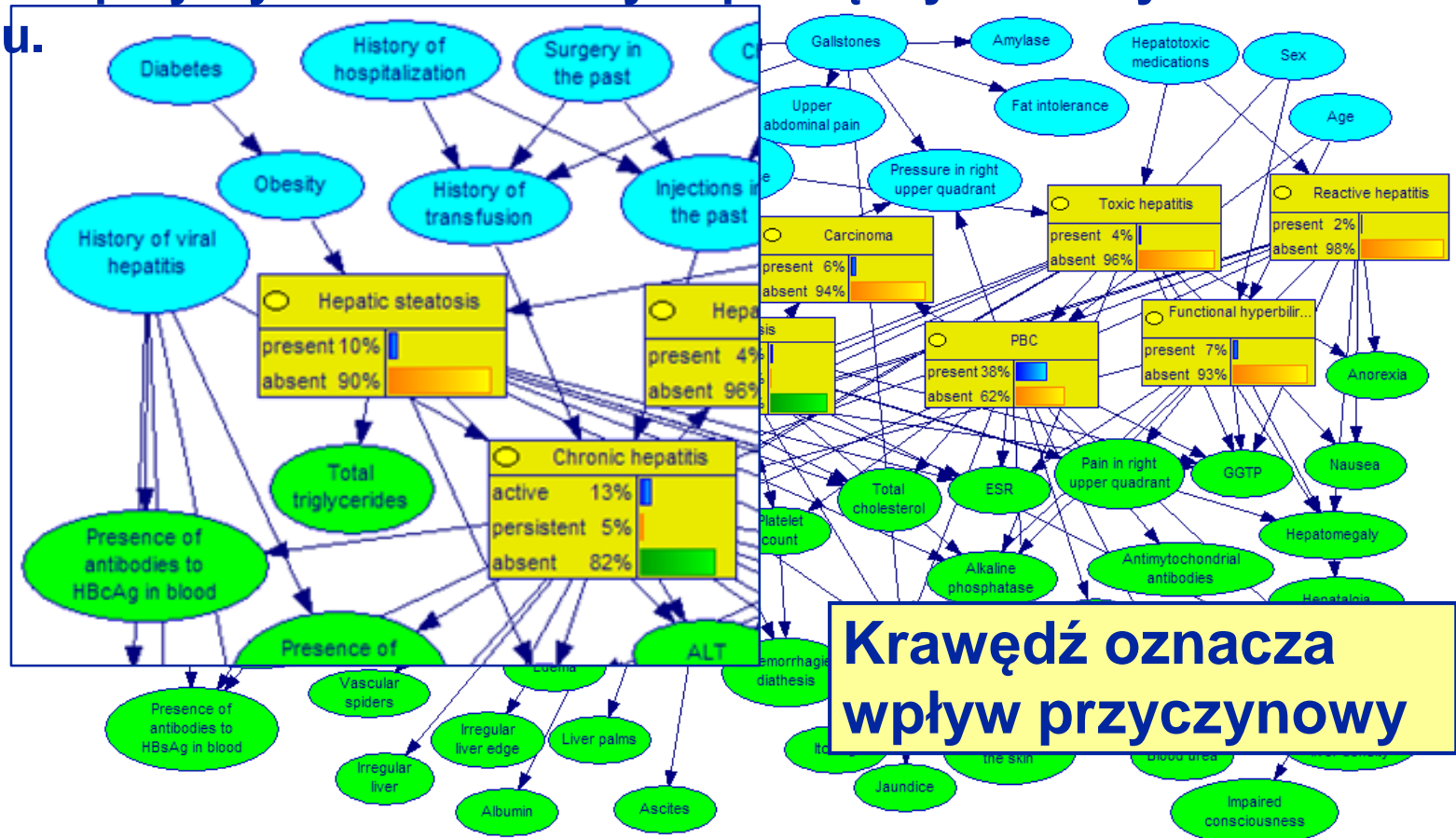
sieć bayesowska



**Obie reprezentacje są równoważne.
 Ale gdy w dziedzinie występują niezależności, sieci bayesowskie są o wiele, wiele bardziej wydajne!**

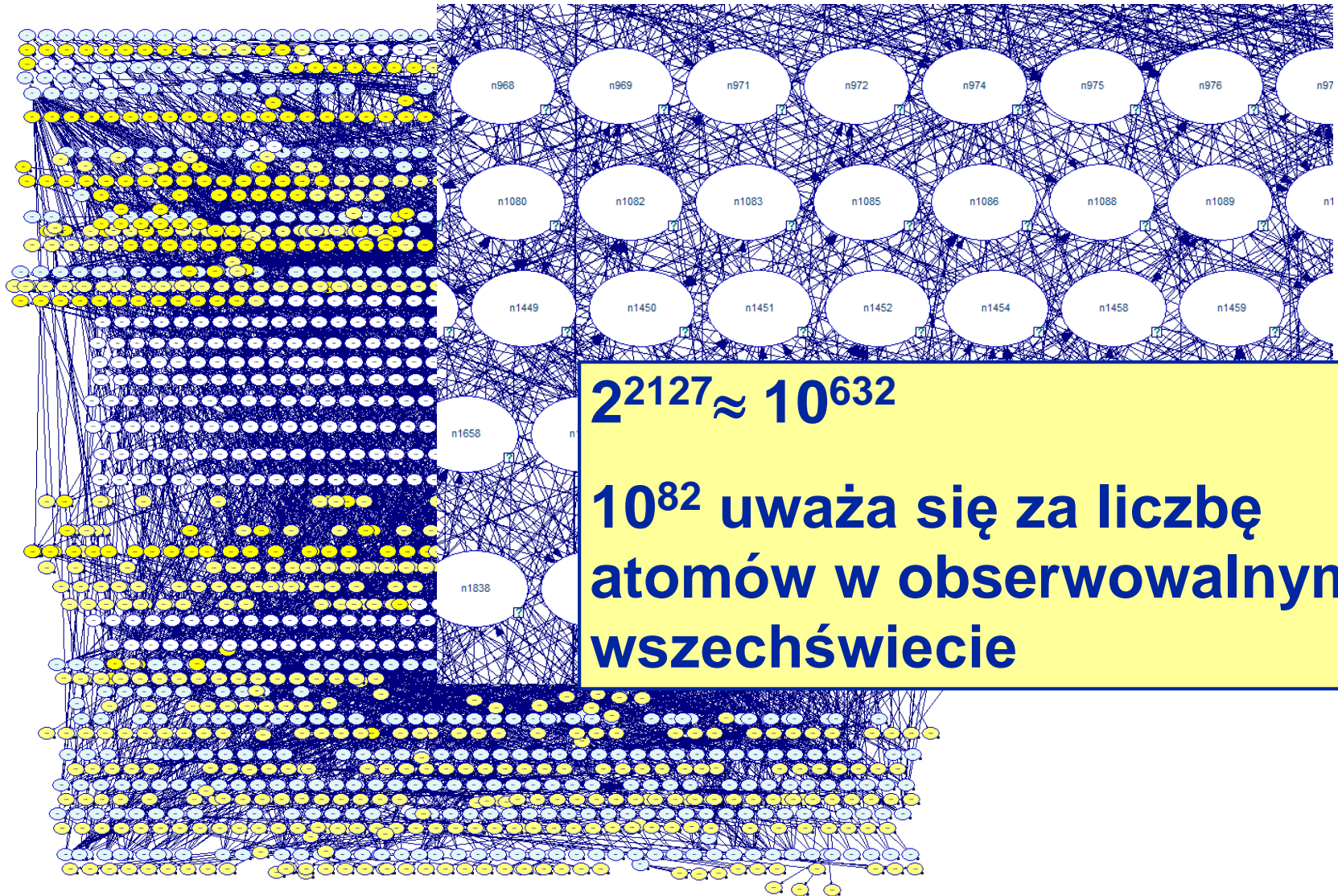
Sieci bayesowskie: alternatywny pogląd (jak się okazuje, całkiem spójny z poglądem teoretycznym!)

