

M Ű E G Y E T E M 1 7 8 2

# **Rejtett változós Bayes-hálóok tanulása szakirodalom felhasználásával**

**Millinghoffer András**

**Budapesti Műszaki és Gazdaságtudományi Egyetem**

**Méréstechnika és Információs Rendszerek Tanszék**

**2004.**

**konzulens: Antal Péter**

## Kivonat

A Bayes-hálók napjaink egyik legfontosabb, absztrakt bizonytalan tudást ábrázoló eszközei, így kutatásuk kiemelkedően fontos. Legfőbb alkalmazási területeik közé tartozik a döntéstámogatás, azaz bizonyos megfigyelések alapján más értékek valószínűségének számítása (becslése), vagy a legvalószínűbb konfiguráció keresése, valamint különböző gépi tanulási módszerek, különösen az a priori ismereteket is felhasználók. Mindkét esetben a sikeres alkalmazások oka az, hogy a Bayes-hálós tudásreprezentáció az igen nagy számú változó együttes eloszlását is hatékonyan ábrázolja, mind valószínűségi-statisztikai, mind számításelméleti és emberi értelmezhetőség szempontjából is.

A Bayes-hálók hatékonyságának egyik fő oka, a rendelkezésre álló a priori tudást már a tulajdonképpeni tanítást megelőzően is hatékonyan meg lehet jeleníteni a hálóban. Sok kutatási területen jelentős szakértői tudás jelenik meg a területtel kapcsolatos szakirodalomban, a megjelentetett publikációkban. E dolgozat fő kutatási irányvonala azt vizsgálja, hogy ezt a rendelkezésre álló szaktudást milyen automatizált, gépi módszerekkel lehet a szakirodalmi cikkek tömegéből hatékonyan kinyerni, és a tanított Bayes-hálókban megjeleníteni, ezzel elősegítve a hálónak a területre vonatkozó tanítását.

A Bayes-hálók általános tanulásának egyik fontos, és a valós alkalmazásoknál is gyakran előforduló esete, amikor a tanításra használt adatok hiányosak, vagy a modell bizonyos változói nem megfigyelhetők. Mivel ez a probléma a szakirodalom alapján történő tanulás esetén is megjelenik, a dolgozat az ekkor felmerülő számításigény-növekedés lehetséges kezelési módjaira is kitér.

A dolgozat egy új szakirodalmi háló tanulási módszert (egy rejtett változós kétszintű Bayes hálós modellen alapulót) vet össze a szakértői modellekkel és egy közvetlen, egyszintű Bayes hálós modellen alapuló tanulási algoritmussal. Az eredmények igazolták, hogy a vizsgált újabb módszer is jól teljesít egy szakértői referenciához képest, illetve csekély, de mérhető javítást jelent az elsőhöz képest.

## **Abstract**

Bayesian networks are frequently applied to represent uncertain knowledge. Main applications include decision support techniques, as well as different machine learning methods, especially the ones using a priori knowledge. In both cases the reason for their successful use is that Bayesian networks efficiently represent the joint distribution of large number of random variables, providing wide range of statistical and computational advantages and supporting human interpretation. Another reason of the Bayesian network's efficiency is that available a priori knowledge can be effectively incorporated in the network, even before the learning phase. In many domains large amount of expert knowledge is available in electronic publications.

This paper examines models and methods based on Bayesian network representation and learning techniques to effectively utilize electronic scientific domain literature as available expert knowledge.

An important theoretical and practical challenge in learning Bayesian networks, when the data used for learning are not complete or some variables of the model are unobservable. Because this problem appears also in the case of learning from literature, this paper considers also relevant methods with increasing computational complexity.

The paper investigates a new Bayesian network model with hidden variables to represent and learn domain models from literature. The result is compared with expert models and with a standard Bayesian network learning method. The evaluation shows that the performance of the new approach is comparable to a gold standard defined by a domain expert and slightly outperforms the performance of standard Bayesian network learning algorithms using fully observable models.

# Tartalomjegyzék

1.	Bevezetés .....	1
2.	A Bayes-hálóknak .....	3
2.1.	Alapvető tulajdonságok .....	3
2.2.	Az adatszerkezet .....	4
2.3.	Kanonikus valószínűségi táblák .....	5
2.4.	Bayes-hálóknak és a tudásmérnökség .....	6
3.	Következtetés Bayes-hálóknakban .....	7
3.1.	A következtetés alapesetei .....	8
3.2.	Következtetés fa gráfokban .....	9
3.3.	Következtetés többszörösen összekötött hálóknakban .....	9
3.3.1.	Sztokasztikus approximáció .....	9
3.3.2.	Csomópontok összevonása .....	10
3.3.3.	A PPTC algoritmus .....	10
4.	Bayes-hálóknak paramétertanulása .....	11
4.1.	A tanításhoz használt hiányos adatokkal kapcsolatos kritériumok .....	12
4.2.	A „maximum likelihood” módszer .....	13
4.3.	A gradiens alapú módszer .....	13
4.4.	Kanonikus CPT-k tanulása .....	15
5.	Bayes-hálóknak struktúratanulása .....	17
5.1.	A struktúratanulás alapjai .....	17
5.2.	Az elvek alkalmazása teljes adatok esetén .....	19
5.3.	Heurisztikák hiányos adatok kezelésére .....	20
5.3.1.	Csomópontonkénti építés .....	20
5.3.2.	Mohó eljárás .....	20
5.3.3.	Kimerítő keresés .....	21
5.4.	A priori tudás beépítése a hálóba .....	21
5.5.	A használt kritériumfüggvény .....	21
5.6.	Kritériumfüggvény az irodalmi modellhez .....	22
6.	Hálóépítés szakirodalom alapján .....	23
6.1.	A kauzalitás megjelenése a szakirodalmi cikkekben .....	24
6.2.	Struktúratanulás szakirodalom alapján .....	25
6.2.1.	Közvetlen modell .....	25

6.2.2.	Kétszintű modell .....	26
6.3.	A tanítási adatok kinyerése a cikkekből.....	27
7.	A struktúratanulás minősítése .....	28
7.1.	A változók között fennálló függési viszonyok .....	29
7.2.	Két hálóstruktúra összevetése .....	30
8.	Az alkalmazási terület .....	31
8.1.	A terület orvosi jellemzése.....	31
8.2.	Az IOTA project.....	33
8.3.	Az irodalmi adatok .....	33
9.	Az alkalmazott algoritmusok megvalósítása .....	33
9.1.	A 'BNet' rendszer.....	33
9.2.	A tanulás minősítése .....	34
9.3.	A paramétertanulás .....	34
9.4.	Struktúratanulás, és heurisztikák .....	36
9.4.1.	Mohó algoritmus.....	36
9.4.2.	Csomópontonkénti építés .....	37
9.5.	Irodalmi modell tanulása.....	37
10.	Mérési tapasztalatok és eredmények .....	39
10.1.	A struktúratanulási heurisztikák összevetése .....	39
10.1.1.	Konklúzió .....	40
10.2.	Irodalmi modell .....	41
10.2.1.	A tanuláshoz felhasznált változók .....	41
10.2.2.	A tanulás érzékenysége a paraméterekre .....	41
10.2.3.	Eredmények.....	46
11.	Konklúzió.....	51
12.	További lehetőségek.....	52
13.	Irodalomjegyzék.....	53
14.	Függelék – tanulás.....	55
14.1.	Paraméterek változása.....	55
14.2.	Szülői halmazok jósága.....	57

# 1. Bevezetés

A mesterséges intelligencia kutatások története során általánosan felmerülő feladat volt a birtokunkban álló ismeretek, az absztrakt tudás ábrázolása. A számos lehetséges és alkalmazott struktúra közül mára a Bayes-hálók váltak a valószínűségi tudás megjelenítésének és kezelésének általánosan elfogadott eszközeivé, mivel a Bayes-hálók rendkívül hatékonyan képesek a valószínűségi változók között fennálló összefüggéseknek mind a kvantitatív, mind a kvalitatív ábrázolására.

A Bayes-hálókat formálisan egy irányított körmentes gráf alkotja, melynek csomópontjai a valószínűségi változókat, élei pedig a változók közti közvetlen függéseket jelölik. Minden csomóponthoz ezen kívül egy-egy feltételes valószínűségi tábla (CPT – conditional probability table) tartozik, mely a csomópont jelképezte változónak a szülei szerinti eloszlását írja le.

A Bayes-hálókat olyan területeken alkalmazzák, mint az orvosi diagnosztika, genetika, gazdasági és banki döntéstámogatás. Ezek mellett a számos más egyéb felhasználási lehetőségük és módjuk is kiemeli őket a többi használt eszköz közül.

A fentiek mellett a Bayes-hálók egyik komoly hátrányossága, hogy igen nagy számú paramétert igényelnek, és a lehetséges gráfstruktúrák mennyisége is szuperexponenciális a csomópontok száma szerint. Ez, együtt azzal a gyakori körülménnyel, hogy a tanításhoz felhasználható adatok hiányosak (ami további jelentős költségnövekedéssel jár), azt eredményezi, hogy valós esetekben a kimerítő (és így biztosan optimális eredményt szolgáltató) tanulás nem járható út.

Az e problémát feloldó eljárások keresése ma is kutatás tárgya. Léteznek eljárások, melyek pusztán a struktúrák kereséséi terének beszűkítésével kívánják elérni a kezelhető léptéket. E módszerek, bár figyelmen kívül hagynak bizonyos struktúrákat, általában elfogadható közelítés adnak.

Egy másik lehetőség (amellyel ez a dolgozat is foglalkozik), ha a rendelkezésre álló szakértői tudást használjuk fel a tanítás során, vagy még inkább az előtt. E területtel foglalkozik a tudásmérnökség tudománya, mely a Bayes-hálók esetében elsősorban a lényeges változók meghatározását, és a köztük lévő kapcsolatok feltérképezését jelenti. Az efféle „kézi” módszerek előnye az egyszerűség, és hogy a kialakított szerkezetek jól ábrázolják a területről alkotott elképzeléseket, így a modell az emberi felhasználók számára is szemléletes marad. Hátrány viszont, hogy így csak egy-egy személy tudására

tehetünk szert, és le kell mondanunk a szakirodalom révén rendelkezésre álló egyéb forrásokról.

A természetes nyelvű szakirodalmi cikkek ilyen célú felhasználása még meglehetősen feltáratlan terület, holott a várható nyereségek kecsegtetőek. A szakirodalmi cikkek tömeges, automatizált felhasználása a Bayes-háló tanítása számára nagy előrelépést jelenthet a tudás ábrázolása és felhasználása terén.

E dolgozat fő témája tehát a természetes nyelvű szakirodalmi cikkek alapján történő hálóstruktúrák alkotása, és azoknak összevetése más referenciahálókkal, elsősorban a szakértők által konstruált, a területről alkotott szaktudást ábrázoló hálókkal.

Bár ezek a módszerek jelentősen könnyíthetik a tanulást, elsősorban a keresési tér szűkítése és egyszerűsítése által, változatlanul fel kell használnunk bizonyos heurisztikus módszereket, a vizsgálandó struktúrák számának csökkentése érdekében.

A fentiek tükrében tehát a dolgozat beosztása a következő lesz:

- A Bayes-háló alapvető bemutatása, adatszerkezetük, kapcsolatuk az adatmérnökséghez
- Következtetés Bayes-hálóban
- A Bayes-háló paramétertannálása, alapesetben teljesen megfigyelhető adatok esetén, illetve egy, hiányos adatok vagy rejtett változók mellett is alkalmazható eljárás, és adaptációja kanonikus valószínűségi táblákra
- A szakirodalmakban fellelhető kauzális összefüggések, és ezek átültetése a vizsgálati területet ábrázoló hálóba, illetve a módszer alkalmazási lehetőségei struktúratannálás esetében
- A struktúratannálás minősítési lehetőségei, különböző hálókat egymással, illetve a szakértői tudással való összevetése
- Az alkalmazási terület és felhasznált adatok leírása
- A megvalósított algoritmusok részletes leírása
- Mérési eredmények és konklúzió
- További kutatási lehetőségek
- Irodalomjegyzék

A dolgozat kutatási iránya több tudományágból is merít, olyan területeket érintve, mint például a tudásmérnökség, a gépi tanulás, vagy az információkivonatolás. E területek

kimerítő és teljes áttekintése meghaladja e dolgozat kereteit, így az érdeklődők számára a következő, áttekintő cikkek ajánlhatók:

- Bayes-hálóok felhasználása a tudásmérnökség területén: (Mahoney96, Neil00)
- gépi tanulás: (Heckerman95b)
- információkivonatolás: (Hirschman02)

## 2. A Bayes-hálóok

### 2.1. Alapvető tulajdonságok

A bayesi hiedelemhálóok, röviden a Bayes-hálóok kutatása a valószínűségi alapon működő következtető rendszerek és a gépi tanulás egyik fontos, ma is fejlődő ágát alkotják. A világ modellezni kívánt elemeit valószínűségi változókkal ábrázoljuk. Az említett változók egy irányított gráf csúcspontjait alkotják, az élek pedig a köztük lévő korrelációs kapcsolatot jelölik. Fontos feltevés, hogy a gráf körmentes, hisz ez a tulajdonság biztosítja a Bayes-hálóok sok elméleti és gyakorlati tulajdonságát

Bár két változó korreláltságát irányítatlan élek is jelölhetnék (ahogy a Markov-hálóokban), hisz a köztük lévő kapcsolat nem feltétlenül ok-okozati, azok irányítottságának fontos technikai és tudásmérnöki jelentősége van, amint azt majd a következtetésről szóló részben látjuk.

A csúcspontok által jelképezett változók feltételes eloszlása a háló fő paramétere (a struktúra mellett), ezek az adott csomópont eloszlását tárolják, annak szülőitől függően. Egy csomópont szülői azon csúcsok, amelyekből él vezet a csomópontba. A fent említett eloszlások lehetnek folytonosak, vagy diszkrét. Jelen tanulmány körében diszkrét változókkal foglalkoztam. A dolgozatban a bayesi értelmezéssel szemben a klasszikus, pontparametризációt és tanulásnál a hozzá kapcsolódó „maximum likelihood” elvet követtem.

A Bayes-hálóok alapvetően az együttes eloszlás hatékony dekompozícióját teszik lehetővé, ami mind a gépi következtetés és a tanulás, mind a tudásmérnökség szempontjából is nagy jelentőségű. Az egyik gyakori alkalmazási terület következtetések végrehajtása, mivel ez a hatékony reprezentáció miatt gyorsan, kis költséggel végrehajtható, azaz néhány változó ismerete mellett a többi értékének valószínűségét ki lehet számolni, vagy közelítő becslést lehet adni rá.



E struktúra több előnyös tulajdonsággal is rendelkezik más döntéstámogató rendszerekhez képest. Ezek az előnyök, a neuronhálókhöz képest például, a következők:

- az egyes csomópontok jól behatárolt, a valósághoz erősen kapcsolódó jelentéssel bírnak, így a háló által hozott döntéseket indoklással is könnyen alátámaszthatjuk a csúcsok közti összefüggésekre hivatkozva
- további előnye ennek, hogy a kialakult hálóstruktúra könnyen értelmezhető, világos leírást ad a megfigyelt jelenségek közti összefüggésekről, aminek segítségével könnyen alkothatunk egyszerű ökölszabályokat
- ennek megfordításaként, ha van némi elképzelésünk a vizsgált jelenségkörrel, annak segítségével jó első közelítést adhatunk a háló struktúrájára
- a fentiek mellett szintén hasznos, hogy hiányos adatok alapján is végrehajthatjuk a következtetést, azaz bárhányat is ismerünk a változókból, a többire adhatunk becslést
- ezeken felül a Bayes-hálóknak nincsenek megkötve a ki- és a bemeneti szerepek, bármely változóra következtethetünk a többi ismeretében

A másik gyakori alkalmazási terület a gépi tanuláshoz kötődik, vagyis a megfigyelt adatok alapján javíthatunk a paramétereken vagy modellstruktúráján. A tanítást mind az egyes valószínűségi táblák szintjén, a struktúrát megtartva (paramétertanulás), mind az élek változtatásával a struktúra szintjén is megtehetjük. Struktúratanulás esetén a paramétertanulás (mint segédalgoritmus) szintén jelen van.

A tanulás fontos jellemzője, hogy a megfigyelési adatok teljesekek vagy hiányosak, ugyanis míg teljes adatokra az analitikus eredmények és dekomponálási lehetőségek miatt viszonylag gyors és egzaktan tekinthető eljárások állnak rendelkezésre, addig hiányos adatok alapján csak jóval lassabban és csak közelítő jelleggel tanulhatunk.

A fenti előnyök mellett a struktúra hátránya a viszonylag bonyolult adatszerkezet, ami a komplikáltabb, lassabb algoritmusokban nyilvánul meg.

A Bayes-hálókat elsődlegesen diagnosztikai rendszerekben alkalmazzák, méretük a néhány csomópontostól akár az ezres nagyságrendig változik.

## **2.2. Az adatszerkezet**

Röviden összefoglalva tehát minden Bayes-háló formálisan a  $B(G, \Theta)$  párossal írható le, ahol  $G$  jelenti a struktúrát, amely egy irányított, körmentes gráf.

A gráf csomópontjai a kezelt változók. Egy csomópont gyermekei azon csúcsok, melyekbe a csomópontból él fut. Hasonlóan egy csomópont szülei azon csúcsok, amelyekből a csomópontba él vezet.

Az általam vizsgált Bayes-hálókbán a csomópontok kivétel nélkül véges sok értéket felvevő, diszkrét valószínűségi változókat jelképeztek. A Bayes-háló „működése” szempontjából elsősorban a ’gyermek-szülők’ halmazoknak van jelentősége. Minden csomóponthoz egy-egy feltételes valószínűségi tábla tartozik, amely a gyermek változó eloszlását tartalmazza a szülők behelyettesítésének függvényében, vagyis minden csomópont mellett

$$n_c \prod_{\forall i: \text{szülő}} n_i \quad [2.1]$$

sok valós paramétert kell tárolnunk (ahol ’ $n_c$ ’ a gyermek, ’ $n_i$ ’ a szülők lehetséges értékeinek száma). A  $B(G, \Theta)$  párosban a második tag, a ’ $\Theta$ ’ jelenti e paraméterek összességét.

### 2.3. Kanonikus valószínűségi táblák

A Bayes-hálóknak számos hasznos tulajdonságuk mellett egyik nagy hátrányuk, hogy viszonylag sok paramétert igényelnek. Míg például a neurális hálózatokban az élekekhez tartozik egy-egy súly addig a Bayes-hálókbán minden csomópontban [2.1] számú súly van, ami a tanulást jelentősen lassítja, a paraméterek nagy száma miatt.

Ezt a komplexitást csökkenthetjük, ha a közös eloszlások helyett (melyeknek minden egyes bejegyzése önálló paraméter) úgynevezett kanonikus eloszlásokat használunk, melyek a hagyományosnál jóval kevesebb paraméterrel leírhatók.

Az ilyen eloszlások alkalmazása esetén általában abból a feltevésből indulunk ki (Pearl88), hogy a gyermek változó értékvételét a szülők, mint egymástól függetlenül fellépő okok indukálják. Ebben az esetben minden szülő hatása leírható egyetlen paraméterrel, a közös hatások pedig az egyes paraméterekből, egyszerű matematikai képletekkel származtathatók.

Az e dolgozat fő kutatási témáját jelentő irodalmi modellek esetében is ilyen kanonikus eloszlásokat használunk majd, konkrétan a ’zajos-vagy’ (noisy-or) eloszlást, mely kétértékű változók kapcsolatát írja le, a következő módon: minden változó kétértékű, melyek a logikai ’0’-nak és ’1’-nek felelnek meg. Bármely szülő ’1’ értéke (mint a logikában) okozhatja a gyermek ’1’ értékét, de ebben az esetben van bizonyos ’p’

valószínűsége, hogy a szülő érke nem vonja maga után a gyermekét. Minden szülő a többitől függetlenül hat, így annak a valószínűsége, hogy a gyermek '0' marad:

$$P(x=0) = \prod_{i:u_i=1} p_i \quad [2.2]$$

Ennek megfelelően a gyermek '1' értékének a valószínűsége:

$$P(x=1) = 1 - \prod_{i:u_i=1} p_i \quad [2.3]$$

Ahol 'x' a vizsgált változót, 'u<sub>i</sub>' pedig a szüleit jelöli. A fenti képletben szereplő 'p<sub>i</sub>' valószínűségeket szokás a 'noisy-or' eloszlás zajparamétereinek is nevezni.

