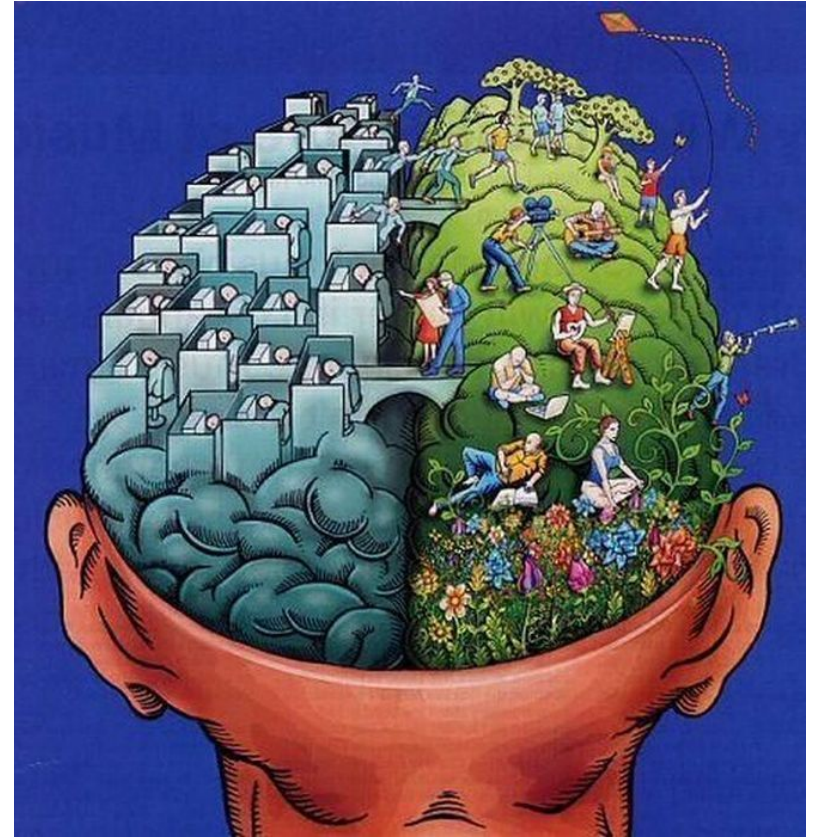


Mesterséges Intelligencia MI

Valószínűségi hálók - alapok

Dobrowiecki Tadeusz
Eredics Péter, és mások



BME I.E. 437, 463-28-99

dobrowiecki@mit.bme.hu,

<http://www.mit.bme.hu/general/staff/tade>

A szükséges apparátus

bináris	Lyuk = Igaz	$P(\text{Lyuk} = \text{Igaz})$	$P(\text{Lyuk})$
többértékű	Időjárás egy a (Napos, Esős, Felhős, Havazás) - ból		$P(\text{Időjárás} = \text{Esős})$
folytonos változó	Hőmérséklet = 22.1 °C,	Hőmérséklet < 22 °C	$P(\text{Hőmérséklet} < 22 \text{ °C})$

$$\begin{aligned} P(X_1 X_2 X_3 \dots X_N) &= P(X_1 X_2 X_3 \dots X_{N-1}) P(X_N | X_1 X_2 \dots X_{N-1}) \\ &= P(X_1 X_2 X_3 \dots X_{N-2}) P(X_{N-1} | X_1 X_2 \dots X_{N-2}) P(X_N | X_1 X_2 \dots X_{N-1}) \\ &= \prod_{i=1 \dots N} P(X_i | X_1 X_2 \dots X_{i-1}) \end{aligned}$$

láncszabály

$P(\text{Fogfájás} | \text{Lyuk}, \text{Fogsz.}) = P(\text{Fogfájás} | \text{Fogsz.})$ **feltételes függetlenség**
ezzel $P(\text{Fogfájás Lyuk} | \text{Fogsz.}) = P(\text{Fogfájás} | \text{Fogsz.}) P(\text{Lyuk} | \text{Fogsz.})$

Összegzés rejtett változókra normalizálás együttes eloszlás

$$P(Y | E = e) = \alpha P(Y, E = e) = \alpha \sum_h P(Y, E = e, H = h)$$

Y: kérdés változó, E: evidencia változó, H: rejtett változó

A szükséges apparátus

Bayes szabály $\rightarrow$ a változók közötti **feltételes függetlenség egyszerűsítheti** a lekérdezett valószínűségek kiszámítását, nagyban **csökkentheti** a megadandó feltételes valószínűségek számát.

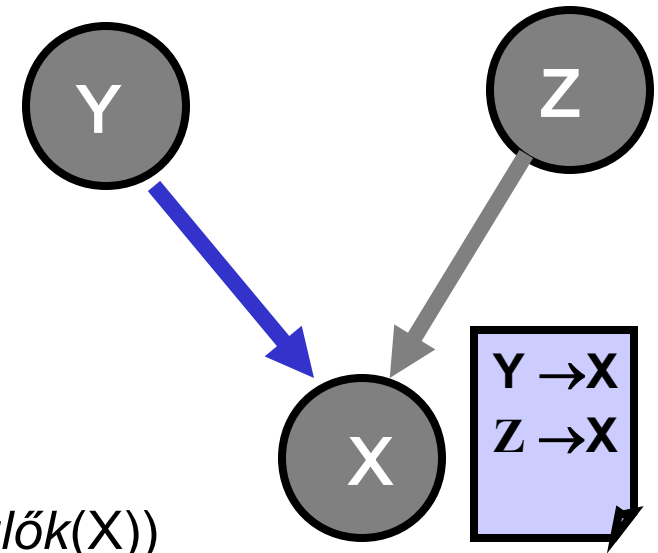
$$\underset{\text{a posteriori}}{\underset{\text{ok}}{P(\text{Fogsz.} \mid \text{Lyuk, Fogfájás})}} = \alpha \underset{\text{tünet1}}{P(\text{Fogfájás, Lyuk} \mid \text{Fogsz.})} \underset{\text{tünet2}}{P(\text{Fogsz.})}$$

$$= \alpha \underset{\text{a priori}}{P(\text{Fogsz.})} \times \underset{\text{tünet1 okból a priori}}{P(\text{Fogfájás} \mid \text{Fogsz.})} \times \underset{\text{tünet2 okból a priori}}{P(\text{Lyuk} \mid \text{Fogsz.})}$$

$$\begin{aligned} &P(\text{Ok} \mid \text{Hatása}_1, \text{Hatása}_2, \dots \text{Hatása}_N) \\ &= \alpha P(\text{Ok}) P(\text{Hatása}_1 \mid \text{Ok}) P(\text{Hatása}_2 \mid \text{Ok}) \dots P(\text{Hatása}_N \mid \text{Ok}) \end{aligned}$$

Valószínűségi háló → egy gráf

1. Csomópontok: valószínűségi változók egy halmaza
2. Csomópontok között: irányított élek halmaza:
Az Y csomópontot az X csomóponttal összekötő nyíl →
az *Y-nak közvetlen befolyása van az X-re*
3. Minden csomópont:
feltételes valószínűségi tábla →
szülők hatása a csomópontra $P(X \mid \text{Szülők}(X))$



4. A gráf nem tartalmaz irányított kört (= **irányított, körmentes gráf = DAG**).

Valószínűségi változó = egy állítás a problémáról

Pl. X = HMG szint magas (két érték: I/N), vagy

X = HMG szint (több érték: Magas / Közepes / Alacsony)

Példa

Otthonunkban egy új **riasztót** szereltünk fel. Ez megbízhatóan észleli a **betöréseket**, de kisebb **földrengés** esetén is jelez.

Két szomszédunk, **János** és **Mária**, megígérték, hogy **felhívnak** a munkahelyünkön, ha meghallják a riasztónkat.

János mindig felhív, ha szól a riasztó, de néha összekeveri a telefoncsörgést a riasztó csengésével és ekkor is telefonál.

Mária viszont, mivel szereti hangosan hallgatni a zenét, néha meg sem hallja a riasztót.

Mi tehát a hívások bekövetkezte vagy hiánya alapján szeretnénk megbecsülni a betörés valószínűségét.

(Bináris) változók (tények):

- Betörés** (megtörtént): Igen/ Nem
- Földrengés** (megtörtént): Igen/ Nem
- Riasztás** (megszólalt): Igen/ Nem
- János telefonál**(t-e): Igen/ Nem
- Mária telefonál**(t-e): Igen/ Nem

Háló topológiája = egy absztrakt tudásbázis

sok esetben érvényes marad, mivel

általános ok-okozati

folyamatokat ír le az adott problémakörben.

A betöréses háló esetén a topológia azt mutatja, hogy a betörés és a földrengés közvetlenül hat a riasztóra, befolyásolja megszólalásának valószínűségét, ellenben János vagy Mária hívásának bekövetkezése csak magán a riasztón múlik – azaz a háló tartalmazza azt a feltevést, hogy ők közvetlenül nem vesznek észre sem a betöréseket, sem a földrengéseket.



Feltételes valószínűségi tábla – FVT

minden egyes csomópontra.