Graficzna część sieci bayesowskiej przedstawia obraz związków przyczynowo-skutkowych pomiędzy zmiennymi modelu.



[Oniśko et al.] 70 zmiennych, 123 krawędzi, 2.415 niezależności, 2.139 parametrów numerycznych (zamiast $2^{70} \approx 10^{21}$!)

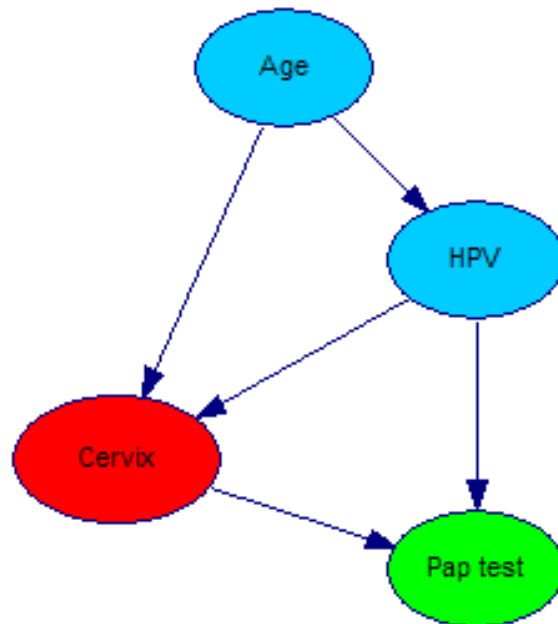
Diagnoza lokomotyw Diesla



[Przytuła et al.] 2.127 zmiennych, 3.595 krawędzi, 2.261.001 niezależności, 12.351 parametrów numerycznych (zamiast $2^{2,127} \approx 10^{632}$!)

Sieci bayesowskie

Sieć bayesowska [Pearl 1988] jest acyklicznym grafem skierowanym składającym się z:

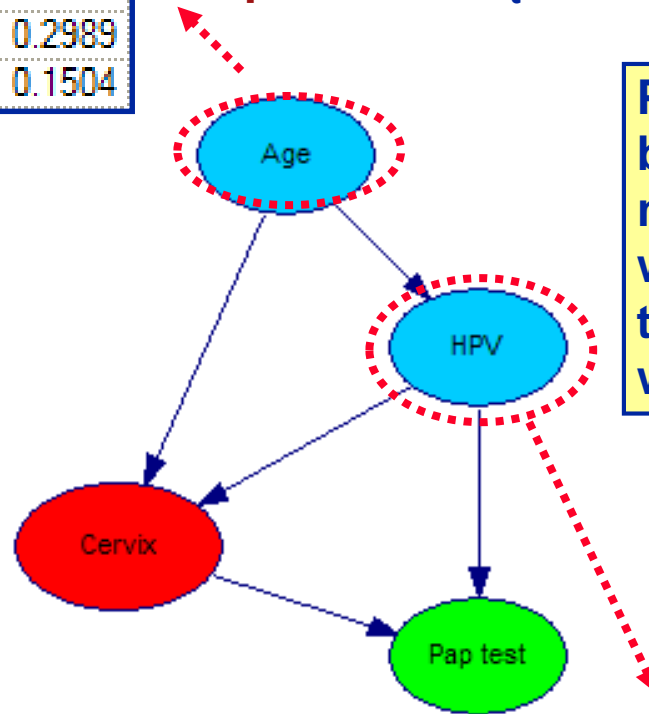


- **części jakościowej**, kodującej zmienne (węzły) i probabilistyczne (zwykle przyczynowe) wpływy między nimi (łuki).
- **części ilościowej**, kodującej łączny rozkład prawdopodobieństwa tych zmiennych.

Sieci bayesowskie: parametry numeryczne

▶ a1_below_20	0.0416
a2_20_29	0.2012
a3_29_45	0.3079
a4_45_60	0.2989
a5_60_up	0.1504

Tabele rozkładu **prawdopodobieństwa a-priori** dla węzłów bez poprzedników (Age)



Proszę zauważyć, że każdy brak łuku (tj. każda niezależność) oznacza jeden wymiar mniej w odpowiedniej tabeli prawdopodobieństwa warunkowego!

