Laboratóriumi vizsgálatok - redundáns információ

Antal Péter

Computational Biomedicine (Combine) workgroup
Intelligent Systems research group,
Department of Measurement and Information Systems,
Budapest University of Technology and Economics

COMBINE
Computational
Biomedicine
Workgroup

Méréstechnika és
Információs Rendszerek
Tanszék

MŰEGYETEM 1782

Áttekintés

- Bevezető
- Bayesi döntéselmélet
- Bayes-hálózatok, döntési hálózatok
- Laboratóriumi tesztek átfogó elemzése
- Szekvenciális labordiagnosztika
- Összefoglalás

Háttér

BME: MIT: Mesterséges intelligencia csoport

- BME
  - Villamosmérnöki és Informatikai Kar (VIK)
    - Méréstechnika és információs rendszerek tanszék (MIT)
      - Mesterséges intelligencia csoport
        - Computational Biomedicine Labor (ComBineLab.hu)
          (Számítógépes orvosbiológiai labor, http://bioinfo.mit.bme.hu/)

Gesztor →

SE-BME
Egészségügyi mérnökképzés
(Biomedical engineering)

Kutatás:
- Genetikai adatelemzés
- Orvosbiológiai adatelemzés
- Klinikai döntéstámogatás
- Gyógyszerkutatás

Partnerek: SE, KULeuven, Manchester University, Rostock University

Oktatás:
- Mesterséges intelligencia
- Gépi tanulás
- Biostatisztika
- Bioinformatika

ComBineLab.hu: témák

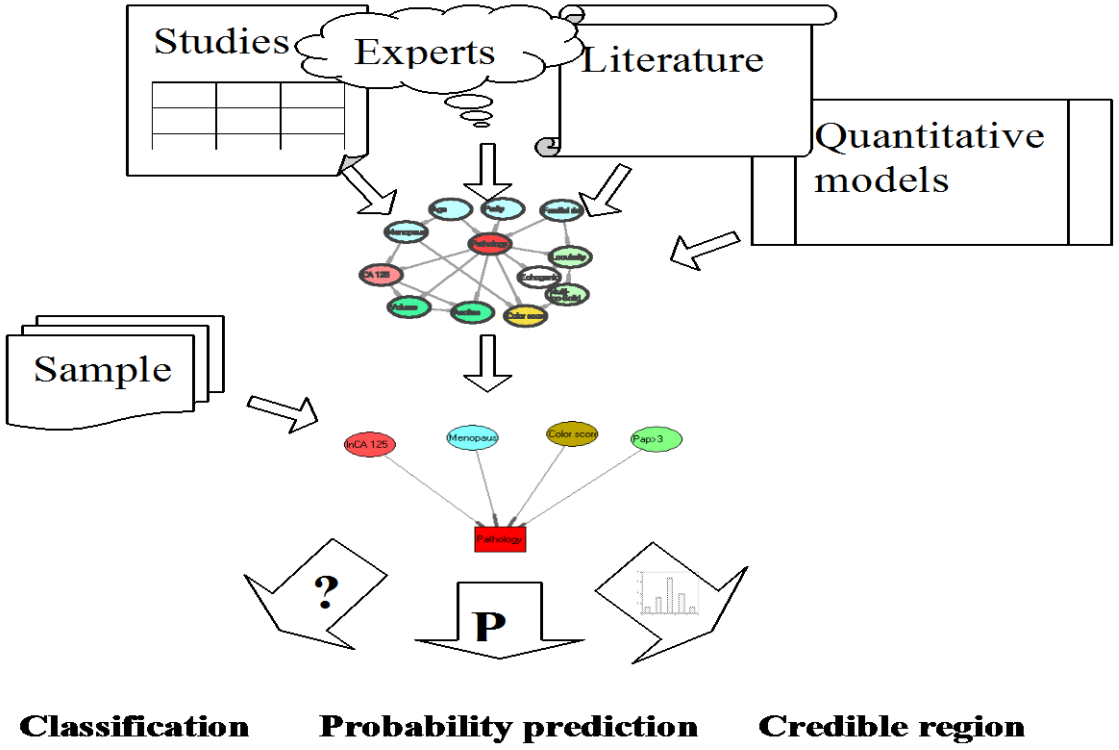
- Tudáselemzés
- Kísérlettervezés
- Adatmérnökség
- Adatelemzés
- Értelmezés
- Alkalmazás

ComBineLab.hu: Adat- és tudásfúzió

- Adatelemzésben: Adat+tudás[➔modell]➔predikció
  - Tudás
    - Szakcikkek
    - Szakterületi fogalmak és relációk (ontológiák)
    - Szakértői vélekedések
    - Szakterületi statisztikák
  - Adat
    - Statisztikai adat
    - Részleges statisztikák
- Döntéstámogatásban: Megfigyelés/beavatkozás+Modell[ek]➔döntés

Petefészekrák nem-invazív diagnosztikája

- International Ovarian Tumor Analysis (IOTA)
  - 1996-1999: ~300 minta (K.U.Leuven)
  - 1998-2004: 68 paraméter, 1066 minta
- Kérdések:
  - Predikciós teljesítmény
  - Biomarkerek
    - Szükségesség, elégségesség, helyettesíthetőség
    - költséghatékonyság

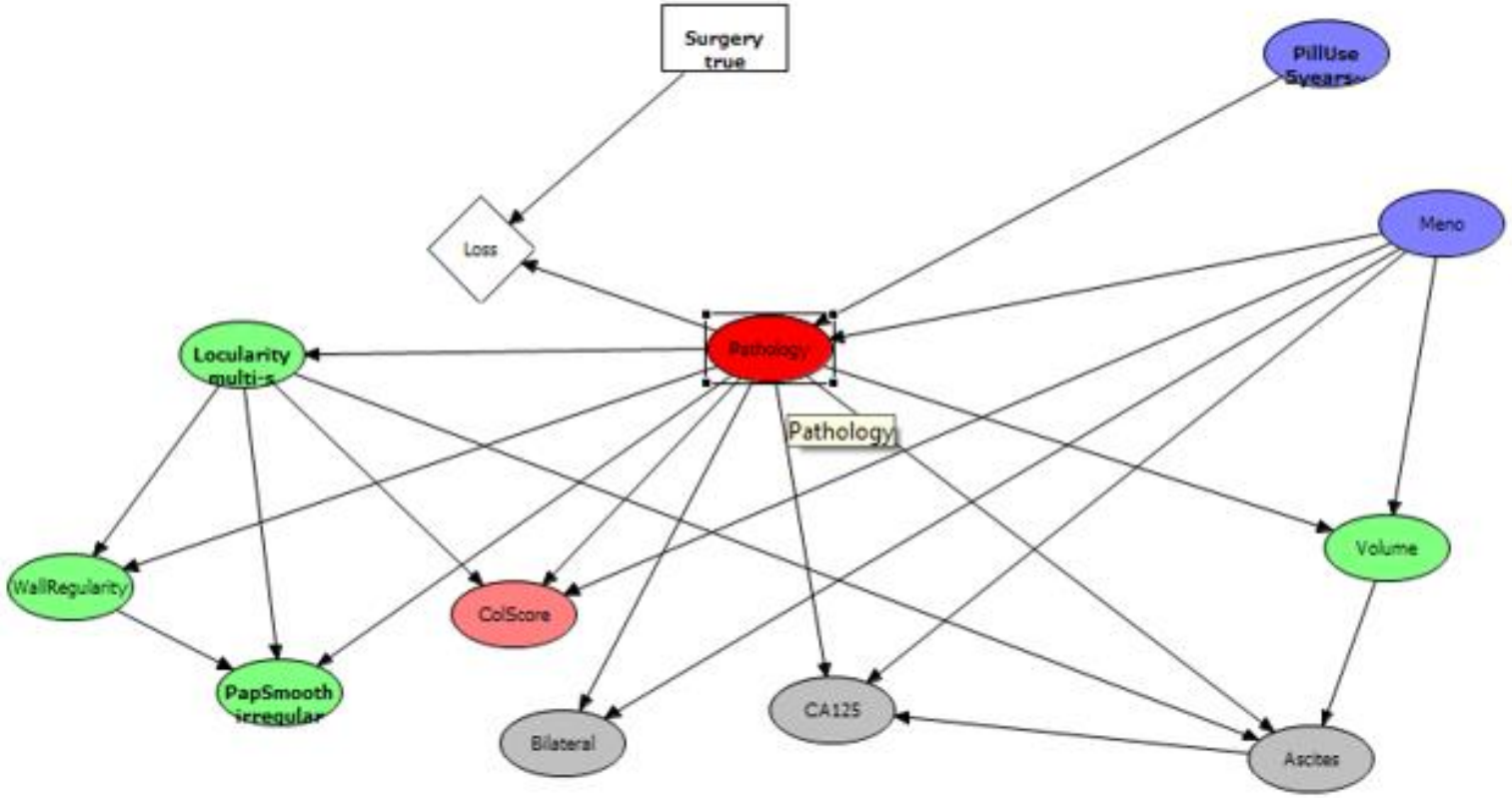


Antal, P. et al. (2000). **How might we combine the information we know** about a mass better? In *Proc. of the 1st Montecarlo Conference on updates in Gynaecology (MCG)* (pp. 1-3).

Timmerman et al. **Terms, definitions and measurements** to describe the sonographic features of adnexal tumors: a consensus opinion from the IOTA group. Utransound Obstet Gynecol, 16:500-505, 2000

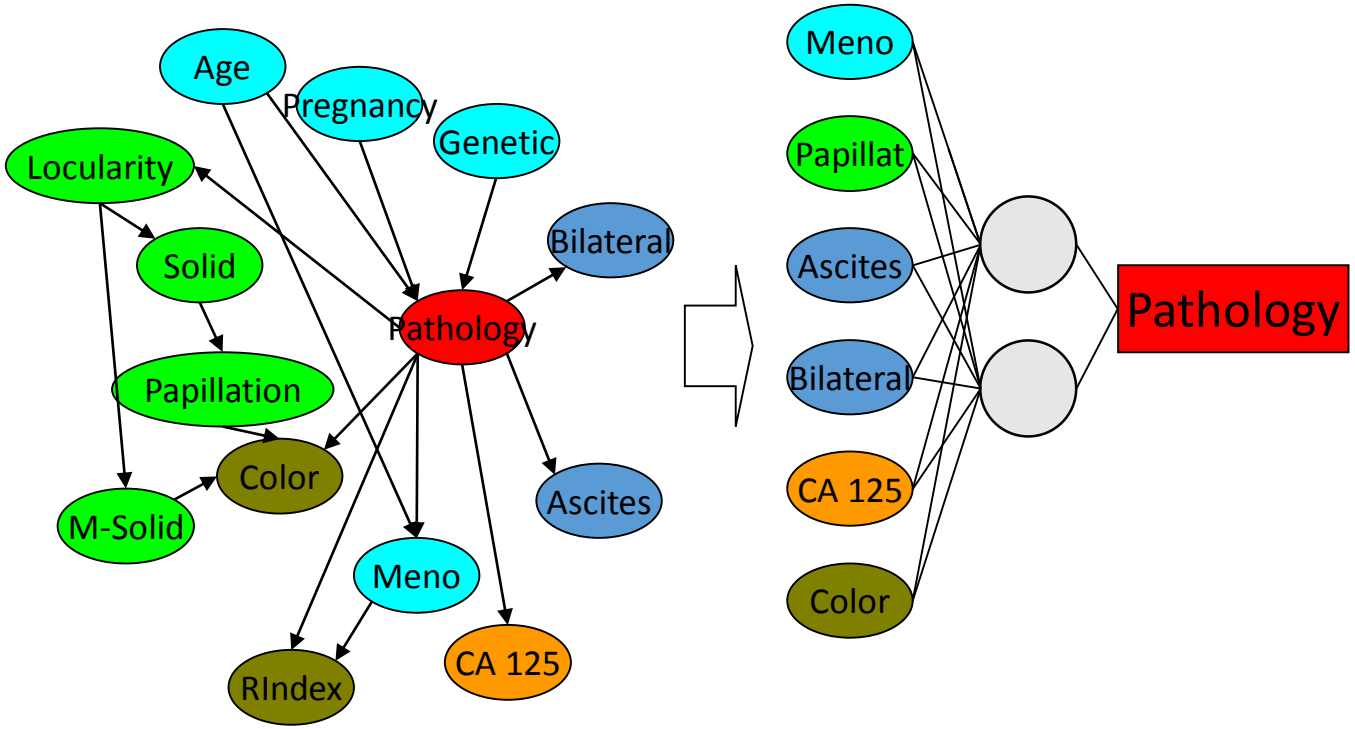
Timmerman, D. (2004). The use of mathematical models to evaluate pelvic masses; **can they beat an expert operator?**. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *18*(1), 91-104.