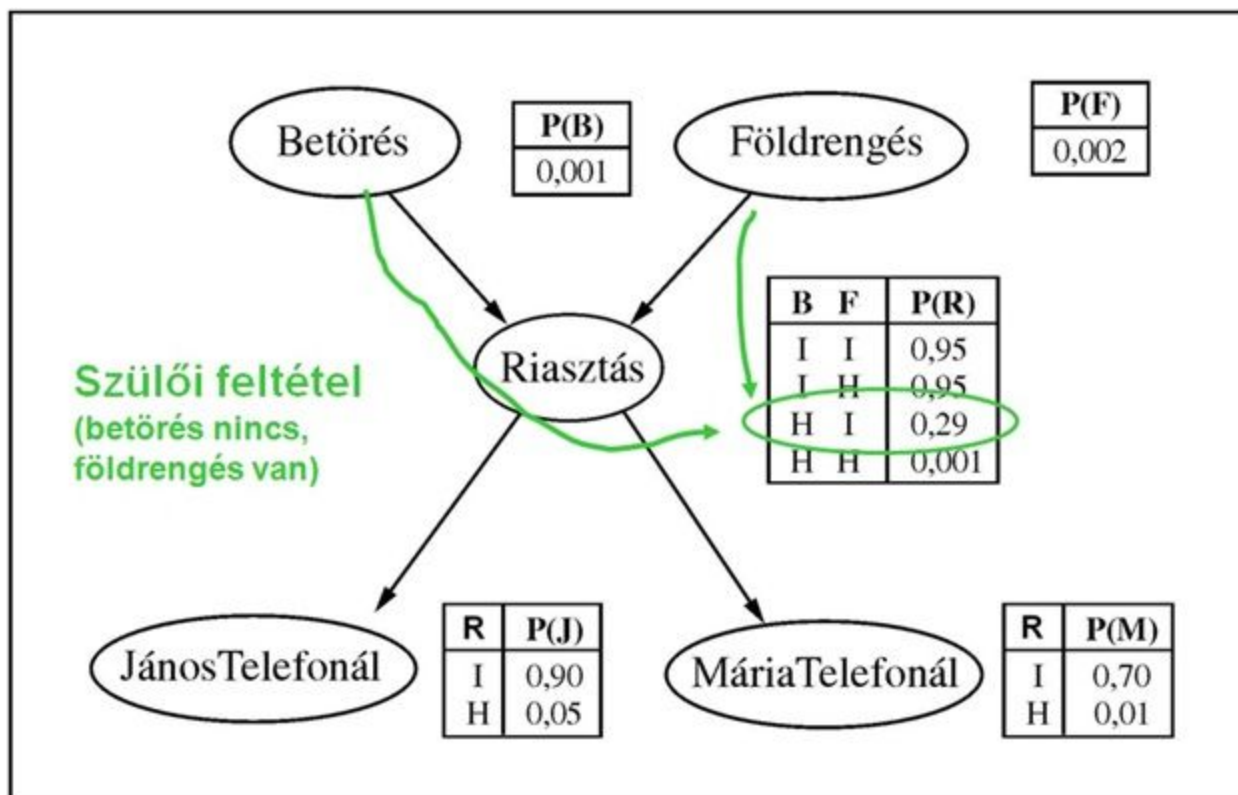
Egy sor a táblázatban az egyes csomóponti értékek feltételes valószínűsége az adott sorhoz tartozó **szülői feltétel** esetén.

Szülői feltétel: a szülő csomópontok értékeinek egy lehetséges kombinációja (**egyfajta elemi esemény a szülők között**).

Általánosabban, ha a változó **bináris**, és ha n **bináris** szülője van:

$2^n - 1$ valószínűség adható meg a feladat alapján.

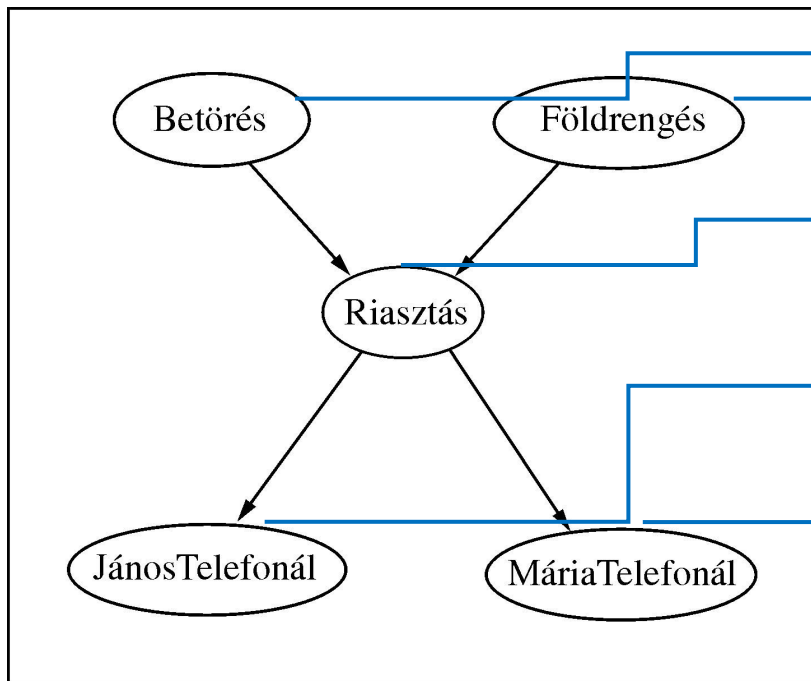
Szülő nélküli csomópont – a változó egyes értékeinek a priori valószínűségei (ha bináris, akkor csak egy).



Együttes eloszlás:

5 változó

= $2^5 - 1 = 31$ db valószínűség



Bayes háló:

2 x 1 db a priori

+

1 x 4 db szülői feltétel

+

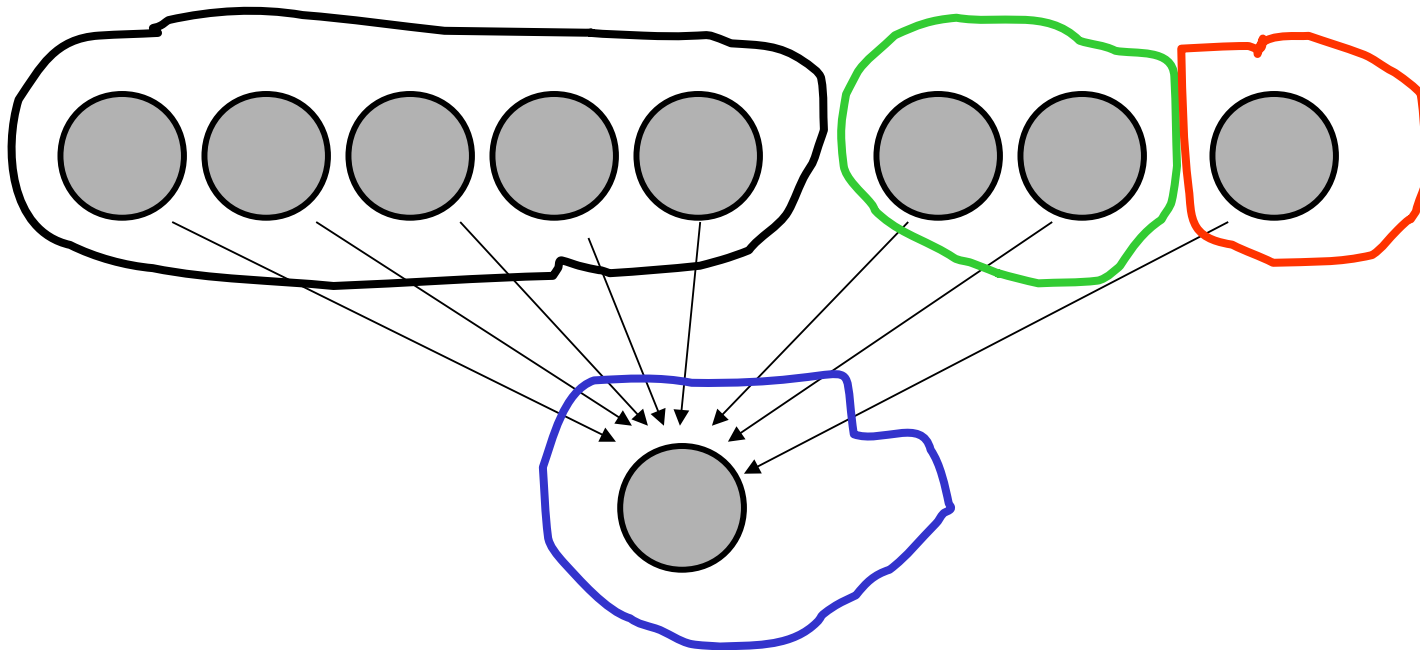
2 x 2 db szülői feltétel

= 2 + 4 + 4 = 10 db

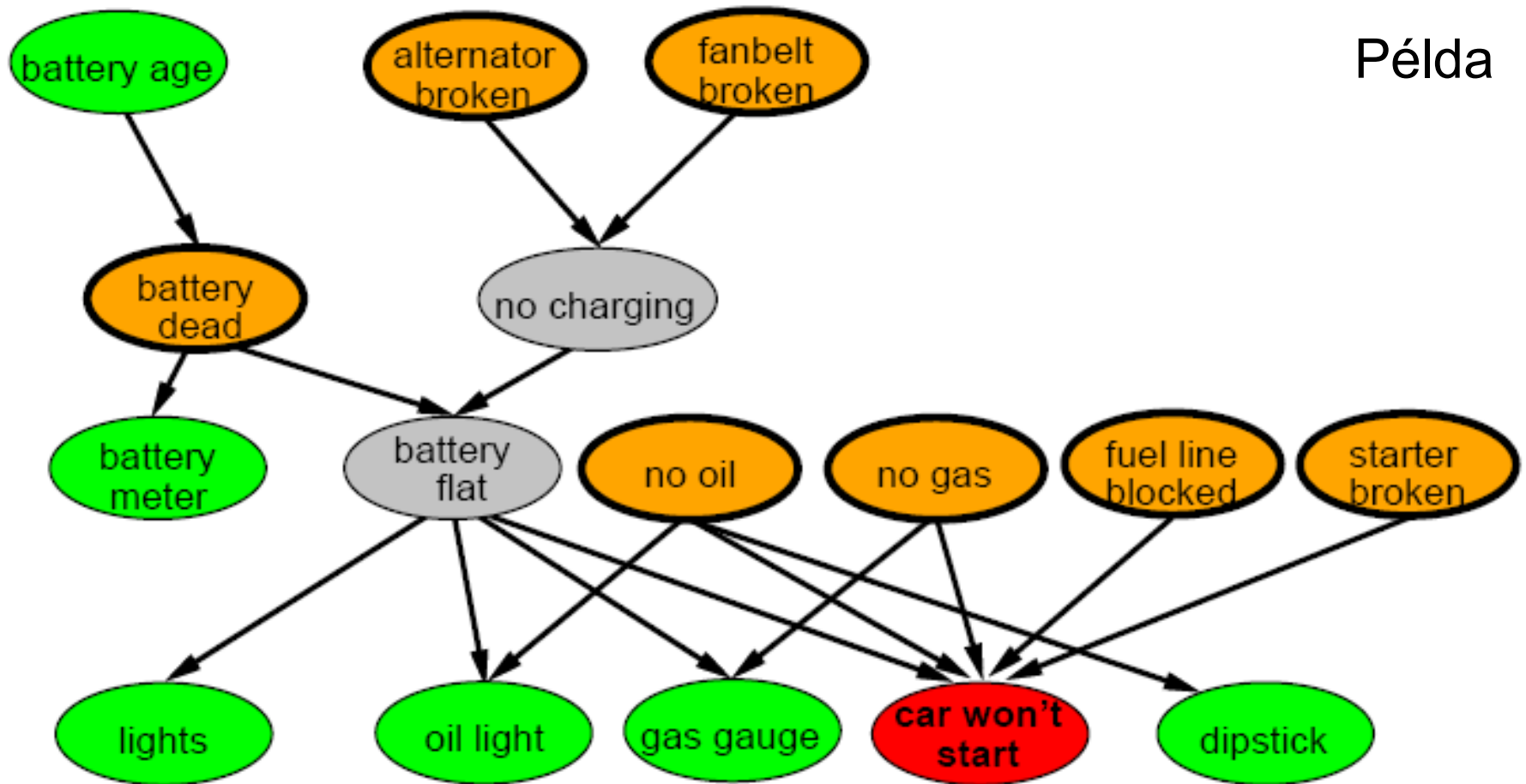
Elég-e?