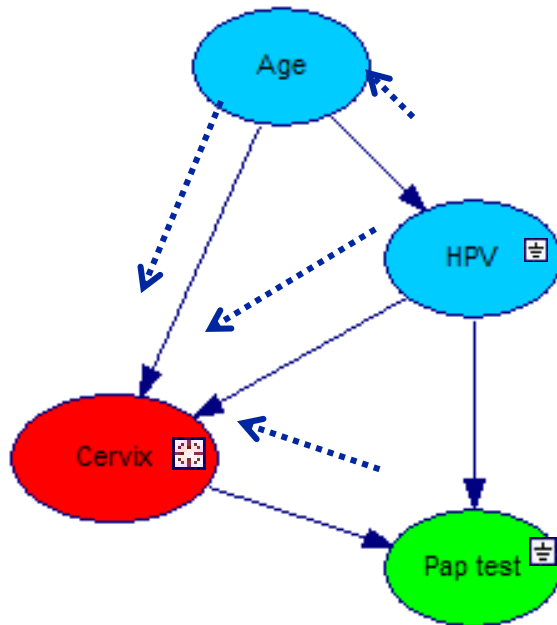
Tabele rozkładów **prawdopodobieństwa warunkowego** dla węzłów z poprzednikami (HPV, Pap test, Cervix)

Age	a1_below_20	a2_20_29	a3_29_45	a4_45_60	a5_60_up
NA	0.8652	0.8387	0.7904	0.8055	0.8851
Negative	0.069	0.0901	0.1782	0.1765	0.1012
▶ Positive	0.0613	0.0667	0.0282	0.0142	0.0082
Qns	0.0045	0.0045	0.0032	0.0038	0.0055

Wnioskowanie w sieciach bayesowskich

Wnioskowanie w sieciach bayesowskich

Najważniejszym typem wnioskowania w sieciach bayesowskich jest aktualizacja prawdopodobieństw wybranych zmiennych (np. diagnozy) na podstawie obserwacji innych zmiennych (np. ustaleń medycznych, wyników testów).



Na przykład:

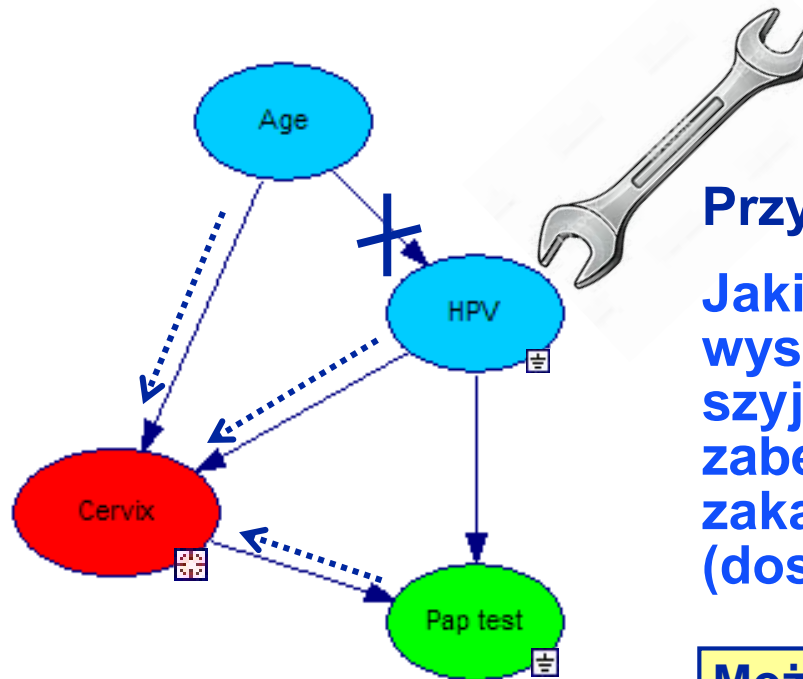
Jakie jest prawdopodobieństwo wystąpienia inwazyjnego raka szyjki macicy u pacjentki z dysplazją dużego stopnia i historią zakażenia wirusem HPV?

$$P(\text{CxCa} \mid \text{HPV}=\text{positive}, \text{HSIL}=\text{yes})$$

Ogólnie, im rzadsza struktura sieci i mniej parametrów, tym szybsze wnioskowanie w sieci bayesowskiej.

Wnioskowanie w sieciach bayesowskich: Interwencja

„Zmiana w strukturze” (pojęcie ekonometryczne) pozwala na przewidywanie skutków interwencji w modelowanym systemie



Przykład:

Jakie jest prawdopodobieństwo wystąpienia inwazyjnego raka szyjki macicy u pacjentki zabezpieczonej przed zakażeniem wirusem HPV (doskonałą) szczepionką?

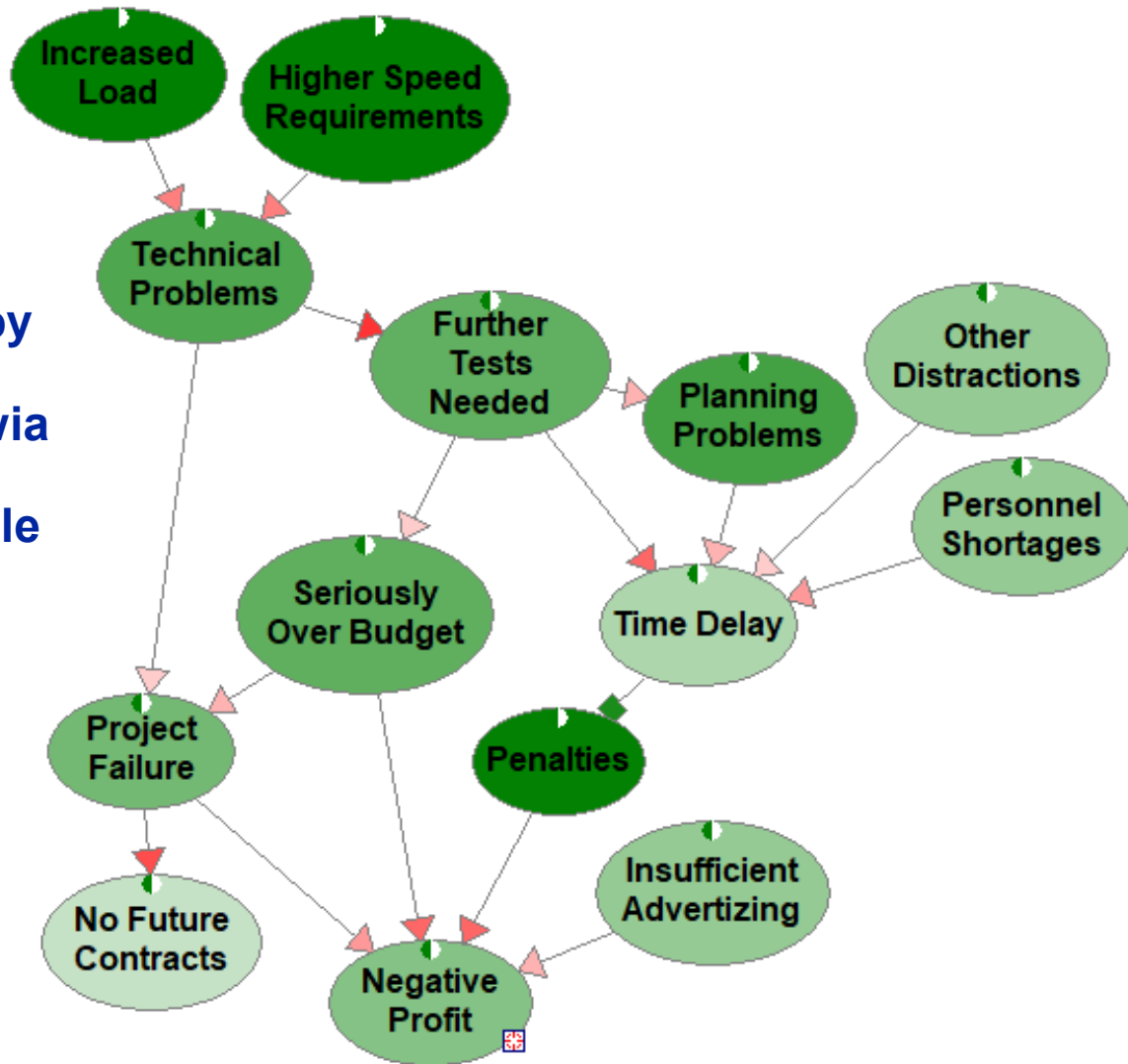
Możemy obliczyć skutki interwencji tylko wtedy, gdy mamy model przyczynowy układu