#### 2.4. Bayes-hálók és a tudásmérnökség

Tehát amint azt láthattuk, a Bayes-hálókat gyakran alkalmazzák szakértői rendszerekben, ahol a vizsgálati területről rendelkezésre álló tudás reprezentálására, illetve pontosítására használják. Ezek miatt alkalmazásuk egyik alapvető feltétele, hogy tudásunkat megfelelően be tudjuk építeni.

A fenti feladatok jó megoldása azért is különösen fontos, mert míg a valószínűségi táblák paramétereinek tanulása meglehetősen egyszerű, addig az adatok statisztikai összefüggéseit reprezentáló élek keresése ennél már jóval költségesebb. Emellett, bár léteznek módszerek új csomópontoknak a hálóba építéséről, ezeket valós mérési adatokkal összerendelni nem lehet. Vagyis lehetséges, hogy a háló paraméterszámát csökkenteni tudjuk új csomópontok felvételével (lásd (Russel95, 00)), ám ezek a változók úgymond rejtettek, azaz értékük sohasem ismert.

Látható tehát, hogy a háló építésekor számos információt kell beépíteni, ugyanis ha ezt elmulasztjuk, nem fogjuk tudni megfelelően leírni a megfigyeléseinket. Ezzel szemben természetesen az is igaz, hogyha túl komplex modellt, azaz túl sok szülőt veszünk figyelembe, a háló túl bonyolult lesz, ebből adódóan lassan fog működni, és könnyen lehet, hogy a vizsgálati terület helyett csak a mérési adatokat írja le, azaz túlilleszkedik. Új modell építésekor amennyire csak lehetséges, fel kell használni a vizsgálat tárgyáról ismert tudást. A tudás megszerzése általában a rendszer fejlesztője és a terület egy szakértője közötti megbeszélés folyamán történik, és tipikusan a következő lépésekből áll:

- Csomópontok definíciója: melyek azok a változók, amelyek fontosak és jól leírják a területet. (Például orvosi alkalmazás esetén, mely vizsgálatok alapján lehet nagy valószínűséggel pontos diagnózist felállítani)

- A változók közötti oksági, vagy időrendi sorrend felállítására is szükség lehet, ami majd a háló topológikus rendezését adja. Erre a struktúranulásnál lehet szükség.
- Változók értékei: milyen értékeket vehetnek fel az egyes változók, azaz melyek azok az értéktartományok, amelyekbe tartozás esetén változik az érték megítélése. (Például a 'magasság' változó tartományai lehetnek: 'alacsony': [-165], 'átlagos': [166-185], 'magas': [186-])
- Élek meghatározása: ez a változók közti szoros vagy laza meghatározottságok (esetleg ok-okozati kapcsolatok) meghatározása. Ez az emberek számára még viszonylag egyszerű feladat, hisz általában a szakértők is amolyan oksági ökölszabályokat alkalmaznak. Emellett már egy ilyen háló is jó első közelítést adhat, vagy jó kiindulási pont lehet a struktúranulás során.
- Általában már itt sor kerül a fentebb meghatározott topológikus (másképp kauzalitási) sorrend felhasználására, vagyis a csomópontokat e szerint építik be a hálóba, mindig a már meglévők közül kiválogatva az új szülőit.
- Valószínűségi értékek (CPT bejegyzések): tekintve, hogy ezek tanulása eléggé könnyű, az emberek számára viszont szinte lehetetlen számokba önteni az általuk használt szabályokat, a háló paramétereit már az adatok alapján a háló maga tanulja.

A fenti fejezetben egy szakértőről és egy modellről beszéltünk, természetesen az információgyűjtésben szinte mindig több szakértő vesz részt (ami nem ritkán okoz ellentmondásokat). Egy háló helyett is gyakran több készül el, amelyek például lehetnek egymás részei is, azaz készítünk egy hálót a legfontosabb csomópontokból, amihez aztán hozzávesszük a közepesen, majd a kevésbé fontosakat is. A csomópontokéhoz hasonlóan az élek osztályozása is lehetséges ily módon.

### **3. Következtetés Bayes-hálókbán**

A Bayes-háló egyik fontos feladata, hogy szakértői rendszerekben a modellezett változók egy adott értékére (vagy a változók eloszlására) adjanak becslést, az ismert behelyettesítésű változók függvényében. Másik tipikus feladat, amikor az ismeretlen változók legvalószínűbb konfigurációját keressük.

A következtetési alapfeladat tehát ez: ismert bizonyos változók értéke, keressük ennek függvényében a többi (ám nem feltétlenül az összes) változó eloszlását, vagy a teljes

eloszlás helyett ezek egy adott konfigurációjának valószínűségét. Ekkor az ismert változókat tényváltozóknak, a keresetteket lekérdezéses változóknak hívjuk.

Egyrészt a kérdés ily általánosan feltehető mivolta adja a Bayes-háló népszerűségét, hisz tetszőlegesen a ki- és a bemeneti szerepek, másrészt hogy bármilyen kevés tudás birtokában is meg tudjuk adni a lehető legpontosabb választ (nem úgy, mint például a neuronhálóknál, ahol elengedhetetlen az összes bemenet ismerete).

A következtetés során fontos szerep jut a „feltételes függetlenség” elvének (lásd (Russel00, Pearl88)), azaz annak, hogy két változó egymástól függetlenné válik, ha bizonyos más változók értékei ismertté válnak. Ahhoz hogy egy változót így elszigeteljünk a háló többi részétől, például az kell, hogy ismerjük az adott változó összes szülőjének, gyermekének és gyermekei szülőinek értékét (azaz az úgynevezett minimális Markov-határát, a Markov-takaróját; lásd (Pearl88)).

### 3.1. A következtetés alapesetei

A legegyszerűbb következtetési feladat, amikor ismert egy vagy több szülő csomópont, és keresett gyermekük eloszlása:

$$P([Q_1] = [q_1] | [E_1..E_m] = [e_1..e_m]) \quad [3.1]$$

ahol 'Q<sub>i</sub>'-k a keresett változók, 'E<sub>i</sub>'-k az ismertek.

Ekkor a feladat (a csomópontok független valószínűségi tábláinak definíciójából adódóan) megoldható bizonyos hálóparaméterek összegzésével.

Ennél a Bayes-háló azonban jóval többre képesek, a fenti módon szerzett ismeretek továbbterjeszthetők az élek iránya mentén lefelé, de kihasználva a Bayes-tételt, felfelé is.

Egy másik fontos feladat, amikor a behelyettesített változók függvényében keressük a szabadok legnagyobb valószínűségű konfigurációját, ám ennek e dolgozat keretében nincs szerepe.

$$[q_1..q_n] = \arg \max P([Q_1..Q_n] | [E_1..E_m] = [e_1..e_m]) \quad [3.2]$$

Ezeket a lehetőségeket kihasználva tehát a következő alaplómódszerek adottak Bayes-hálóknál történő következtetésre:

- Okozati: ismert a hatás, ennek függvényében keressük ennek valamilyen okozatát, azaz egy belőle irányított úton elérhető változó eloszlását.
- Diagnosztikai: az előző fordítottja, ismert az okozat, keresett az ok.

- Okok közötti: ismert egy adott változó egyik oka, keressük e mellett a másik ok valószínűségét.

Az alaplómódszerek kombinált felhasználásából megkaphatjuk a legáltalánosabb, úgynevezett kevert módszert, amikor például ismert egy változó egy oka (őse) és okozata (leszármazottja), és ezek függvényében keressük a változó jellemzőit.

### **3.2. Következtetés fa gráfokban**

Ha a Bayes-háló fa gráf alakú, vagyis bármely két pontja között legfeljebb egy út létezik, a következtetés viszonylag egyszerű. Ha egy bizonyos csomópont eloszlását akarjuk kiszámítani, a lépések a következők:

1. bontjuk fel a csomópontok halmazát a keresett változó őseire és leszármazottaira (esetleg azokra, amelyek nincsenek a változóval semmilyen úton összekötve)
2. számítsuk ki az ősök hatását úgy, hogy kiszámítjuk a hatást a közvetlen szülőkére, majd ezeket a keresettre
3. hasonlóan a leszármazottak hatása is kiszámítható a fenti két lépésben: először az összes hatása a gyermekekre, majd azoké a keresett változóra

Vagyis látható, hogy a lekérdezési változóból rekurzívan haladunk a bizonyítékok felé, amit úgy is megfogalmazhatunk, hogy a bizonyítékok hatását terjesztjük a keresési alany felé (lásd (Russel00)).

### **3.3. Következtetés többszörösen összekötött hálókbán**

A feladat már nem ilyen egyszerű, ha a hálónak többszörösen összekötött részei is vannak. Ennek az esetnek a kezelésére két fő módszer terjedt el: vagy valamilyen közelítéssel becslést alkalmazunk, vagy a háló gráfját alakítjuk át fa alakúra, hogy a következtetés itt már végrehajtható legyen.

#### **3.3.1. Sztochasztikus approximáció**

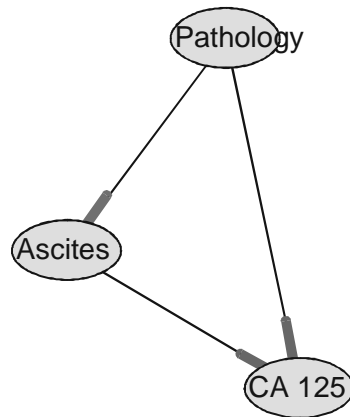
Elvi szinten ez a legegyszerűbb megoldás, és mint neve is mutatja, nem egzakt számításokon, hanem közelítésen alapul.

Az eljárás lényege, hogy a hálóból kiindulva generálunk bizonyos méretű esethalmazt. A csomópontokon a topológikus sorrendben végighaladva behelyettesítünk minden egyes változót az ismert értékek, illetve a már behelyettesített szülők függvényében. Miután a halmaz elkészült, egyszerűen a megfigyelni kívánt esetek részarányával becsüljük azok valószínűségét.

A módszer nyilvánvaló előnye az egyszerűség, hátránya azonban a bizonytalanság, ami főleg kis valószínűségű események esetén jelent problémát. Mivel a kis valószínűségű esetek a generálás során is csak ritkán fordulnak elő, az eljárást sokáig kell folytatni a pontos becsléshez.

### 3.3.2. Csomópontok összevonása

A nem fa gráfokban a problémát az okozza, hogy a bizonyítékok több úton is „terjedhetnek” (mint a [3.1] ábrán 'CA125' felé).



[3.1] ábra - bizonyíték terjedése

Az egyik lehetséges megoldás ekkor az, ha bizonyos csomópontokat összevonunk (úgynevezett megacsomópontokba), ezzel megszüntetve a többszörös utakat. Az új csomópontok ekkor azon változók együttes eloszlását reprezentálják, amelyekből létrehoztuk őket. Az újra fává alakított hálóban a következtetés így már végrehajtható.

### 3.3.3. A PPTC algoritmus

A PPTC (probability propagation in trees of clusters) algoritmus is a fenti ötletből kiindulva oldja meg a feladatot, ám az eljárás optimalizálása érdekében több köztes lépést és struktúrát is felhasznál.

Az eljárás lényege, hogy a Bayes-hálót egy olyan fa (vagy esetleg erdő) gráffá alakítja, melyben a csomópontok az eredeti háló csomópont-halmazai, majd ebben hajtja végre a következtetést. A fában emellett igaznak kell lennie, hogy két csomópont közti út bármely elemének tartalmaznia kell a két csomópont metszetét. Az átalakítás lépései röviden a következők:

1. a háló éleinek elhagyjuk az irányítottságát, majd minden csomópont szüleit páronként összekötjük

2. a gráfot „háromszögesítjük”, azaz addig adunk hozzá éleket, amíg igaz nem lesz, hogy tetszőleges, három élnél hosszabb körben van legalább két összekötött csúcs
3. a kapott gráfban megkeressük a klikkeket (a maximális, teljes részgráfokat), ezek lesznek a fa csúcsai
4. behúzzuk a fa éleit: az üres gráfból indulunk (ahol a csúcsok tekinthetők egy-egy fának), majd a fákat páronként egyesítjük.

A fa csúcsai felett aztán úgynevezett potenciálokat definiálunk (ami a csúcs változóhalmazai eloszlásának általánosítása: itt nem szükséges, hogy az összeg 1 legyen). Ebben a struktúrában a fákra már fentebb vázolt következtetéshez hasonló módon lehet a bizonyítékok hatását terjeszteni. A részletes leírást lásd a (Huang96)-ban. A módszer az egzakt számítást alkalmazók közül a leghatékonyabb, de költsége még így is exponenciális. Bár a fa gráfban már a háló paramétereinek számában lineáris költséggel végrehajtható a következtetés, az eredetihez képest ez mégis exponenciális, hisz a megacsomópontok valószínűségi tábláinak paraméterszáma a létrehozó csomópontok táblái méretének szorzata.

Mint azt majd később láthatjuk, a következtetési mechanizmusra a paramétertárolás során is szükség van. Az általam használt 'BNet' rendszernek ez az algoritmus már a része volt, amit a vizsgálatok elvégzéséhez fel is használtam.

## 4. Bayes-hálók paramétertárolása

A Bayes-hálók tanítási problémája formálisan a következő képletre vezethető vissza:

$$B(\hat{G}, \hat{\Theta}) = \arg \max P(B(G, \Theta) | D) \quad [4.1]$$

ami a Bayes-tétel szerint:

$$P(B | D) = \frac{P(D | B)P(B)}{P(D)} \quad [4.2]$$

Ezt az általános elvet a legnagyobb valószínűség (maximum likelihood) elvének nevezik. Azaz keresett az a háló, mely megfigyeléseink valószínűségét maximálja. Mint látható, a kifejezés értéke függ az egyes konfigurációk előzetes, a priori valószínűségétől. A tanítás egyik alaptípusa, amikor ismert és rögzített struktúra mellett csak az egyes csomópontok valószínűségi tábláinak paramétereit próbáljuk a megfigyelési adatokhoz illeszteni. Ebben az esetben minden paraméterezéshez azonos előzetes valószínűséget rendelünk.



Ez a tanulási mód több szempontból is rendkívüli fontossággal bír, ezek a következők:

- Gyakori, hogy készen kapjuk az optimális struktúrát, vagy annak egy jó közelítését, hisz az élek jelképezhetik az egyes változók közti oksági függéseket, amelyekről az emberi szakértőknek is viszonylag pontos és helyes elképzeléseik lehetnek. Ezzel szemben valószínűségi táblák kitöltését már nem várhatjuk el, viszont a pontosan kalibrált paraméterek nélkül a struktúra nem sokat ér.
- A paramétertalanulás a struktúratanulás során is megjelenik szubrutinként, hisz a hálók modellezési képességének minősítéséhez jól beállított paraméterekre van szükség.

A tanulás során felhasználhatjuk, hogy a maximalizálni kívánt valószínűség felírható a következő módon, táblánként csak lokális (a gyermekre és szülőire vonatkozó) információ felhasználásával:

$$P(X_1 \dots X_n | D) = \prod_i P(X_i | \Pi_i, D) \quad [4.3]$$

ahol ' $\Pi_i$ ' az ' $X_i$ ' változó szülői halmaza (nyilván  $\Pi_i \subseteq \{X_j\}_{j=0}^{i-1}$ , vagyis a szülők csak a topológikus sorrendben előrébb álló csomópontok lehetnek).

Ezt felhasználva, minden egyes csomópont paraméterezését külön-külön állíthatjuk be. Mint már említettük, bármiféle tanulási algoritmus esetében kardinális kérdés, hogy a felhasznált adatok teljesek-e. Ennek megfelelően e dolgozat keretében is két algoritmusról esik szó, egyről, mely teljes adatokra, és egy másik, gradiens alapúról, mely hiányos adatokra alkalmazható.

#### **4.1. A tanításhoz használt hiányos adatokkal kapcsolatos kritériumok**

Mielőtt a tanulás konkrét módszereiről beszélünk, szót kell ejtenünk az azokhoz felhasznált adatokról.

A tanulás lényegében arról szól, hogy a Bayes-hálóval megpróbáljuk reprodukálni a megfigyelt adatok eloszlását. Ahhoz azonban, hogy ezt megtehesük, a felhasznált mintának is megfelelőnek kell lennie. Teljes megfigyelések esetén természetesen ez a probléma nem merülhet fel, de ha az adatok hiányosak, tudnunk kell, hogy az adatvesztés milyen módon történt.

Annak a feltétele, hogy a hiányok ne befolyásolják a tanulás sikerességét, a véletlen adatvesztés (missing at random) feltételének kell teljesülnie (Gelman95). A feltétel formálisan a következő egyenlőséggel írható le:

$$P(I | x, y) = P(I | x, y_{obs}) \quad [4.4]$$

ahol:

- 'x' jelképezi a teljesen megfigyelt adatokat, azaz azokat, amelyek egy sorban sem hiányoznak
- 'y' azokat, amelyek helyenként hiányoznak, ezen belül
  - 'y<sub>obs</sub>' a megmaradtakat
  - 'y<sub>mis</sub>' a hiányzóakat
- 'I' pedig egy indikátorváltozó, ami azt jelzi, hogy történt-e adatvesztés

Vagyis a képlet szerint az adatvesztés nem függhet 'y<sub>mis</sub>'-től, a hiányzó adatok halmazától. Ez közelítőleg azt jelenti, hogy az adatok hiánya nem lehet informatív, nem hordozhat információt az eltűnt adatokra.

#### 4.2. A „maximum likelihood” módszer

Paramétertaniás esetében, ha hiánytalan minták állnak rendelkezésre, a tanulás nem több mint a valószínűségi táblák beállítása a mintából származó gyakoriság értékekre (Friedman96). Teljes adatok esetén tehát pusztán annyi a teendő, hogy végigolvasva az adatokat, számon tartjuk, az egyes gyermek-szülői konfigurációk hányszor fordulnak elő, hisz ezeknek a gyermekcsomópont táblájának egy-egy értéke felel meg. Ennek megfelelően az eljárás pszeudokódja a következő lehet:

```

for(minden mintára){
    if(a sor teljes){
        for(minden csomópontra){
            ++(a konfiguráció paramétere);
        }
    }
}
for(minden táblára) normalizál;
[4.1] táblázat - paramétertaniás teljes adatokból

```

Az algoritmus nyilvánvaló előnye, hogy gyors: az adatok egyszeri végigolvasása is egzakt eredményt szolgáltat. Hátránya viszont, hogy semmilyen módon nem kezeli a hiányos adatokat.