Általános eset:

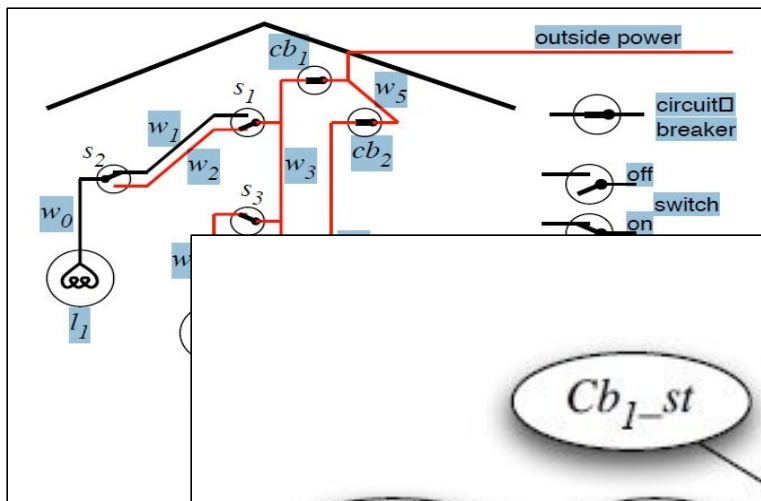
Mi van, pl. ha 5 db szülő csomópont bináris értékű
 2 db szülő csomópont 3-as értékű
 1 db szülő csomópont 4-es értékű és
 az eredmény csomópont egy 5-ös értékű
 változó?



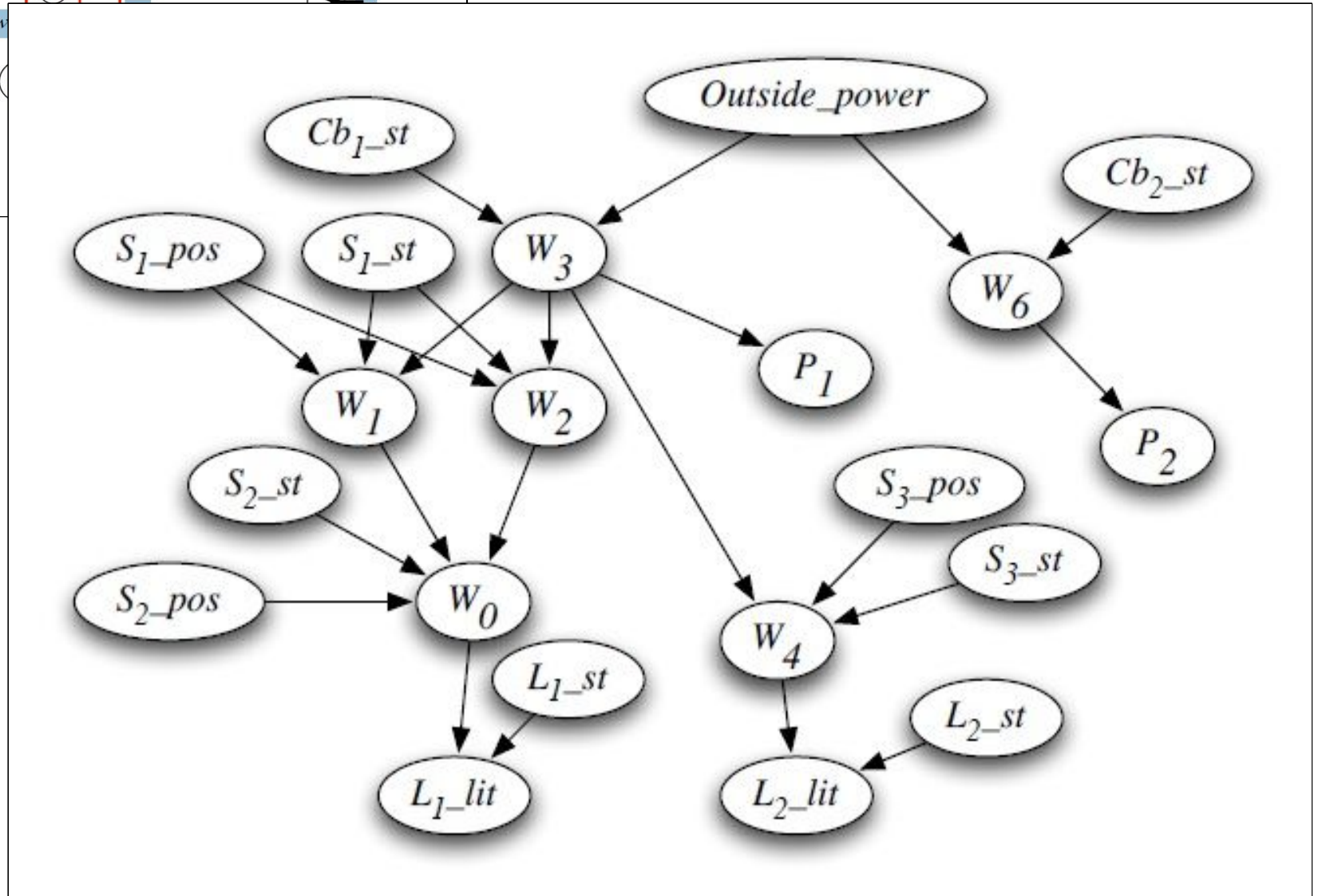
Példa

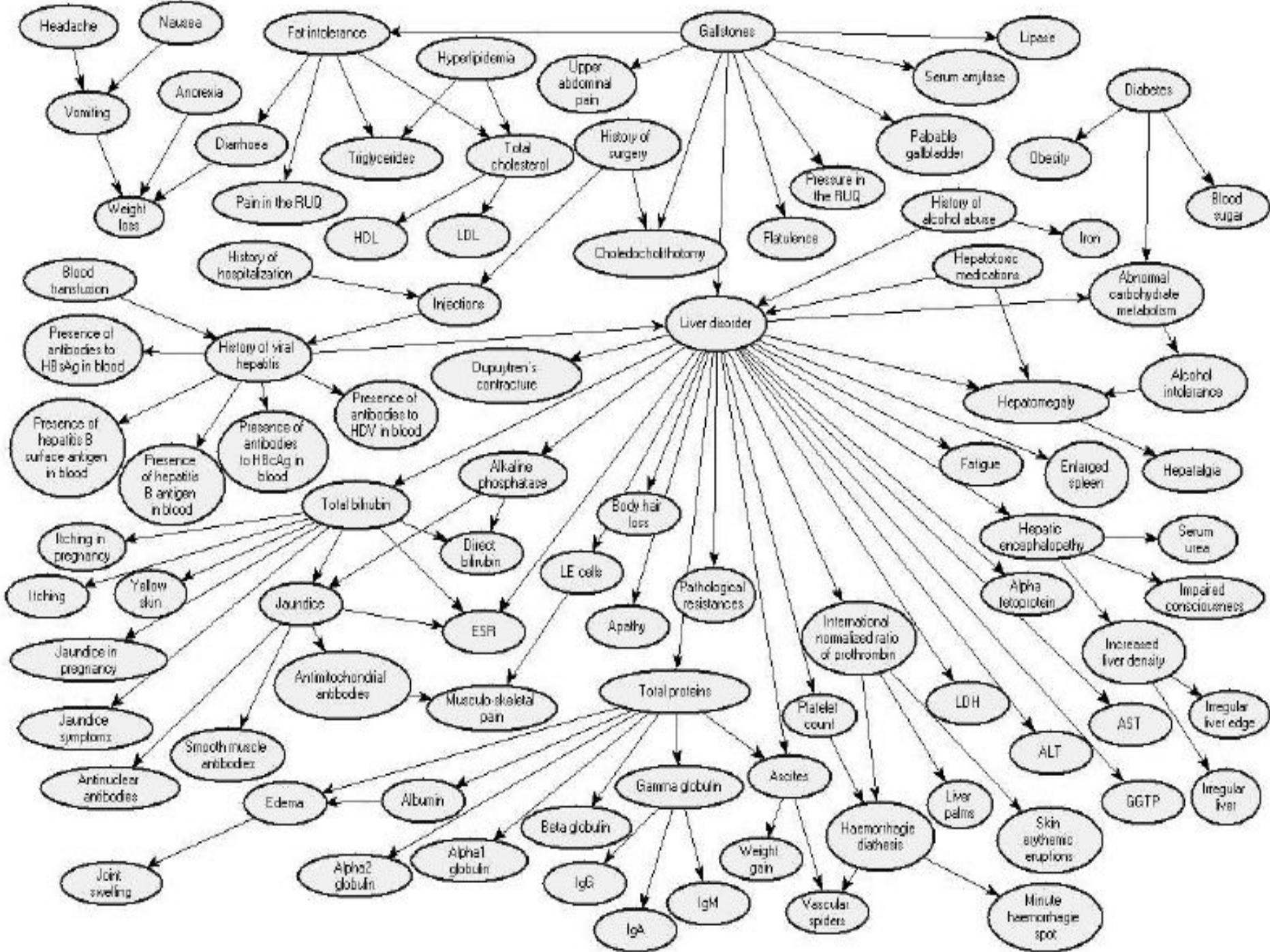


16 (bináris) változó = $2^{16}-1 = 16383$ valószínűség
háló = $7 + 4 \times 2 + 4 \times 4 + 1 \times 32 = 63$ valószínűség (x 260)



2097151 : 71 x 30.000





A valószínűségi hálók szemantikája

- (1) a háló = **együttes valószínűségi eloszlás** egy leírása,
- (2) a háló = **feltételes függetlenségekről szóló állítások** együttese.

A két szemlélet **ekvivalens**

az első abban segít, hogy hogyan **hozzunk létre** egy hálót
a második abban, hogy hogyan **tervezzünk** következtetési eljárásokat.