$P(\text{CxCa} \mid \text{HPV}=\text{negative}, \text{HSIL}=\text{yes})$

Rozszerzona rodzina bayesowskich modeli graficznych

Jakościowe sieci bayesowskie

Jakościowy interfejs do sieci bayesowskich wymaga niewielkiej liczby prawdopodobieństw numerycznych i umożliwia szybkie budowanie i analizę modeli. Doskonale sprawdza się podczas grupowych sesji decyzyjnych.



Systemy oparte na równaniach i modele graficzne

$$\begin{aligned} \text{classsize} &= (\text{nstud} * \text{cload}) / (\text{nfac} * \text{tload}) \\ \text{facsal} &= (\text{oinc} + \text{tuition} * \text{nstud}) / (\text{nfac} * (1 + \text{overh})) \\ \text{stratio} &= \text{nstud} / \text{nfac} \end{aligned}$$

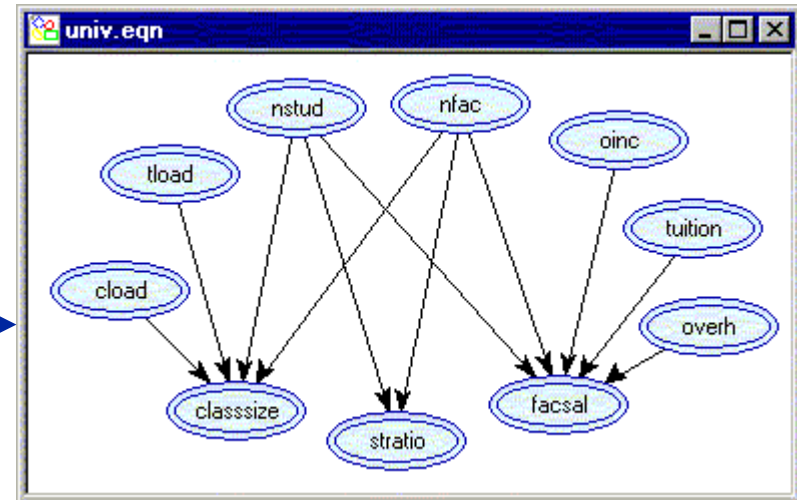
← Równania podstawowe

$$\begin{aligned} \text{cload} &= 15 \\ \text{tload} &= 6 \\ \text{nstud} &= 22102 \\ \text{nfac} &= 3006 \\ \text{oinc} &= 30000000 \\ \text{tuition} &= 12000 \\ \text{overh} &= 0.48 \end{aligned}$$

← Równania dla zmiennych egzogenicznych

Razem określają strukturę modelu →

Wyjaśnienie asymetrii
zapropozowane przez Herberta
Simona (wczesne lata 1950)



Systemy oparte na równaniach : Odwracanie przyczynowości

$$\text{classsize} = (\text{nstud} * \text{cload}) / (\text{nfac} * \text{tload})$$

$$\text{facsal} = (\text{oinc} + \text{tuition} * \text{nstud}) / (\text{nfac} * (1 + \text{overh}))$$