Döntési hálózatok



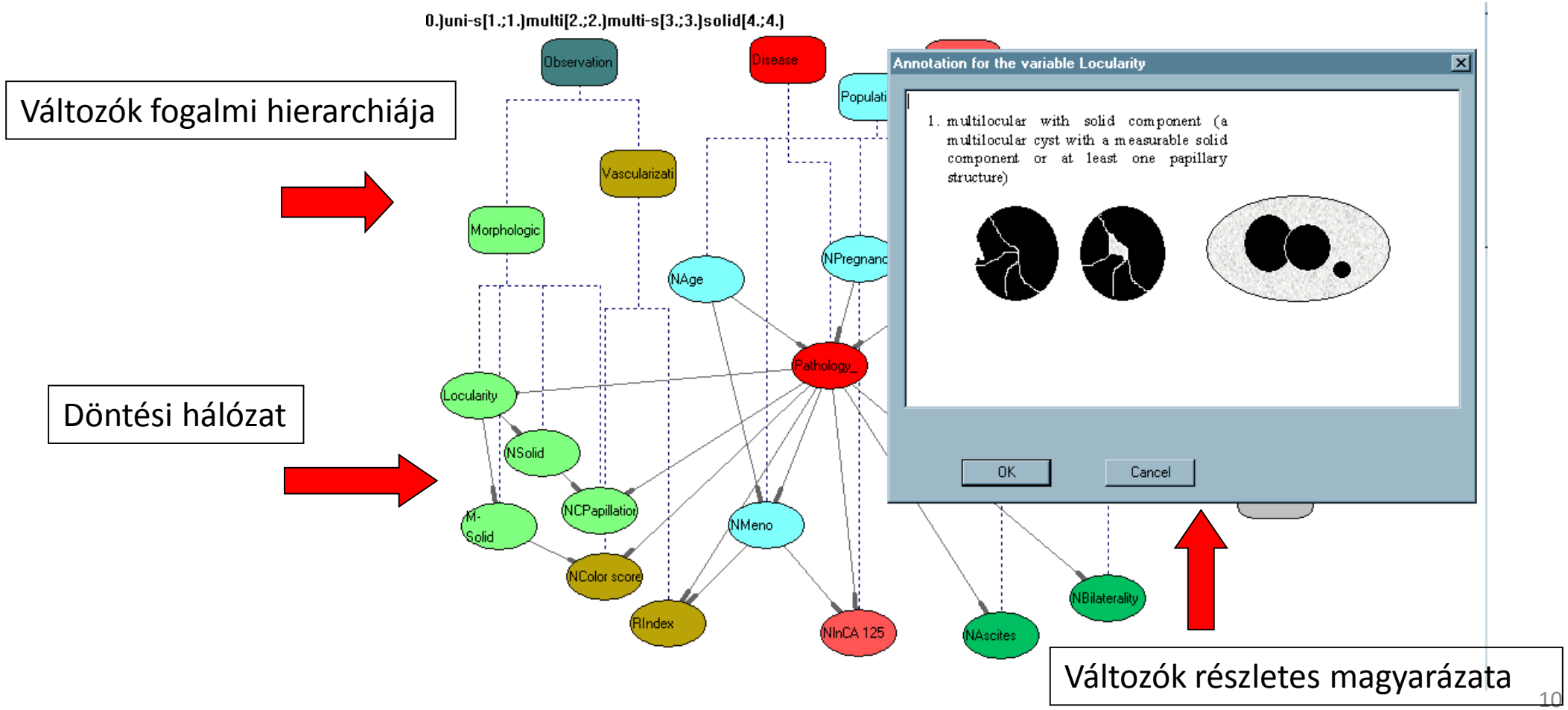
Antal, P., Fannes, G., Timmerman, D., Moreau, Y. and De Moor, B., 2004. Using literature and data to learn Bayesian networks as clinical models of ovarian tumors. *Artificial Intelligence in medicine*, *30*(3), pp.257-281.

„Tudásgazdag” neurális hálózatok



P. Antal, G. Fannes, D. Timmerman, Y. Moreau, B. De Moor: Bayesian Applications of Belief Networks and Multilayer Perceptrons for Ovarian Tumor Classification with Rejection, *Artificial Intelligence in Medicine*, vol. 29, pp 39-60, 2003

Bizonyíték-alapú, értelmezhető döntési hálózatok

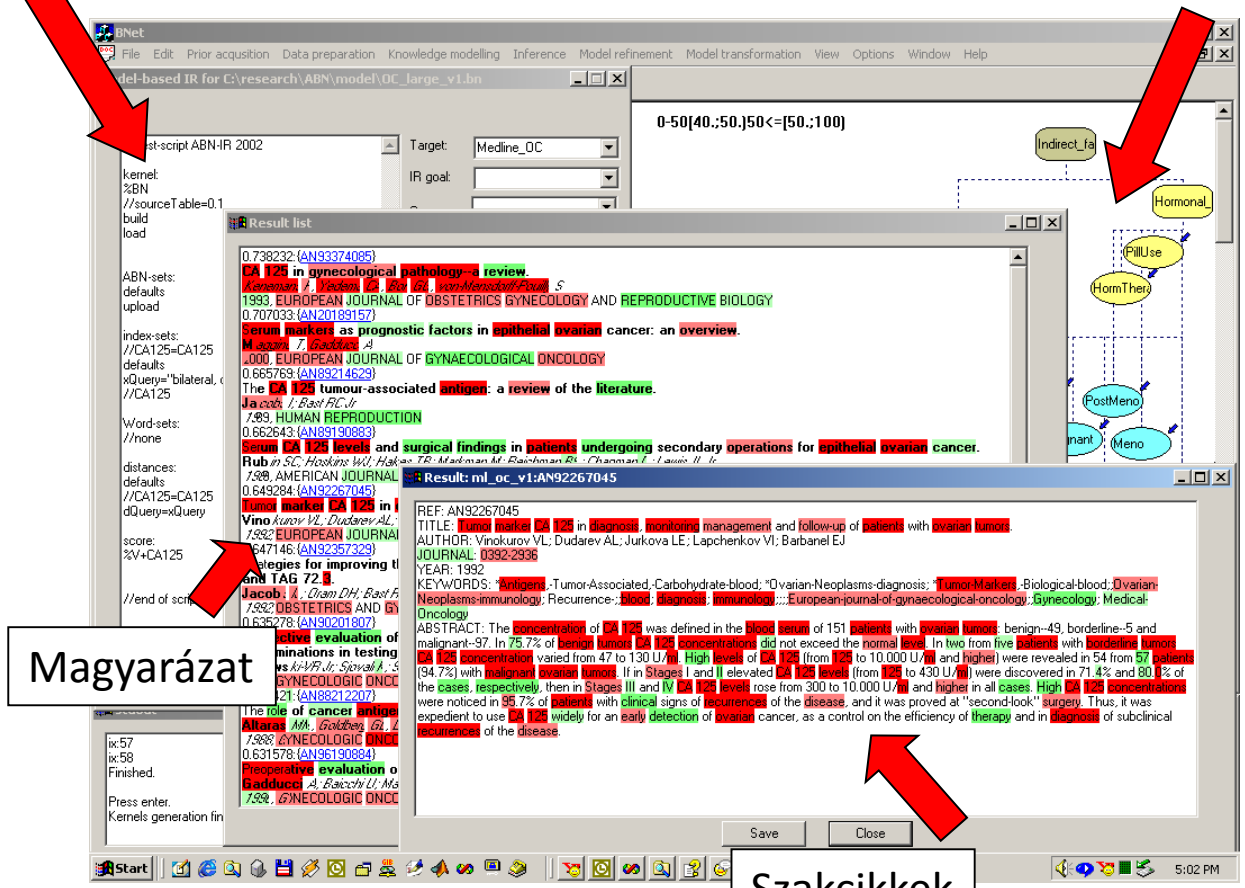


Antal, P., Mészáros, T., De Moor, B., & Dobrowiecki, T. (2001). Annotated Bayesian Networks: a tool to integrate textual and probabilistic medical knowledge. In *Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001* (pp. 177-182). IEEE.

Érthető, megmagyarázott döntések

Kérdés

Bizonyíték-alapú döntési hálózat



Magyarázat

Szakcikkek

Antal, P., De Moor, B., Timmerman, D., Mészáros, T., & Dobrowiecki, T. (2002). Domain knowledge based information retrieval language: an application of annotated Bayesian networks in ovarian cancer domain. In *Proceedings of 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002)* (pp. 213-218). IEEE.

MI az egészségügyben 2019 – IBM Watson

Clinical Therapeutics/Volume 38, Number 4, 2016

Review Article

IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research

CrossMark

[1]IBM Almac

News room > News releases >

At ASCO 2017 Clinicians Present New Evidence about Watson Cognitive Technology and Cancer Care

- Watson matched tumor board treatment recommendations in up to 96% of cases; reduced clinical trial screening time by 78%, studies find
- Prostate cancer is latest add to Watson for Oncology; the tech will be available to support 80 percent of the incidence of cancer by year-end
- Nine new adopters of Watson oncology offerings around the globe expands Watson's reach to 55 organizations worldwide

How Watson for Oncology Is Advancing Personalized Patient Care

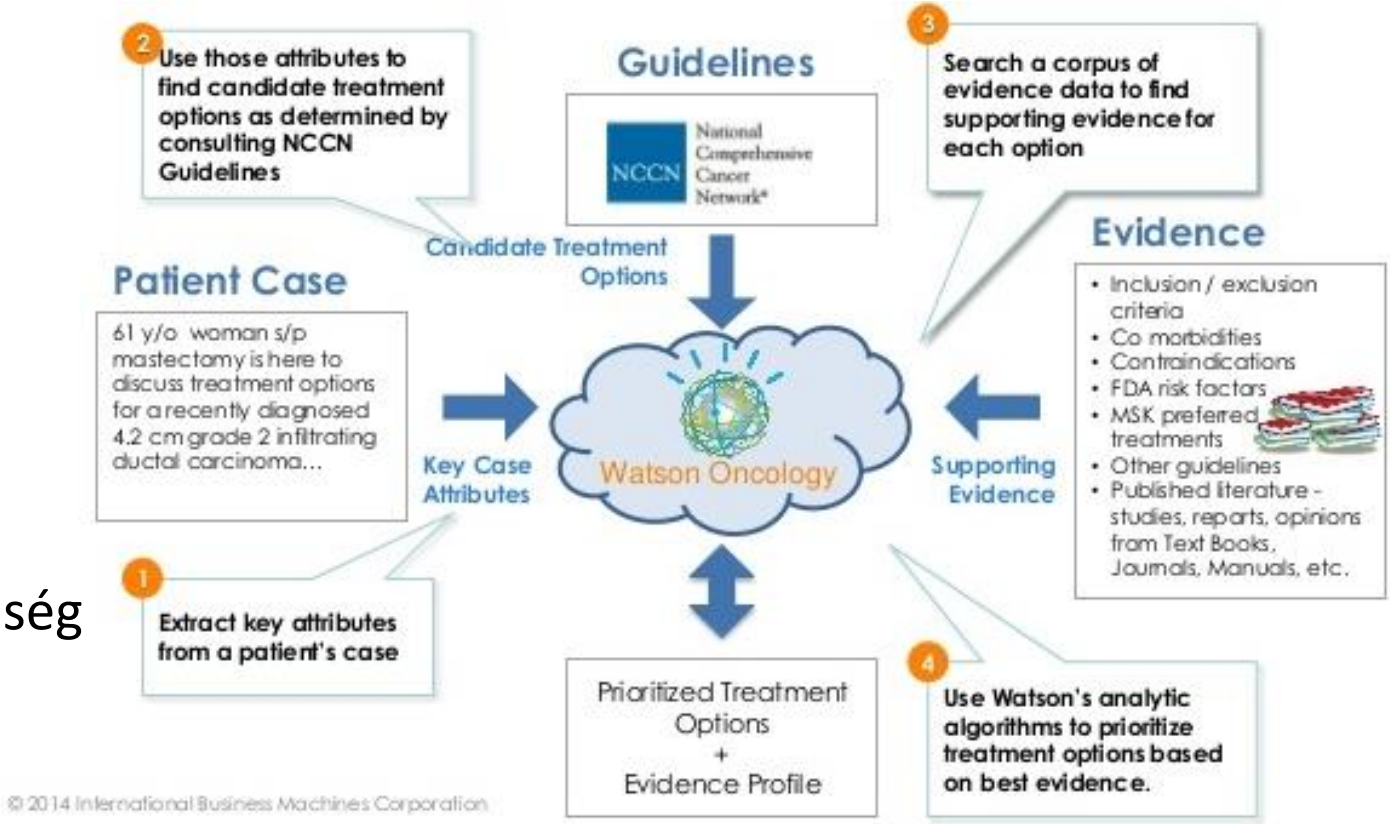
By Jo Cavallo

June 25, 2017

MI az egészségügyben: lehetőségek és korlátok

- +
  - Felügyelt tanulás
    - Adat + közvetlen hasznosság/veszteség
  - Megerősítéses tanulás
    - Autonómia + eseti hasznosság/veszteség
  - Természetes nyelv feldolgozása
- -
  - Információ beszerzés (szabványos adat)
  - Értelmezhetőség, magyarázat
  - Taníthatóság
  - Alapvető és hosszútávú hasznosság/veszteség
  - Felfedező vs. szakértői rendszer

Együttműködő ember-gép szakértői csapat



Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, *56*(04), 24-31.

Matematika alapok

Adatok és tudás normatív fúziója

- **Vélekedések és döntési preferenciák egységes kerete**
  - **Thomas Bayes (c. 1702 – 1761)**
  - Bayes-szabály
  - Bayes-hálók
  - Bayesi döntéselmélet (Bayesi statisztika)
  - ...