Feltételes függetlenség

Ha a szülő (szülői feltétel) ismert, nem érdekes az ő:

$$P(X_i | X_{i-1}, \dots, X_1) = P(X_i | \text{Szülő}(X_i)), i=1 \dots n,$$

ha $\text{Szülő}(X_i) \subseteq \{X_{i-1}, \dots, X_1\}$.

Ez utóbbi könnyen teljesíthető a csomópontok olyan sorszámozásával,
ami konzisztens a gráf implicit részleges rendezésével.

$$P(X_i | \text{Szülő}(X_i), \text{Ősök}(X_i)) = P(X_i | \text{Szülő}(X_i))$$

Az együttes valószínűségi eloszlásfüggvény leírása

A valószínűségi hálóban található információk alapján:
az együttes valószínűségi eloszlás bármely bejegyzése kiszámítható.

Egy bejegyzés értéke (globális szemantika):

$$P(X_1 \dots X_n) = \prod_{i=1 \dots n} P(X_i | \text{Szülők}(X_i)), i=1 \dots n$$

Az együttes valószínűségi eloszlás minden bejegyzése felbontható a FVT megfelelő elemeinek a szorzatára.

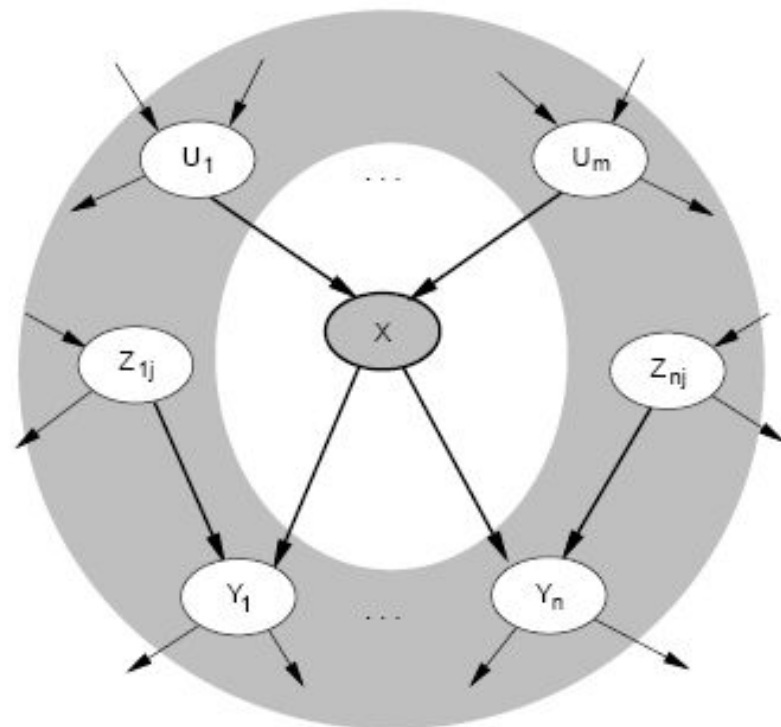
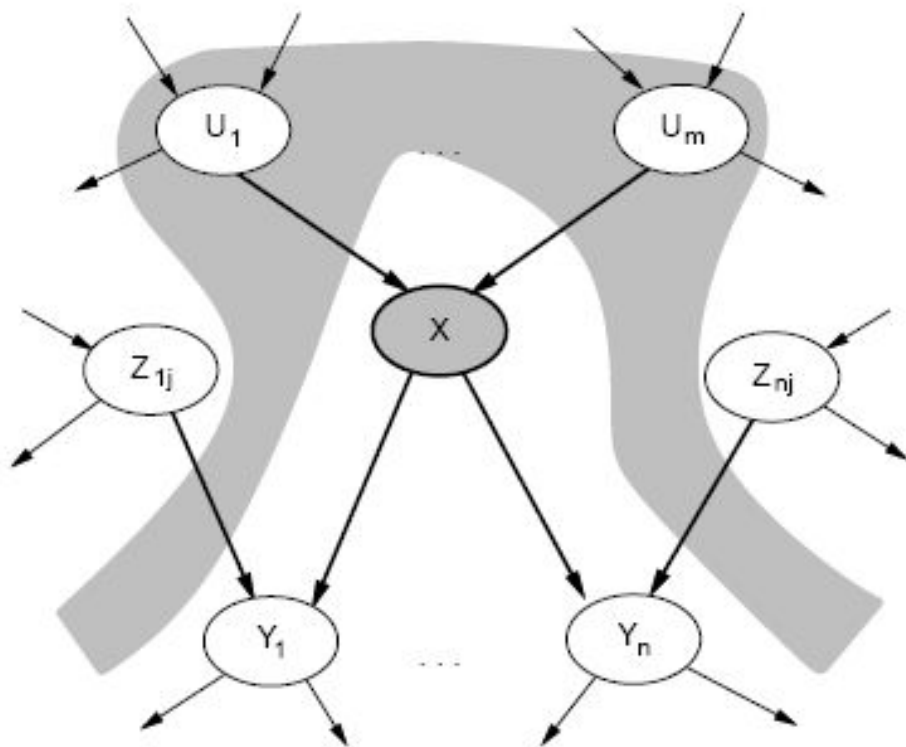
FVT-k: az együttes valószínűségi eloszlás dekomponált leírása.

Egy elemi esemény: pl. a riasztó megszólal, de sem betörés, sem földrengés nem volt, azonban János is és Mária is telefonál:

$$\begin{aligned} P(J \ M \ R \ \neg B \ \neg F) &= P(J | \ M \ R \ \neg B \ \neg F) \ P(M \ R \ \neg B \ \neg F) \\ &= P(J | R) \ P(M | R \ \neg B \ \neg F) \ P(R \ \neg B \ \neg F) \\ &= P(J | R) \ P(M | R) \ P(R | \neg B \ \neg F) \ P(\neg B) \ P(\neg F) \\ &= .90 \times .70 \times .001 \times .999 \times .998 = .00062 \end{aligned}$$

Lokális (topológiai) szemantika:

- a. Minden csomópont (X) feltételesen független a nem leszármazottjaitól (Z), ha a **szülők** adottak (U).
- b. Minden csomópont (X) feltételesen független minden mástól, ha a **Markov-takarója** adott (szülők+gyerekei+gyerekeinek szülők).
- c. **d-elválasztás** ...



Valószínűségi hálók építése

Egy valószínűségi háló csak akkor helyes reprezentáció, **ha minden csomópont feltételesen független a sorrendben őt megelőzőktől, feltéve, hogy a szülők értékei ismertek.**

Az X_i csomópont szülei halmazának tartalmaznia kell mindazokat a változókat az $\{X_1, \dots, X_{i-1}\}$ közül, amik közvetlenül befolyásolják X_i -t.

Tételezzük fel, hogy a hálót már teljesen befejeztük, csupán a *MáriaTelefonál* szüleit kell már megválasztani.

A *MáriaTelefonál*-t egyértelműen befolyásolja, hogy történt-e *Betörés* vagy *Földrengés*, de nem közvetlenül - csak a riasztó által.

Hasonlóan, ha a riasztó állapota ismert, akkor János hívásának bekövetkezte vagy elmaradása már nincs hatással Mária telefonálására.

$P(\text{MáriaTel.} \mid \text{JánosTel.}, \text{Riasztás}, \text{Földrengés}, \text{Betörés}) =$

$P(\text{MáriaTel.} \mid \text{Riasztás})$

Az általános eljárás egy háló fokozatos megépítésére:

1. Határozzuk meg a problémát leíró változókat (miről érdemes beszélni).
1. Határozzunk meg **egy sorrendet**.
3. Ameddig maradt még érintetlen változó:
 - a) Válasszuk a következő X_i változót és adjunk egy csomópontot a hálózhoz.
 - b) Legyen a $Szülő_k(X_i)$ a csomópontok azon minimális halmaza, amik már szerepelnek a hálóban és a feltételes függetlenség tulajdonságát teljesítik – húzzuk be a kapcsolatokat.
 - c) Definiáljuk X_i csomópont feltételes valószínűségi tábláját.

Mindegyik csomópontot csak korábbi csomópontokhoz csatlakoztathatunk
= a háló körmentes lesz.

A valószínűségi háló nem tartalmaz redundáns valószínűségi értékeket, kivéve soronkénti bejegyzéseket a feltételes valószínűségi táblában.

Tömörség és a csomópontok sorrendje

Egy valószínűségi háló gyakran sokkal **tömörebb**, mint az együttes valószínűségi eloszlásfüggvény.

A valószínűségi háló tömörsége a **lokálisan strukturált** (vagy **ritka**) rendszerek egy példája.

Lokálisan strukturált rendszer: egy komponense csak korlátos számú más komponenssel van kapcsolatban közvetlenül, függetlenül a komponensek teljes számától. Lokális struktúrák: **inkább a lineáris, mint az exponenciális komplexitás** növekedés.

Valószínűségi háló n változóból: legtöbb esetben egy változót csak k számú más változó befolyásol, bináris változók = $n \times 2^k$ érték. Együttes valószínűségi eloszlás = $2^n - 1$ érték.

Példa: 20 cs.pont ($n = 20$), max. 5 szülője legyen ($k = 5$), valószínűségi háló: **640** érték, együttes valószínűségi eloszlás: > **egy millió**.

A gyakorlatban előforduló információcsökkenés miatt lép fel, hogy a valós problémák igen strukturáltak, amit a hálók könnyűszerrel képesek kihasználni.