#### 4.3. A gradiens alapú módszer

A keresés célja, hogy a 'P(B(·, Θ) | D)' mennyiséget maximalizáljuk ('B' a háló, 'D' az adat). Ez a (Russel00) szerint haladva, a Bayes-tétel miatt a következőképpen alakul:

$$P(B|D) = \frac{P(D|B)P(B)}{P(D)} \quad [4.5]$$

A gradiens-módszer alap elképzelése szerint tehát a következőt kell tennünk: ki kell számítanunk ' $P(D|B(\Theta))$ ' gradiensét a ' $\Theta$ ' paraméter szerint, ehhez pedig az ezt alkotó súlyok (feltételes valószínűségi paraméterek) szerinti parciális deriváltakat. Ezek számítása helyett azonban jóval könnyebben kiértékelhető kifejezést kapunk, ha ' $P(D)$ ' helyett ' $\log P(D)$ '-t deriváljuk, amikor a képletek a következőképpen alakulnak:

$$\frac{\partial \ln P(D)}{\partial w_i} = \frac{\partial \ln \prod_j P(D_j)}{\partial w_i} = \sum_j \frac{\partial \ln P(D_j)}{\partial w_i} = \sum_j \frac{\partial P(D_j)/\partial w_i}{P(D_j)} \quad [4.6]$$

vagyis az egyes adatsorok hatása külön-külön számítható. A ' $\Sigma$ '-n belüli kifejezésre kell tehát végrehajtani a ' $w_i$ ' szerinti deriválást (ahol  $w_i = P(X = x_i | U = u_i)$ )

$$\frac{\partial P(D_j)/\partial w_i}{P(D_j)} = \frac{\frac{\partial}{\partial w_i} (\sum_{x,u} P(D_j | x,u) P(x,u))}{P(D_j)} = \frac{\frac{\partial}{\partial w_i} (\sum_{x,u} P(D_j | x,u) P(x|u) P(u))}{P(D_j)} \quad [4.7]$$

Kihhasználva, hogy ' $w_i$ ' valójában az összegzésnek csak az egyik tagjában szerepel, a ' $\Sigma$ ' elhagyható:

$$\frac{\partial P(D_j)/\partial w_i}{P(D_j)} = \frac{P(D_j | x_i, u_i) P(u_i)}{P(D_j)} \quad [4.8]$$

amit tovább alakítva, kihasználva, hogy ' $P(x|u)=w_i$ ', valamint a Bayes-tételt felhasználva:

$$\frac{\partial P(D_j)/\partial w_i}{P(D_j)} = \frac{P(x_i, u_i | D_j) P(D_j) P(u_i)}{P(x_i, u_i) P(D_j)} = \frac{P(x_i, u_i | D_j)}{P(x_i | u_i)} = \frac{P(x_i, u_i | D_j)}{w_i} \quad [4.9]$$

Látható, hogy a gradiens tagjainak kiszámításához csupán két értéket kellett felhasználnunk, az egyes szülő-gyermeki konfigurációk valószínűségét, a megfigyelést feltéve, valamint az e konfigurációkhoz tartozó paraméterértéket (ami a gyermek valószínűsége, a szülőket feltéve). Ezek szerint tehát a paraméterek módosítását akár a modell működése közben valós időben (online) is elvégezhetjük, azokon minden egyes új megfigyelés alkalmával finomíthatunk.

Ezzel szemben persze arra is lehetőség van, hogy mint esetünkben, több megfigyelés hatását egyszerre vegyük figyelembe. Ekkor választhatunk, hogy az egyes adatsorok módosító hatását egyszerre vagy külön-külön érvényesítsük, azaz taníthatunk batch és nem-batch módon is. Az eljárások között implementációjukat nézve nincs sok eltérés, például a batch tanulás pszeudokódja:

```

for('epoch' sokszor){
    for(minden sorra){
        for(minden csomópontra){
            kiszámít gradiens; /* P(x,u|D)/w */
            tárol változtatás;
        }
    }
    for(minden táblára){
        hozzáad módosítás;
        normalizál tábla;
    }
}

```

#### [4.2] táblázat - gradiens alapú paramétertárolás

A nem-batch módszerben a módosításokat  $(P(x, u | D) / w)$  egyenként érvényesítjük, azok összegzése helyett. Ez lassabb működést okoz a nagyobb számításigény miatt, viszont összességében kezdetben gyorsabb konvergenciát okoz (mivel minden adatot amint lehet, figyelembe veszünk). Hátránya azonban, hogy az optimum közelében lehet, hogy nem is tart egy pontos értékhez, hisz a gradiens képletéből a ' $\Sigma$ '-t elhagyni olyan mintha zajt adnánk a rendszerhez, ami miatt a konvergencia csak egy tartományon belülre tart majd.

#### 4.4. Kanonikus CPT-k tanulása

Kanonikus eloszlások esetén a gradiens alapú tanulási eljárás szintén alkalmazható, csupán figyelembe kell vennünk, hogy egy adott paraméter mely táblabejegyzésekben szerepel, és a láncszabályt alkalmazva megkaphatjuk az eloszlás-paraméterek szerinti gradienst.

Nézzük példaként a használt 'noisy-or' eloszláshoz tartozó gradiensvektor egy elemét. Tehát a levezetés hasonlóan indul, mint az előző esetben:

(jelölések:  $D$  – adat,  $B$  – a háló egésze,  $w$  – egy adott szülőhöz tartozó zajparaméter (a [2.2] és [2.3] képletbeli ' $p_i$ '),  $x$  – gyermek -,  $u$  – szülő csomópontok)

$$\frac{\partial \ln P(D|B)}{\partial w} = \frac{\partial \ln \prod_D P(D|B)}{\partial w} = \sum_D \frac{\partial \ln P(D|B)}{\partial w} = \sum_D \frac{\partial P(D|B) / \partial w}{P(D|B)} \quad [4.10]$$

a ' $\Sigma$ '-n belüli kifejezést továbbvezetve (az adott csomópont szerinti gyermek-szülők konfigurációk szerinti összegzéssel):

$$\frac{\partial P(D|B) / \partial w}{P(D|B)} = \frac{1}{P(D|B)} * \frac{\partial}{\partial w} * \sum_{x,u} P(D|x,u,B)P(x,u|B) = \quad [4.11]$$

$$= \frac{1}{P(D|B)} * \frac{\partial}{\partial w} * \sum_{x,u} P(D|x,u,B)P(x|u,B)P(u|B) \quad [4.12]$$

Az első különbség itt adódik, a zajparaméter ugyanis a 'Σ' több elemében is szerepel, annak 'P(x|u,B)' tagjában, ami miatt tehát egy tanulási epoch alatt a paraméterekhez jóval több módosító érték adódik, ami intuitíve sugallja a tanulás gyorsabb voltát.

Kihasználva, hogy 'w' a 'Σ'-nak csak a 'P(x|u,B)' tagjában szerepel, és ennek a 'w' szerinti deriváltját 'Δ'-val jelölve:

$$= \frac{1}{P(D|B)} * \sum_{x,u} P(D|x,u,B)P(u|B) * \Delta = \sum_{x,u} \frac{1}{P(D|B)} P(D|x,u,B)P(u|B) * \Delta = \quad [4.13]$$

amit a Bayes-tétel segítségével tovább alakítva:

$$= \sum_{x,u} \frac{1}{P(D|B)} * \frac{P(x,u|D,B)P(D|B)}{P(x,u|B)} * P(u|B) * \Delta = \sum_{x,u} \frac{P(x,u|D,B)P(u|B)}{P(x,u|B)} * \Delta \quad [4.14]$$

Vagyis a végleges képlet, az [4.10] 'Σ'-ját is visszaírva:

$$\frac{\partial \ln P(D|B)}{\partial w} = \sum_D \sum_{x,u} \frac{P(x,u|D,B)}{P(x,u|B)} * \Delta(x,u) \quad [4.15]$$

ahol

$$\Delta = \Delta(x,u) = \frac{\partial P(x|u,B)}{\partial w} \quad [4.16]$$

így annak értéke 'w<sub>i</sub>' szerinti deriváláskor:

gyermek (x) értéke	szülő (u <sub>i</sub> ) értéke	Δ értéke
0	0	0
1	0	0
0	1	$\prod_{j:j \neq i, u_j=1} w_j$
1	1	$-\prod_{j:j \neq i, u_j=1} w_j$

[4.3] táblázat -  $\frac{\partial P(x|u,B)}{\partial w}$  értéke 'x' és 'u<sub>i</sub>' függvényében

A tábla alapú függőségi modellhez tartozó képletnek [4.9] és a kanonikus lokális függőségi modell képletének [4.15] az összehasonlításából látható, hogy a gradiens módszer esetén a tömörebb reprezentáció nagyobb számítási komplexitást igényel.

## 5. Bayes-hálók struktúratanulása

### 5.1. A struktúratanulás alapjai

Struktúratanulás során az alapfeladat ugyanaz, ám itt a keresést a lehetséges struktúrák fölé is kiterjesztjük. Mivel tetszőleges két csúcstól futhat el az egyik vagy másik irányba, illetve hiányozhat az él, még azzal a megkötéssel is, hogy a gráfnak irányítottkör-mentesnek kell lennie, a lehetséges szerkezetek száma rendkívül nagy. Azt is hozzávéve ehhez, hogy az adatok hiányos volta miatt csak a lassabb gradiens-módszer alkalmazható paramétertanulásra, látható, hogy az összes lehetséges struktúra nem értékelhető ki, hisz már a teljes hálók száma is  $n!$  (a lehetséges topológikus rendezések száma), amihez még a nem teljes hálók is hozzáadódnak.

A fenti probléma orvoslására több módszer is kínálkozik, az általam vizsgált orvosi probléma esetében ez a következő volt: tekintve, hogy korábbi vizsgálatok eredményeképpen az orvosszakértő segítségével már kialakítottak egy sorrendezést a változók felett (amely reprezentálhatja az azok közötti ok-okozati összefüggést vagy esetleg valamilyen időbeli sorrendet megjelenésükben), az egyes csomópontok lehetséges szülőinek köre korlátozott (hisz szülő csak korábbi változó lehet, illetve szülő nem lehet a gyermek okozata).

Tehát ezeket a korlátozásokat némiképp kibővítve, a szakértői tudást is felhasználva elérhető volt egy változósorrend, ami a majdani háló topológikus rendezését adhatja (így minden csomópont lehetséges szülőinek köre az őt megelőzőkre korlátozódik) (Antal04).

Az alábbi táblázat tartalmazza ezt a sorrendet. A legfontosabb változókról az „Alkalmazási terület” című fejezetben található leírás.

FamHist	1
Meno	2
Age	3
PillUse	4
Parity	5
HormTherapy	6
Pathology	7
Papillation	8
Locularity	9

Echogenicity	10
TAMX	11
ColScore	12
Volume	13
Ascites	14
Bilateral	15
CA125	16

[5.1] táblázat - a használt változók kauzális sorrendje

A sorrendezés adta egyszerűsítéssel azonban még mindig nem oldottuk meg a problémát, hisz még a maradék struktúrák mennyisége is kezelhetetlen. A következő megközelítés tehát a szülői halmazok méretének maximalizálása, amivel az előzőleg

$$\prod_{i=1}^n 2^{i-1} \quad [5.1]$$

(ahol 'i' a csomópontok sorszáma: [0,n]) méretűre csökkentett mennyiséget, tovább apaszthatjuk:

$$\prod_{i=1}^n \binom{i}{k} \quad [5.2]$$

ahol 'k' a maximális szülőszám

Ezt a lépést egyébként még az az elv is indokolja, amellyel majd a kritériumfüggvény megválasztásánál is találkozunk, nevezetesen, hogy általános esetben lehetőség szerint kicsi, egyszerű hálót próbáljunk kialakítani.

Ezek után tehát lássuk a struktúranulás pszeudokódját:

```

for(minden kiértékelendő struktúrára){
    tanít paramétert;
    kiszámít kritérium;
    if( kritérium jobb, mint az eddigiek ){
        eltárol struktúra;
    }
}

```

[5.2] táblázat - általános struktúranulás

Mint az látszik, a tanulás elve meglehetősen egyszerű: minden megvizsgált struktúránál először optimalizáljuk a paramétereket, majd kiértékeljük a kritériumfüggvényt, végül a legmagasabb értéket elért hálót tartjuk meg.

A tanulás formálisan a ' $P(B(G,\cdot)|D)$ ' maximalizálását jelenti, vagyis keressük azt a struktúrát, amely valószínűsége az adatokat feltéve maximális, ez pedig jól közelíthető az optimális paraméterezésű struktúra valószínűségével:  $P(B(G, \hat{\Theta})|D)$ .

## 5.2. Az elvek alkalmazása teljes adatok esetén

Ha a megfigyelési adatok teljesek, már a paraméterek illesztésénél is láthattuk, hogy a tanulás jóval gyorsabb és egyszerűbb algoritmusokkal is végrehajtható. Tekintve, hogy a struktúratanulás esetén számos hálóra kell a tanítást elvégezni, a gyorsaság kritikus lehet.

Az előbbi mellett egy további körülmény is segít, ha adataink hiánytalanok: ekkor a tanítást csomópontonként elvégezhetjük. Ez a szeparálhatóság (a csomópontonkénti) megmarad a struktúratanulásnál is, vagyis a már fentebb ismertetett algoritmust a következőképpen alakíthatjuk át:

```
for( minden csomópontra ){
    for( minden lehetséges szülői halmazra /*  $\Pi$  */ ){
        kiértékel  $\Pi$ ;
        if(  $\Pi$  jobb mint az eddigiek ) eltárol  $\Pi$ ;
    }
}
```

[5.3] táblázat - csomópontonkénti struktúratanulás

A fontos újítás tehát az, hogy itt a hálót csomópontonként építjük össze, vagyis végigmenve a kitűzött változósorrenden, a csomópontokat egyesével kötjük be a háló már készen lévő, felső részébe, megkeresve az optimális szülői halmazt, vagyis a maximalizálni kívánt valószínűség a következőképp alakul:

$$P(B(G,\cdot)|D) = \prod_i P(\Pi_i | X_i, D) \quad [5.3]$$

ahol ' $\Pi_i$ ' az ' $X_i$ ' változó szülői halmazát jelenti, nem annak egy adott behelyettesítését. Nagyon fontos, hogy ekkor az új változók felvétele nincs hatással a már készen lévő háló jóságára, aminek eredményeként az egy sorrend feletti összes struktúra végigkeresésének költsége a következő lesz:

$$\sum_{i=0}^n \sum_{j=0}^i \binom{i}{j} \quad [5.4]$$

ahol a csomópontok sorszámai '0'-tól 'n'-ig terjednek.



Azonban még ilyen költségek mellett sem valósítható meg, hogy nem csak adott topológikus rendezéseket vizsgálunk, hanem ezek teljes terét végigkeressük, vagyis hiánytalan adatok mellett sem lehetséges kimerítő keresést alkalmazni a struktúrák vizsgálatához.

### **5.3. Heurisztikák hiányos adatok kezelésére**

Ha az adatok hiányosak, az előbb leírt dekompozíció nem hajtható végre, azaz a csomópontok egy részhalmaza feletti optimális struktúrát egy új változóval kiegészítve, a háló már nem feltétlenül lesz optimális.

Röviden ez azt jelenti, hogy minden egyes megvizsgált struktúrát újra kell tanítani, és kiértékelni, a korábbi számítások eredményeit nem tudjuk újra felhasználni. Ez, és a paramétertanulás lassúsága együttesen már lehetetlenné teszi a teljes struktúráter végigkeresését. Emiatt vissza kell térnünk az egy sorrend feletti kereséshez, és le kell mondanunk az algoritmus egzakt voltáról, heurisztikus módszereket kell alkalmaznunk. Ilyen heurisztikákból több is rendelkezésre áll, néhány lehetőség leírása a következő bekezdésekben található.

#### **5.3.1. Csomópontonkénti építés**

Bár a teljes adatokra kidolgozott módszert itt nem „jogos” alkalmazni, hisz az egyes csomópontokhoz tartozó optimális szülői halmazok függhetnek a sorrendben hátrébb lévő változóktól, közelítésként mégis alkalmazható.

Tehát az eljárás hasonlít a teljes adatok esetére kidolgozotthoz, azzal az eltéréssel, hogy itt a paraméterillesztést valamilyen bonyolultabb eljárással, például (mint esetünkben is) a gradiens alapúval kell végezni.

#### **5.3.2. Mohó eljárás**

Az ismert elv szerint a keresési térben lépéseket teszünk, igyekezve mindig a lehető legtöbb nyereséggel járó kiválasztani.

Az itt tehető lépések tehát új éleknek a hálózathoz vételét jelentik. Minden egyes ciklusban sorra vesszük a hálózathoz adható éleket, és megvizsgáljuk a kapott struktúrát. A legtöbb nyereséggel járó élt megtartjuk, és az új hálóból kiindulva folytatjuk a keresést.

Az eljárás leállítására több kritériumot is figyelembe vehetünk:

- az eddig megtett lépésszámot
- a kapott háló komplexitását – ez némiképp hasonlít az előzőhöz, de valamivel intelligensebb annál (például kiköthetjük, hogy a háló összefüggő legyen)

- elért kritérium-értéket
- következőként megteendő lépés nyereségét (ha az már túl kicsi, vagy már rontana a mutatón)

### 5.3.3. Kimerítő keresés

A fenti módszerek mellett természetesen lehet alkalmazni a nyers erőre építő kimerítő keresést is. Az eredmény biztos, ám ezzel a [5.2] képlet szerinti Ordo(exponenciális) számítás jár együtt. A dolgozatban elsősorban ellenőrző szereppel használjuk ezt a módszert, annak vizsgálatára, hogy egy adott heurisztika alkalmazása milyen gyengítést jelent a tanulás eredményességére nézve

### 5.4. A priori tudás beépítése a hálóba

Mint azt már a [2.4] fejezetben láthattuk, a rendelkezésünkre álló szakértői tudást felhasználva, jó első közelítést kaphatunk az optimális hálóra. A fentiekben olyan eljárásokat tekintettünk át, amelyek vagy eleve üres hálóból indultak, vagy nem igazán vették figyelembe, hogy milyen volt a kiindulási háló.

Ha azonban tudjuk, hogy olyan hálóból indulunk ki, amelyet szakértői tudás alapján konstruáltunk, vagyis a háló vélhetően eleve jó közelítést ad, több dolgot is tehetünk, hogy az általunk kialakított háló az eredetihez közel legyen (a struktúráját tekintve):

- Csak azokat a hálókat vesszük sorra, amelyek csak legfeljebb bizonyos számú élben térnek el az eredetitől.
- A keresés során használt kritériumfüggvényben büntetjük az eltérést a szakértői hálótól, például valamilyen ' $\Delta^n$ ' (' $\Delta$ ' 1-nél kisebb pozitív konstans, ' $n$ ' az eltérések száma) alakú szorzótényező alkalmazásával a ' $P(\text{adat|háló})$ ' tagra.

Az értékelésnél természetesen elhagyjuk ezt a büntetőtagot.

Ezeket a módszereket kívül magát a keresést is implementálhatjuk úgy, hogy az csak bizonyos hálókat vegyen sorra.

### 5.5. A használt kritériumfüggvény

Az alap feladat tehát a ' $P(D|B)$ ' valószínűség maximalizálását jelenti, azonban e mennyiség kritériumként való használata nem feltétlenül vezet a várt eredményre. Információelméleti megfontolásokból látható, hogy olyan szülői halmazt keresünk, melynek a gyermekkel vett kölcsönös információja maximális (Sarkar96). Mivel a változók felvételével az információtartalom csak monoton nőhet, ilyen kritérium mellett

a teljes háló optimális. Ez a feltétel tehát a nagy hálókat preferálja, ami a következők miatt nemkívánatos:

- A bonyolult háló magas mutatói vélhetően nem a jelenségkör jó modellezését jelentik, hanem sokkal inkább az adatokét. Ez a más tanuló rendszereknél is ismert túlilleszkedés jelensége.
- A nagy hálókat megérteni is nehezebb, vagyis a túlságosan összetett struktúra már nem segíti annyira a vizsgált terület megértését, a döntések indoklása sem lesz olyan világos.

A fentiek miatt tehát a kritériumfüggvényt módosítani kell: ki kell egészíteni valamilyen büntetőtaggal, ami az egyszerűbb szerkezetű hálókat részesíti előnyben. E kívánalmaknak tesz eleget a (Friedman97)-ben leírt, Schwarz által bevezetett BIC (Bayesian information criterion) függvény, amely a használt tanítási minta logaritmusával, és a háló 'méretével' egyenesen arányos büntetést alkalmaz. Továbbá a valószínűség helyett annak logaritmusát használja, vagyis a kritériumfüggvény a következő:

$$BIC(B, D) = \sum_{i=1}^N \log P(D_i | B) - \frac{\log N}{2} |\Theta| \quad [5.5]$$

Amiben az első tag a minta valószínűségének logaritmus, a hálót feltéve, a második tag pedig a büntetés, ahol 'N' a tanítóhalmaz sorainak a száma, '|\Theta|' pedig a hálóparamétereké, azaz a csomópontokban található valószínűségi táblák méretének összege.