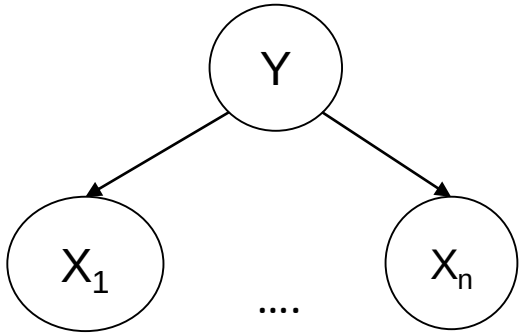
(G.E.P.Box: „all models are wrong, but some are useful”)

$$p(X \mid Y) = \frac{p(Y \mid X)\,p(X)}{p(Y)} = \frac{p(Y \mid X)\,p(X)}{\sum_X p(Y \mid X)\,p(X)}$$

$$p(Modell \mid Adat) \propto p(Adat \mid Modell)\,p(Modell)$$

$$p(Cause \mid Effect) \propto p(Effect \mid Cause) \times p(Cause)$$

Naív Bayes-hálózat

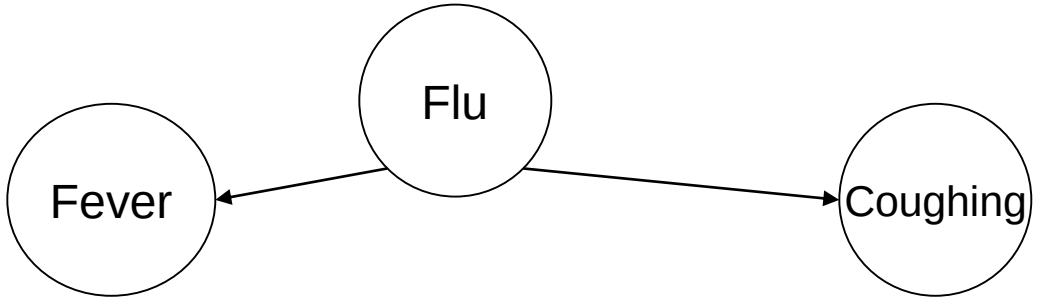


Együttes eloszlás dekomponálása:

$P(Y,X_1,..,X_n) = P(Y)\ P(X_1|Y)\ldots\ P(X_n|Y)$

Diagnostic inference:

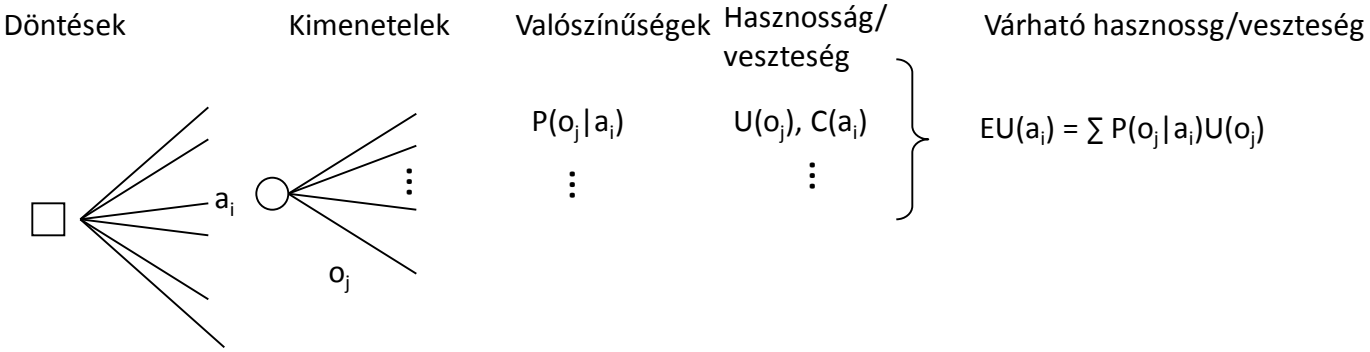
$P(y \mid x_1,..,x_n) = P(y)\ P(x_1|y)\ldots\ P(x_n|y)\ /\ P(x_{i1},..,x_{ik}) \propto P(y)\ P(x_1|y)\ldots\ P(x_n|y)$



$$p(Flu = present \mid Fever = absent, Coughing = present)$$
$$\propto p(Flu = present)\, p(Fever = absent \mid Flu = present)\, p(Coughing = present \mid Flu = present)$$

Döntéselméleti háttér

- Döntési helyzet:
  - Akciók a_i
  - Kimenetelek o_j
  - Valószínűségek a kimenetelekhez $p(o_j \mid a_i)$
  - Hasznosság/veszteség a kimenetelekhez $U(o_j \mid a_i)$
  - Maximumális várható hasznosság $EU(a_i) = \sum_j U(o_j \mid a_i)\, p(o_j \mid a_i)$
    - Optimális döntés $a^* = \arg\max_i EU(a_i)$



Döntési hálózatok

Döntési (naív) hálózat

P(Betegség)

()	Nincs	Van
()	0.3	0.7

fixed row height sample size/probability

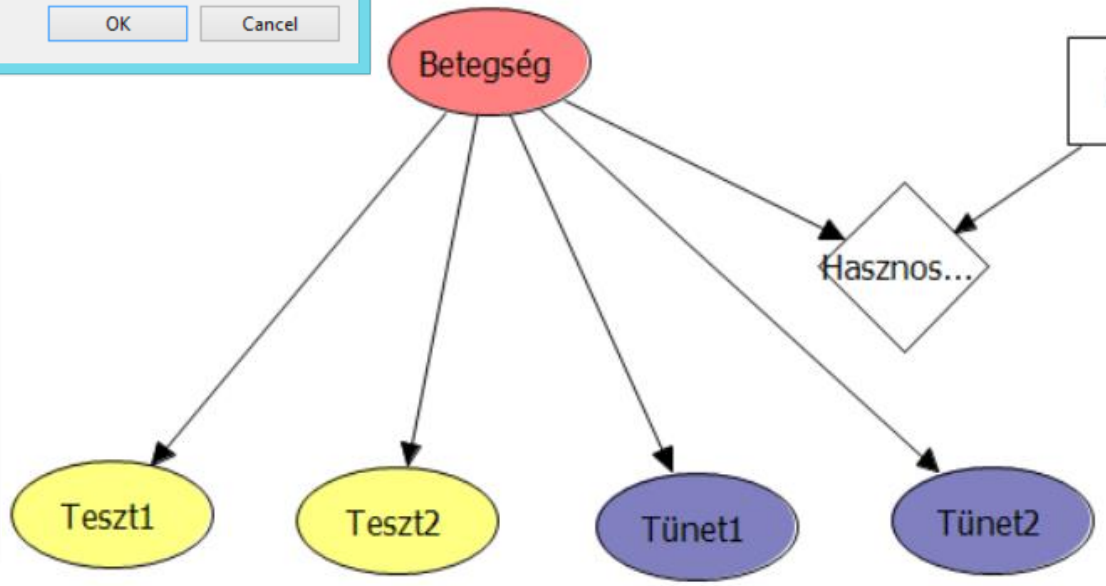
OK Cancel

P(Teszt1|Betegség)

(Betegség)	Normal	Alacsony	Magas
(Nincs)	0.55	0.45	0.0
(Van)	0.05	0.45	0.5

fixed row height sample size/probability

OK Cancel



Hasznosság(Betegség, Dö...