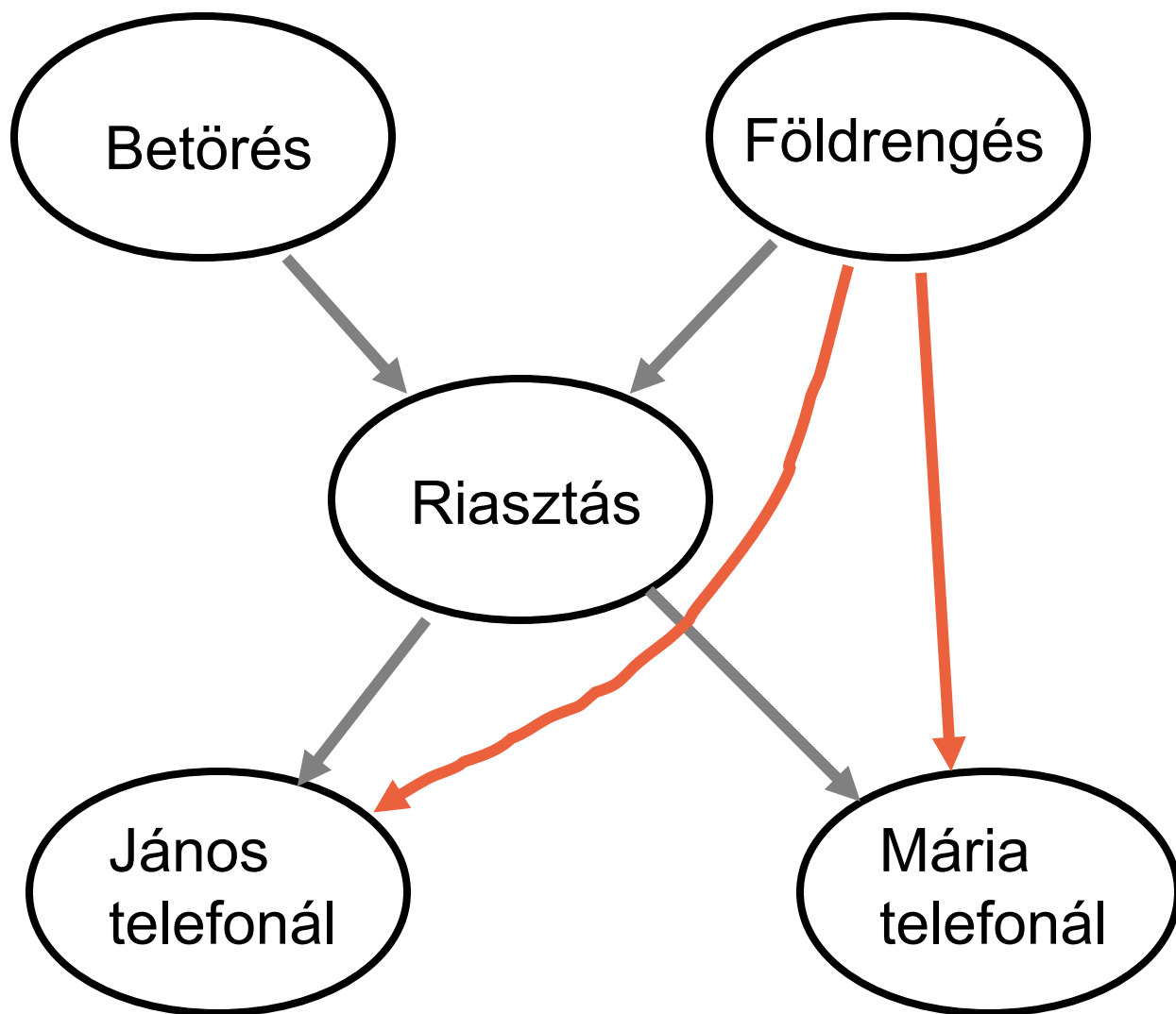
Tömörség és a csomópontok sorrendje

Bizonyos tárgytartományokban létezhetnek olyan **jelentéktelennek tűnő függőségek**, amiket feltétlenül modellezni kell egy új kapcsolat felvételével.

De ha ezek a függőségek ténylegesen jelentéktelenek, akkor lehet, hogy nem éri meg a háló komplexitását megnövelni a pontosság kismértékű növelésének érdekében.

Pl.:

hogya földrengés van, akkor Mária és János akkor sem telefonálna, ha hallanák a riasztót, mivel feltételezik, hogy a földrengés okozta.



$$2 + 4 + 2 + 2 = 10 \text{ (eddig)}$$

$$2 + 4 + 4 + 4 = 14$$

Megéri-e?

Tömörség és a csomópontok sorrendje

A konstrukciós eljárás működése miatt előbb a „közvetlen befolyásolók”-at kell a hálózathoz adni, ha azt szeretnénk, hogy szülőknek tudjuk őket választani az általuk befolyásolt csomópontnál.

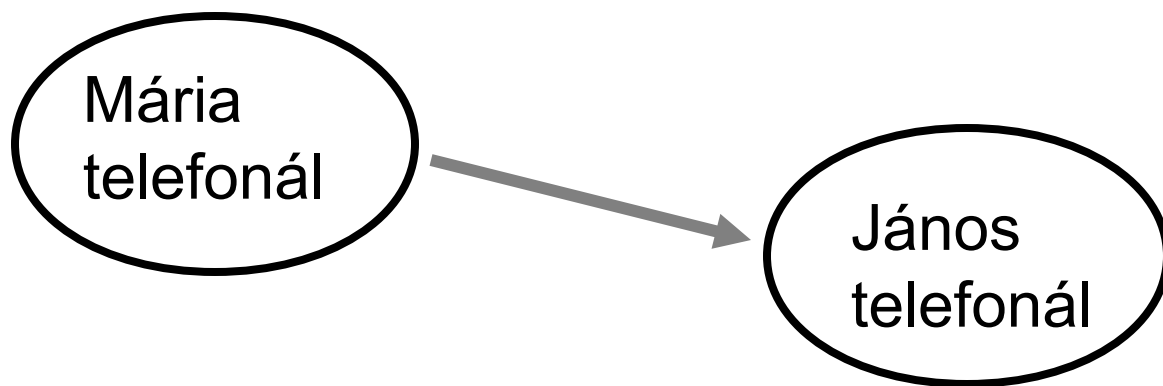
Ezért a helyes sorrend a csomópontok hozzáadásánál: először az „alapvető okokat” adjuk a hálózathoz, majd a változókat, amiket befolyásolnak, és ezt addig folytatjuk, amíg el nem érjük a „leveleket”, amiknek már nincs közvetlen okozati hatása más változókra (azaz a kauzalitás sorrendjében).

Mi történik, ha történetesen egy rossz sorrendet választunk?
pl.

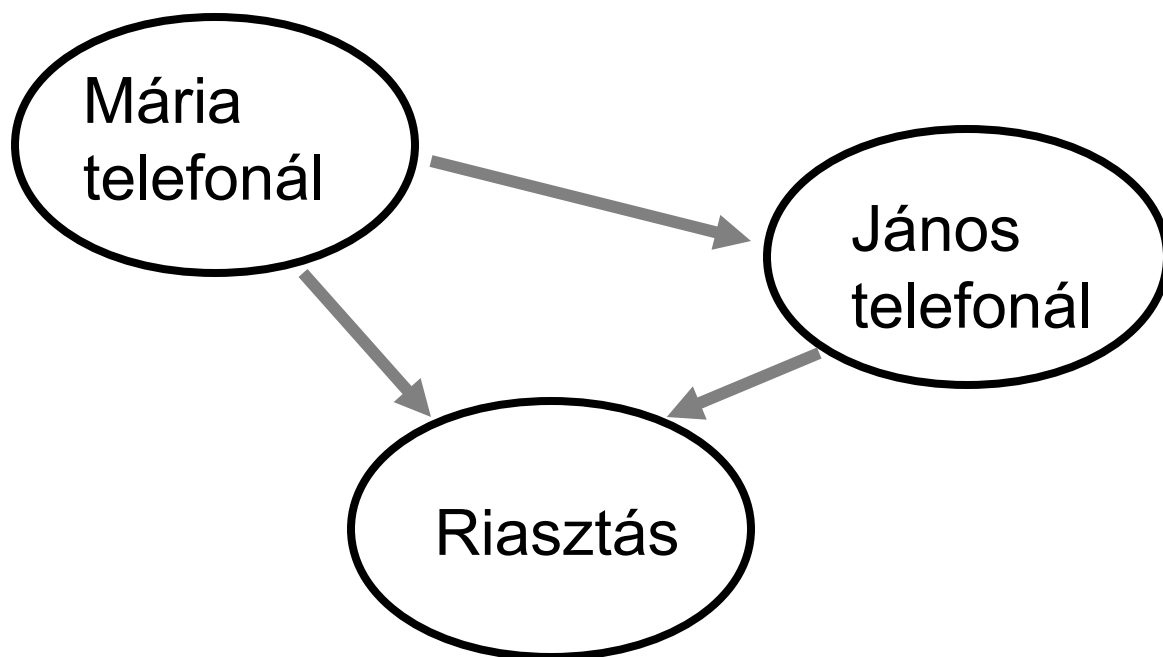
MáriaTelefonál → JánosTelefonál → Riasztás → Betörés → Földrengés.