### 5.6. Kritériumfüggvény az irodalmi modellhez

Az előző fejezet általános elvei azonban nem emelhetők át minden változtatás nélkül az irodalmi modellbe. A jószágfüggvény első tagja, mely a logaritmus valószínűséget adja, itt is alkalmazható, a büntetőfaktor viszont már nem. Egyrészt itt az elsődleges cél nem feltétlenül a teljes valószínűségi eloszlás minél pontosabb lemásolása, másrészt az '|\Theta|' paraméterszám sem az összes CPT bejegyzés számát jelenti, mivel az alkalmazott kanonikus eloszlások miatt itt a táblázatértékek nem függetlenek egymástól (a szakirodalmi hálókkal részletesen a [6] fejezet foglalkozik).

$$BIC(B, D) = \sum_{i=1}^N \log P(D_i | B) - \frac{\log N}{2} |E| \quad [5.6]$$

A képletben 'E' jelöli a gráf éleinek számát.

Ekkor a valós paraméterszám drasztikusan csökken és a büntető tag viszonylagos kicsinysége miatt szerepe elhanyagolható, így az irodalmi modell tanulásánál a következő jósági mutatót használtam:

$$BIC(B, D) = \sum_{i=1}^N \log P(D_i | B) \quad [5.7]$$

## 6. Hálóépítés szakirodalom alapján

Általában bármilyen kutatási vagy modellezni kívánt területen központi kérdés az egyes jelenségek, megfigyelések közti ok-okozati, azaz kauzalitási kapcsolatok felderítése. A Bayes-hálós modellek esetében az olyan modelleknél, amelyek pusztán a változók együttes eloszlását ábrázolják pontosan, az olyanok értékesebbek, amelyek tükrözik a változók közti kauzalitást is. Az ilyen hálók sokkal hasznosabbak, hisz nem csak adott konfigurációk valószínűségének becslésére képesek, hanem például egymástól távol (több élnyire) lévő változók közti összefüggések felderítésére is. Az ilyen hálók sokkal alkalmasabbak az általuk hozott döntések indoklására is, ami a döntéstámogató rendszerek egyik elengedhetetlen funkciója.

A Bayes-hálók tanításának elsődleges célja és módja, hogy a háló paramétereit (a CPT-k bejegyzéseit és magát a struktúrát) úgy módosítsuk, hogy a háló által leírt valószínűségi eloszlás a valós adatokéhoz közelítsen. A valós adatokkal kapcsolatban azonban gyakran felmerül a hiányosság problémája, ami mind a paraméter- mind a struktúranulást jelentősen lelassítja, az algoritmusok időigényét szuperexponenciálissá téve.

E probléma orvoslására több lehetőség is van, például a struktúranulás során a már fentebb említett heurisztikák alkalmazása. Egy másik lehetőség, ha a modellezni kívánt területről rendelkezésre álló információt valamiképpen belefoglaljuk a hálóba. A legegyszerűbb lehetőség, ha egy szakértőt felkérünk, hogy írja le a terület fontos összefüggéseit, majd ez alapján konstruálunk egy hálót. Az így kapott struktúra aztán lehet végleges, vagy jelentheti további struktúranulás kiindulópontját.

A szakértői tudás feldolgozásának egy másik módja, ha a rendelkezésre álló cikkekben tárolt tudást használjuk fel a struktúra megalkotásánál, lehetőleg automatizált módon. Ahhoz azonban, hogy a szakirodalmi cikkeket felhasználhassuk, bizonyos feltételezéseket el kell fogadnunk.

## 6.1. A kauzalitás megjelenése a szakirodalmi cikkekben

Ezen egész dolgozat fő célja olyan módszerek vizsgálata, amelyek lehetővé teszik, hogy a tárgyterület fogalmi közti valós összefüggésekre az adott területtel foglalkozó cikkek feldolgozása alapján következtethessünk. Az alapötlet szerint tehát az egyes valószínűségi változóknak a cikkekben való együttes előfordulási gyakorisága alapján következtethetünk a köztük lévő valódi összefüggés gyakoriságára. Röviden szólva azok a változók fognak együtt megjelenni, amelyek közt a valóságban is szoros a kapcsolat. Ebből az igencsak erős feltételezésből juthatunk a lentebb kifejtett módszerekre, melyek tehát általánosan azon alapulnak, hogy egy a kulcsszavaknak az irodalmi cikkekben való megjelenését jól leíró modellstruktúra a valós világ kapcsolatait is helyesen adja vissza.

Bár az egyes írók publikációs szokásai, és az azok mögött meghúzódó kognitív mechanizmusok bonyolult és kevésbé ismert területet alkotnak, bizonyos egyszerű és elfogadható meglátásokból kiindulva a feltevés indokolható.

Általánosan elmondható, hogy minden szakirodalmi cikk, az adott tárgyterület néhány, nem túl sok fogalmáról szól, melyek szorosabban vagy lazábban kapcsolódnak egymáshoz. Mivel az emberi olvasók számára egy területen belül főleg az oksági kapcsolatok az érdekesek, és a legtöbb kutatás számára is a kulcsváltozók közti mechanizmusok, belső összefüggések feltárása a cél, ezért várhatóan a cikkek többsége is ilyen kapcsolatokról szól majd.

Másik, a publikálóval szemben támasztott elvárás, hogy a cikkében bevezetett fogalmakat az olvasók számára érthetővé tegye. Ezt úgy mondhatjuk, hogy egy-egy, a cikkben központi fogalmat (mely nem feltétlenül kell, hogy új legyen), más az azonos területről jól ismert változó segítségével definiálni kell. Eszerint tehát egy cikk említette fogalmak alapvetően két szerepet tölthetnek be: lehetnek kulcsfogalmak (magyarázott fogalmak), amelyekről a szerző saját szándéka szerint a cikknek szólnia kell, vagy lehetnek magyarázó fogalmak, melyek a kulcsváltozókat vezetik be.

Ez a két „kényszer” tehát azt sugallja, hogy egy cikkben az író által szándékolt változókon kívül még az azokhoz fogalmilag szorosan kapcsoló, vagy azzal kauzális kapcsolatban álló egyéb változók fognak megjelenni. Természetesen fontos tény, hogy csak a szakterületre releváns és a modellépítésben felhasznált fogalmakat kap meg tanítási adatként a tanulási algoritmus, azaz még a tanulási fázis előtt kiszűrődnek a modellbe nem illő fogalmak.

## **6.2. Struktúratanulás szakirodalom alapján**

Mivel tehát feltevésünk szerint a valós változók közti kauzalitás a szakirodalmi cikkekben való előfordulások közti korrelációkban is megőrződik, így a cikkekből nyert előfordulásokat felhasználhatjuk a modell tanulásánál előzetes struktúrák készítéséhez.

Az alapötletet felhasználva azt mondhatjuk, hogyha alkotunk egy úgynevezett irodalmi modellt, amely tehát azt írja le, hogy az egyes változók említése egy cikkben mely mások megjelenését vonja maga után, akkor a valós változókat ugyanilyen struktúrába rendezve, egy jó hálót kapunk. Ezt a hálóstruktúrát ezután meg is tarthatjuk, a továbbiakban már csak a paramétereket finomítva, vagy felhasználhatjuk további struktúratanulás kiindulásaként.

A módszer nyilvánvaló előnye, hogy a cikkek kivonatolása történhet teljesen automatikusan és egy ilyen eljárás költségei elenyészők. Emellett az egyszerűbb megközelítésekben (mint azt lejjebb látni fogjuk) a tanításhoz használt adatok teljes volta garantálható, ami az algoritmusok gyorsaságán jelentősen javíthat. További gyorsító tényező még, hogy az ilyen irodalmi modellek változói tipikusan kétértékűek, ellentétben a valósakkal, melyek ennél jóval több értéket is felvehetnek, ez pedig jelentősen csökkentheti a tanulandó paraméterek számát.

A fentiek fényében elmondható tehát, hogy az efféle módszerek megoldást kínálhatnak minden olyan területen, amelyeken az adatgyűjtés csak nagy költségek árán, esetleg egyáltalán nem kivitelezhető, de a területről viszonylag gazdag szakirodalom hozzáférhető.

### **6.2.1. Közvetlen modell**

Az alapelv egy közvetlen megvalósítását kapjuk, ha az irodalmi és a valós változók között egy az egyhez megfeleltetést létesítünk. Egy ilyen modell intuitíve azt írja le, hogy egy adott változó megjelenése milyen valószínűséggel okozza a változó gyermekeinek (a tőle közvetlenül függő változóknak) megjelenését. Ez a modell így azt a feltevést sugallja, hogy egy cikk írása közben a szerző igyekszik végigkövetni a teljes kauzalitási láncot, vagyis ekkor minden modellváltozó két szerepet lát el egyszerre, mindegyikük 'magyarázott' és 'magyarázó' is egyidejűleg.

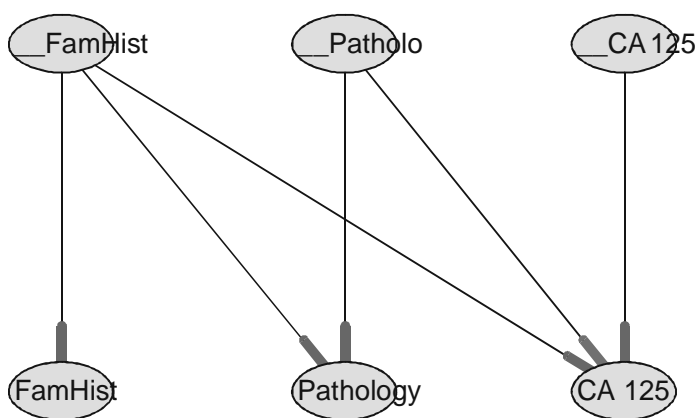
Ezzel a módszerrel a tanított háló struktúrája közvetlenül örökíthető a valós modellbe, vagyis a kapott szerkezet azonnal összevethető a referenciamoddellel.

### 6.2.2. Kétszintű modell

A fenti modell hiányossága, hogy mivel nem tesz különbséget változók között azok szerepei szerint (azaz aszerint, hogy azok a szerző szándéka miatt, vagy a 'magyarázási kényszer miatt' kerültek a cikkbe). Így a modell logikája szerint, ha egy cikkben megjelenik egy változó, akkor annak minden leszármazottja is meg kell, hogy jelenjen a modell szerinti valószínűséggel, függetlenül attól, hogy a szülője miért került be.

A fenti ellentmondást egy összetettebb irodalmimodell-struktúra oldhatja fel. Ekkor minden valós változónak két irodalmi változó felel meg:

- egy, mely a változó cikkbeli említését jelenti
- és egy másik, amely a szerzőnek a változó említésére vonatkozó szándékát jelképezi



[6.1] ábra - tipikus irodalmimodell-struktúra

A fenti elrendezésben az alsó sor változói ábrázolják a cikkben megjelenő, a felső soréi pedig a szándékolt változókat. A modell szerint mindkét sorban a változók egymástól függetlenek, élek csak a felső szint felől futnak lefelé. A modellben így a felső változók eloszlásai mutatják, hogy a szerzők maguktól milyen valószínűségekkel írnak az egyes változókról, ami persze 1 valószínűséggel maga után vonja a megfelelő alsó változó megjelenését. Ezeket a kényszereket a megfelelő változók élek jelölik. A további élek mutatják azokat a változókat, amelyek még maguk után vonhatják a változó megjelenését.

Ennek az elrendezésnek az előnye, hogy nem mossa össze az egyes változók szerepeit, így elvileg jobb eredményeket kell szolgáltatnia. Hátránya azonban, hogy kétszerannyi csomópontot használ, amelyek közül a felső sorbeliek nem is megfigyelhetők, így a tanulás már jóval költségesebb kell, hogy legyen.

szülők			gyermek - CA 125	
FamHist	Pathology	CA 125	0	1
0	0	0	1,0	0,0
0	0	1	0,0	1,0
0	1	0	0,5	0,5
0	1	1	0,0	1,0
1	0	0	0,5	0,5
1	0	1	0,0	1,0
1	1	0	0,25	0,75
1	1	1	0,0	1,0

[6.1] táblázat

**irodalmi modell egy csomópontjának lehetséges paraméterezése**

A táblázatban szerepelnek a 'CA 125' változó eloszlásának paraméterei a szülői konfigurációk szerint. Látható, hogy a saját szülő ('\_\_CA 125') megjelenése 1 valószínűséggel implikálja a gyermek '1' értékét. A többi szülő hatása a 'noisy-or' eloszlást követi, mindegyikük zajparamétere '0,5'.

Az ábrából leolvasható még az alkalmazott jelölési konvenció: a saját szülőket gyermekeiktől a '\_' prefixum különbözteti meg.

A kauzális mechanizmusok gyakori önálló tényezőkénti vizsgálata és azok publikálására miatt feltehetjük továbbá, miszerint az egy változó megjelenésének okai egymástól függetlenek (amit a struktúra is sugall), és hatásukat is függetlenül fejtik ki. Ilyen esetben a csomópontok általános valószínűségi táblája helyett alkalmazhatunk egy kanonikus eloszlást is (mint a fenti táblázatban), mely jelentősen kevesebb paraméterével gyorsíthatja a tanulást. Az e dolgozat keretében végzett mérések során is éltünk ezzel a lehetőséggel, a tanított modellekben a [2.3]-ban már leírt 'noisy-or' táblákat használtuk.

Az így nyert struktúra azonban nem használható fel közvetlenül a valós modellben. Ezt a problémát áthidalhatjuk, azzal az egyszerű megoldással, ha az egymás feletti (az ugyanahhoz a fogalomhoz kapcsolódó) csomópontokat összevonjuk, a gráfelméletben is gyakran alkalmazott eljárással

**6.3. A tanítási adatok kinyerése a cikkekből**

Az irodalmi modellek tanításához felhasználható adatoknak az egyes változók (pontosabban a változóhoz tartozó kulcsszavak, szinonimák) cikkekre vonatkozó jellemző voltát legegyszerűbben a cikken belüli gyakoriságuk jellemezheti. Az így



kapott természetes számokat azonban nehéz közvetlenül felhasználni, érdemes ezért ezeket egy véges intervallumba leképezni.

Másik további javítási lehetőség lenne, a szavaknak nem csak tisztán az előfordulási számait figyelembe venni, hanem például az egymáshoz való elhelyezkedésüket, a cikken belüli eloszlásukat, vagy akár a természetes nyelv által meghatározott relációikat. E módszerek kutatása, vagy akár tárgyalása azonban meghaladja e dolgozat kereteit, így itt csupán a felhasznált adatok létrehozásának egy leírását adjuk, a számítások részletes leírása (Antal04)-ben található.

A számítás több lépcsőből áll, ezek a következők:

1. meghatározzuk a kulcsszavak súlyozott gyakoriságát a

$$w_{ij} = f_{ij} \ln \left( \frac{L}{n_j} \right) \quad [6.1]$$

képlet szerint (ahol ' $w_{ij}$ ' a 'j.' szó 'i.' dokumentumra vonatkoztatott súlya, ' $f_{ij}$ ' az előfordulások száma a dokumentumban, ' $L$ ' az összes dokumentum száma, ' $n_j$ ' a 'j.' szót tartalmazó dokumentumok száma.

2. az így kapott számértékek, minden dokumentumra egy vektort adnak, amelyek segítségével a ' $D_x$ ' és ' $D_y$ ' dokumentum hasonlóságát a

$$\text{sim}(D_x, D_y) = \cos(W_x, W_y) \quad [6.2]$$

kifejezés adja, mely tehát a két vektor (' $W_x$ ', ' $W_y$ ') által bezárt szög koszinusza

3. ahhoz, hogy egy kulcsszónak egy dokumentumra vonatkozó fontosságát megkapjuk, a hasonlósági mutatót kiszámoljuk a dokumentumra és egy minden változóhoz külön-külön rendelkezésre álló annotációra, egy rövid, a változó nevét, szinonimáit, stb. tartalmazó leírásra
4. az egyszerűség kedvéért az így kapott [0, 1]-beli számot diszkretizálni is szokás: a változót [0, 0,1)-ben tekintjük '0'-nak, azaz nem jellemzőnek, [0,1, 1]-ben '1'-nek, azaz jellemzőnek

## 7. A struktúranulás minősítése

Természetes elvárás, hogy a tanított hálók által megjelenített tudást minősíteni tudjuk, azaz a háló tanulásának sikerességéről fogalmat alkothassunk. Ez részben már a háló tanítása során is megtörténik, hisz a struktúranulás során az elért mutatóérték alapján

választjuk ki nyertes hálóstruktúrát, emellett azonban más finomabb módszerek iránt is merülhet fel igény.

A fenti „módszer” egyik velejárója, hogy nem pusztán a hálók struktúráját, hanem a tanult paraméterek jóságát is összeveti, amely jobban függhet maguktól a tanuláshoz felhasznált adatoktól. Másrészt ez megnehezíti a hálónak a szakértői tudással való összemérését, hisz míg egy szakértő tudását, a terület összefüggéseiről alkotott elképzeléseit viszonylag egyszerűen lehet kvalitatív módon egy hálóstruktúrával ábrázolni, addig ahhoz olyan számszerű paramétereket, amelyeket egy Bayes-háló igényel, hozzárendelni már jóval nehezebb.

### **7.1. A változók között fennálló függési viszonyok**

Mivel a Bayes-hálók struktúrája jeleníti meg az ábrázolt valószínűségi változók közti közvetlen függéseket, így a struktúrák minősítésénél a csúcspontok közti topológikus viszonyok azonossága, különbsége lehet mérvadó.

Két valószínűségi változó egymáshoz való viszonyát statisztikai alapon jól leírja korreláltságuk. A Bayes-hálókbán azonban fontos, hogy a struktúra ne csupán a korreláltságot, hanem az ok-okozati összefüggéseket is megjelenítse. A problémát itt az okozhatja, hogy két változó korreláltságát nem pusztán a köztük lévő közvetlen, esetleg egy láncolaton keresztül ható ok-okozati összefüggés okozhatja, hanem egy harmadik változó is mely mindkettő szülői vagy ősei között megtalálható. Ez főleg a kutatási terület intuitív megértését gátolhatja, hisz a szakértők számára itt elsősorban az oksági mechanizmusok felderítése az érdekes.

A fentiek figyelembevételével tehát minden változópár a köztük lévő kapcsolat szerint a következő osztályok egyikébe sorolható:

- közvetlen függés (kauzális él): az egyik csomópont a másik szülője
- oksági láncolat (kauzális út): az egyik csomópont a másik őse, azaz irányított út vezet köztük
- zavaró tényező/megzavart: a két csomópontnak létezik egy közös őse
- a két előző pont együtt: a csomópontok között vezet irányított út, és létezik egy közös ős, amelyből diszjunkt utak vezetnek a csomópontokhoz
- nincs függés: a két csomópont két külön komponensben van

## 7.2. Két hálóstruktúra összevetése

Ha két háló minden csomópontpárjára megvizsgáljuk, hogy az mely kategóriába esik az egyik, illetve a másik hálóban, a különbözőség egy jó fokmérőjét kapjuk, megtekintve, hogy az egyik háló adott típusú kapcsolataiból hány őrződik meg a másikban, illetve milyenek helyettesítik azokat.

Ezek a számok egy mátrixban ábrázolhatók, melynek tehát (i, j)-edik eleme azt mutatja, hogy az egyik háló i. típusú kapcsolatai közül hányból lesz a másik hálóban j. típusú kapcsolat. Intuitíve is érezhető, hogy két háló akkor hasonlít egymásra, ha ebben a mátrixukban a főátlóbeli elemek képviselik a nagyobb súlyokat.

A mátrixos formánál könnyebben értelmezhető, bár kevésbé részletes mutatót kapunk, ha a mátrix elemeit megfelelő súlyozással összeadjuk:

$$DIFF(B_1, B_2) = \sum_{i,j} w_{i,j} * M_{i,j} \quad [7.1]$$

ahol 'B<sub>1</sub>', 'B<sub>2</sub>' a két hálóstruktúra, 'w<sub>i,j</sub>' az 'M<sub>i,j</sub>' mátrixelemhez rendelt súly.

A mérések során alkalmazott súlyozás a következő volt:

különbség a referenciaháléhoz képest	súly
független - megzavart	1,0
független - kauzális út	1,0
független - kauzális él	1,0
megzavart - kauzális út	0,0
megzavart - kauzális él	0,0
kauzális él - kauzális út	0,0

[7.1] táblázat - struktúrabeli különbségek büntető súlyai

A változók fix sorrendezése miatt az egyes élek (illetve utak), nem jelenhetnek meg az egyik hálóban a másikhoz képest fordítva.