$$\text{stratio} = \text{nstud} / \text{nfac}$$

$$\text{cload} = 15$$

$$\text{tload} = 6$$

$$\text{nstud} = 22102$$

~~$$\text{nfac} = 3006$$~~

$$\text{stratio} = 10$$

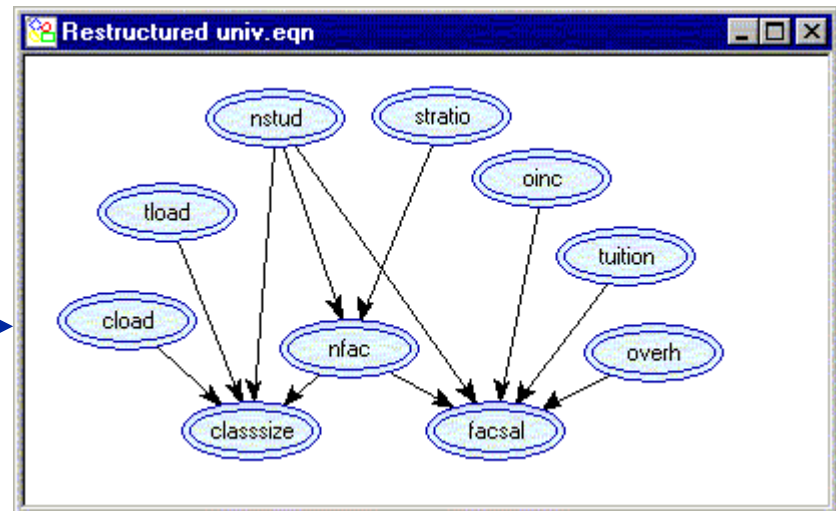
$$\text{oinc} = 30000000$$

$$\text{tuition} = 12000$$

$$\text{overh} = 0.48$$

Ustawienie *stratio* jako
egzogenicznego kosztem *nfac*

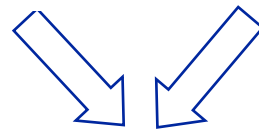
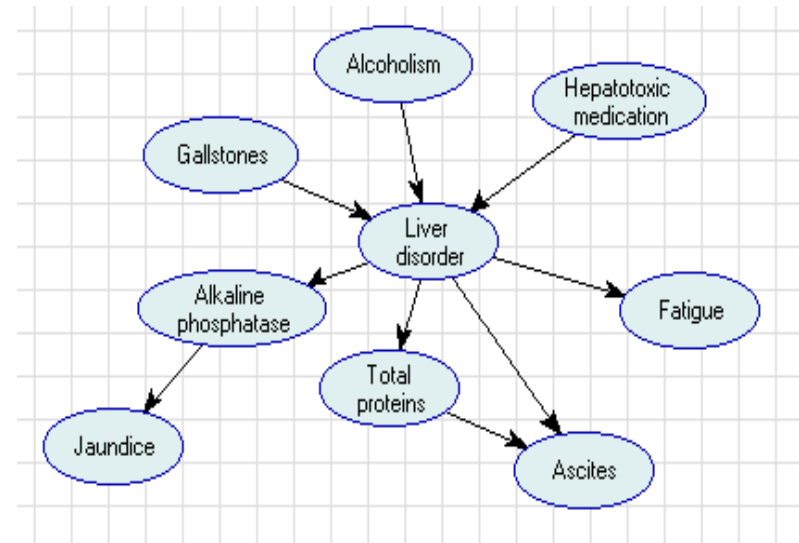
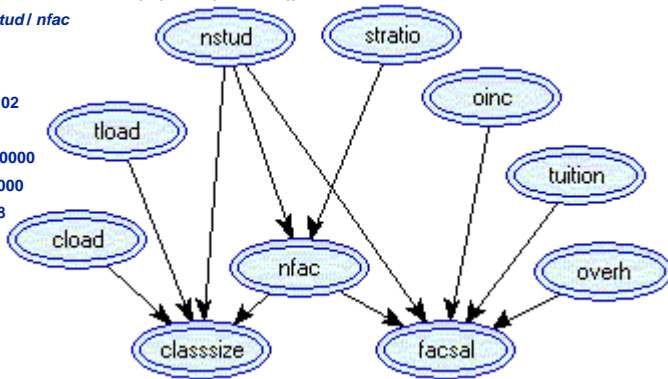
Nowa struktura modelu →



Rodzina grafów skierowanych (szerszy obraz)

(znane również jako “sieci wpływu,” “diagramy przyczynowe,” itd.)

$classsize = (nstud * cload) / (nfac * tload)$
 $facsal = (oinc + tuition * nstud) / (nfac * (1 + overh))$
 $stratio = nstud / nfac$
 $oload = 15$
 $tload = 6$
 $nstud = 22102$
 $nfac = 3006$
 $oinc = 30000000$
 $tuition = 12000$
 $overh = 0.48$



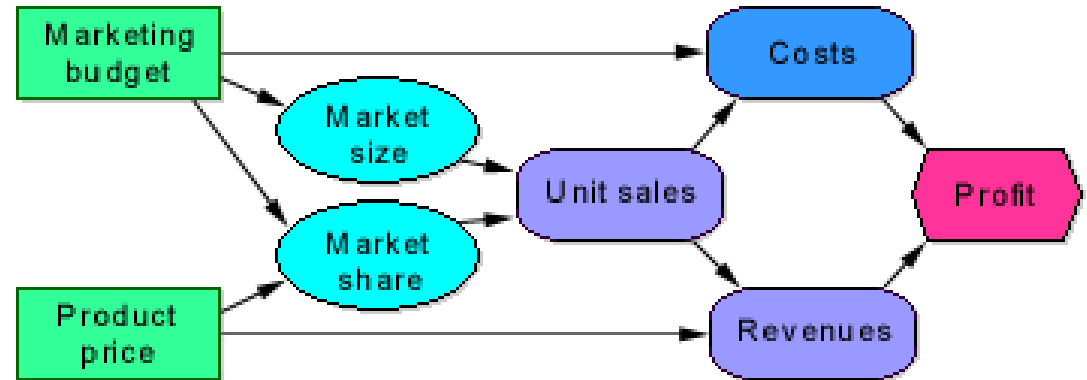
Zarówno układy równań, jak i rozkłady prawdopodobieństwa łącznego można przedstawić za pomocą acyklicznych grafów skierowanych.

Arkusze obliczeniowe

ave. error	max. error	ave. rel. error		
0.08936	0.8002	0.4048		
0.06576	0.6	0.34581		
0.02682	0.2102	0.25562		
0.0158	0.11538	0.19176	0.1276	0.40891
0.00749	0.07541	0.15928	0.12924	0.35773
0.006	0.05357	0.10524	0.05523	0.21613
0.00299	0.02477	0.06739	0.04723	0.1467
0.00213	0.01465	0.07098	0.01874	0.08993
0.07545	0.46004	0.49267	0.01126	0.0627
0.07424	0.69	0.44543		
0.0233	0.12914	0.36243		
0.01917	0.19157	0.3057	0.00636	0.4253
0.00876	0.06715	0.1857	0.00206	0.01178
0.00636	0.04253	0.14596	0.00193	0.01383
0.00206	0.01178	0.07837	0.08227	0.481
0.00193	0.01383	0.05761	0.05043	0.46004
0.08227	0.481	0.61467		
0.05043	0.46004	0.58405		
0.02341	0.1276	0.40891		
0.01983	0.12924	0.35773		
0.008	0.05523	0.21613		
0.00667	0.04723	0.1467		
0.00233	0.01874	0.08993		
0.00193	0.01126	0.0627		

- Można je również postrzegać jako grafy.
- Grafy pokazywałyby zależności przyczynowe między komórkami (zmiennymi).
- Oczywiście, w przypadku każdego praktycznego arkusza kalkulacyjnego otrzymalibyśmy w zasadzie spaghetti połączeń.
- Układy równań współbieżnych i arkusze kalkulacyjne nie są najlepszym rozwiązaniem.
- **Grafy skierowane wydają się ofiarować lepszy interfejs użytkownika!**

Wizualne arkusze obliczeniowe



- Naprawiają prawie wszystko, co było nie tak z arkuszami obliczeniowymi.
- Bardzo dobre, ale zdecydowanie można je jeszcze ulepszyć!