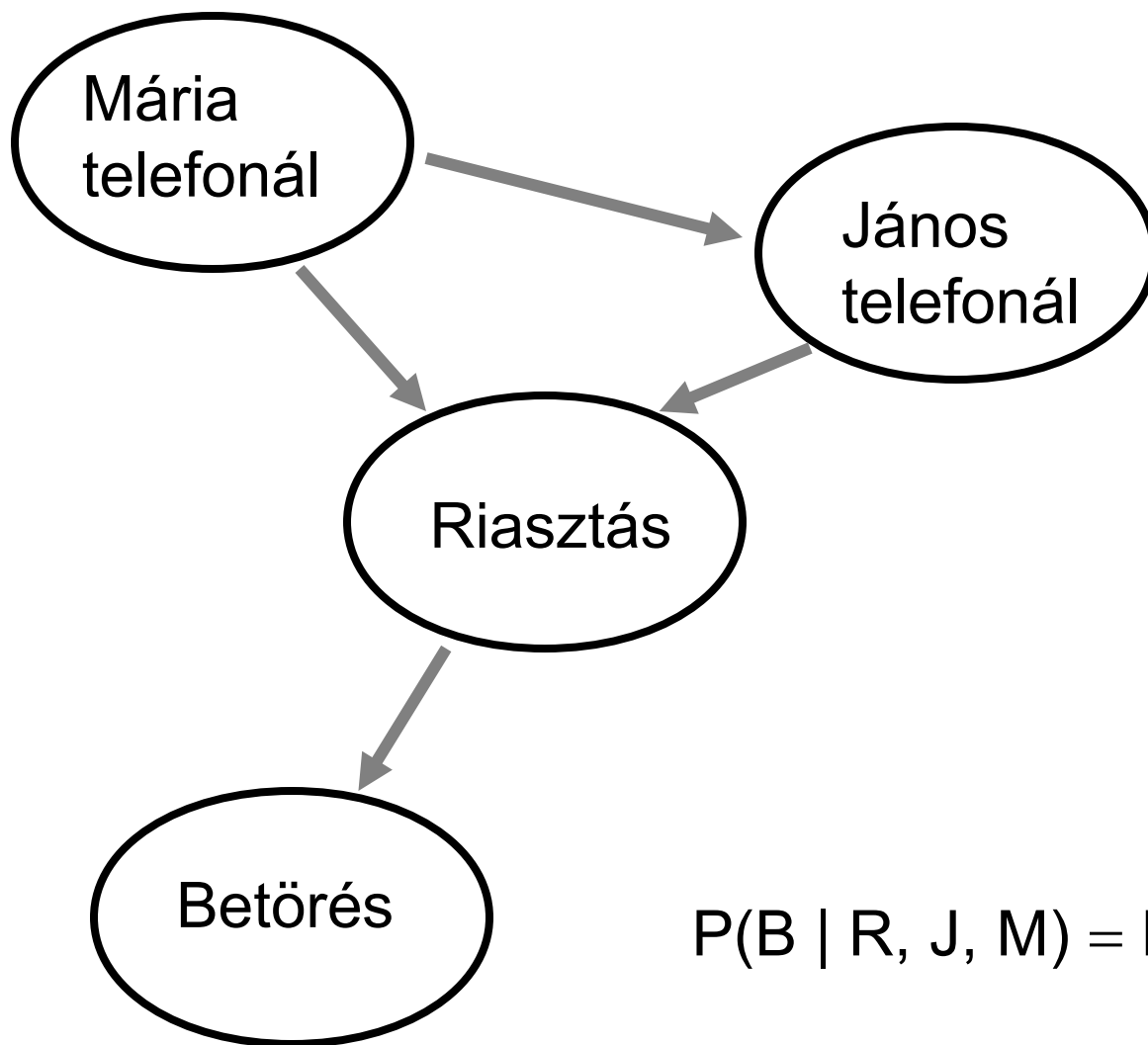
Mária
telefonál



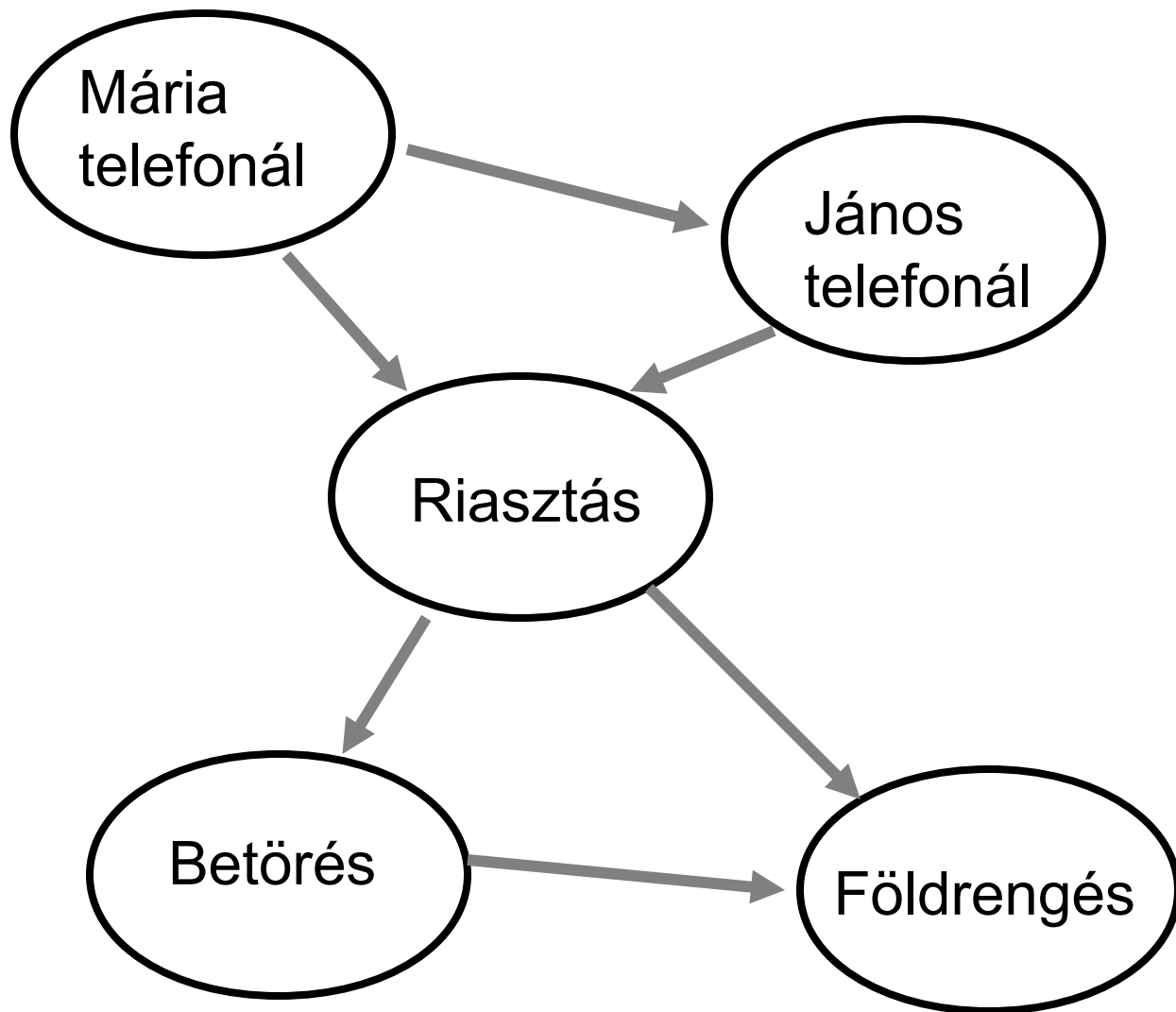
$P(J \mid M) = P(J)$? Nem (azaz van nyíl)



$P(R \mid J, M) = P(R \mid J)?$ $P(R \mid J, M) = P(R)?$ Nem



$P(B \mid R, J, M) = P(B \mid R)?$ Igen



$$\begin{aligned} &1 + 2 + 4 \\ &+ 2 + 4 \\ &= 13 \end{aligned}$$

$$P(F \mid B, R, J, M) = P(F \mid B, R)? \quad = P(F \mid R)?$$

Igen Nem

Mi történik, ha történetesen egy nagyon rossz sorrendet választunk?