A módszer egyik nagy előnye, hogy pusztán a struktúrákat veti, össze, így a tanulás eredményességét egy afféle etalonnak tekintett, szakértő által konstruált hálóstruktúráról való eltérés alapján mérhetjük le. Ez azért is hasznos, mert a kutatási terület Bayes-hálós ábrázolásánál nem csak változók eloszlásának a reprodukálása, hanem a területet jellemző fontos oksági kapcsolatoknak a hálóban való megjelenése is lényeges szempont.

## 8. Az alkalmazási terület

Vizsgálataimat a konzulensem által is kutatott területen végeztem, ez pedig az orvosbiológia, azon belül is a petefészekrák preoperatív diagnosztikája (Timmerman00). A téma több okból is figyelemre érdemes, egyrészt az orvoslás területén belül elfoglalt fontos szerepe miatt, másrészt olyan technikai jellegű okok miatt, amelyek a probléma számítógépes feldolgozását segítik elő. Ezek:

- A területről áll rendelkezésre mérési adat, így a tanulás reális alapokon végezhető.
- A probléma jellegéből adódóan, az egyes orvosi mérési, vizsgálati eredmények jól leírhatók egymástól függő valószínűségi változókkal.
- A témáról részletes szakértői tudás áll rendelkezésre, amelyet könnyen lehet szabályok és ok-okozati összefüggések formájában megfogalmazni, így a Bayes-hálók eredményei jól összevethetők a szakértők tudásával.
- A feladat jól skálázható, kisebb modellek is már viszonylag jó közelítéseket adnak, de elég adat áll rendelkezésre akár 50 csomópontos hálók konstruálásához is.
- A területen elérhető adatok, valamint a priori információk nagy mennyisége a feladatot ideális mintaalkalmazássá teszi az „integrált adat és tudás” témában.

Az orvosi probléma tehát röviden összefoglalva a következő: adottak betegek vizsgálati eredményei, ezek alapján kell eldönteni, hogy a betegség jó- vagy rosszindulatú (az adatbázisban nincsenek egészséges páciensek adatai, vagyis a Bayes-hálós vizsgálat már csak a biztosan beteg személyek adataira terjed ki).

### 8.1. A terület orvosi jellemzése

A petefészekrák diagnosztikájának kutatása azért is kiemelkedően fontos, mert a páciensek mintegy kétharmadánál már csak az előrehaladott állapotban sikerül diagnosztizálni a betegséget, ez pedig a kezelési esélyeket nagyban rontja.

A daganatok leggyakrabban a petefészek felszínén alakulnak ki. Azt, hogy ezek rosszindulatú mivolta mire vezethető vissza, több elmélet is próbálja magyarázni, melyek többek között a következő jellemzőket veszik figyelembe:

- Ovulációk száma
- Gonadotropinok szintje
- Karcinogén anyagok

- Genetikai rendellenességek

Ezek mellett még befolyásolhatja a kockázatot a szülések száma, a terméketlenség, hormonális kezelések, hasonló betegség a családban, életkori jellemzők, korábbi méheltávolítás, a tumor kétoldali mivolta.

A fentiekén túl elérhetőek még további orvosi mérések is, például a daganat alaki tulajdonságai, tumormarkerek szintjei (CA 125). Némely faktor hatása kvalitatívan ismert, mások azonban jelentős szubjektivitást tartalmaznak.

Az alábbi táblázat tartalmazza a legfontosabb változók nevét, rövid leírását és lehetséges értékeit:

Age	Életkor.	(. ,40), [40-50), [50-60), [60-70), [70, .)
Ascites	Folyadék a Douglas-üregben.	<3mm, 3mm<=
Bilateral	Bilaterális daganat.	Nominális (2)
CA125	Tumor marker.	Nominális: <35, [35-65), 65<=
ColScore	Véráram erőssége, bősége a Color Doppler ultrahang (CDI) eljárás alapján.	Nominális (4)
Echogenicity	Ciszta tartalma ultrahang alapján	Nominális (5)
HormTherapy	Hormonális kezelés.	Nominális (5)
Locularity	Tumor alakta (1 üregű ciszta, 1 üregű ciszta tömör résszel, sok üregű ciszta, sok üregű ciszta tömör résszel, tömör	Nominális (5)
Meno	Klimax	Nominális (2)
Papillation	Tömör cisztába benyúló 3 mm-nél nagyobb képződmények.	Nominális (2)
Parity	Szülések száma	0,1,2,3,4<=
Pathology	Daganat patológiai besorolása	Nominális (2)
PillUse	Fogamzásgátló tabletták szedésének hossza hónapban.	Numerikus

[8.1] táblázat - a legfontosabb változók leírása

Az informatikai feladat tehát olyan rendszer fejlesztése, amely képes a területet jól modellező háló tanítására, olyan hálóra tehát, amely a vizsgálati eredmények alapján

képes a betegség (mint az fentebb látszik) kétértékű osztályozására. Természetesen semmi sem szab gátat annak, hogy a diagnózist finomabb felosztású változóval (esetleg változókkal) reprezentáljuk így pontosítva azt.

E dolgozat vizsgálatának tárgya azonban még nem egy az orvosi munkát segítő rendszer fejlesztése, hanem csupán az ilyen feladatot ellátó rendszerek informatikai tulajdonságainak és megvalósíthatóságuk kérdésének vizsgálata.

## **8.2. Az IOTA project**

A „International Ovarian Tumor Analysis Consortium” (IOTA) egy olyan testület, mely a petefészekrák operáció előtti diagnosztikáját tanulmányozza. A project keretében létrejött egy adatbázis, mely egyrészt a terület egységes leírását adja, másrészt valós (de anonim) mérési adatokat is szolgáltat a kutatások számára.

A tárgyterületről, illetve az 'IOTA' projectről magáról az (Antal04) és a (Timmerman00) cikkek adnak részletesebb leírást.

## **8.3. Az irodalmi adatok**

Konzulensem még korábban elkészített egy irodalmi adathalmazt, mely a [6.3] fejezet szerinti eljárást követve készült, számos, a vizsgálati területtel kapcsolatos cikk alapján. Vizsgálataim során én is ennek az adatbázisnak egy 500 rekordot tartalmazó részét használtam, mely (néhány cikk eltéréssel) az 1998-2002. között publikált petefészekrákhoz kapcsolódó orvosi szakcikket tartalmazza. További részletek (Antal04)-ben találhatóak.

# **9. Az alkalmazott algoritmusok megvalósítása**

## **9.1. A 'BNet' rendszer**

Konzulensem, Antal Péter, aki e kutatások során is a segítségemre volt, már korábban készített egy, a Bayes-háló kezelésére szolgáló számítógépes keretrendszert. Az általam megvalósított és tesztelt algoritmusokat én is ennek a programnak a segítségével futtattam, felhasználva a benne már eleve meglévő, olyan alapvető eljárásokat, mint a Bayes-háló adatszerkezetének (struktúrájának és paraméterezésének) kezelése, a háló grafikus megjelenítése és a hálóban történő következtetés.

Az általam vizsgált algoritmusokat (paraméter- és struktúranulás) magam valósítottam meg, és ezekkel a rendszert a magam számára kiegészítettem.

## 9.2. A tanulás minősítése

A modelleknek mind a paraméter- mind a struktúratanulása során egységesen a 'BIC' függvényt alkalmaztam, a struktúra bonyolultságát büntető tag figyelembe vétele nélkül. Vagyis a függvény pontos értéke:

$$BIC(B, D) = \sum_{i=1}^N \log P(D_i | B) \quad [9.1]$$

Azaz a sorok háló szerinti valószínűsége természetes alapú logaritmusának összege. Vagyis az eljárás a következő pszeudokódot követi:

```
score = 0;
for( minden adatsorra ){
    for( minden komponensre ){
        beállít( ismert értékek );
        score += log( valószínűség(konfiguráció) );
    }
}
return score;
```

[9.1] táblázat - BIC függvény kiszámítása

Amennyiben a háló több komponensből állt, a következtetést részenként külön végeztem el és az eredményeket összeszorozva adódott a változók adott behelyettesítésének valószínűsége. A valószínűségek ily módon lehetséges szorzatra bonthatóságát az összes következtetés esetében alkalmaztam.

## 9.3. A paramétertanulás

Az irodalmi modell tanulása során, tekintve, hogy a kétszintű modell esetében a változók fele nem megfigyelhető, a már leírt gradiens alapú módszert alkalmaztam, a következő kód szerint:

```
while( score_új/score_régi<threshold
|| score_új/score_régi>1.0 ){
    for( minden adat sorra ){
        for( minden csomópontra ){
            kiszámít gradiens;
            /*  $\sum P(x,u|D,B) \Delta(x,u) / P(x,u|B)$  */

            csomópont->súlymódosítás += gradiens;
            csomópont->változott++;
        }
    }
    for( minden csomópontra ){
        for( minden súlyra ){
```

```

        súly += mu*súlymódosítás/változott
    }
    újraszámol tábla;
}
kiszámol score;
}

```

**[9.2] táblázat - irodalmi modell paramétertannulása**

Vagyis minden adatsorra, csomópontonként a szülői halmaz nem megfigyelhető változóinak minden behelyettesítésre kiszámítjuk a konfigurációnak az ismert értékek szerinti feltételes valószínűségét, majd az ennek segítségével megkapott gradiens értékek összegével arányosan módosítjuk a 'noisy-or' paramétereket:

$$\frac{\partial \ln P(\underline{D} | B)}{\partial w} = \sum_D \sum_{x,u} \frac{P(x,u | D, B)}{P(x | u, B)} * \Delta(x, u) \quad [9.2]$$

$$w[i + 1] = w[i] + \mu * \left( \frac{\partial \ln P(\underline{D} | B)}{\partial w[i]} / MOD \right) \quad [9.3]$$

A képletben fontos szerep jut a 'MOD' változónak, amely azt a számot jelöli, ahányszor az adott epochban a megfelelő súly gradiense értékes eredményt adott, hisz mint az a gradiens képletéből (a 'Δ' tagból - [4.3] táblázat) látszik, egy-egy súly gradiense csak bizonyos behelyettesítések mellett vehet fel nemnulla értéket. Ezzel a korrekciós taggal így elérhető, hogy a tanulás nagyjából azonos mértékben érintse az összes súlyt, és ne csak néhány módosulhasson számottevően.

A képletben a 'μ' a bátorsági faktort jelöli, ennek a segítségével hangolható, hogy a tanulás kisebb vagy nagyobb lépésekkel haladjon. Tipikus értékei '0,05' körül mozognak, a '0,15'-ös 'μ' már meglehetősen durva lépésköznek mondható.

A tanulás egy másik paramétere a 'threshold', mely a tanulási ciklus kilépési feltételében szerepel. A tanulás addig folytatódik, amíg a háló előző és új 'BIC' értékének aránya 'threshold'-nál kisebb (vagyis amíg a konvergencia bizonyos gyorsaságot elér), vagy '1'-nél nagyobb (vagyis, ha még nincs konvergencia). A 'threshold' nyilvánvalóan '0' és '1' között veheti fel értékét, alacsony beállítás mellett a tanulás általában túl hamar leáll, a túl magas beállítás a kilépési feltétel teljesülését veszélyeztetheti. A mérések során a '0,95'-ös érték jó választásnak bizonyult.

Ha egy háló paraméterezése nem volt adott, hanem valamilyen alapbeállítást kellett alkalmazni, akkor ez hagyományos tábláknál mindig az egyenletes eloszlás volt, 'noisy-or' táblák esetében pedig a kezdeti zajparaméterek a '0,5' értéket kapták.



Ezek a 'noisy-or' paraméterek mindig korlátozva voltak oly módon, hogy értékük a [0,01-0,99] intervallumba essen, így megakadályozva, hogy a következtetés során bizonyos konfigurációk valószínűsége '0'-nak adódjon.

#### 9.4. Struktúratanulás, és heurisztikák

A Bayes-hálók struktúratanulása általánosan a következő képletet követi:

- Bizonyos heurisztika szerint valahány struktúra vizsgálata
- Struktúráként: paramétertárolás, jóságfüggvény-számítás
- Legjobb struktúra kiválasztása

A következőkben részletezett két eljárás is ehhez igazodik. A paramétertárolással kapcsolatban érdemes megjegyezni, hogy teljes adatok esetén a tanulás csomópontonként külön-külön végrehajtható, amivel jelentős megtakarításokat lehet elérni. Mivel a vizsgált esetekben ez a feltétel nem állt fenn, a paraméterek tanítását mindig el kellett végezni, az alapbeállításokról kezdve.

A struktúratanulás során fontos szerep jut a tudásmérnökség kapcsán említett változósorrendnek ([5.1] táblázat), mivel az új struktúrák építése során él csak a sorrendben előbb álló csúcsból mehet egy másik felé.

Tekintettel a lehetséges struktúrák nagy számára, a struktúratanulás egy általános és fontos paramétere a szülői halmaz maximális mérete, vagyis általában nem megengedett, hogy egy csomópontnak tetszőleges számú szülője legyen. Ez a korlátozás a tanulás minőségének biztosítása miatt is fontos, ezzel ugyanis megakadályozható a túllilleszkedés jelensége, mely esetleg bekövetkezhet, ha a rendszerek túl nagy szabadságot adunk.

##### 9.4.1. Mohó algoritmus

A mohó algoritmus a következő kódot követi:

```
do{
    for( minden lehetséges élváltoztatásra ){
        új struktúra;
        paramétertárolás( struktúra )
        score( struktúra )
    }
    if( talált új struktúrát ){
        megtesz( legjobb lépés );
    }
}while( volt( javító lépés ) )
```

[9.3] táblázat - struktúratanulás mohó algoritmussal

Vagyis egy alap struktúrából kiindulva, megvizsgálja az összes lehetséges változtatási lehetőséget, azaz azokat a struktúrákat, amelyek egy él hozzáadásával vagy törlésével előállhatnak, majd a legjobb új hálóból folytatja az eljárást. Ezt addig folytatja, amíg a háló már nem javítható tovább, vagy esetleg, amíg egy adott iterációs küszöböt el nem ér.

Az eljárás egyetlen lényegi paramétere a kiinduló struktúra.

#### **9.4.2. Csomópontonkénti építés**

Az eljárás az összes csomóponton végiglépdel a megadott sorrendben, és kiválasztja hozzá a lehető legjobb szülői halmazt, a csomópontot a sorrendben megelőző változók közül. A következő csomóponton végzett tanulást már az újonnan tanult beállítások mellett végzi, vagyis az előző csomópontok már az optimalizált szülői halmazzal szerepelnek a struktúrában.

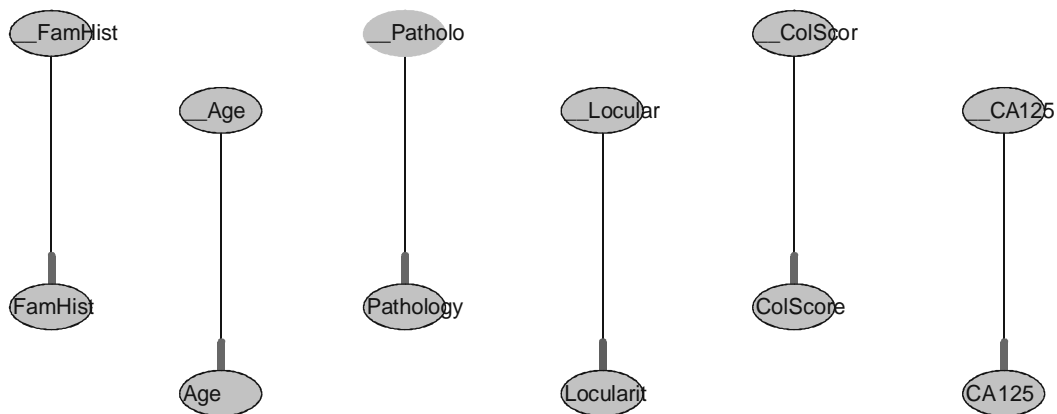
```
for( minden csúcspontra ){
    for( minden szülői halmazra ){
        új struktúra;
        paramétertanulás( struktúra );
        score( struktúra );
    }
    hozzáad( struktúra, legjobb szülői halmaz)
}
```

[9.4] táblázat - struktúratanulás csomópontonkénti építéssel

Az eljárás csupán a már fentebb is említett szülőszám korlátozással paraméterezhető.

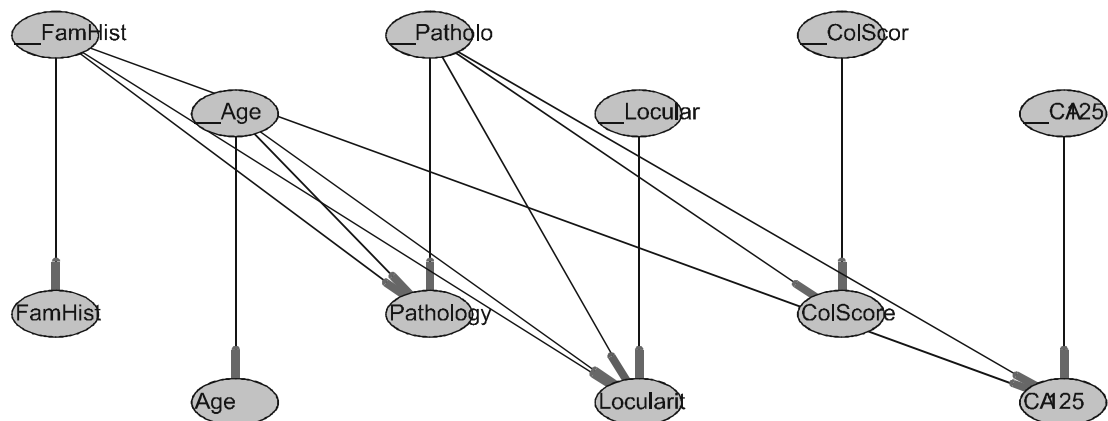
#### **9.5. Irodalmi modell tanulása**

A kétszintű modell tanításához bármelyik struktúratanulási módszer alkalmazható, az annak megfelelő paraméterezéssel. A struktúrák keresésének talán a legfontosabb paramétere a kiindulási háló, mely ez esetben az „üres” háló lehet, vagyis az, amelyben az egymáshoz tartozó „szándék” és a valós megjelenést szimbolizáló csúcs között megy él, és más összeköttetés ezeken kívül nincs.



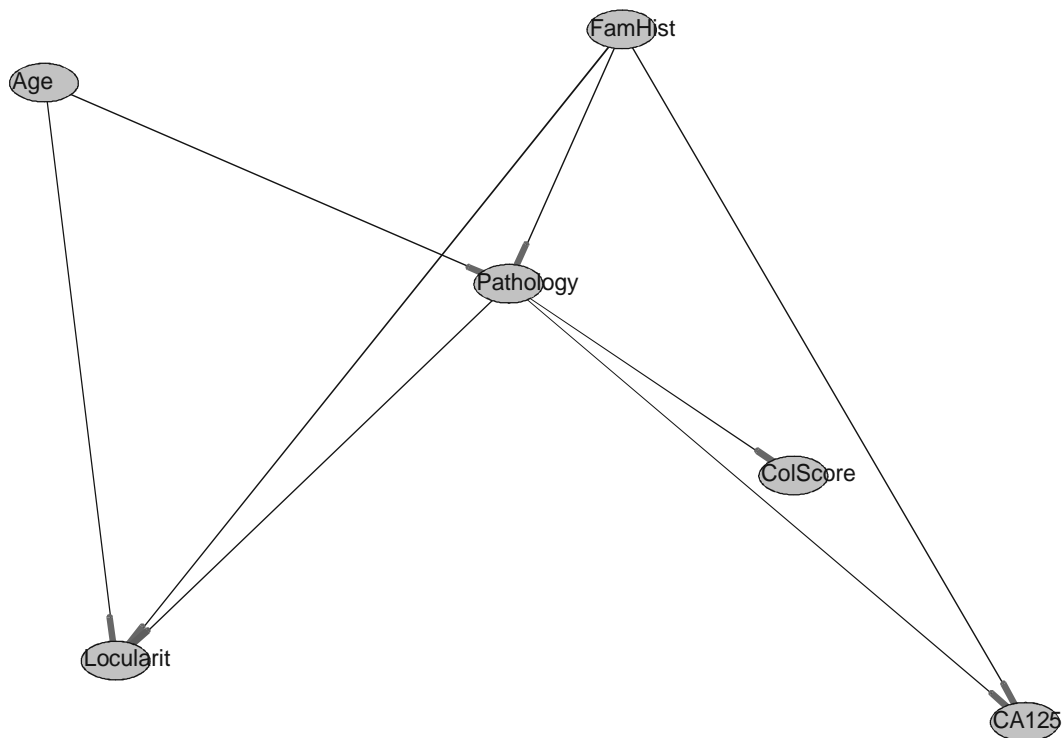
[9.1] ábra - „üres” háló

A modell paramétertánulása során felhasználható a gyermek csomópont és annak saját szülője közti logikai implikáció jellegű összefüggés (szülő=1  $\rightarrow$  gyermek=1). Emiatt ha a gyermek a '0' értéket veszi fel, biztos, hogy a szülő is, így pedig gyakran az elvben nem megfigyelhető szülő értéke is megadható, aminek segítségével a tanulási algoritmus gyorsítható.