Moim ulubionym programem do wizualnych arkuszy obliczeniowych jest Analytica
(<http://www.lumina.com/>)

Przykład modelu opartego na równaniach strukturalnych przekształconego w sieć bayesowską

Model ogrzewania i chłodzenia budynków.

Dwa równania podstawowe, zmienne/rozkłady ciągłe.

Równania wiążące temperatury przed i za tłumikiem:

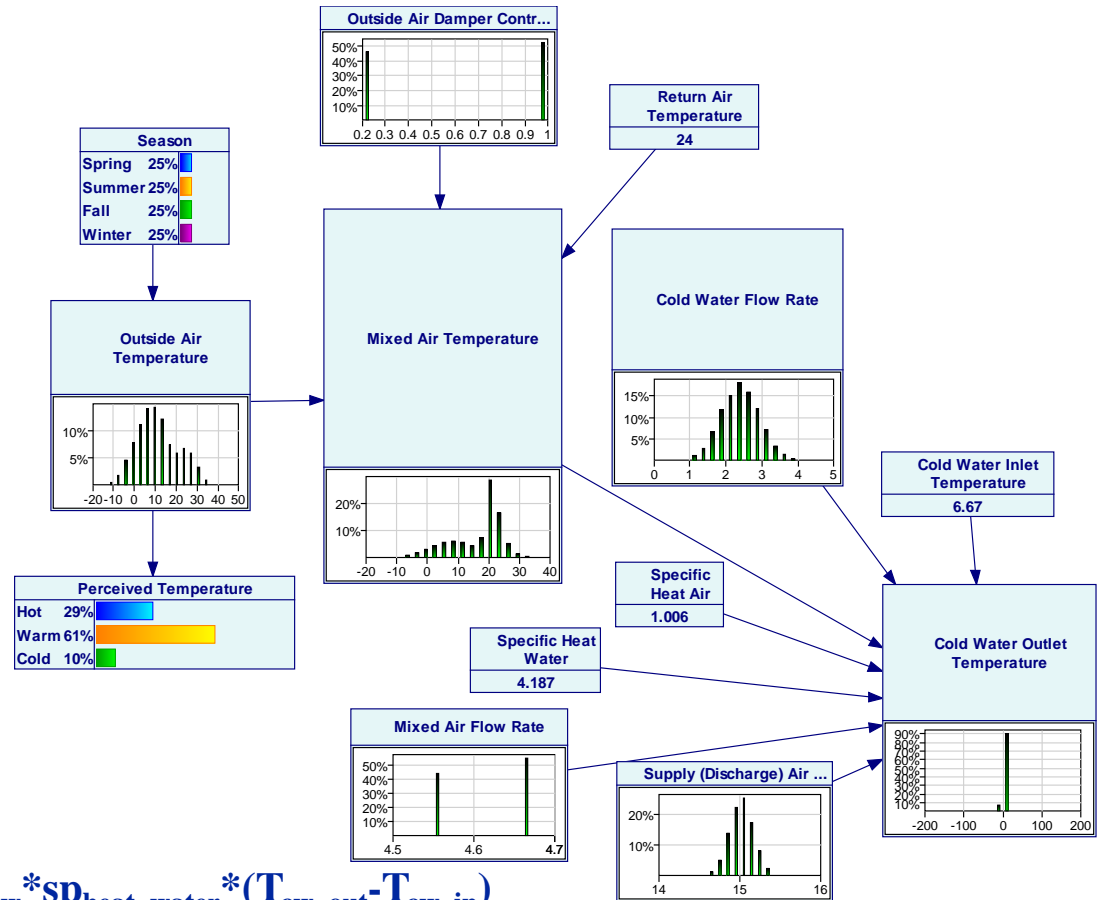
$$T_{ma} = T_{oa} * u_d + T_{ra} * (1 - u_d)$$

Jeśli jest tylko chłodzenie ($u_{hc}=0$)

$$m_{flow_ma} * sp_{heat_air} * (T_{sa} - T_{ma}) = m_{dot_cw} * sp_{heat_water} * (T_{cw_out} - T_{cw_in})$$

Jeśli jest tylko ogrzewanie ($u_{cc}=0$)

$$m_{flow_ma} * sp_{heat_air} * (T_{sa} - T_{ma}) = m_{dot_hw} * sp_{heat_water} * (T_{hw_out} - T_{hw_in})$$