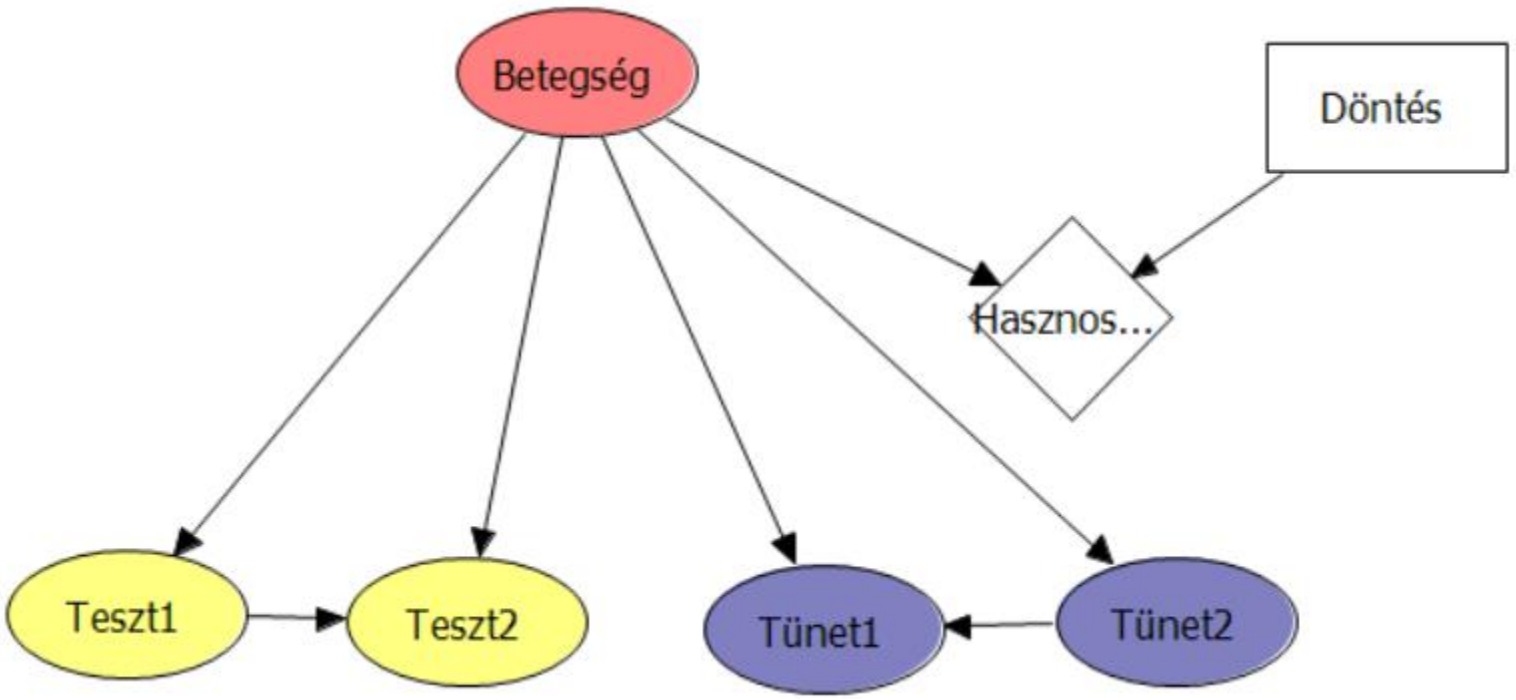
Döntés

Betegség	Nincs	Van
Nincs	0	-100
Van	-5	10

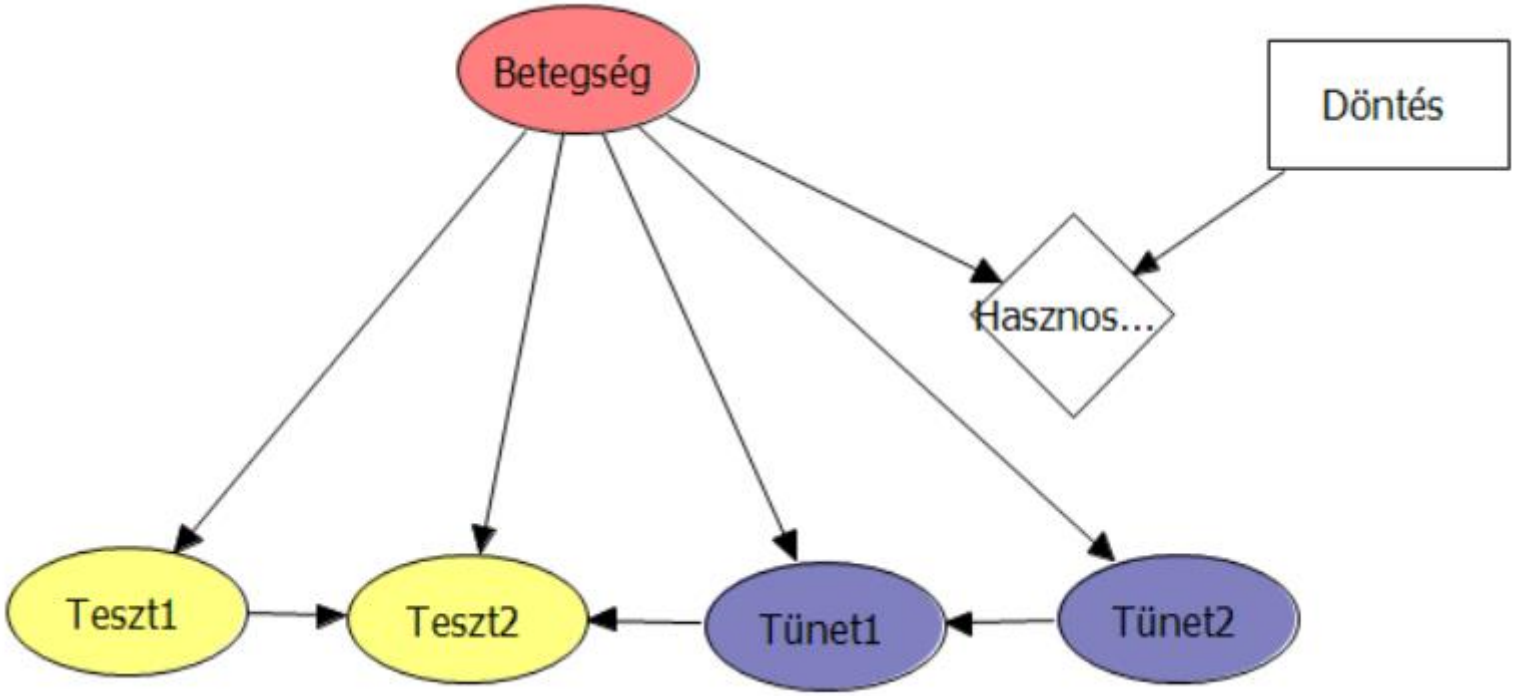
fixed row height matrix view

OK Cancel

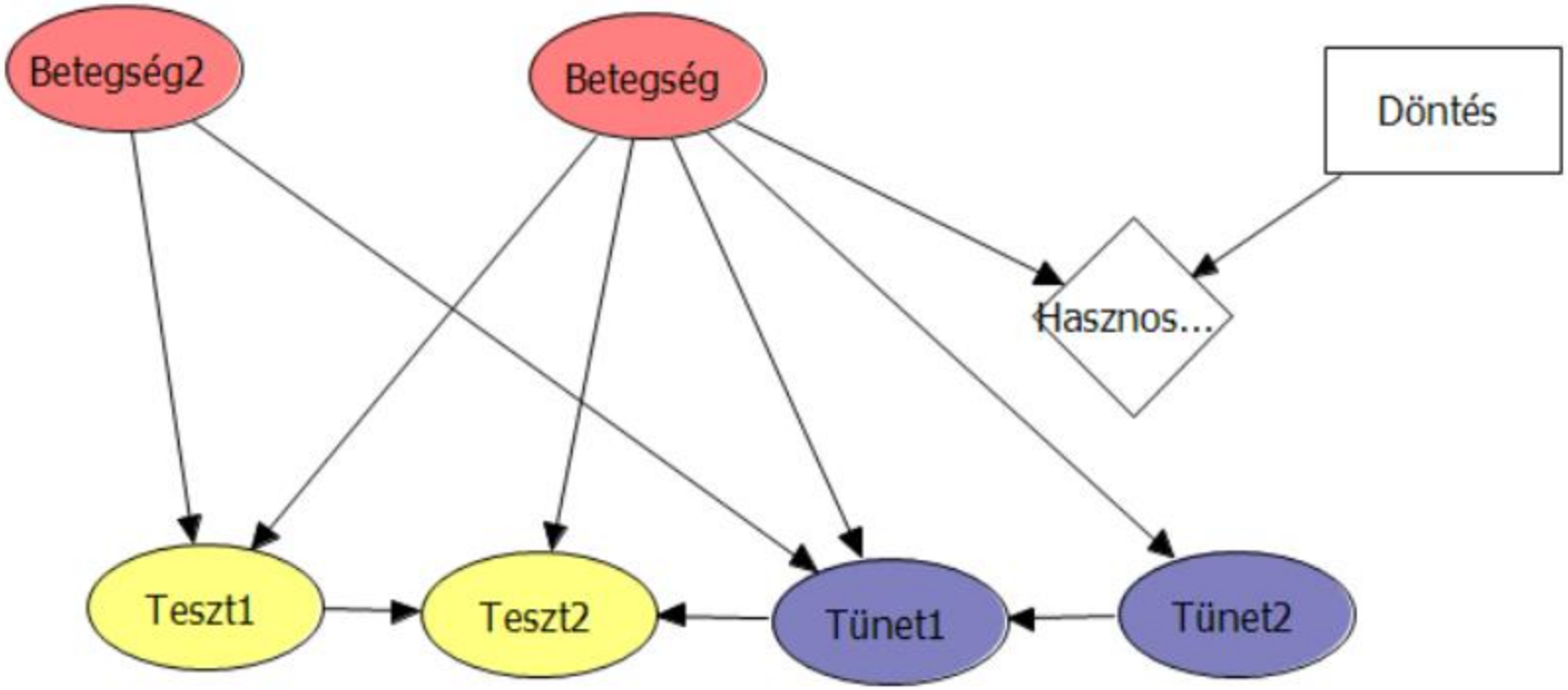
Döntési hálózat: redundáns következmények



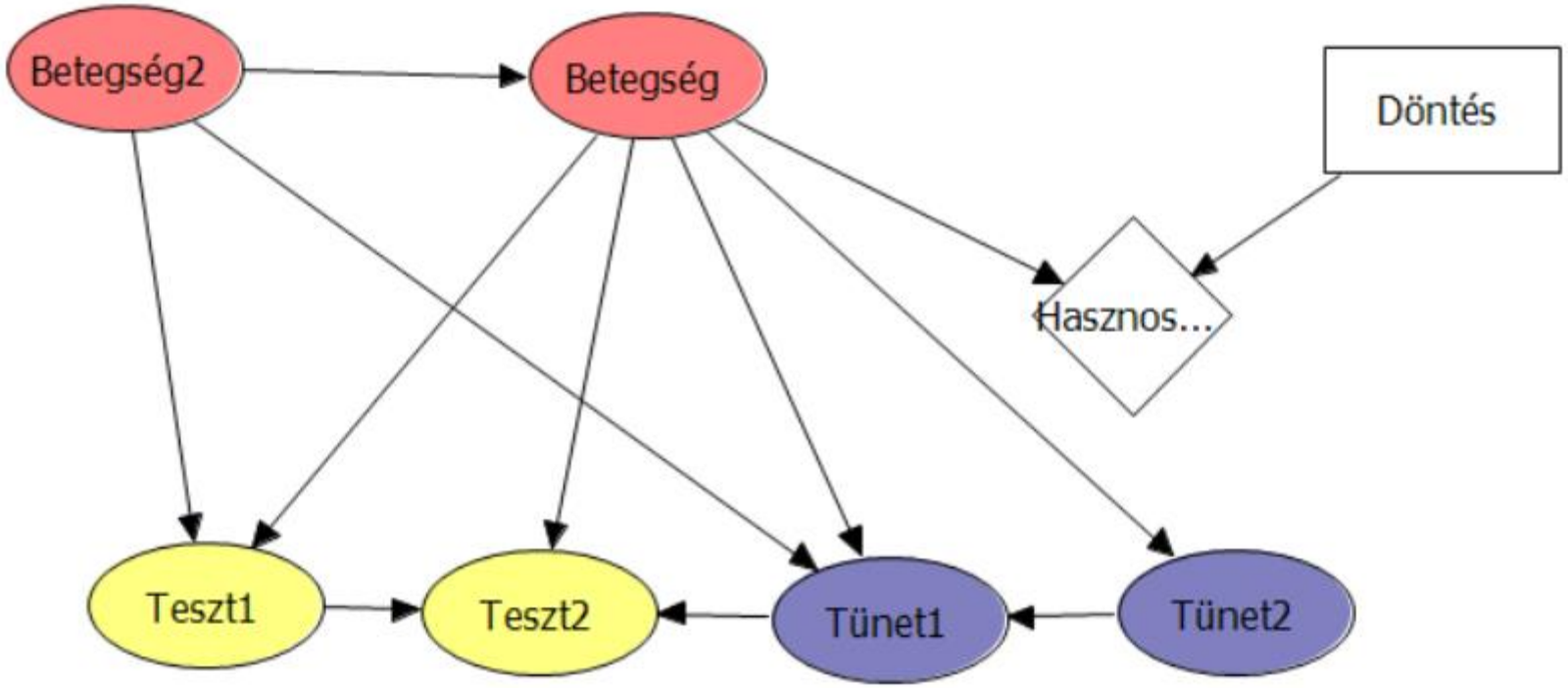
Döntési hálózat: redundáns következmények+



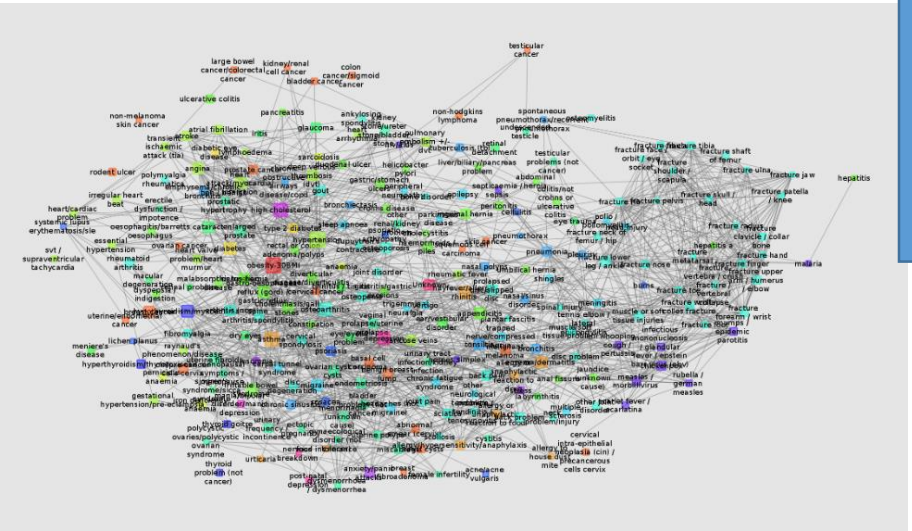
Döntési hálózat: több ok

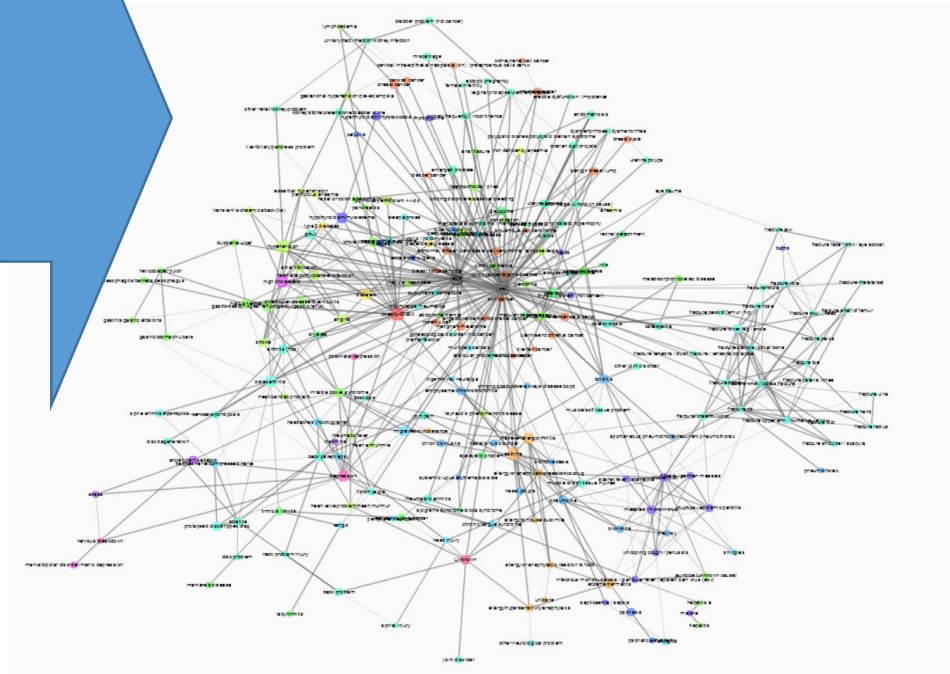


Döntési hálózat: több összefüggő ok



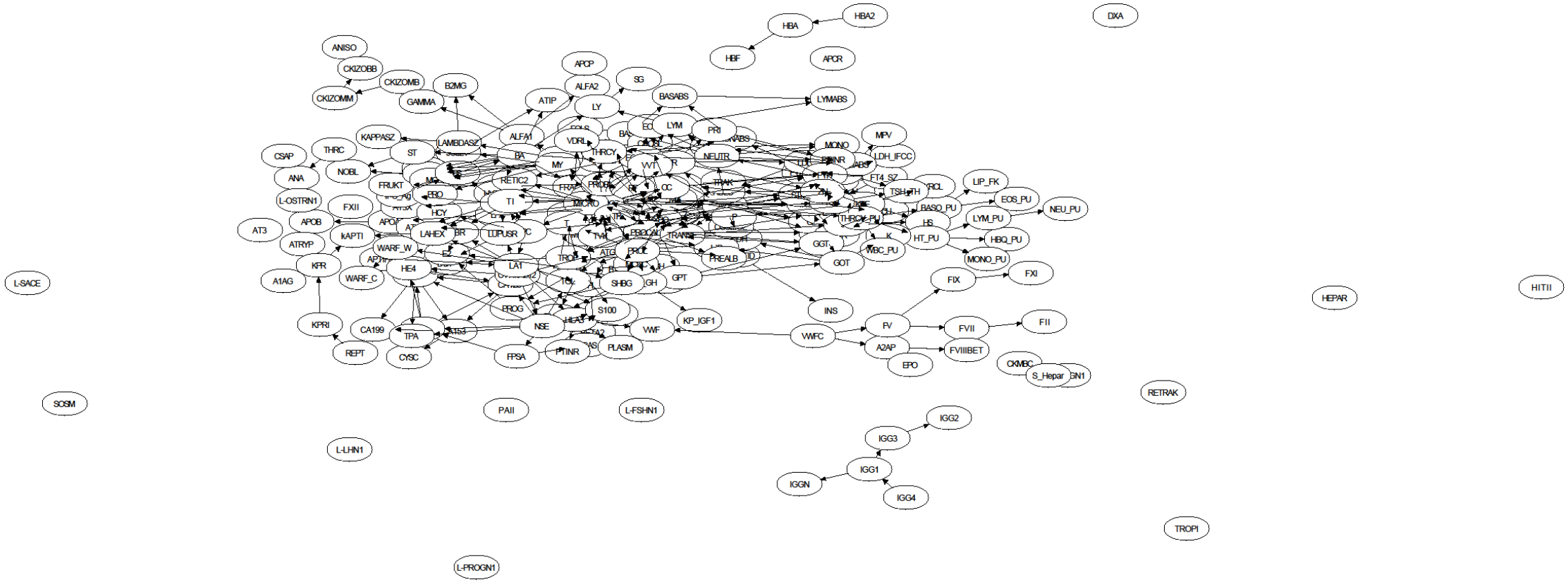
Multimorbid („redundáns”) betegségek





Marx, P., Antal, P., Bolgar, B., Bagdy, G., Deakin, B. and Juhasz, G., 2017. Comorbidities in the diseasome are more apparent than real: What Bayesian filtering reveals about the comorbidities of depression. *PLoS computational biology*, *13*(6), p.e1005487.