[9.2] ábra - tanított háló

A tanulás végeztével a modellt még a [6.2]-ben leírt módon át kell alakítani, hogy ne tartalmazza a „szándékváltozókat”:



[9.3] ábra - a valósmodell-alakba konvertált háló

## 10. Mérési tapasztalatok és eredmények

### 10.1. A struktúrat tanulási heurisztikák összevetése

A valós modellen végzett általános mérések elsősorban a területtel kapcsolatos előkutatásoknak tekinthetők. Ezekre a vizsgálatokra még a diplomamunka készítése előtt, egy TDK dolgozat írása kapcsán került sor. Az eredmények itt történő felhasználásának célja ennek megfelelően az volt, hogy a különböző alkalmazható heurisztikák közül egy olyat tudjak kiválasztani, amely az adott változókból alkotott modellen helyesen és megbízhatóan működik.

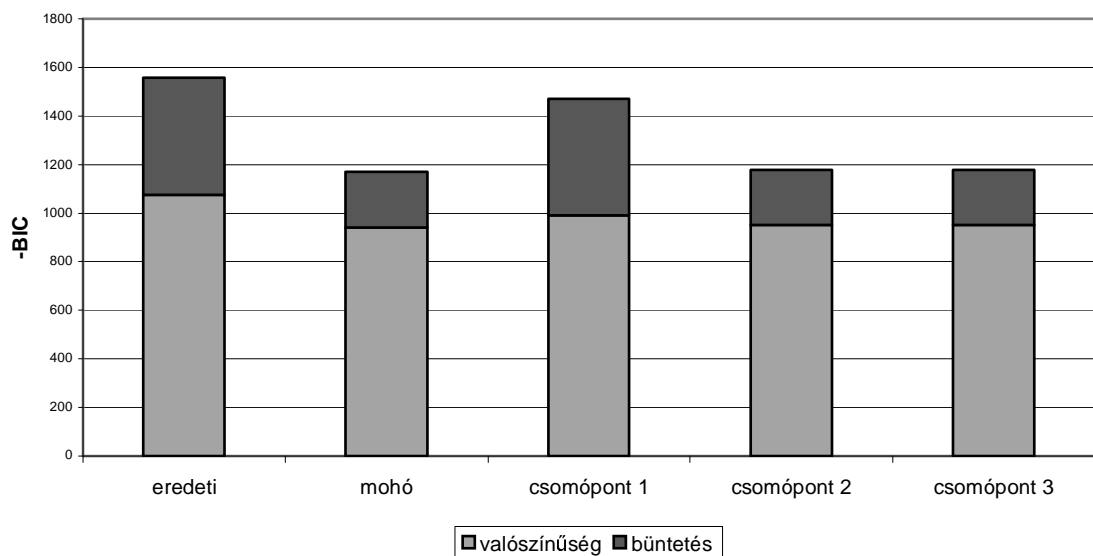
Az [5.3] fejezetben felsorolt heurisztikák közül a mohó algoritmust, valamint a csomópontonkénti építést vizsgáltam meg, ez utóbbit '1', '2' és '3' maximális szülőihalmaz-méret mellett.

A csomópontonkénti módszerek a vártak megfelelően a szülőihalmaz-méret növekedtével egyre jobbnak bizonyultak, ami nem is meglepő, hisz az  $e$  paraméter növelésével kapott új keresési tér az előzőt mindig tartalmazza. Lényeges megfigyelés volt még, hogy a szülők lehetséges számának további növelésével már nem javultak az

eredmények, ami tehát azt sugallhatja, hogy ennek a kutatási területnek és változóhalmaznak a háromszülős modellek már meglehetősen pontos leírását adják.

A vizsgálatok kiterjedtek még a mohó keresésre is, mely adott esetekben a csomópontonkénti építéssel közel azonos jótságot ért el, ennek ellenére ez a módszer gyengébbnek nevezhető, mivel (főleg talán a keresési tér átfésülésekor meglátogatott kevesebb struktúra miatt), erősebb mértékben érzékeny nemcsak az adatvesztés nagyságára, hanem az eltűnt bejegyzésekre magukra is. Vagyis ez azt jelenti, hogy többször lefuttatva az algoritmust ugyanolyan veszteségű különböző adathalmazokra (melyek természetesen ugyanabból a teljes mintából készültek véletlen sorsolással), meglehetősen nagy szórást tapasztalhatunk az eredményekben.

Az alábbi diagram az egyes módszerek sikerességét mutatja 10%-os adatvesztés mellett:



[10.1] ábra

**struktúratanulási heurisztikák összevetése valós adatokon**

### 10.1.1. Konklúzió

A fenti eredményekből tehát látszik, hogy a csomópontonkénti építés 3 szülővel már megbízhatóan adja a legjobb eredményt a heurisztikák közül, a szülői halmaz további növelésével pedig már nem érhető el javulás. Bár a mohó algoritmus is elérhet ilyen eredményt, de emellett a tanulás bizonytalanabb marad.

Összességében ez, és a több megvizsgált struktúra nyújtotta biztonság is, a csomópontonkénti építést tünteti fel jobb fényben, amiért tehát a további vizsgálatok során ennek a módszernek az alkalmazására szorítkoztam.

## **10.2. Irodalmi modell**

Az irodalmi modellek taníthatóságának kutatása tehát elsősorban azt a célt szolgálja, hogy olyan módszereket fejleszthessünk ki, amelyek segítségével egy adott szakterületről rendelkezésre álló szaktudást a terület szakirodalmából nyerjük ki, ahelyett, hogy ehhez szakértők segítségét kellene felhasználnunk.

A mérések fő vonulata tehát azt vizsgálta, hogy milyen hasonlóságok, illetve különbségek adódnak az automatizált módon tanult és az etalonnak tekintett szakértői háló között. A vizsgálatok során két irodalmi modellt hasonlítottunk össze: a konzulensem által megvalósított egyszintű és az általam megvalósított kétszintű modellt.

A kétszintű modellek tanulása során történő paraméterváltozásokra, valamint az egyes változók különböző szülői halmazainak jóságára (illetve az azok közti különbségekre) a [14.1] és a [14.2] táblázat ad illusztrációt.

### **10.2.1. A tanuláshoz felhasznált változók**

A tanuláshoz felhasznált változók felsorolása már az [5.1] fejezetben megtörtént, ez a 16 változó egy bővebb, 35-ös halmaznak azon részhalmaza, melynek az elemei a terület kulcsváltozójához, a patológiához (Pathology) szorosabban kötődnek. Ez okozta, hogy a tanítások eredményeként kapott hálók viszonylag sűrűk lettek a terület bővebb modelljeihez képest.

A paraméterérzékenységet, illetve az algoritmusok robusztusságát egy még kisebb, 6 csomópontból álló gráfon vizsgáltam, ezek:

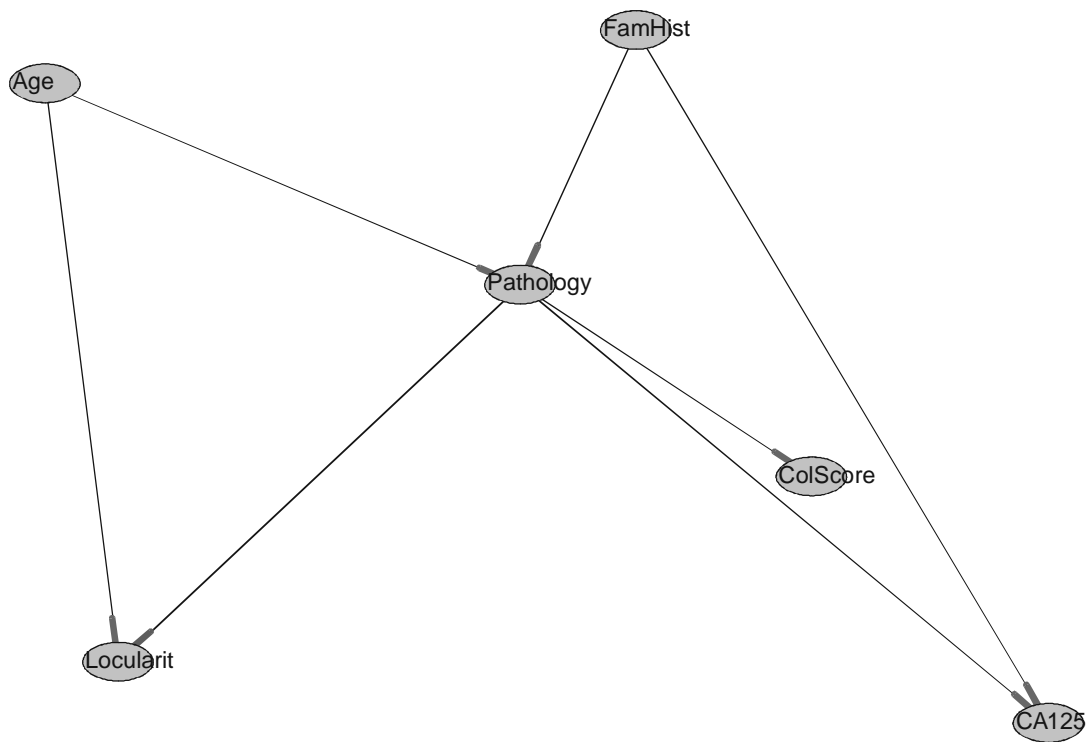
Age  
FamHist  
Pathology  
Locularity  
ColScore  
CA125

[10.1] táblázat - a robusztusság vizsgálatához használt változóhalmaz

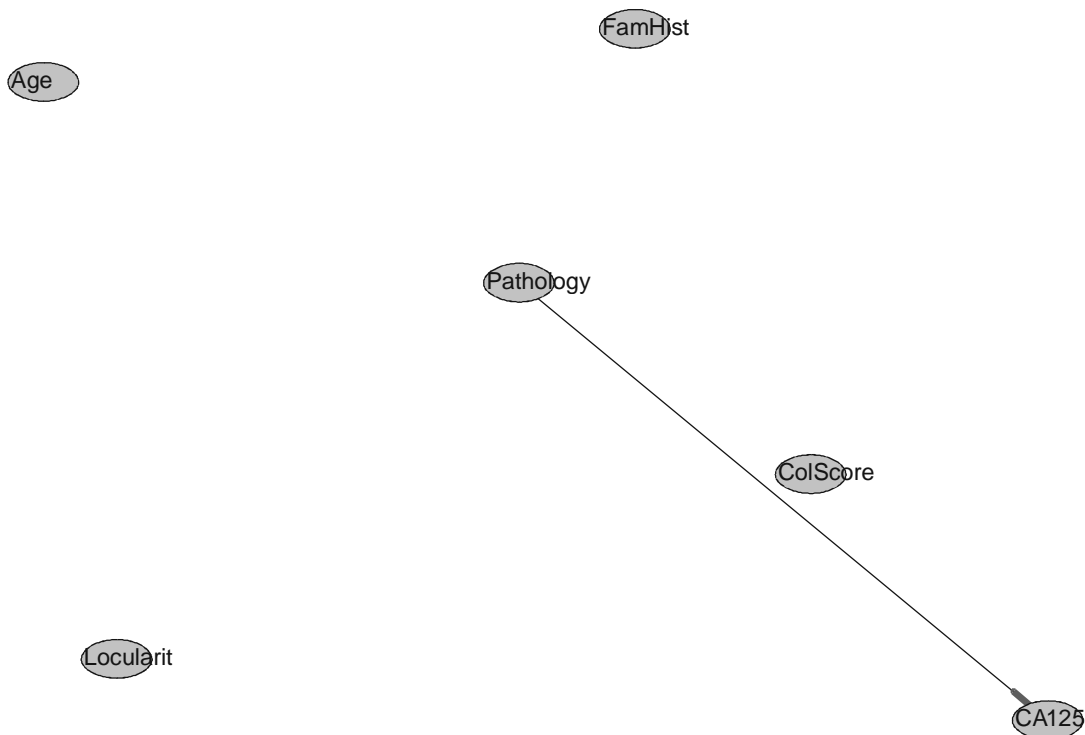
### **10.2.2. A tanulás érzékenysége a paraméterekre**

A kétszintű modell tanítása során az algoritmus összesen 4 értékkel paraméterezhető: az alkalmazott büntetőfaktorral, a szülők számával, valamint a paramétertanulás 'μ' és 'threshold' paraméterével.

A fentiek közül a **büntetőfaktor** 3 értéket vehet fel, a konstans 0-t ([5.7]-es képlet), azt, amelyik a csomópontok táblái bejegyzéseinek számát, vagy amelyik a 'noisy-or' zajparaméterek számát tartalmazza ([5.5]-ös, [5.6]-os képlet). Míg a CPT bejegyzéseken alapuló tanulás jelentősen csökkenti az élek számát, és így egyértelműen rontja a tanulást, addig a másik két beállítás között nincs mérhető különbség, a zajparaméterek alacsony száma miatt.



[10.2] ábra - tanulás alacsony, vagy '0' büntetés mellett



[10.3] ábra - tanulás erős büntetéssel

A **szülők száma** természetesen erősen befolyásolja a tanulás kimenetelét, ennek a paraméternek a növelése esetén az látható, hogy a háló általában új élekkel bővül, azaz viszonylag kevés az olyan eset, amikor például az egyszülős tanulás során legjobbnak bizonyult szülő helyett a kétszülős tanulás két új élt talál. Ez egyértelműen jelzi, hogy a módszer a szülőszám-növeléssel javul.

Konkrét adatok ezzel a paraméterrel kapcsolatban a következő fejezetben láthatók.

A fennmaradó két paraméter a paramétertanuláshoz kapcsolódik, és afféle finomhangoló szerepet tölt be. Általánosan elmondható, hogy az érzékenyebb paraméter a 'threshold', vagyis a 'BIC' érték javulási arányának küszöbe a kilépési feltételben. Ennek alacsonyan tartása túl korai leállást eredményez, ami meg is látszik a tanult hálón. A ' $\mu$ ' paraméter az előző mellett inkább csak kiegészítő szerepet játszik, a tanulás végső eredményére nincs kihatással kellően magas ('1'-hez közeli) 'threshold' mellett. A magas 'threshold' melletti nagy ' $\mu$ ' megakadályozhatja a tanulás kívánt mértékű konvergenciáját. A 'threshold' növelése természetesen a tanulás lassulásával jár.

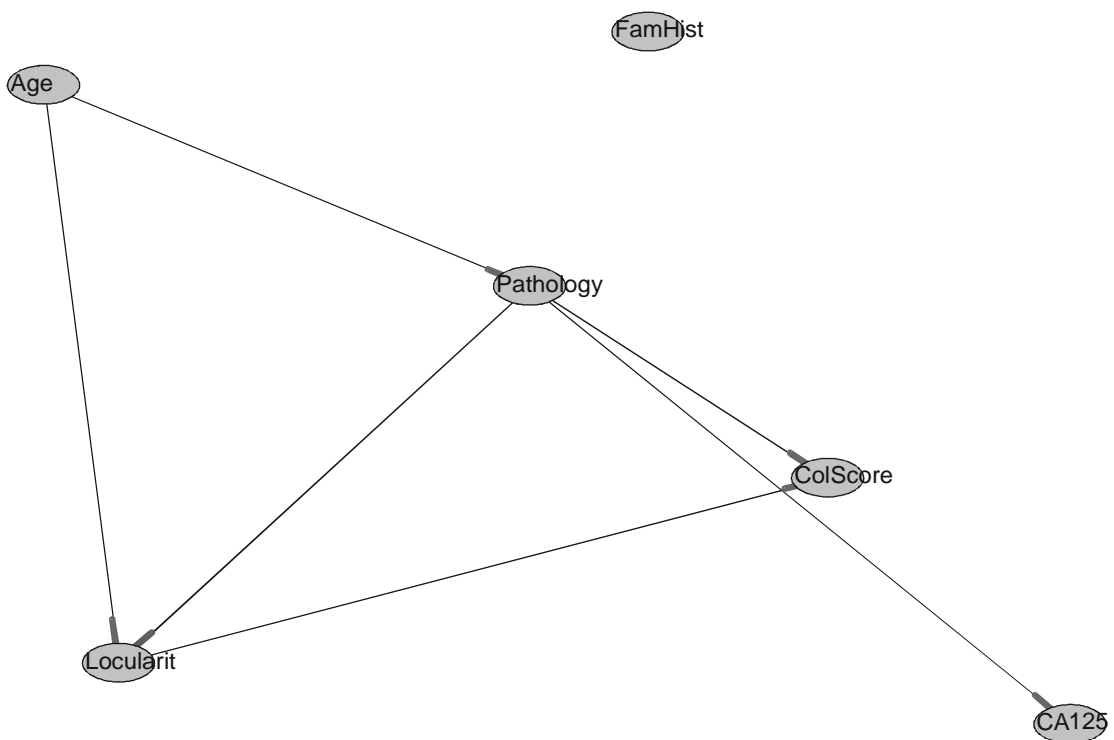


E két paramétert 5 beállítás mellett vizsgáltam, ezek:

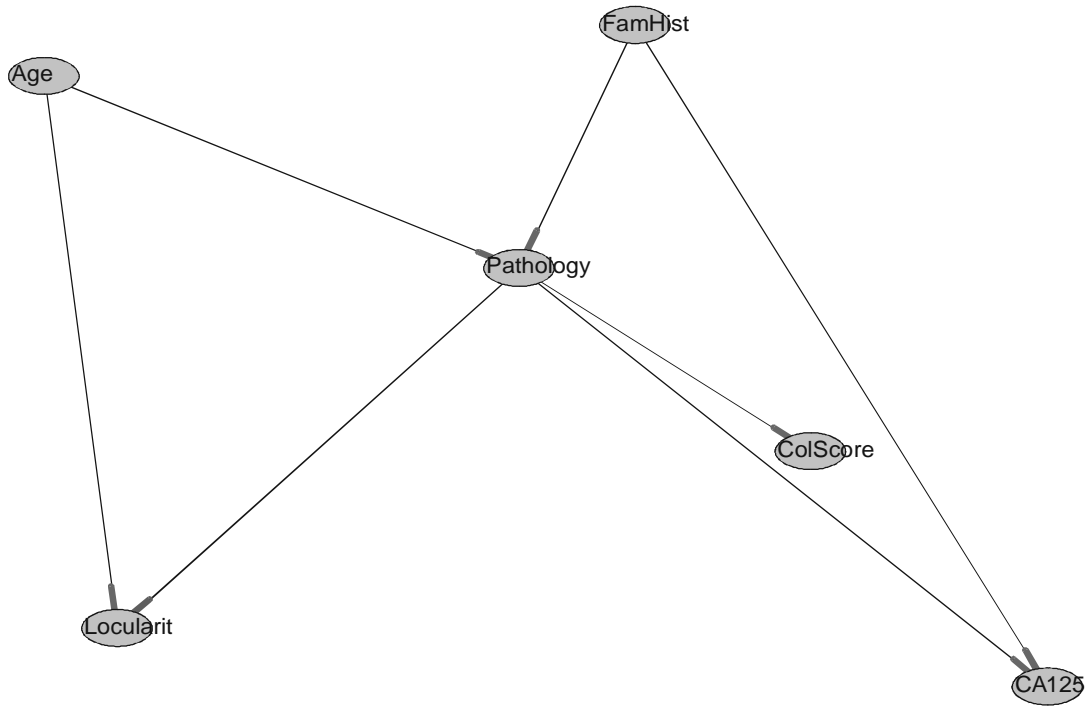
	$\mu$	threshold
1.	0,10	0,80
2.	0,10	0,90
3.	0,05	0,95
4.	0,02	0,975
5.	0,01	0,9875