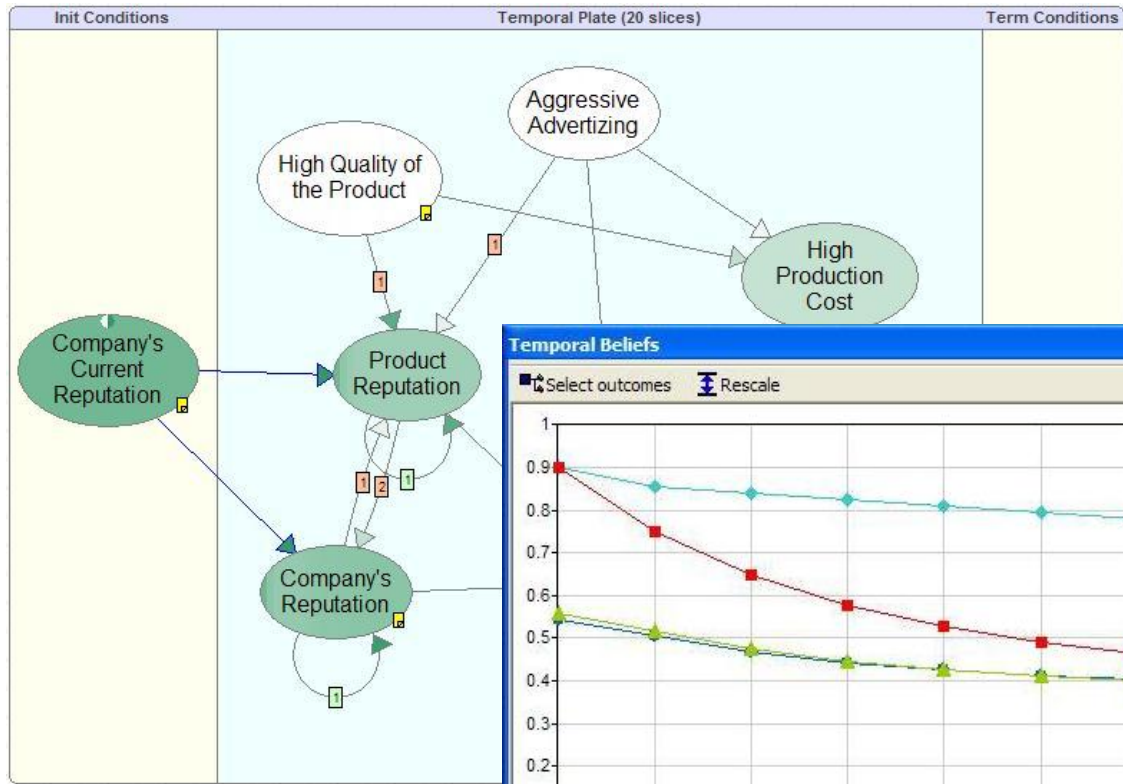


Zalety grafów skierowanych

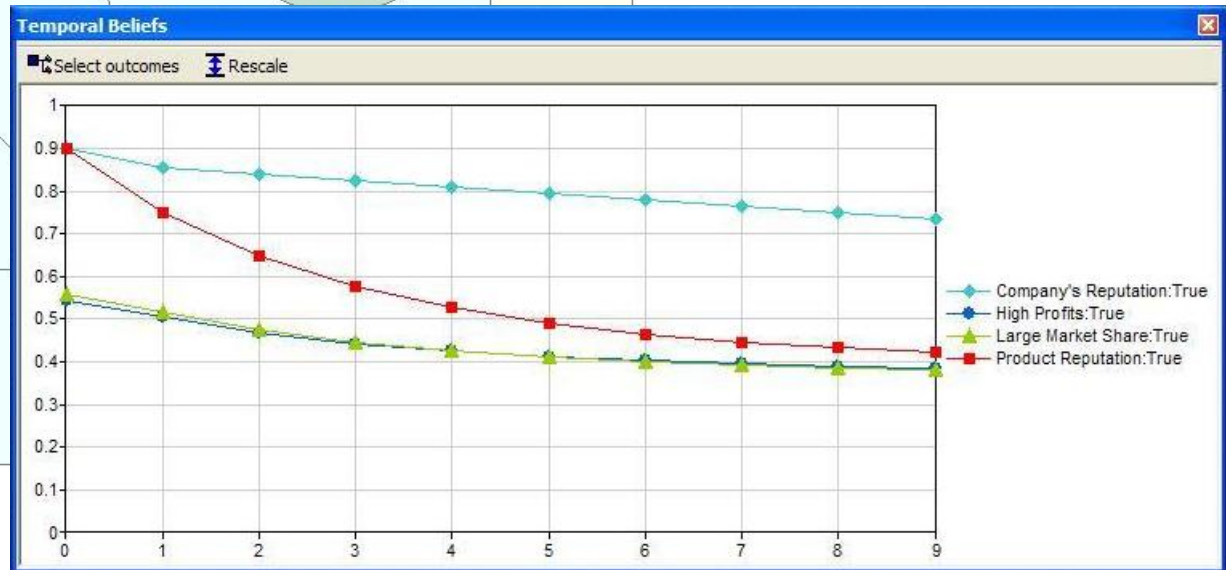
- Mogą być zbudowane tak, aby odzwierciedlały strukturę przyczynową modelu (pomaga to w zrozumieniu problemu)
- Mogą uwzględniać reprezentację niepewności
- Mogą być rekonfigurowane w razie potrzeby
- Mają solidne podstawy teoretyczne: Mamy tu do czynienia z teorią prawdopodobieństwa i teorią decyzji
- Możemy rozmawiać (prawie) tym samym językiem ze statystykami, filozofami i naukowcami

Wnioskowanie temporalne: dynamiczne sieci bayesowskie

Dynamiczne sieci bayesowskie pozwalają śledzić rozwój systemu w czasie i wspomagają podejmowanie decyzji w złożonych środowiskach, w których liczy się nie tylko efekt końcowy, ale także trajektoria systemu.

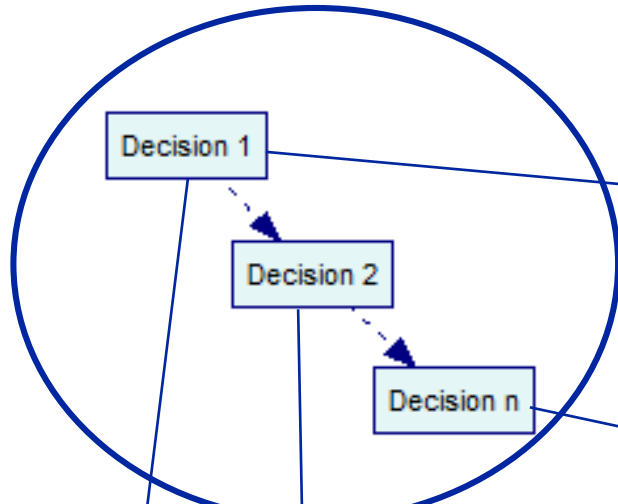
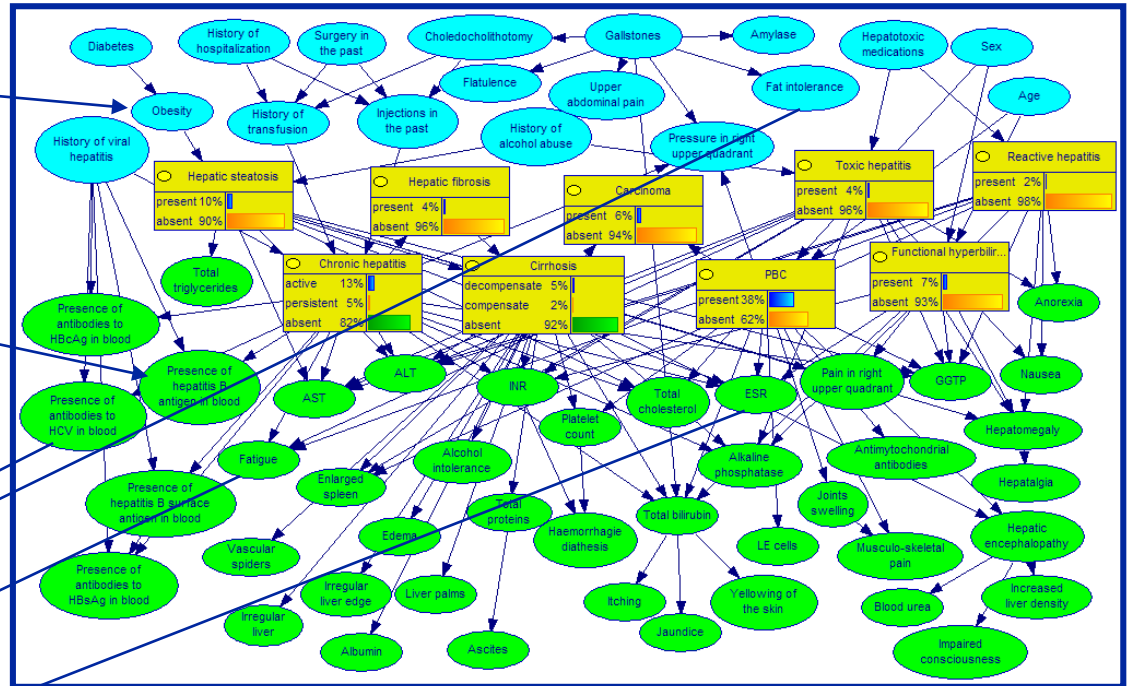