Redundáns laboratóriumi tesztek



Guenfoud, Zeyneb, and Péter Antal. "Bayesian exploration of dependencies of laboratory tests and evaluation of test redundancy." *2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS)*. IEEE, 2018.

Bayes-hálózatok

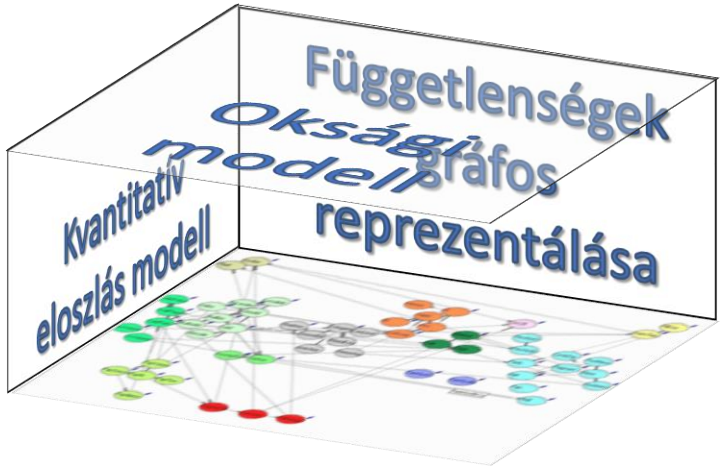
Bayes-hálózatok

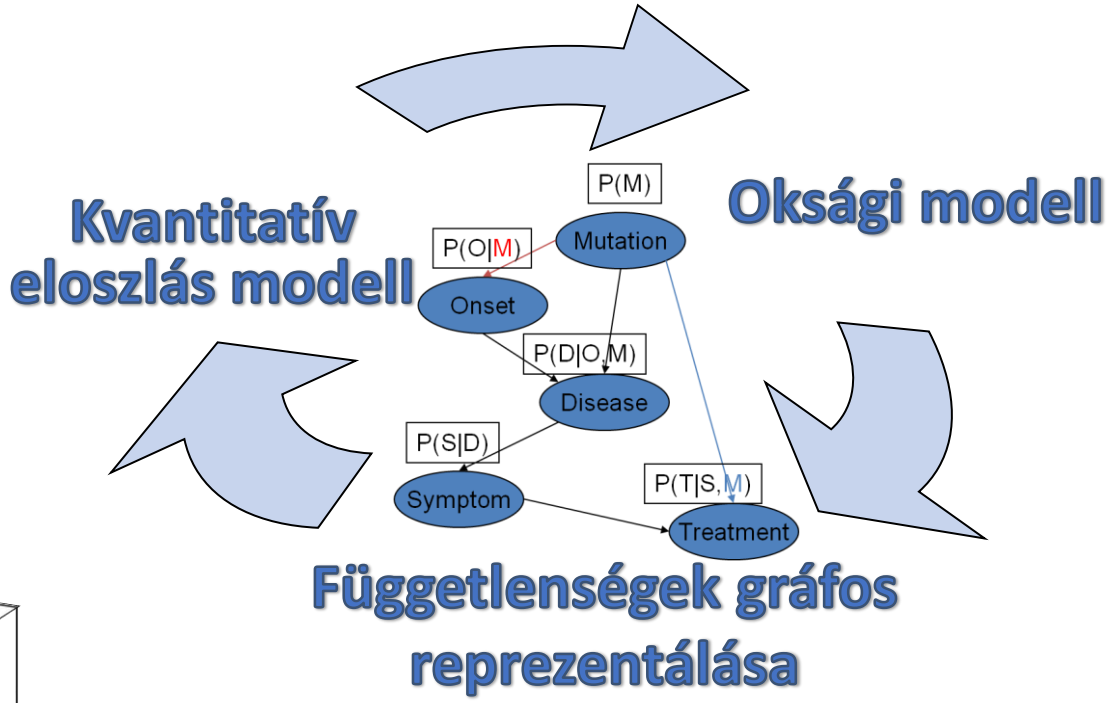
Irányított körmentes gráf (DAG)

- Csomópont – véletlen változók
- él – közvetlen függés (oksági kapcsolat)

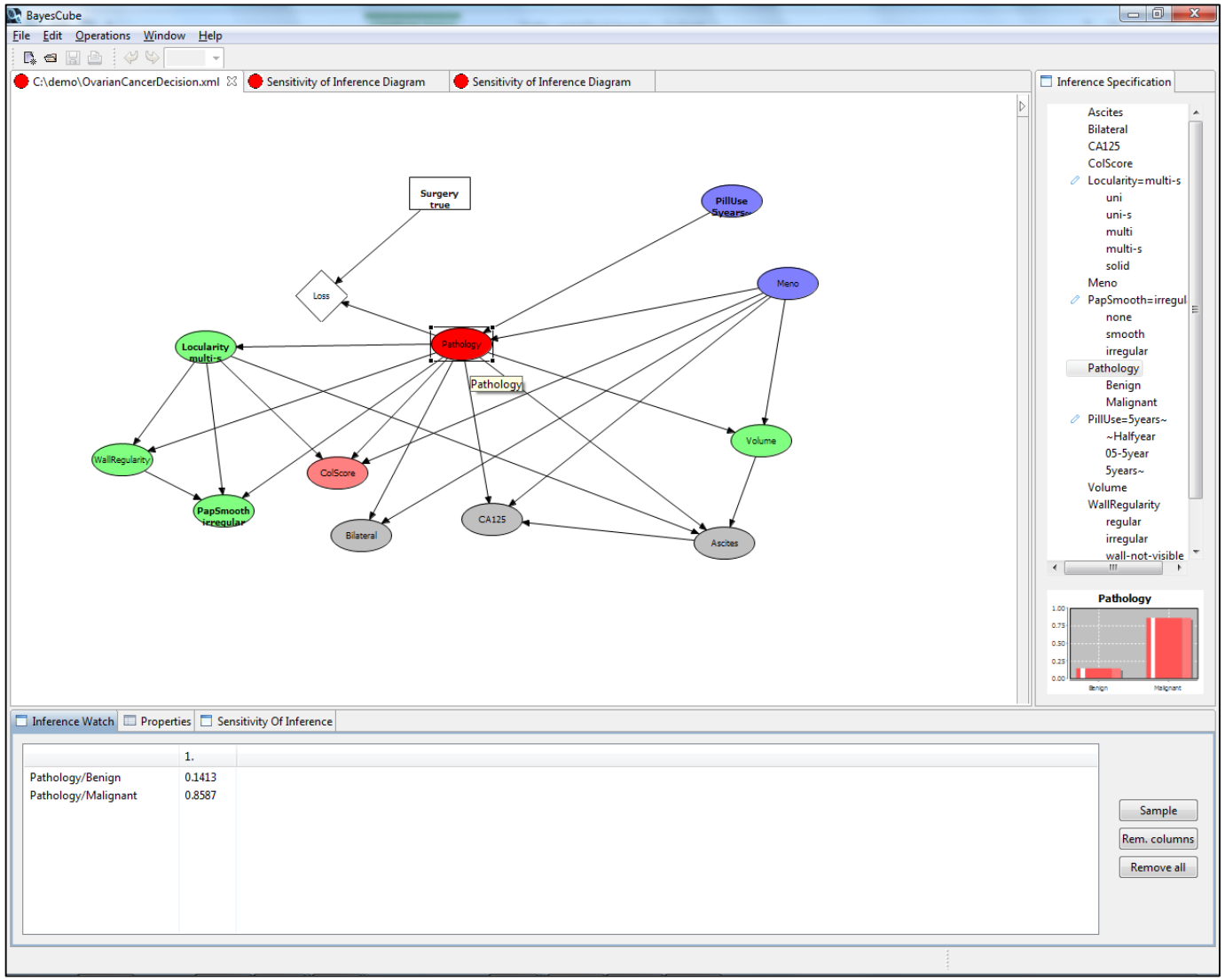
Lokális modell - $P(X_i|Pa(X_i))$

Három értelmezés:





Döntési hálózatok: következtetés

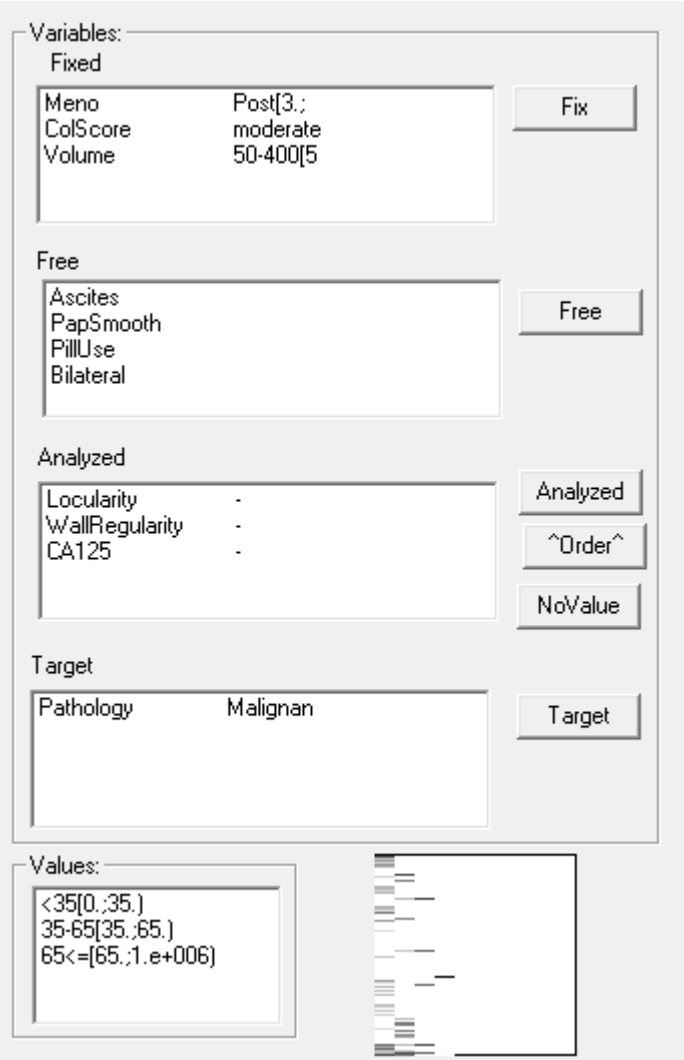


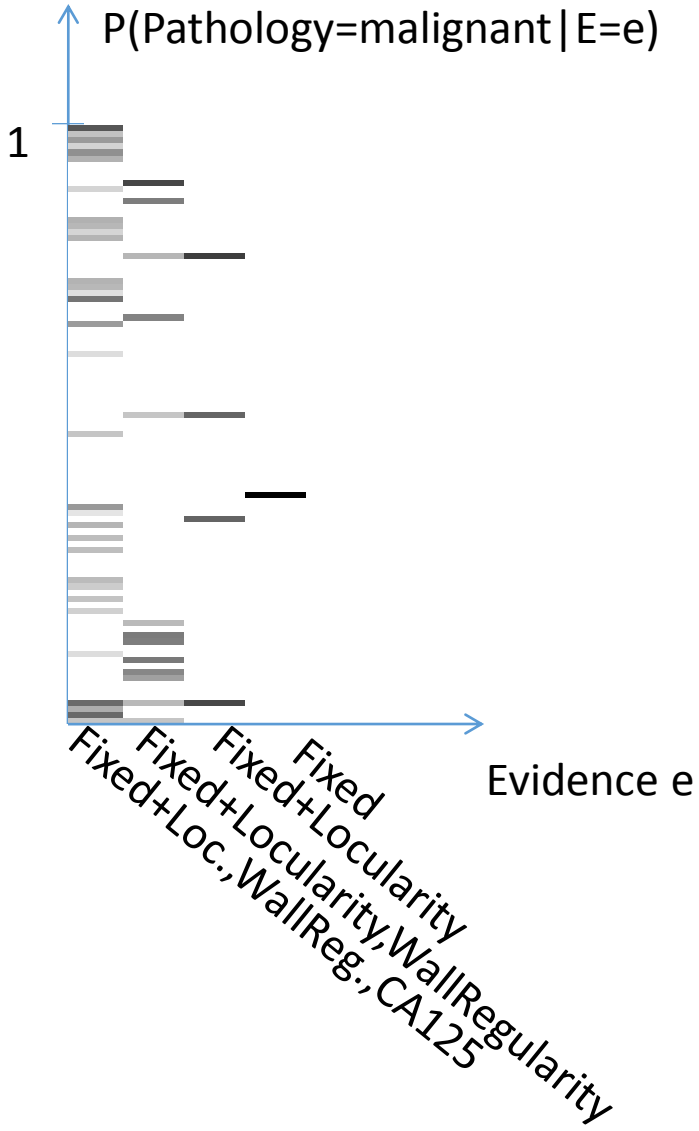
D.Timmerman, P.Antal.,2001: ~400 paraméter