[10.2] táblázat - ' $\mu$ ' és 'threshold' vizsgált beállításai

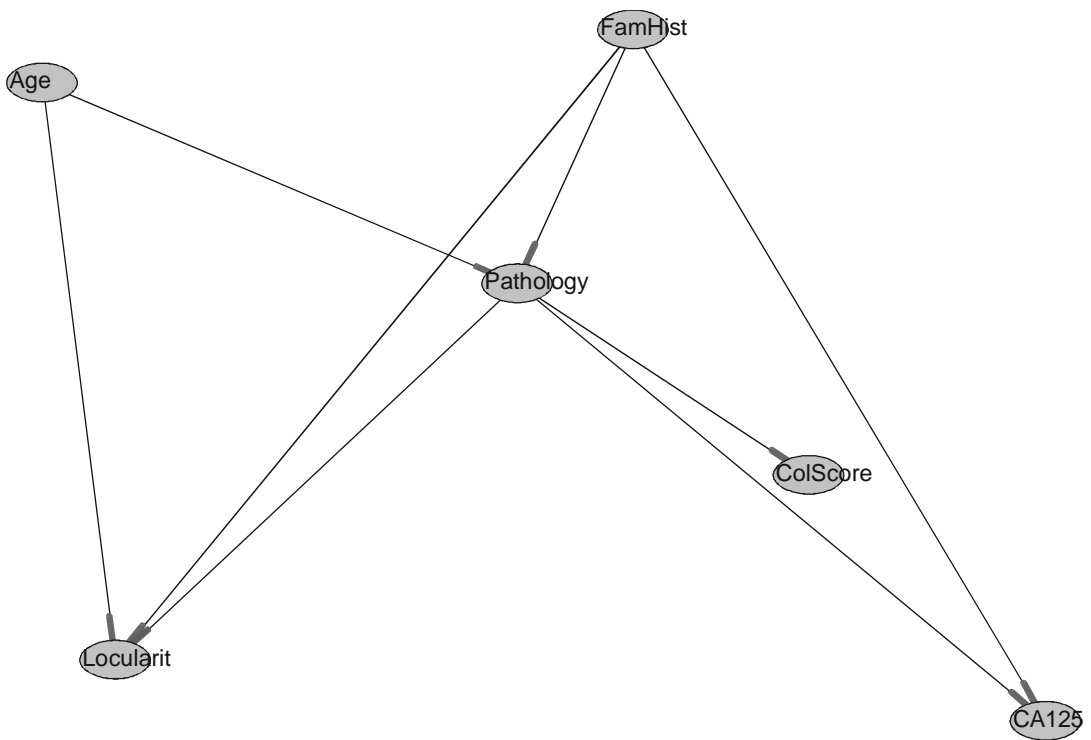
Az eredményül kapott hálók:



[10.4] ábra - ' $\mu=0,1$  threshold=0,8' vagy ' $\mu=0,1$  threshold=0,9'



[10.5] ábra - ' $\mu=0,05$  threshold=0,95'



[10.6] ábra - ' $\mu=0,02$  threshold=0,975' vagy ' $\mu=0,01$  threshold=0,9875'

Látható, hogy az 1. és 2. eset viszonylag nagy eltérést produkál a többihez képest, ez a túl kicsi 'threshold' beállítással magyarázható. A 3. és a 4., 5. eset közti egyetlen él

különbség sugallja, hogy a tanulás elérte a lehetséges optimumot. Összességében a 3. eset tekinthető az arany középút választásnak, mely elfogadható tanulási idő mellett hoz jó eredményt.

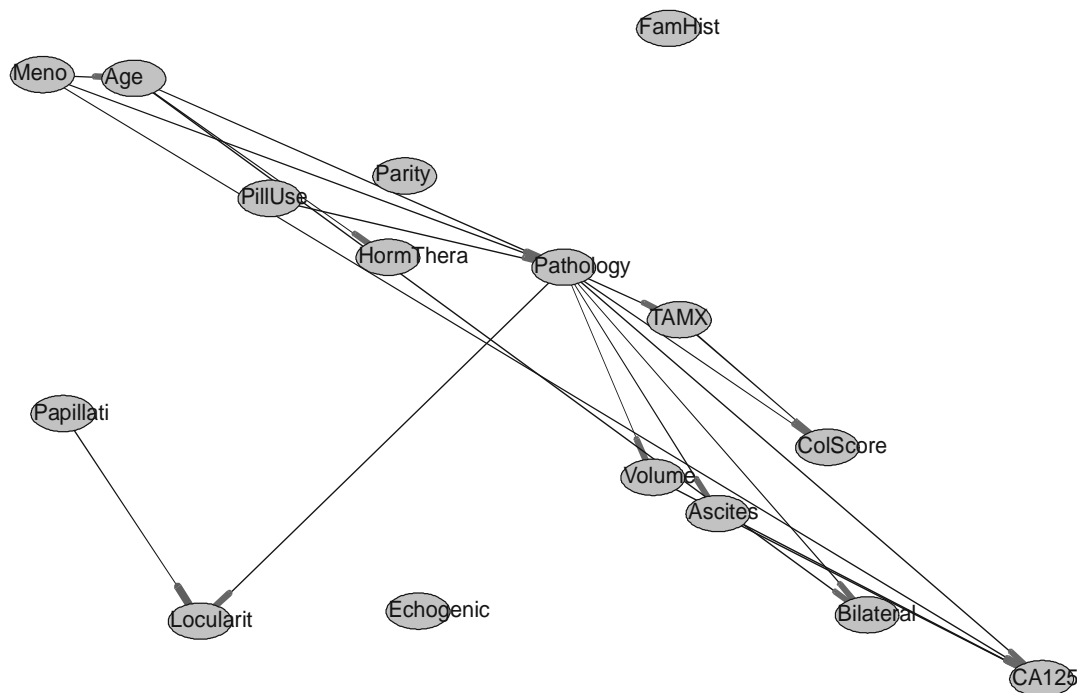
A fentiek fényében tehát a továbbiakban a kétszintű modellek mindig büntetőfaktor nélkül, ' $\mu=0,05$ ' és ' $\text{threshold}=0,95$ ' beállításokkal lettek tanítva.

### 10.2.3. Eredmények

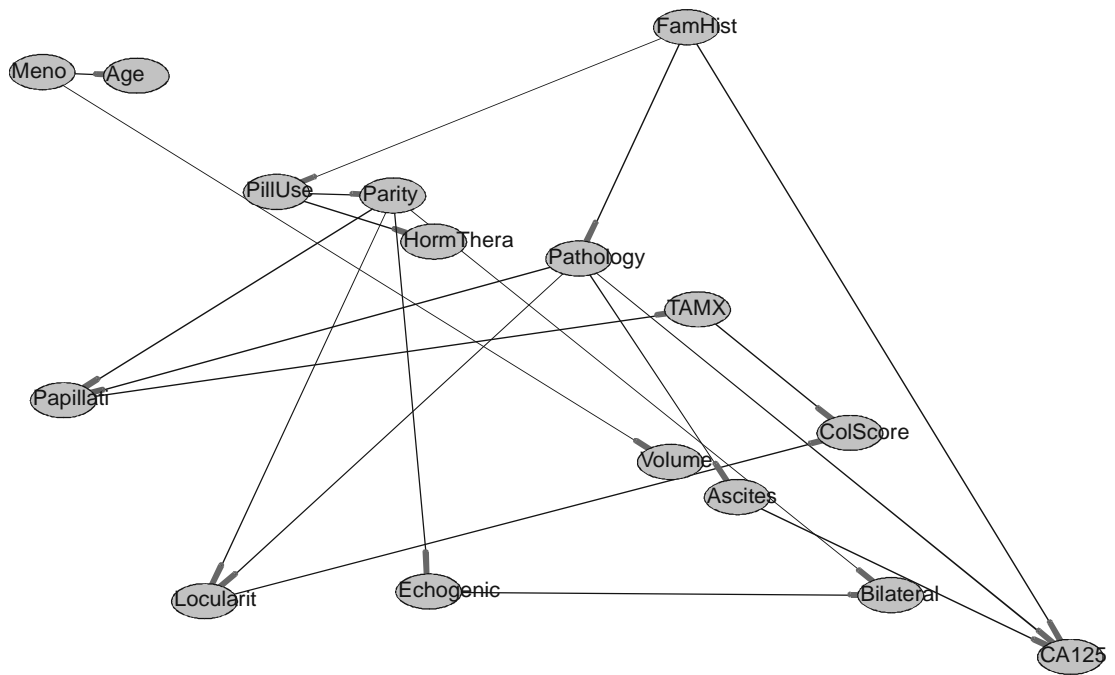
A 16 változón összesen öt modellt vettem össze:

- a szakértői hálót, mint viszonyítási alapot
- a konzulensem által tanított egyszintű modellt
- és a kétszintű modell egy-, két- és háromszülős változatát

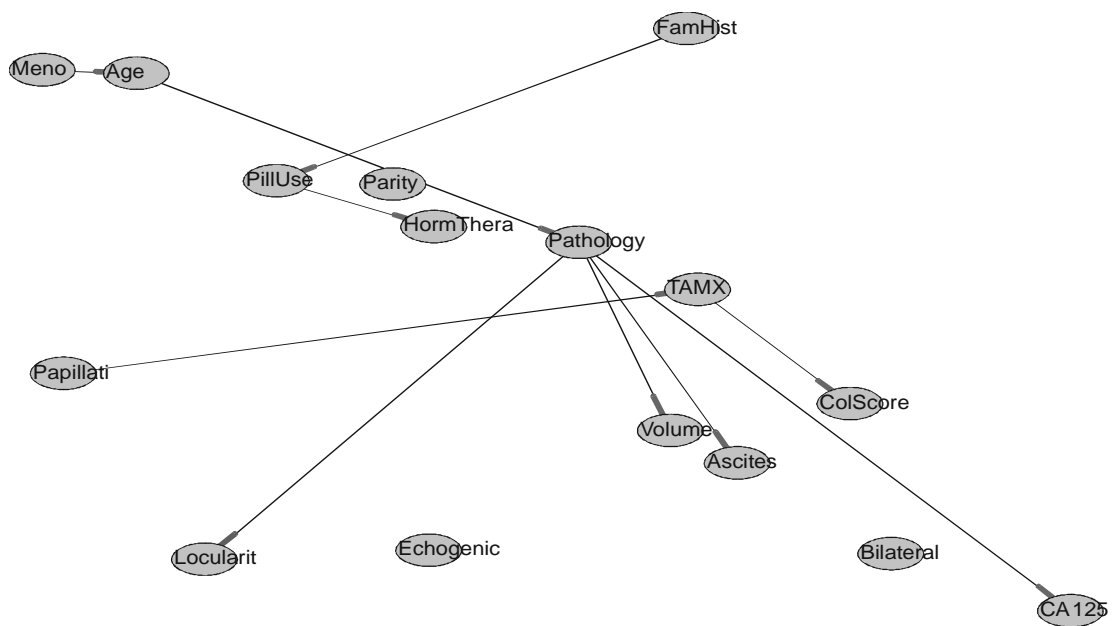
Az alábbi ábrák mutatják az egyes modelleket:



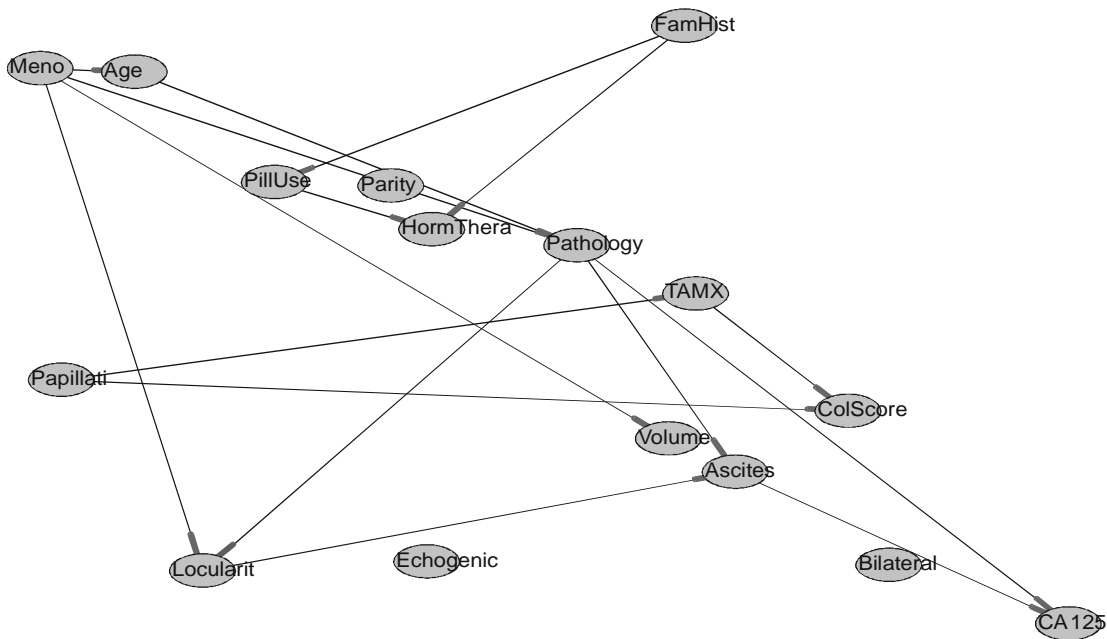
[10.7] ábra - szakértői modell



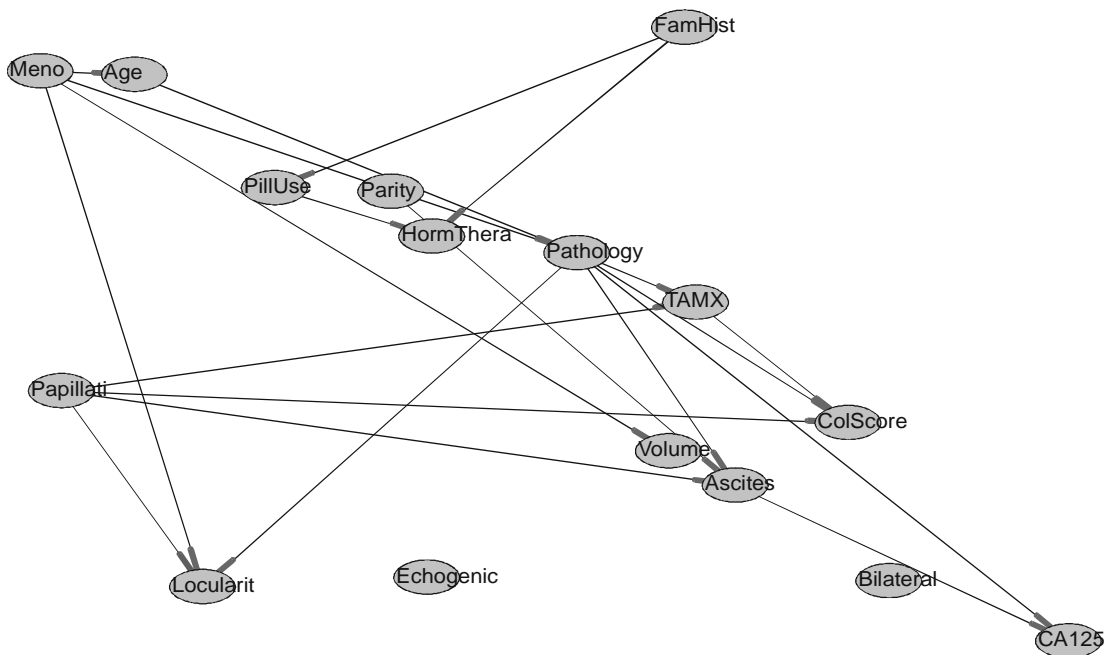
[10.8] ábra - egyszintű modell



[10.9] ábra - kétszintű modell - 1 szülő



[10.10] ábra - kétszintű modell - 2 szülő



[10.11] ábra - kétszintű modell - 3 szülő

Tekintve, hogy a kétszintű modell a szülőszám növelésével csak javulhat, a továbbiakban csak a háromszülős kétszintű modellt, illetve az egyszintűt vizsgáljuk.

A hálóknak egy jó „kézi” összevetési módszere, ha a legfontosabb változók szülői halmazait vizsgáljuk. Ezek a változók a következők: 'Pathology', 'Papillation', 'ColScore', 'Ascites', 'CA125'. Egy összefoglaló táblázatban jól láthatók a különbségek:

GYERMEK CSOMÓPONT	SZÜLŐI HALMAZ – MODELL		
	SZAKÉRTŐI	EGYSZINTŰ	KÉTSZINTŰ, 3 SZÜLŐ
Pathology	<b>Meno</b> <b>Age</b> PillUse Parity	<i>FamHist</i>	Meno Age
Papillation	<b>Pathology</b>	<i>Parity</i> Pathology	-
ColScore	<b>Pathology</b> Papillation Locularity TAMX	Locularity TAMX	Pathology Papillation TAMX
Ascites	<b>Pathology</b> Locularity Volume	Pathology	<i>Parity</i> Pathology <i>Papillation</i>
CA125	<b>Pathology</b> <b>Ascites</b> Volume Meno Age	<i>FamHist</i> Pathology Ascites	Pathology Ascites

[10.3] táblázat - szülői halmazok a vizsgált modellekben

A táblázat 'szakértő' oszlopában félkövérek a nagyon, normál szedésűek a közepesen fontos kapcsolatok. A másik két oszlopban normál szedésűek a szakértői modellel egyező, dőlték az azzal nem egyező elemek. Egy egyszerű minősítés készíthető, ha ezután mindkét modellre összeszámoljuk, hogy a szakértői modellhez képest hány ott meglévő él hiányzik, illetve hány került be tévesen.

Ebben az összevetésben az egyszintű modell 14, a kétszintű 11 „hibapontot” szerez, ami azt sugallja, hogy az utóbbi általában jobban teljesít az egyes csomópontok közvetlen környezetének meghatározásában.

A [7] fejezetben leírt mátrixokon alapuló módszer a fenti kézi összehasonlításnál pontosabb és részletesebb eredményekkel szolgálhat, hisz itt nem csak néhány kiemelt változó közvetlen környezete (szülői halmaza), hanem az összes csomópont-csomópont kapcsolat alapján történik az összevetés.

Az etalon ezúttal is a szakértői háló, pontosabban annak két változata, az egyik, amelyben csak a nagyon ('H'), a másik, amelyben a nagyon és a közepesen fontos élek szerepelnek ('HM').