Zainspirowane układami równań różniczkowych (podwaliny pod to położyli Iwasaki i Simon na początku lat 90.)

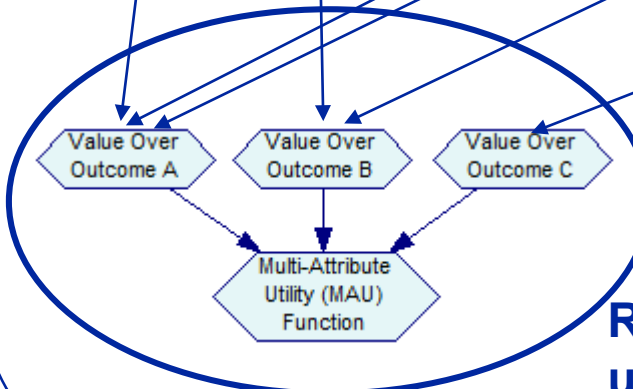


Wspieranie decyzji: Diagramy wpływu

Sieć bayesowska (model systemu)



Reprezentacja decyzji



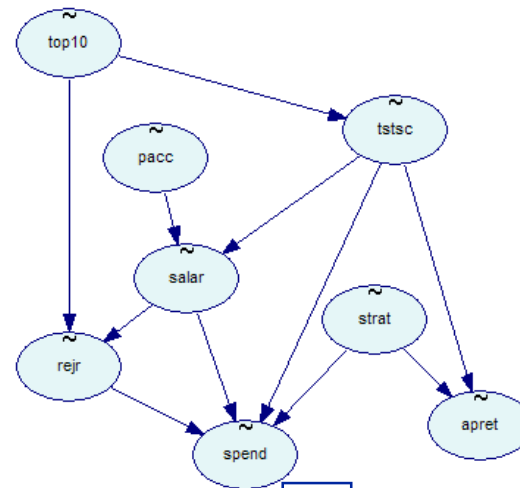
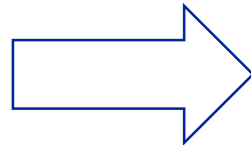
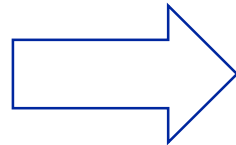
Reprezentacja wartości (funkcje użyteczności, potencjalnie wieloatrybutowe)

Uczenie z danych / eksploracja danych

Istnieją algorytmy pozwalające analizować dane, odkrywać w nich wzorce przyczynowo-skutkowe i budować modele na podstawie tych danych.

spend	apret	top10	rej	tstsc	pacc	strat	salar
9855	52.5	15	29.474	65.063	36.887	12	60800
10527	64.25	36	22.309	71.063	30.97	12.8	63900
7904	37.75	26	25.853	60.75	41.985	20.3	57800
6601	57	23	11.296	67.188	40.289	17	51200
7251	62	17	22.635	56.25	46.78	18.1	48000
6967	66.75	40	9.718	65.625	53.103	18	57700
8489	70.333	20	15.444	59.875	50.46	13.5	44000
9554	85.25	79	44.225	74.688	40.137	17.1	70100
15287	65.25	42	26.913	70.75	28.276	14.4	71738
7057	55.25	17	24.379	59.063	44.251	21.2	58200
16848	77.75	48	26.69	75.938	27.187	9.2	63000
18211	91	87	76.681	80.625	51.164	12.8	74400
21561	69.25	58	44.702	76.25	26.689	9.2	75400
20667	65	68	22.995	75.625	28.038	11	66200
10684	61.75	26	8.774	66	33.99	9.5	52900
11738	74.25	32	25.449	66.875	27.701	12	63400
10107	74	43	11.315	71	29.096	16.2	66200
7817	65.75	36	33.709	64.25	52.548	17.7	54600
7050	26	11	0	55.313	55.651	18.8	59500
9082	83.5	73	64.668	77.375	43.185	13.6	66700
11706	60	56	16.937	73.75	39.479	12.7	62100
7643	49.25	23	36.635	62.813	39.302	18.7	57700
25734	90	77	67.758	80.938	44.133	10	80200
20155	86	84	69.31	79.688	48.766	17.6	74000
29852	94.5	84	75.009	81.313	51.363	10.6	74100
7980	68.5	34	9.122	63.875	35.294	16.3	53100

data



struktura

Success	0.2		
Failure	0.8		
Good	0.4	0.1	
Moderate	0.4	0.3	
Poor	0.2	0.6	

parametry numeryczne

GeNIe

Środowisko do budowy graficznych modeli decyzyjnych (<https://www.bayesfusion.com/>).

Interfejs jakościowy:
QGeNIe

Uczenie i odkrywanie:
SMiner

System do budowy modeli: **GeNIe**.
Zaimplementowany w Visual C++ w otoczeniu Windows.

Wrapery: **SMILE.NET**, **jSMILE**, **rSMILE**,
PySMILE, **SMILE.COM**, **Pocket SMILE**

Pozwalają na dostęp do **SMILE** z aplikacji innych niż kompilator C++

Silnik wnioskujący: **SMILE** (Structural Modeling, Inference, and Learning Engine).

Niezależna od platformy biblioteka klas C++ dla modeli graficznych.