Kérdéstípusok az orvosi döntéstámogatásban

- Diagnosztikai következtetés
  - P(Diagnózis|Passzív megfigyelések)
  - Legkisebb várható veszteségű diagnózis passzív megfigyelések esetén
- Optimális információgyűjtés
  - További információ hatása a következtetésre: P(Diagnózis|megfigyelések, új megfigyelés)
  - További információ hasznossága
- Terápiás következtetés
  - P(Kimenetel|Megfigyelés, Beavatkozás)
- Kontrafaktuális következtetés
  - P(ElképzeltKimenetel|Megfigyelés, Beavatkozás,Kimenetel, ElképzeltBeavatkozás)

További információ hasznossága





További információ értéke

Vezessük be a következő jelölést:

- Evidenciák (biztos ismeretek): *E*
- Akciók: *a* , legjobb akció *E* esetében: α
- Kimenetelek: S_i
- Potenciális további evidencia: E_i, lehetséges értékei e_{jk}

$$EU(\alpha|E) = \max_a \sum_i U(S_i)\ P(S_i|E, a)$$

- Legjobb akció további e_{jk} mellett: $\alpha_{e_{jk}}$

$$EU(\alpha_{e_{jk}}|E, E_j = e_{jk}) = \max_a \sum_i U(S_i)\ P(S_i|E, a, E_j = e_{jk})$$

- További evidencia értéke: Várható hasznosságok különbözősége

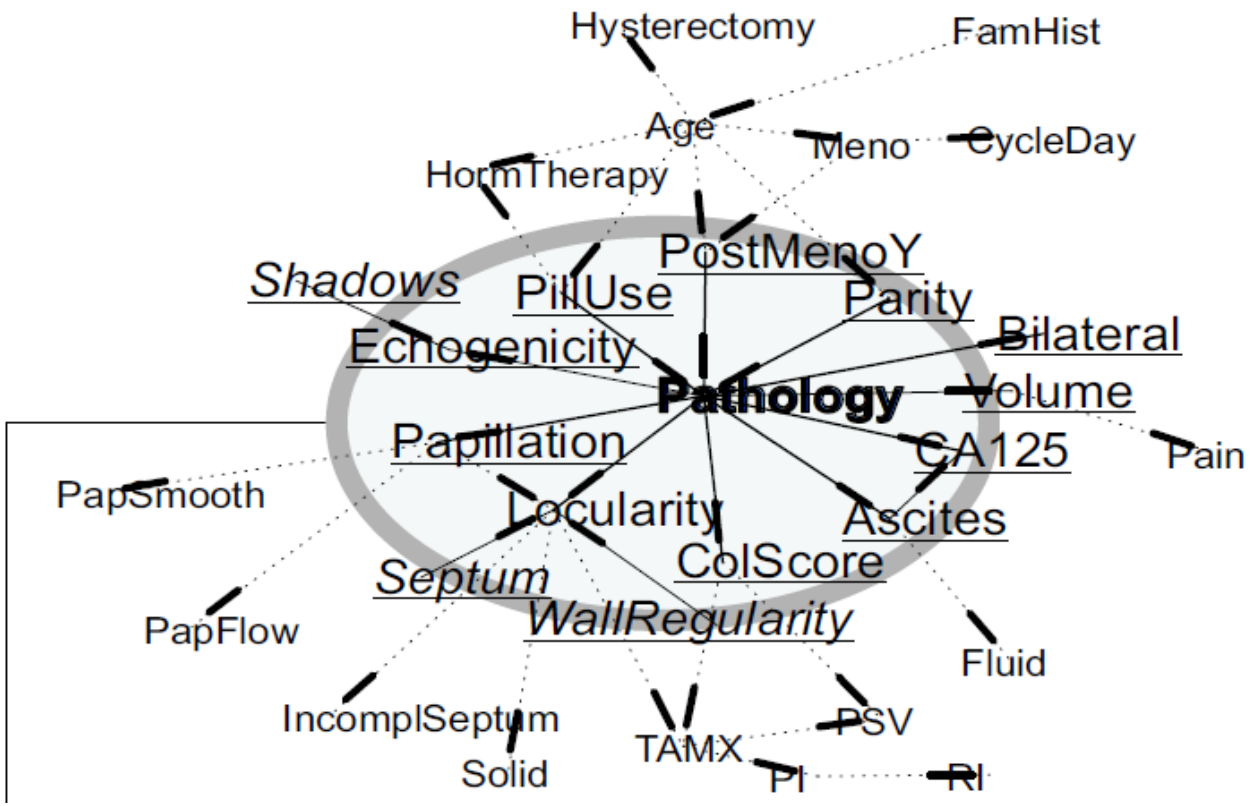
$$VPI_E(E_j) = \left(\sum_k P(E_j = e_{jk}|E) EU(\alpha_{e_{jk}}|E, E_j = e_{jk})\right) - EU(\alpha|E)$$

Diagnosztikai tesztek strukturális tulajdonságai

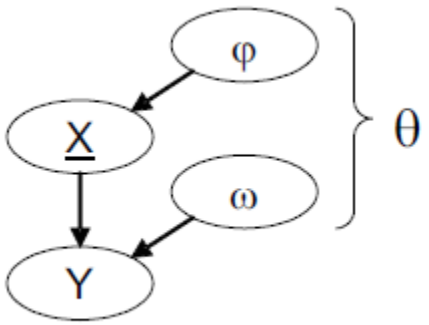
Biomarkerek, tesztek fontos jellemzői

„Maximum prediktivitás, minimum redundancia”

- Predikciós erő
- Közvetlenség
- Bizonytalanság
  - Létezési
  - Hatáserősség
- Okozatiság



A jegykiválasztási probléma

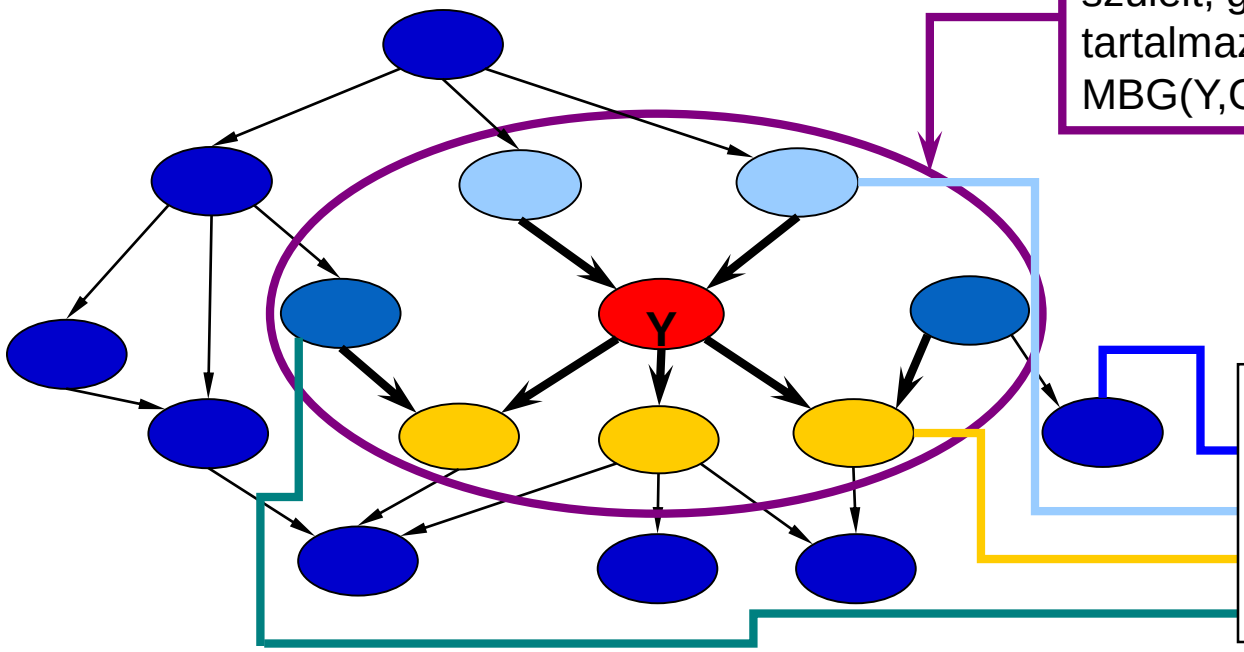


- Egy X_i jegy **erősen releváns**, ha létezik x_i, y_i, és $s_i = x_1, \ldots x_{i-1}, x_{i+1}, \ldots, x_n$ úgy, hogy
  - $p(x_i, s_i) > 0$ és
  - $p(y|x_i, s_i) \neq p(y|s_i)$.

Egy X_i jegy **gyengén releváns**, ha létezik x_i, y_i, és valamely s_i úgy, hogy

- $p(x_i, s_i) > 0$ és
- $p(y|x_i, s_i) \neq p(y|s_i)$.

Relevancia a Bayes hálókban



MGB : *G* azon algráfja, amely a Y szüleit, gyerekeit és azok egyéb szüleit tartalmazza a **Markov Takaró Gráf** MBG(Y,G).

MBG(Y,G)-ben egy változó X_i :

- (1) nem fordul elő
- (2) szülő
- (3) gyermek
- (4) (tisztán) egyéb szülő

irreleváns

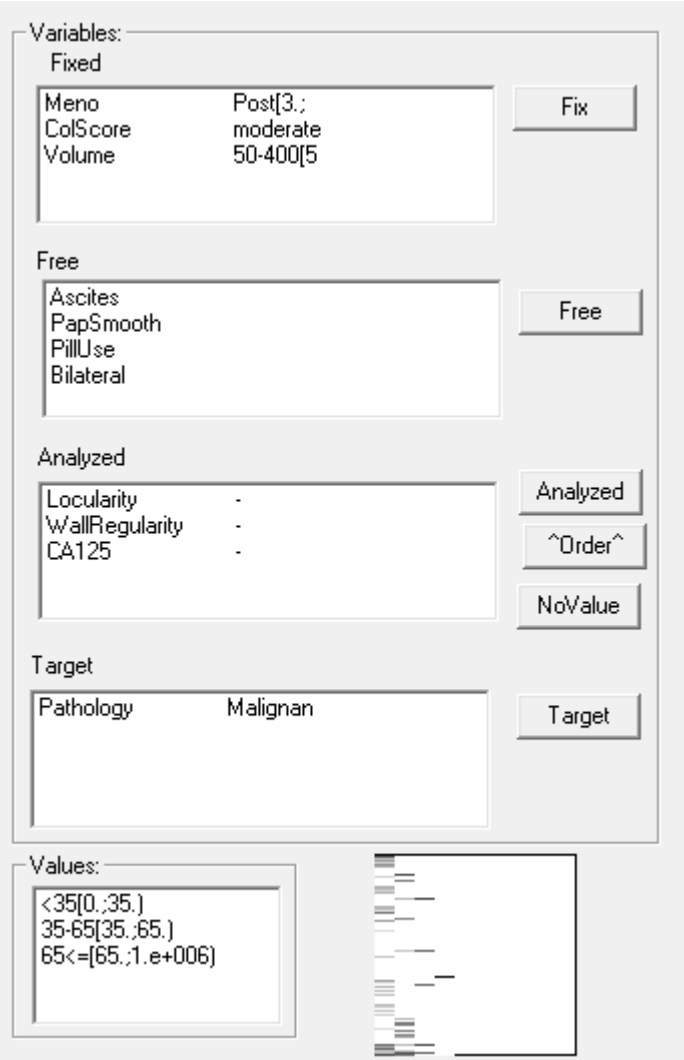
releváns

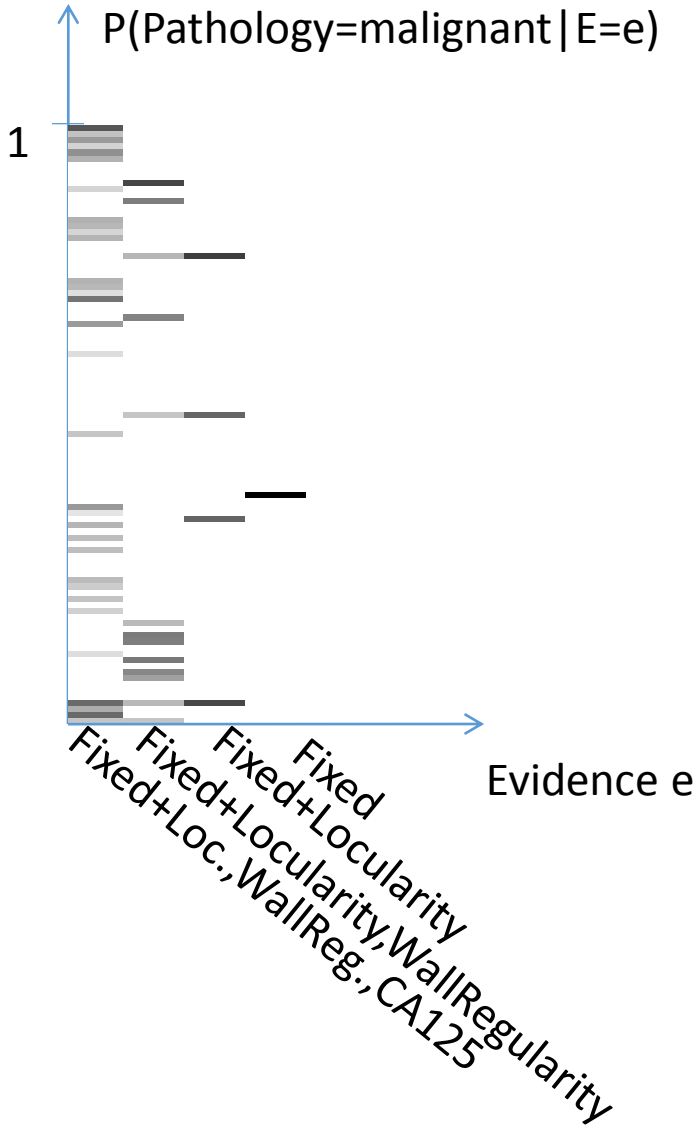
Markov Burkoló Halmaz (MBS) az a csomóponthalmaz, amely valószínűségi értelemben elszigeteli egy célcsomópontot(/halmazt) a többitől.