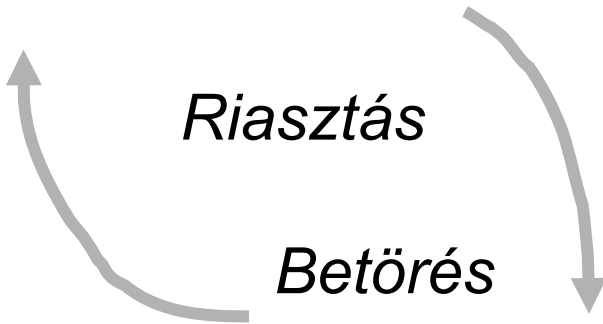
MáriaTelefonál

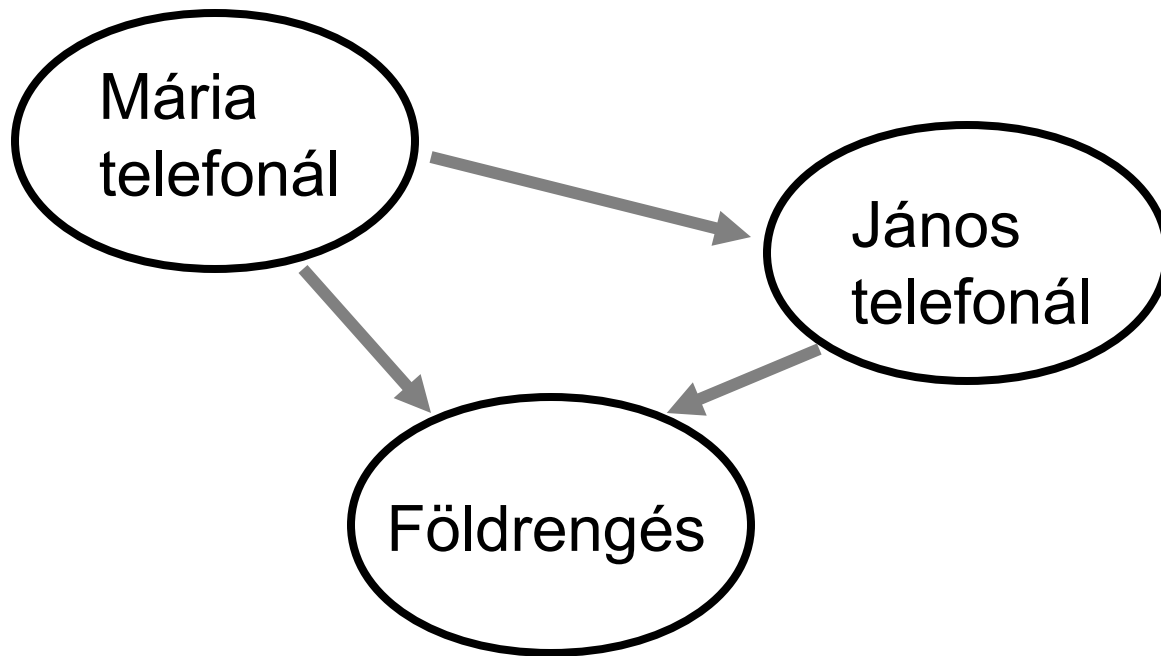
JánosTelefonál

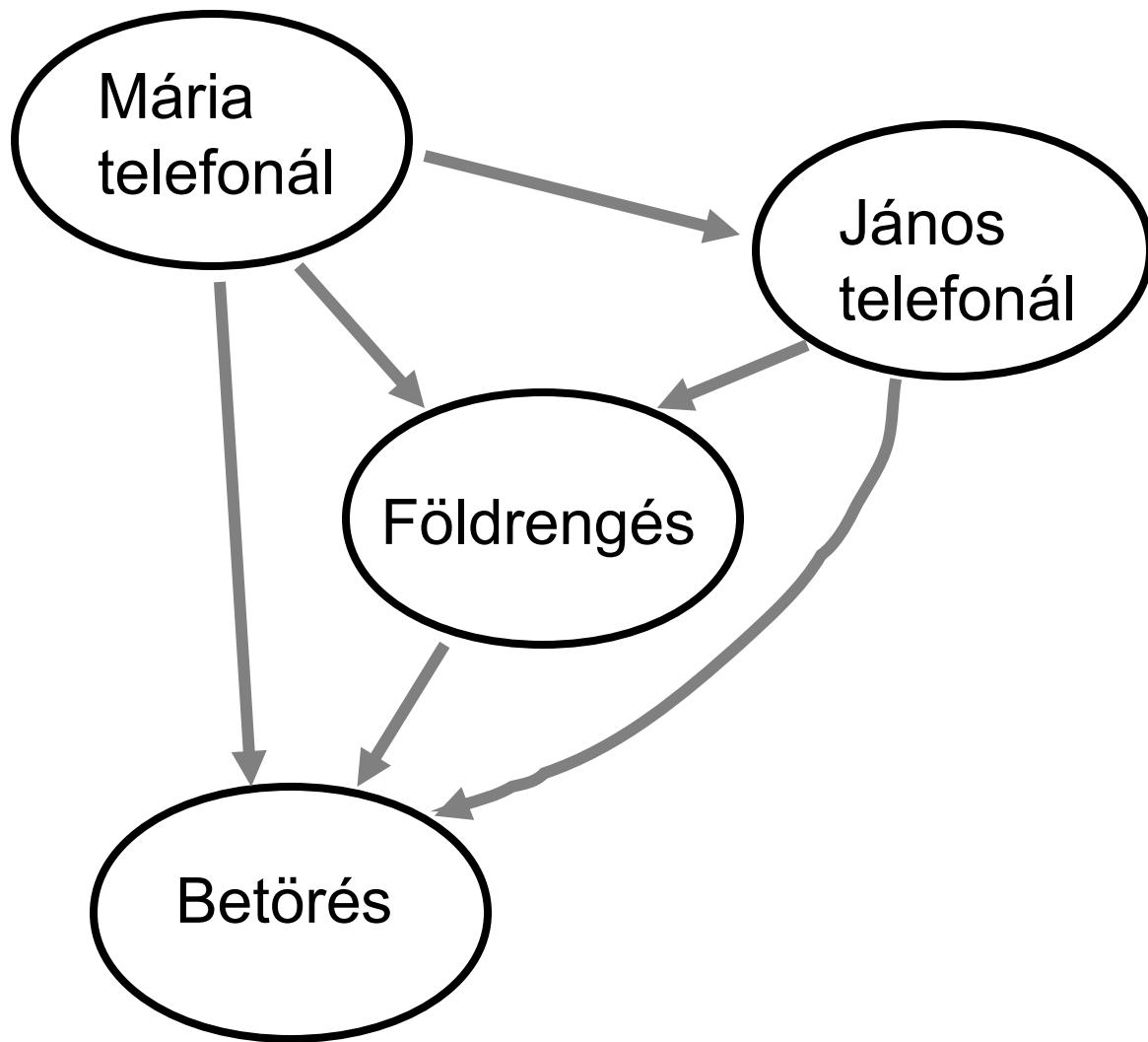
Földrengés.

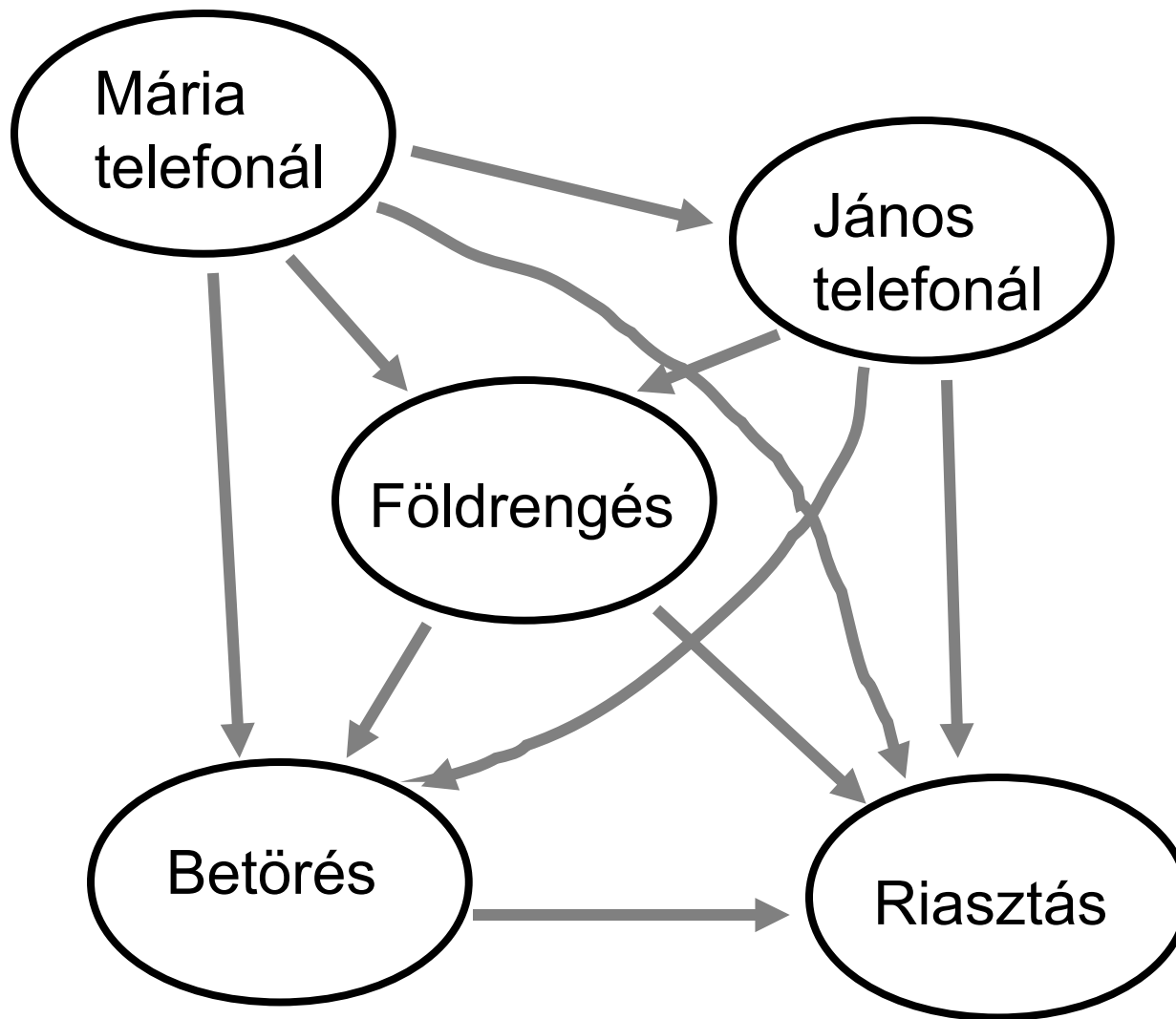
Riasztás

Betörés









$$\begin{aligned} &1 + 2 + 4 \\ &+ 8 + 16 \\ &= 31 \\ &= 2^5 - 1 \end{aligned}$$

Sok (elméletileg exponenciális számú) **valószínűség**

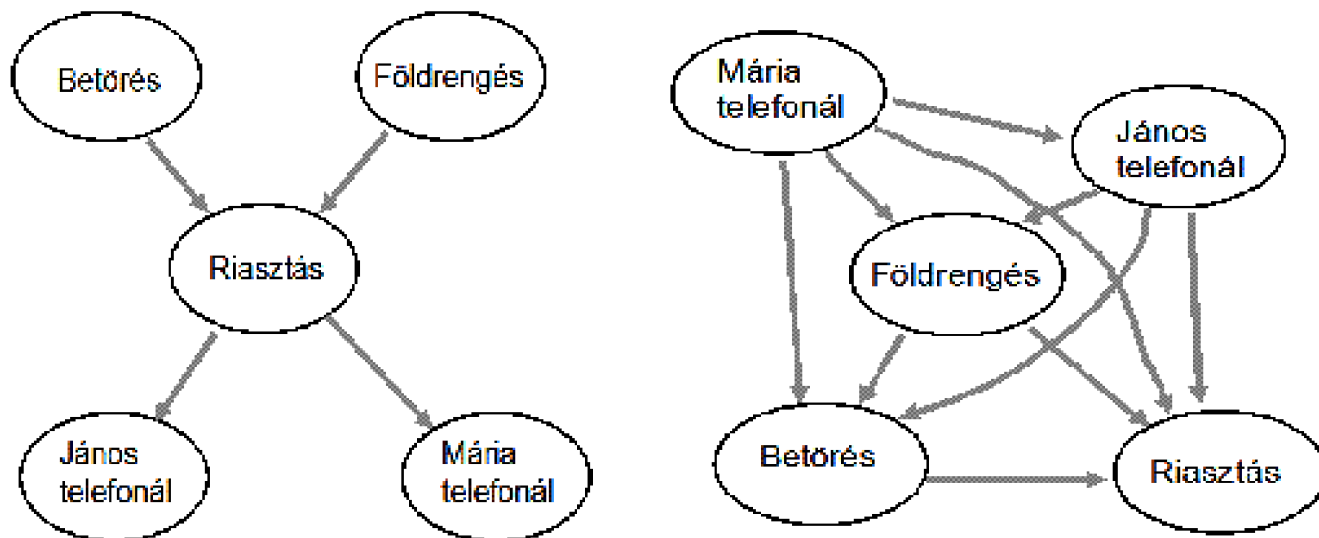
De a legrosszabb következmény:

néhány kapcsolódás megfoghatatlan viszonyt reprezentál,
ami nehéz és **nem természetes valószínűségi ítéleteket igényel**, például

$P(\text{Földrengés} \mid \text{Mária} \dots, \text{János} \dots) ?$

$P(\text{Betörés} \mid \text{Földrengés}) ? \dots \dots \dots$

Okozati (kauzális) modell: kevesebb, megbízhatóbb érték, az értékeket gyakran könnyebb elérni



Feltételes eloszlások hatékony reprezentációja

K db. szülő $O(2^k)$!