Markov-határ: szükséges és elégséges változóhalmaz

Diagnosztikai tesztek kvantitatív tulajdonságai

További információ hasznossága





További információ értéke

Vezessük be a következő jelölést:

- Evidenciák (biztos ismeretek): *E*
- Akciók: *a* , legjobb akció *E* esetében: α
- Kimenetelek: S_i
- Potenciális további evidencia: E_i, lehetséges értékei e_{jk}

$$EU(\alpha|E) = \max_a \sum_i U(S_i)\ P(S_i|E, a)$$

- Legjobb akció további e_{jk} mellett: $\alpha_{e_{jk}}$

$$EU(\alpha_{e_{jk}}|E, E_j = e_{jk}) = \max_a \sum_i U(S_i)\ P(S_i|E, a, E_j = e_{jk})$$

- További evidencia értéke: Várható hasznosságok különbözősége

$$VPI_E(E_j) = \left(\sum_k P(E_j = e_{jk}|E) EU(\alpha_{e_{jk}}|E, E_j = e_{jk})\right) - EU(\alpha|E)$$

Adaptív labordiagnosztikai mérések

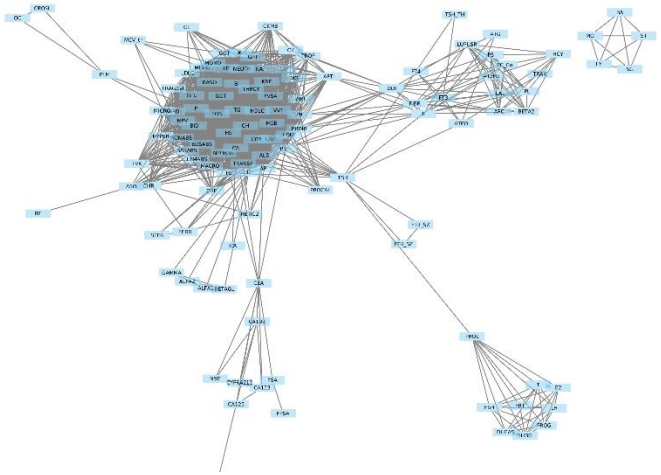
Diagnosztikai laboratóriumi tesztek elemzése

- Adat
  - Forrás: SE Központi Laboratórium
  - 225 változó
  - 4 év: ~200,000 beteg
  - Normál tartomány ➔ szemi-kvantitatív kódolás
    - 0: ismeretlen (adott esetben: feltételezhetően nem kóros/nem gyanús)
    - 1: mért és normál tartománybeli
    - 2: mért és normál tartomány alatti
    - 3: mért és normál tartomány feletti

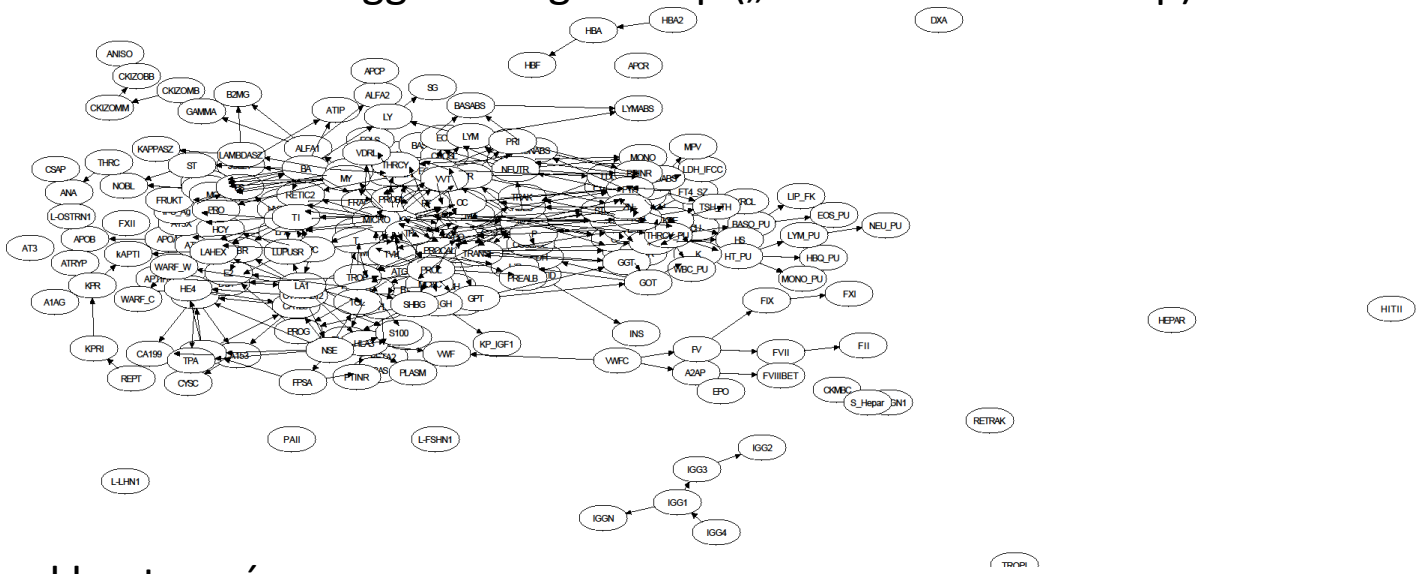
Guenfoud, Zeyneb, and Péter Antal. "Bayesian exploration of dependencies of laboratory tests and evaluation of test redundancy." *2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS)*. IEEE, 2018.

Diagnosztikai laboratóriumi tesztek elemzése II.

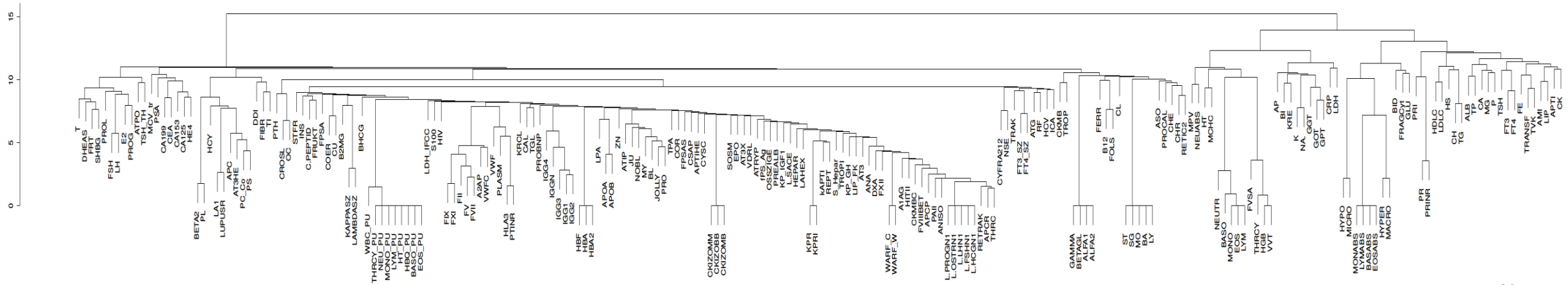
Páronkénti associációs hálózat



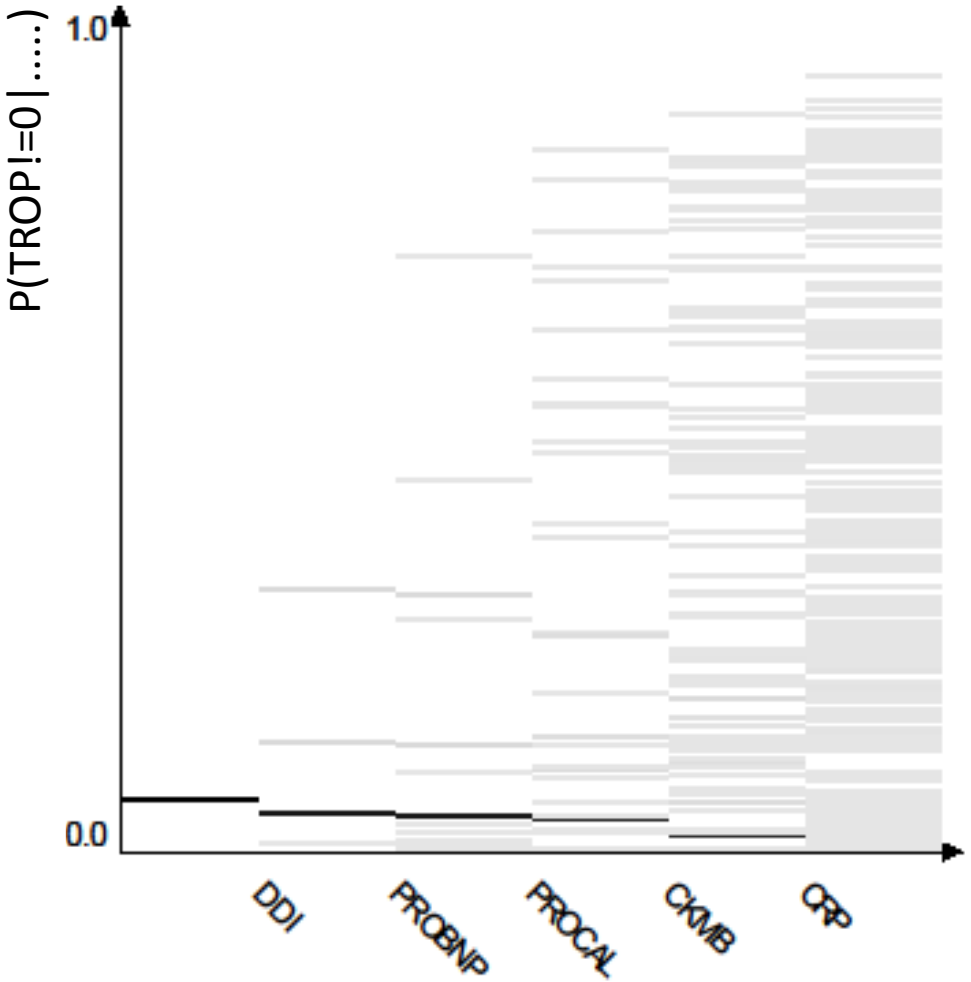
Függetlenségi térkép („multi-marker” térkép)



Hierarchikus klaszterezés

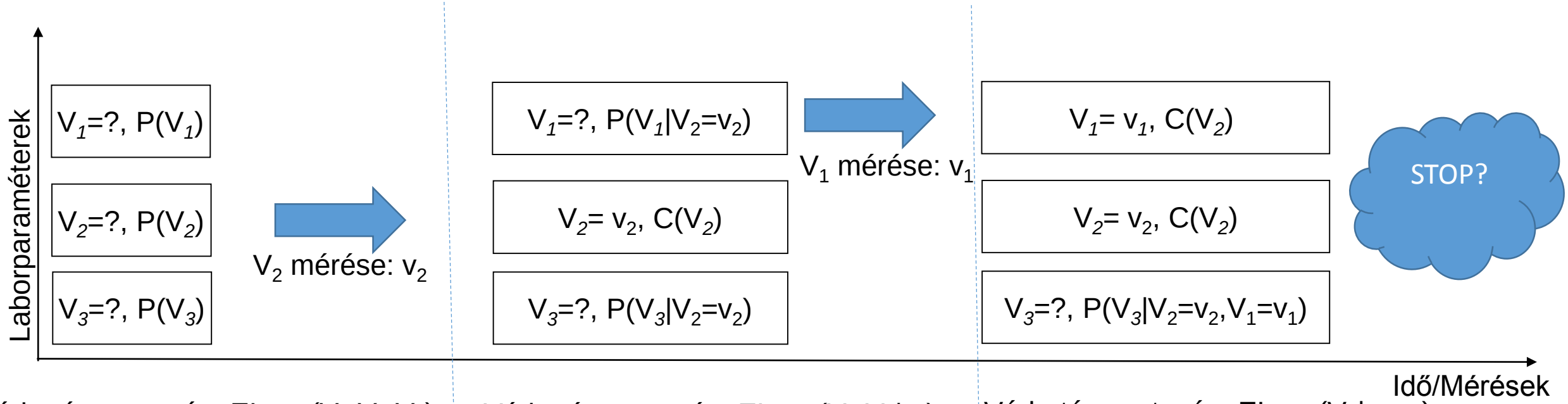


Szekvenciális labordiagnosztika



Troponin normál értéktartományon kívüliségének feltételes valószínűsége egyes szcenáriókban.

Szekvenciális labordiagnosztika: leállás



Várható veszteség: $ELoss(V_1,V_2,V_3)$

V_i (perfekt) megmérésének várható nyeresége:
$ELoss(\underline{V}_{-i},|V_i)$- $ELoss(V_1,V_2,V_3)$

V_i megmérésének költsége: $C(V_i)$

V_i mérési ideje: $T(V_i)$

Várható veszteség: $ELoss(V_1,V_3|v_2)$

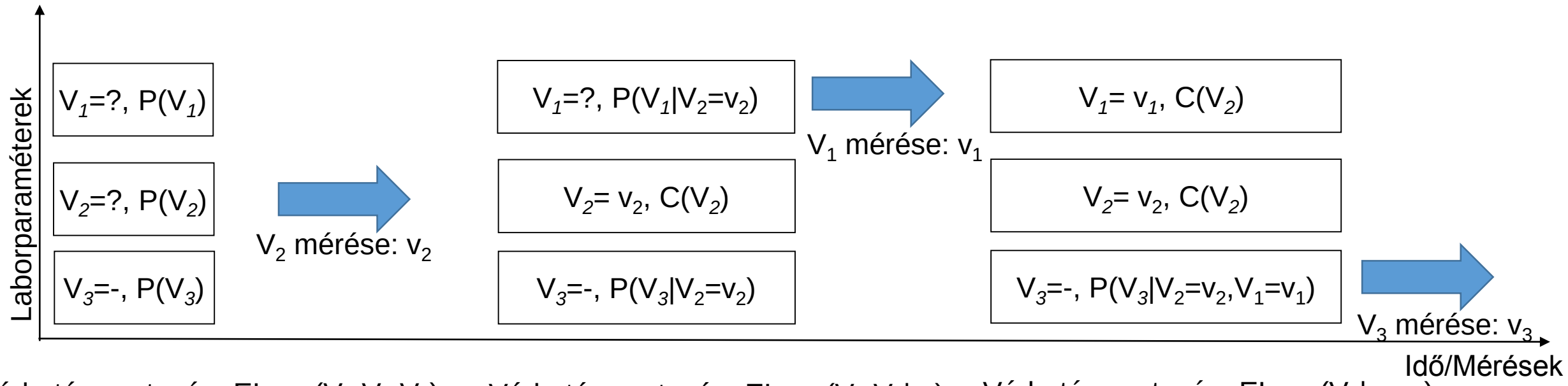
V_j (perfekt) megmérésének várható nyeresége:
$ELoss(\underline{V}_{-i,j}|V_i,V_{ij})$ - $ELoss(V_1,V_3|v_2)$

Várható veszteség: $ELoss(V_3|v_1,v_2)$

V_3 (perfekt) megmérésének várható nyeresége:
$ELoss(.|v_1,v_2,v_3)$ - $ELoss(V_3|v_1,v_2)$

„Optimal stopping" problem

Szekvenciális labordiagnosztika: kiterjesztés



Várható veszteség: $ELoss(V_1,V_2,V_3)$

V_i (perfekt) megmérésének várható nyeresége:
$ELoss(\underline{V}_{-i},|V_i) - ELoss(V_1,V_2,V_3)$

V_i megmérésének költsége: $C(V_i)$

V_i mérési ideje: $T(V_i)$

Várható veszteség: $ELoss(V_1,V_3|v_2)$

V_j (perfekt) megmérésének várható nyeresége:
$ELoss(\underline{V}_{-i,j}|V_i,V_{ij}) - ELoss(V_1,V_3|v_2)$

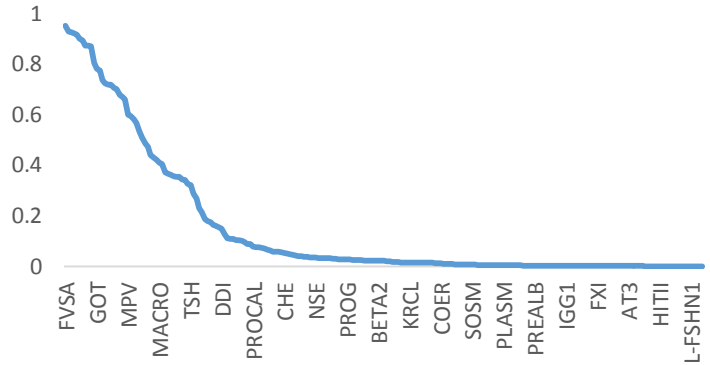
Várható veszteség: $ELoss(V_3|v_1,v_2)$

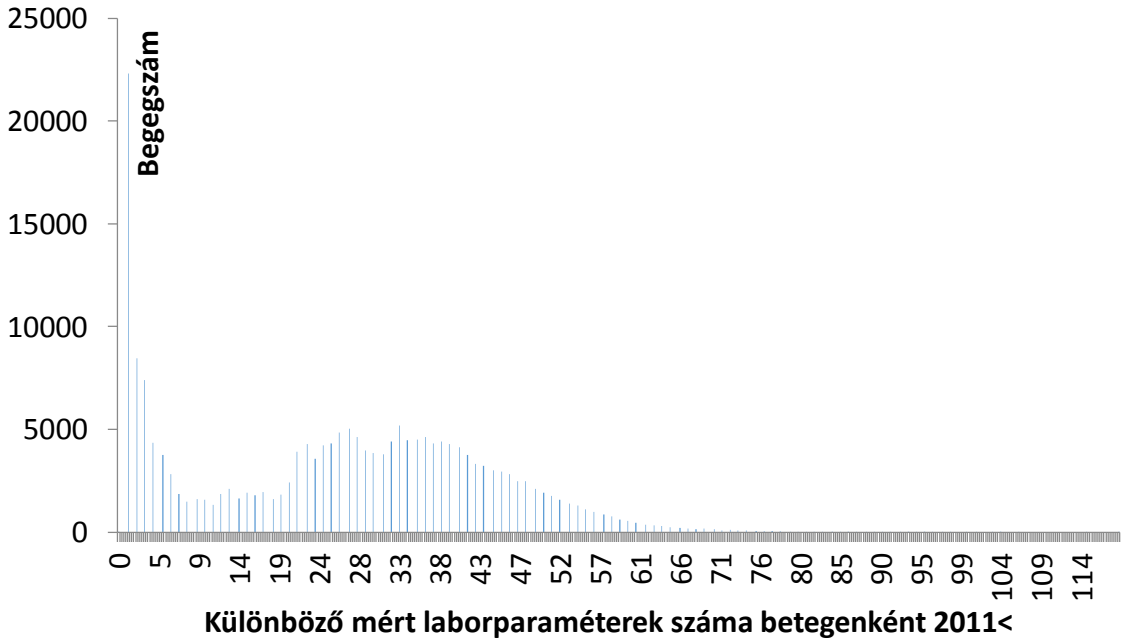
V_3 (perfekt) megmérésének várható nyeresége:
$ELoss(.|v_1,v_2,v_3) - ELoss(V_3|v_1,v_2)$

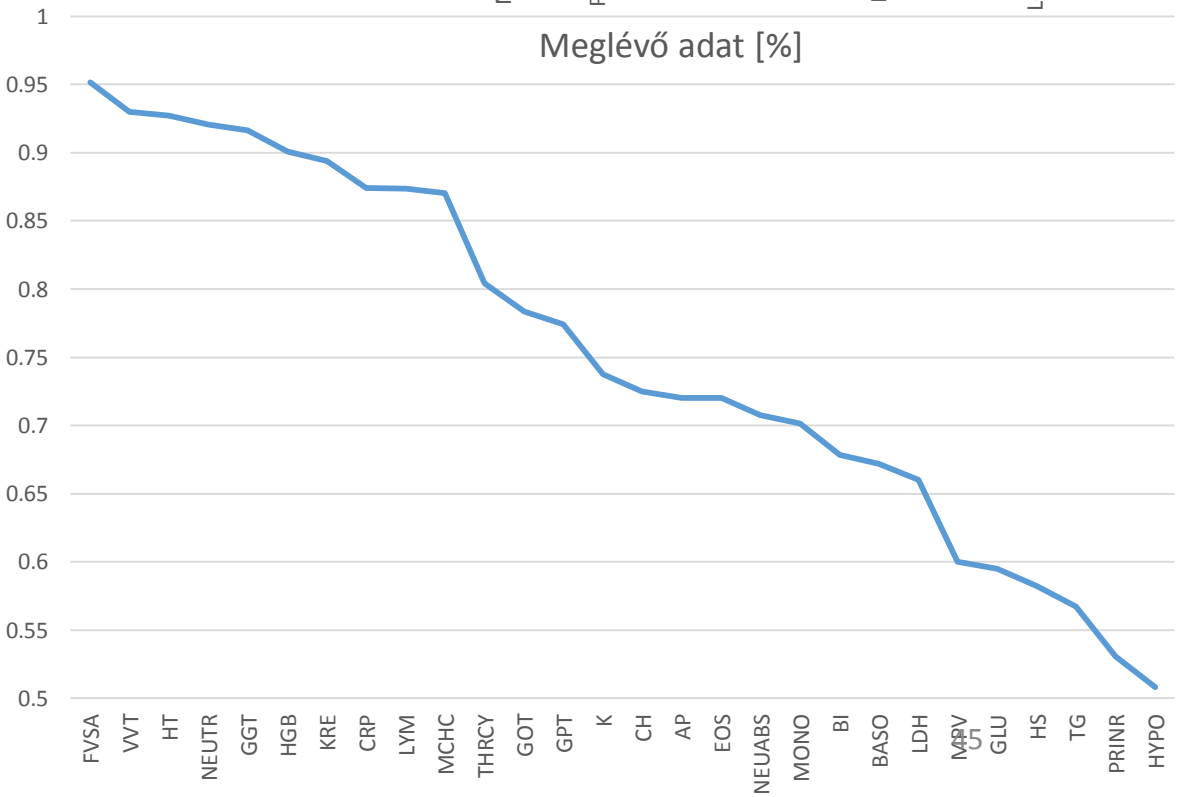
Budgeted learning

Modellek tanulása erősen hiányos adatból

- Erősen hiányos adat
  - 202958 résztvevő, 225 változó: 14.78% elérhető



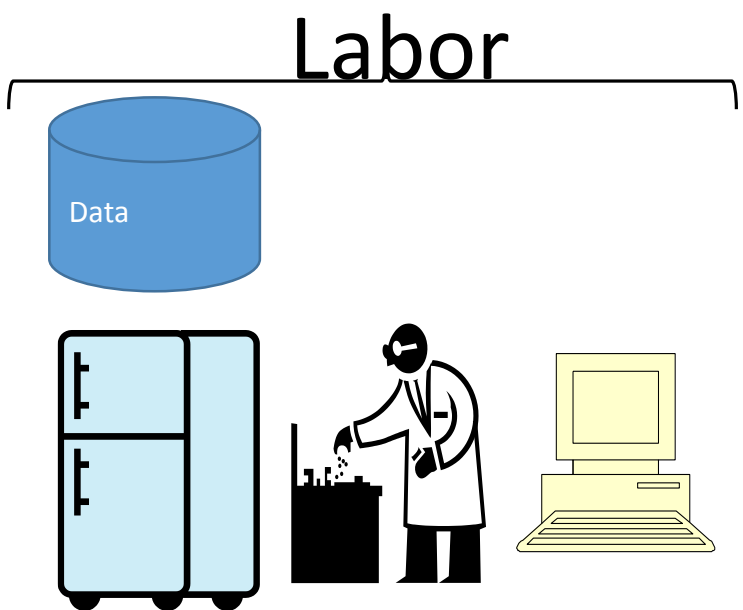




Protokoll: orvos, labor, rendszer, adatbázis

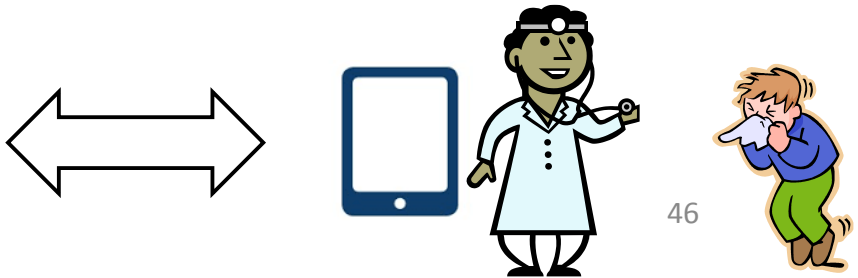
Szempontok, kihívások és egy javasolt architektúra

1. Adatbázistól független döntéstámogató rendszer
2. Labor általi (~orvostól független) költséghatékonysági kontroll (+/-!)
3. Orvoskontroll biztosítása (betegadatok teljessége és védelme)
4. Labormérések indításának segítése adott minta esetén
5. Labor kapacitásának automatikus figyelembevétele



Folyamatos tájékoztatás a keletkező eredményekről és következő mérendő(k)ről

Finomítási lehetőség: leállás-kiterjesztés



Adaptív mérési stratégiák

- Tesztek költségén alapuló (drágábbak felé haladó)
- Idői alapú
  - Legrosszabb (teljes) mérés is adott idői korlátba elkészüljön
  - Az optimalizált adaptív mérés várható értéke maradjon egy korlát alatt
  - Orvosi visszajelzésre várás ütemezése
    - Leállásra?
    - Kiterjesztésre?
- Tesztek várható költségén alapuló
- Várható nyereségen alapuló optimalizálás

➔ „**Myopic information gathering policies**”

Összefoglaló

- Átfogó, redundanciát kezelő valószínűségi modellek
  - nagy mennyiségű adatból tanulhatóak
  - egy és több lépéses döntéstámogatásban felhasználhatóak.
- (Minimális) Szükséges és elégséges diagnosztikai tesztek együttese
  - Markov-határ/takaró/burkoló
- Diagnosztikai teszt relevanciája
  - Erősen releváns: önálló információt hordoz, nem helyettesíthető
  - Gyengén releváns: hordoz információt, de helyettesíthető
- Diagnosztikai teszt (döntéselméleti) értéke
  - Várható hasznosság javulása

Köszönöm a figyelmet!