Kanonikus eloszlások

determinisztikus, numerikus szülő-gyerek kapcsolat

zajos-VAGY szülők = az összes ok (+ esetleg „szivárgás” csomópont), minden okra külön-külön **kudarc (gátlás) valószínűség**, gátlások függetlenek,

L – láz, F – influenza, G – megfázás, M - malária

$$P(\neg L \mid \mathbf{G}, \neg F, \neg M) = 0.6$$

$$P(\neg L \mid \neg G, \mathbf{F}, \neg M) = 0.2$$

$$P(\neg L \mid \neg G, \neg F, \mathbf{M}) = 0.1$$

8 érték helyett 3

N szülő esetén 2^N érték helyett N

G	F	M	P(L)	P($\neg L$)
H	H	H	0.0	1.0
H	H	I	...	0.1
H	I	H	...	0.2
H	I	I		0.1×0.2
I	H	H	...	0.6
I	H	I		0.6×0.1
I	I	H		0.6×0.2
I	I	I		$0.6 \times 0.2 \times 0.1$

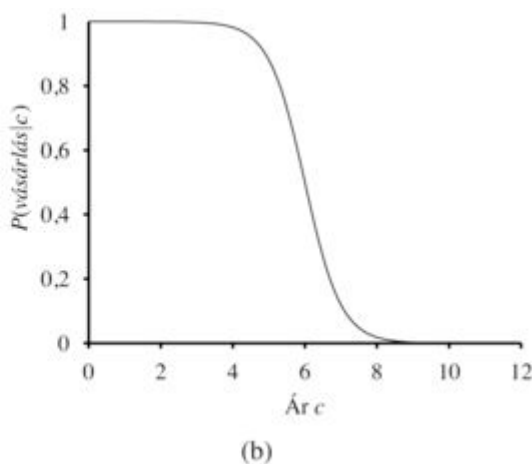
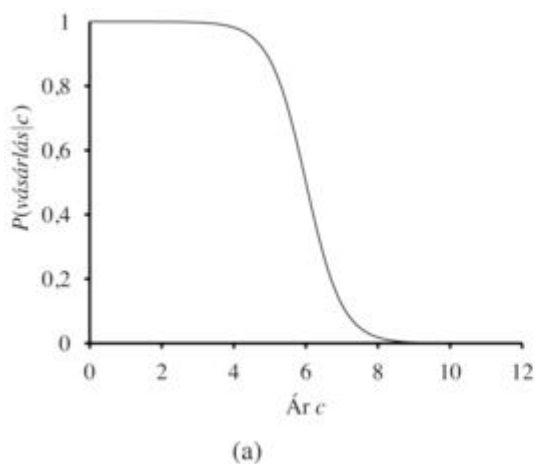
Feltételes eloszlások hatékony reprezentációja

Kanonikus eloszlások – folytonos változók

- **lineáris Gauss:** diszkrét (sz) → folytonos (gy)

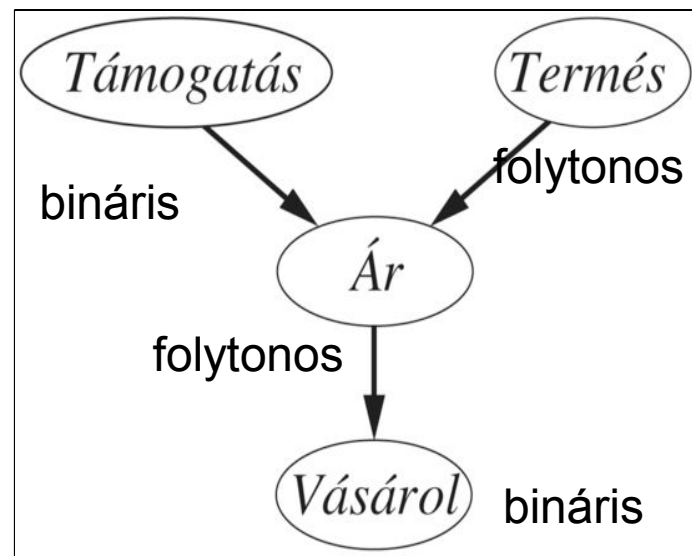
- **probit:** folytonos → diszkrét, „lágymű” küszöb

$$P(\text{Vásárlás} \mid \text{Ár} = a) = \Phi((-a + \mu)/\sigma)$$



$$\frac{1}{1 + \exp\left(-2\frac{-c + \mu}{\sigma}\right)}$$

- **logit:** szigmoid



$$P(a \mid t, \text{Támogatás}) = N(a_t + b_t, \sigma_t^2)$$
$$P(a \mid t, \neg \text{Támogatás}) = N(a_h + b_h, \sigma_h^2)$$

Következtetés valószínűségi hálóokban

Az alapvető feladat:

kiszámítani az a posteriori valószínűséget a **lekérdezéses változókra**, ha a **tény** vagy **bizonyíték változóknak** az értékei adottak:

$$P(\text{Lekérdezéses} \mid \text{Bizonyíték})$$

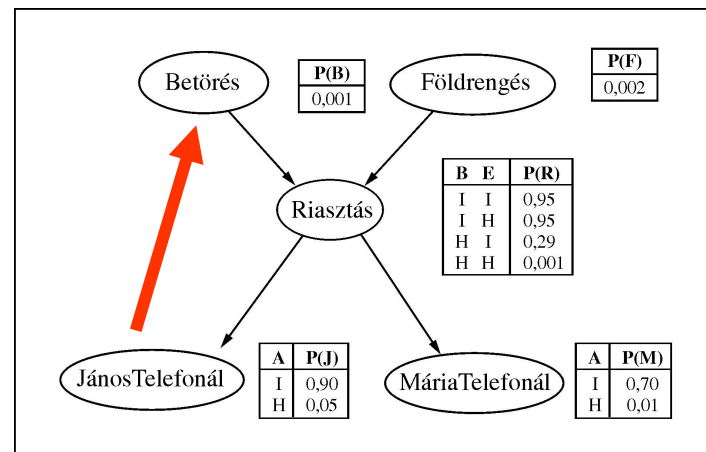
A riasztós példában, pl.:

$$P(\text{Betörés} \mid \text{MáriaTelefonál és JánosTelefonál})$$

Általában az ágens az érzékeléséből (vagy egyéb következtetésből) kap értékeket a tény változókhoz, és más változók lehetséges értékeiről kérdez, hogy el tudja dönteni milyen cselekvéseket végezzen.

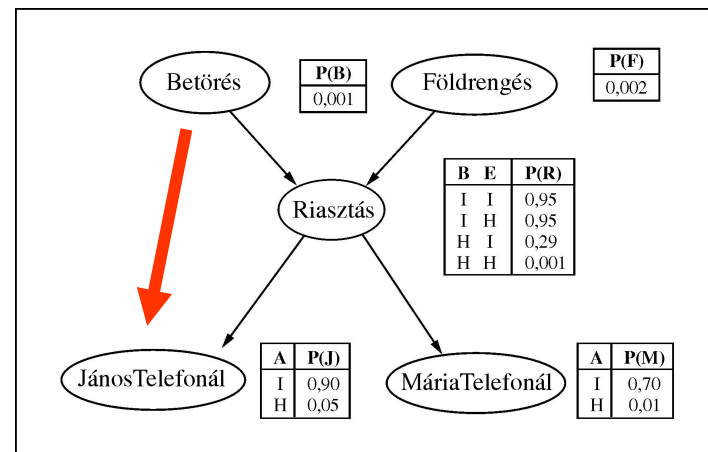
Diagnosztikai következtetés (hatásról az okra)

Ha adott a *JánosTelefonál*, akkor
kiszámíthatjuk pl., hogy
 $P(\text{Betörés} | \text{JánosTelefonál}) = 0,016$.

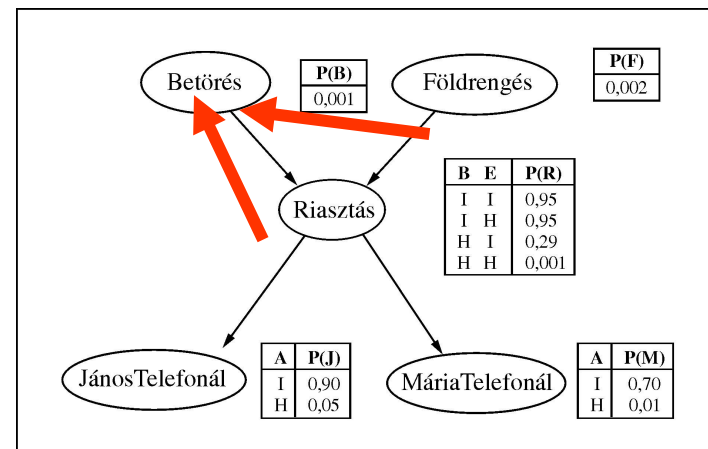


Okozati következtetés (okról a hatásra)

Ha adott a *Betörés*, akkor
kiszámíthatjuk pl., hogy
 $P(\text{JánosTelefonál} | \text{Betörés}) = 0,67$.



Okok közötti következtetés (következtetés egy közös hatás okai között)



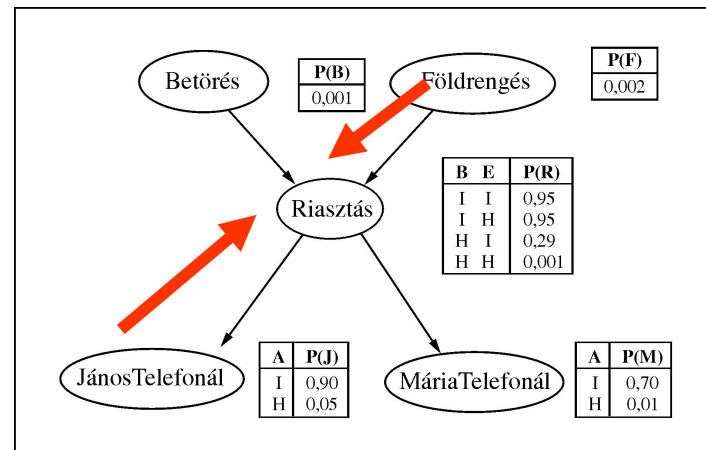
Ha adott a *Riasztás*,
akkor $P(\text{Betörés} | \text{Riasztás}) = 0.376$.

Ha azonban hozzávesszük azt a tényt is,
hogy a *Földrengés* igaz,
akkor $P(\text{Betörés} | \text{Riasztás}, \text{Földrengés}) = 0.003$.

Bár a betörések és a földrengések függetlenek, az egyik jelenléte a másik valószínűségét csökkenti.
Ez a fajta következtetési mód a **kimagyarázás**.

Kevert következtetések (a fentiek kombinált használata).

Ha a *JánosTelefonál* okozat igaz és a *Földrengés* ok hamis, akkor



$$P(Riasztás | JánosTelefonál, \neg Földrengés) = 0,03.$$

Ez a diagnosztikai és okozati következtetés együttes felhasználása.

Hasonlóan,

$$P(Betörés | JánosTelefonál, \neg Földrengés) = 0,017.$$

Ez a diagnosztikai és az okok közötti következtetés kombinálása.

Érzékenységi vizsgálat elvégzése, annak érdekében, hogy megértsük, hogy a modell mely vonatkozásának van a legnagyobb hatása a lekérdezéses változókra nézve.