(A mátrixokban a sorok jelölik a szakértői hálót, az oszlopok az irodalmi modellt, vagyis például a 'független' sor 'kauzális út' oszlopának eleme jelöli, hogy a szakértő szerinti független párok közül hány lett az irodalmi modellben kauzális úttal összekötött; a vizsgálatban rendezett párok szerepelnek, azaz a 'Pathology-Papillation' és a 'Papillation-Pathology' pár külön számít):

szakértő\modell	független	megzavart	kauzális út	kauzális él	$\Sigma$
független	14	14	12	12	52
megzavart	6	14	0	2	22
kauzális út	44	48	24	14	130
kauzális él	14	6	4	12	36
$\Sigma$	78	82	40	40	

[10.4] táblázat - HM vs. egyszintű - mutató: 102

szakértő\modell	független	megzavart	kauzális út	kauzális él	$\Sigma$
független	44	0	0	8	52
megzavart	14	8	0	0	22
kauzális út	82	18	20	10	130
kauzális él	8	4	2	22	36
$\Sigma$	148	30	22	40	

[10.5] táblázat - HM vs. kétszintű - mutató: 112

A ' $\Sigma$ ' sor alapján a két irodalmi modell előzetesen egymással is összevethető, látszik, hogy az egyszintű modell „összekötöttebb” eredményt ad, vagyis ugyanannyi 'kauzális él' kapcsolat mellett (40-40), a kétszintű modellben jóval több a független csomópont-pár (148 a 78-cal szemben). Ez az egyszintű modellt tünteti fel jobb fényben, hisz a tárgyterület vizsgálata során kedvezőtlen, ha sok változó marad elszigetelve egymástól. Ezzel szemben viszont kevesebb a közös őstől 'megzavart' kapcsolat a kétszintű hálóban (30 - 82), ami hasznos tulajdonság lehet, ha tisztán kauzális kapcsolatokat akarunk feltérképezni.

A [7.1] képlet szerinti mutató az egyszintű modell javára dönt, vagyis azt mondhatjuk, hogy ez általános értelemben közelebb áll a szakértők 'HM' hálójához. A mátrixok főátlóbeli elemeit ha összegezzük, a kétszintű modell jobbnak bizonyul 94 – 64 arányban, azaz ebben több kapcsolat marad sértetlenül, ami az egyes élekre is igaz, mint az a 'kauzális él – kauzális él' elemekből is látszik. Emellett megjegyzendő, hogy a 'kauzális' (jobb alsó 2x2-es) részmatrixok elemösszege megegyezik, vagyis pusztán a kauzalitás megőrződését tekintve a két modell megegyezik.

szakértő\modell	független	megzavart	kauzális út	kauzális él	$\Sigma$
független	32	42	34	24	132
megzavart	16	6	0	2	24
kauzális út	30	34	4	8	76
kauzális él	0	0	2	6	8
$\Sigma$	78	82	40	40	

[10.6] táblázat - H vs. egyszintű - mutató: 146

szakértő\modell	független	megzavart	kauzális út	kauzális él	$\Sigma$
független	98	8	8	18	132
megzavart	4	10	6	4	24
kauzális út	46	12	8	10	76
kauzális él	0	0	0	8	8
$\Sigma$	148	30	22	40	

[10.7] táblázat - H vs. kétszintű - mutató: 84

Ha az összehasonlítást a 'H' szakértői hálóval végezzük, amely tehát kevesebb élt tartalmaz, akkor, mint az várható is, a kétszintű, kevésbé összekötött modell ér el jobb mutatót, és az előbb tapasztalt különbség a főátló-összegekben is markánsabbá válik (124 – 48). Itt már a fentebb említett 'kauzális' részmatrixok elemösszegei is eltérést mutatnak, 26 – 20 arányban a kétszintű modell „nyer”, külön érték továbbá, hogy benne megjelenik az összes szakértői él (lásd a mátrix 'kauzális-él' sorát).

## 11. Konklúzió

Az előző fejezetben két tanulási módszernek a szakértői tudást megjelenítő hálóval való összevetését láthattuk. Ahhoz, hogy eldönthessük, vajon az új módszerek fejlesztése kifizetődő-e, azok költségét és időigényét, illetve az általuk kínált eredményeket kell megvizsgálunk.

Az eljárások költségkímélő volta nyilvánvaló, hisz végrehajtásukhoz nem szükségesek a vizsgálati terület valódi megfigyelési adatai, amelyeknek összegyűjtése adott esetben rendkívül költséges, vagy akár megvalósíthatatlan is lehet. Noha persze a tanult struktúra paramétertanulásához mindenképpen szükségesek mérési adatok, ezek mennyisége jóval kevesebb lehet, mint ha struktúratanuláshoz is fel akarnánk használni



őket. A valós adatokkal szemben az irodalmi modellek tanító adatai könnyen és olcsón hozzáférhetőek (a legtöbb cikk ingyenesen letölthető az internetről).

A fent említett szempontok általában igazak az időigénnyel kapcsolatban is, azaz némi előfeldolgozástól eltekintve az adatgyűjtés elhanyagolható időt igényel, míg a valós adatok esetleg évek alatt keletkeznek megfelelő mennyiségben. Hasznos továbbá, hogy az emberi szakértővel való konzultáció is elhagyható, vagy legalábbis jelentősen csökkenthető, amit a valós adatokból történő tanulás esetén, a tanulás időigényes volta miatt esetleg nem tehetünk meg.

Elmondható tehát, hogy a vizsgált új módszerek már csak várható gyorsaságuk és olcsóságuk miatt is egy megvizsgálendő lehetőséget jelentenek. Természetesen figyelembe kell venni, hogy a kapott eredmények mennyire felhasználhatók, hisz ezek nélkül az új tanulási módszerek nem sokat érnének.

A [10] fejezet vizsgálatai bizonyítják, hogy az irodalmi modellek tanulásuk során az eddig jónak elfogadotthoz (a szakértői hálózathoz) közel álló eredményeket szolgáltatnak. Különösen a kétszintű irodalmi modellek figyelemreméltóak, hisz a számszerű eredmények mellett olyan hasznos és statisztikai szempontokból lényeges tulajdonságokat is megtartanak, mint például a csomópontok egymástól való függetlensége, illetve a csomópontok közvetlen szülői környezete. Ezek mellett a [10.2.2] fejezetben látott robusztusság is értékes jellemzője az algoritmusnak, hisz így tanulása bizonyos szintig objektívnek tekinthető, mely csak kis mértékben függ a felhasználó által választott paramétereiktől.

Összességében kijelenthető tehát, hogy a vizsgált, viszonylag egyszerű módszerek is jó, a lényeges szempontokat megőrző eredményeket adnak, ami mindenképpen további vizsgálatukat, és továbbfejlesztési lehetőségeik kutatását indokolja.

## **12. További lehetőségek**

Miután láthatóvá vált, hogy az irodalmi modellek kutatása hasznos eredményeket szolgáltat, az elsődleges továbbfejlesztési lehetőségek mindenképpen a keresési tér kibővítését, illetve új modell típusok vizsgálatát sugallják.

A keresési tér bővítése történhet egyrészt új, több struktúrát érintő, és ezáltal biztosabb keresési heurisztikák alkalmazásával, melyek kihasználva a keresés [10.2.2]-ben látott robusztus mivoltát, a paramétertárolás gyorsítása által vizsgálhatnak át nagyobb struktúrákat.

A másik lehetőség, ha az eddig fix változószorrenden változtatásokat is engedünk keresés közben, így növelve a lehetséges gráfok számát. Az új sorrendek közül külön figyelmet érdemel az eredeti fordítottja, mely a már említett „publikációs szabályok” egy eddig figyelmen kívül hagyott lehetőségét testesíti meg: a fogalmak egymással való magyarázatának láncolata nem csak az ok felől haladhat az okozat felé, hanem fordítva is.

A másik lehetséges út, ha az irodalmi modell erősen korlátozott struktúrájának adunk nagyobb szabadságot, azaz nem csak a párosgráf-struktúrát engedélyezzük, hanem a 'szándék' változók között is megengedünk éleket. Ez valamiképpen a publikáló személy szándékainak belső egymásrahatását jelenítheti meg. Logikus feltevés, hogy a különböző fogalmakról való írás szándéka nem független egymástól, vagyis az egymáshoz közel álló változók nem csak az egymást magyarázó szerepben jelennek meg, hanem a köztük lévő kapcsolat miatt is „vonzzák” egymást.

További, bár az eddigi vonulattól már valóban távol álló lehetőség a tanító adatok kinyerésének fejlesztése. Bár az eddig alkalmazott mutató-értékek is sikeres alkalmazásokat tettek lehetővé, kifinomultabb képletekkel talán növelhető a hatékonyság. A jelenlegi formula nem érzékeny a kulcsszavak szövegben való elhelyezkedésére (a szó nagyjából mindenhol előfordul, vagy csak a dokumentum egy részében), sem a szöveggörnyezet szemantikájára, bár a természetes nyelvi elemzés már valóban messzire mutat.

### 13. Irodalomjegyzék

- (Antal04) P. Antal, G. Fannes, Y. Moreau, D. Timmerman, B. De Moor: **Using Literature and Data to Learn Bayesian Networks as Clinical Models of Ovarian Tumors**, Artificial Intelligence in Medicine, vol 30, pp 257-281, 2004
- (Cooper92) G. F. Cooper, E. Herskovits: **A Bayesian Method for the Induction of Probabilistic Networks from Data**, Machine Learning, vol 9, pp 309-347, 1992.
- (Friedman96) N. Friedman, Z. Yakhini: **On the Sample Complexity of Learning Bayesian Networks**, Proc. of the 12<sup>th</sup> Conf. on Uncertainty in Artificial Intelligence, pp 274-282, 1996.

- (Friedman97) N. Friedman: **Learning Belief Networks in the Presence of Missing Values and Hidden Variables**, Proc of the 14<sup>th</sup> international conference on machine learning, pp 125-133, 1997.
- (Gelman95) A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin: **Bayesian Data Analysis**, Chapman & Hall, 1995.
- (Heckermann95a) D. Heckerman: **A Tutorial on Learning with Bayesian Networks**, Technical Report, vol MSR TR-95-06, 1995.
- (Heckerman95b) D. Heckerman, D. Geiger: **Likelihoods and parameter priors for Bayesian networks**, Technical Report vol MSR-TR-95-54, 1995.
- (Heckerman95c) D. Heckerman, D. Geiger, D. Chickering: **Learning Bayesian networks: The Combination of Knowledge and Statistical Data**, Machine Learning, vol 20, pp 197-243, 1995.
- (Hirschman02) L. Hirschman1, J. C. Park, J. Tsujii, L. Wong, C. H. Wu: **Accomplishments and challenges in literature data mining for biology**, Bioinformatics Review, vol 18, pp 1553-1561, 2002.
- (Huang96) C. Huang, A. Darwiche: **Inference in Belief Networks: A procedural guide**, International Journal of Approximate Reasoning, vol 15, pp 225-263, 1996.
- (Mahoney96) S. Mahoney, K. Blackmond Laskey: **Network Engineering for Complex Belief Networks**, Proc. 12th Conf. on Uncertainty in Artificial Intelligence, pp 389-396, 1996.
- (Neil00) M. Neil, N. E. Fenton, L. Nielsen: **Building large-scale Bayesian Networks**, The Knowledge Engineering Review, pp 257-284, vol 15, 2000.
- (Pearl88) J. Pearl: **Probabilistic Reasoning in Intelligent Systems**, Morgan Kaufmann, San Francisco, 1988.
- (Russel95) S. Russell, J. Binder, D. Koller, K. Kanazawa: **Local learning in probabilistic networks with hidden variables**, In proceedings of the 14<sup>th</sup> international joint conference on artificial intelligence, pp 1146-1152, 1995.
- (Russel00) S. Russel, P. Norvig: **Mesterséges intelligencia modern megközelítésben**, Panem – Prentice Hall, pp 523-561, 692-698, 2000.
- (Sarkar96) S. Sarkar, I. Murthy: **Constructing Efficient Belief Network Structures With Expert Provided Information**, Knowledge and Data Engineering, vol 8, pp 134-143, 1996.

(Timmerman00) D. Timmerman, L. Valentin, T. H. Bourne, W. P. Collins, H. Verrelst, I. Vergote: **Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group**, *Ultrasound Obstetrics Gynecology*, vol 16, pp 500-505, 2000.

## 14. Függelék – tanulás

Az alábbi táblázatok a kétszintű irodalmi modell hálóinak tanulási előrehaladását ábrázolják. Tekintve, hogy a teljes állományok közlésére itt hely hiányában nincs mód, az alábbi részletek csak illusztrációnak tekinthetők.

### 14.1. Paraméterek változása

Az alábbi táblázat egy három csomópontot ('FamHist', 'Pathology', 'CA 125') tartalmazó háló tanulását ábrázolja (nem számítva a természetesen jelen lévő „szándékváltozókat”, pl. '\_\_FamHist'-et). A táblázatban a '>>' végződésű sorok jelölik az éppen vizsgált csomópontot, melyhez a '<<' kezdetű sorban felsorolt csomópontok kapcsolódnak szülőként. Az ezt követő táblázat tartalmazza a háló paramétereit (a '\_\_[név]' oszlopok a felső sor változóinak paraméterét, vagyis a '0' érték a priori valószínűségét, a '[prefixum]-[prefixum]' oszlopok pedig a jelzett csomópontok közti él zajparaméterét – itt nem tüntettük fel az esetleges '\_\_' előjelet).

#### [14.1] táblázat

#### paramétermódosulások az irodalmi modell tanulása során

NODEWISE STRUCTURE LEARNING  
mu: 0.05, threshold: 0.95

FamHist >>

<<

__Pathol	__FamHis	__CA125	score:	-1039.720771
0.4790	0.5885	0.5609	score:	-919.377261
0.4605	0.6622	0.6119	score:	-833.476261
0.4441	0.7262	0.6555	score:	-768.840638
0.4296	0.7828	0.6929	score:	-719.177460
0.4167	0.8329	0.7250	score:	-681.064629
0.4053	0.8768	0.7523	score:	-652.599525

structure: 1

Pathology >>

<<

__Pathol	__FamHis	__CA125	score:	-1039.720771
0.4790	0.5885	0.5609	score:	-919.377261
0.4605	0.6622	0.6119	score:	-833.476261
0.4441	0.7262	0.6555	score:	-768.840638

```

0.4296 0.7828 0.6929 score: -719.177460
0.4167 0.8329 0.7250 score: -681.064629
0.4053 0.8768 0.7523 score: -652.599525
structure: 2

```

```

Pathology >>
<< __FamHist
Pat - Fam __Pathol __FamHis __CA125 score: -1031.611469
0.4667 0.4814 0.5910 0.5609 score: -910.118887
0.4356 0.4649 0.6676 0.6119 score: -823.190391
0.4065 0.4502 0.7348 0.6555 score: -757.449376
0.3789 0.4372 0.7952 0.6929 score: -706.585684
0.3527 0.4256 0.8502 0.7250 score: -667.287853
0.3277 0.4153 0.9003 0.7523 score: -638.075001
structure: 3

```

```

CA125 >>
<<
Pat - Fam __Pathol __FamHis __CA125 score: -1031.611469
0.4667 0.4814 0.5910 0.5609 score: -910.118887
0.4356 0.4649 0.6676 0.6119 score: -823.190391
0.4065 0.4502 0.7348 0.6555 score: -757.449376
0.3789 0.4372 0.7952 0.6929 score: -706.585684
0.3527 0.4256 0.8502 0.7250 score: -667.287853
0.3277 0.4153 0.9003 0.7523 score: -638.075001
structure: 4

```

```

CA125 >>
<< __FamHist
Pat - Fam __Pathol __FamHis CA1 - Fam __CA125 score: -1033.389677
0.4789 0.4889 0.5915 0.5089 0.5737 score: -905.226531
0.4575 0.4789 0.6685 0.5162 0.6379 score: -813.305382
0.4360 0.4699 0.7362 0.5220 0.6956 score: -743.622897
0.4142 0.4618 0.7971 0.5263 0.7485 score: -689.926078
0.3922 0.4544 0.8525 0.5293 0.7975 score: -649.377295
0.3698 0.4478 0.9031 0.5308 0.8434 score: -621.485227
structure: 5

```

```

CA125 >>
<< __Pathology
Pat - Fam __Pathol __FamHis CA1 - Pat __CA125 score: -1183.304381
0.4775 0.4864 0.5914 0.5692 0.5737 score: -1018.325904
0.4575 0.4740 0.6685 0.6276 0.6379 score: -898.340190
0.4393 0.4626 0.7364 0.6778 0.6956 score: -806.099459
0.4226 0.4522 0.7977 0.7211 0.7485 score: -733.570934
0.4071 0.4427 0.8539 0.7583 0.7975 score: -676.685326
0.3926 0.4339 0.9056 0.7892 0.8434 score: -633.817193
0.3788 0.4258 0.9532 0.8135 0.8867 score: -606.477342
structure: 6

```

```

CA125 >>
<< __FamHist __Pathology
Pat - Fam __Pathol __FamHis CA1 - Fam CA1 - Pat __CA125 score: -1187.088127
0.4720 0.4864 0.5912 0.5170 0.5663 0.5737 score: -1022.351920
0.4459 0.4739 0.6680 0.5317 0.6225 0.6379 score: -902.009628
0.4212 0.4626 0.7355 0.5440 0.6711 0.6956 score: -809.073422
0.3976 0.4521 0.7963 0.5541 0.7132 0.7485 score: -735.579622
0.3746 0.4426 0.8516 0.5619 0.7494 0.7975 score: -677.440749
0.3521 0.4338 0.9022 0.5673 0.7798 0.8434 score: -632.900127
0.3299 0.4256 0.9477 0.5705 0.8039 0.8867 score: -602.761741
structure: 7
structures (nodewise): 7

```

## 14.2. Szülői halmazok jósága

Az alábbi táblázat a 16 csomópontos háló tanulását mutatja, '[gyermek] >> << [szülők] [eredmény]' formátumban. A táblázat itt sem teljes, csak néhány fontos változónak a legjobb szülői halmazait ábrázoltuk, az azok jósági mutatói közti különbségek érzékeltesére.

[14.2] táblázat - szülői halmazokhoz tartozó BIC mutatók

Pathology	<< Meno Age	score -973.114452
Pathology	<< Meno Age Parity	score -973.114452
Pathology	<< Age	score -973.192569
Pathology	<< Age Parity	score -973.192569
Pathology	<<	score -973.46146
Pathology	<< Parity	score -973.46146
Pathology	<< Meno	score -973.726599
Pathology	<< Meno Parity	score -973.726599
Pathology	<< FamHist Age PillUse	score -986.391634
Pathology	<< FamHist Meno Age	score -986.45604
Locularity	<< Meno Pathology Papillation	score -967.828072
Locularity	<< Meno Pathology	score -968.224783
Locularity	<< Meno Parity Pathology	score -968.224783
Locularity	<< Pathology Papillation	score -969.082433
Locularity	<< Parity Pathology Papillation	score -969.082433
Locularity	<< Meno Age Pathology	score -969.172187
Locularity	<< Pathology	score -969.404724
Locularity	<< Parity Pathology	score -969.404724
Locularity	<< Age Pathology Papillation	score -969.685067
Locularity	<< Age Pathology	score -970.022993
Locularity	<< Age Parity Pathology	score -970.022993
Locularity	<< FamHist Papillation	score -970.99961
Locularity	<< FamHist Parity Papillation	score -970.99961
Locularity	<< FamHist PillUse Papillation	score -971.328024
Locularity	<< FamHist HormTherapy Papillation	score -971.371086
Locularity	<< Papillation	score -972.014785
Locularity	<< Parity Papillation	score -972.014785
Locularity	<< FamHist	score -972.170225
Locularity	<< FamHist Parity	score -972.170225
Locularity	<< FamHist PillUse	score -972.499469
Locularity	<< FamHist PillUse Parity	score -972.499469
Locularity	<< FamHist HormTherapy	score -972.542087
Locularity	<< FamHist Parity HormTherapy	score -972.542087
Locularity	<< FamHist PillUse HormTherapy	score -972.985471
Locularity	<<	score -973.114452
TAMX	<< Pathology Papillation	score -967.210499
TAMX	<< Parity Pathology Papillation	score -967.210499
TAMX	<< Pathology Papillation Echogenicity	score -967.210499
TAMX	<<	score -967.828072
TAMX	<< Parity	score -967.828072
TAMX	<< Echogenicity	score -967.828072
TAMX	<< Parity Echogenicity	score -967.828072
TAMX	<< Age Pathology Papillation	score -967.9669
TAMX	<< FamHist	score -968.029079
TAMX	<< FamHist Parity	score -968.029079
TAMX	<< FamHist Echogenicity	score -968.029079
TAMX	<< FamHist Parity Echogenicity	score -968.029079
TAMX	<< HormTherapy	score -968.097005
TAMX	<< Parity HormTherapy	score -968.097005
TAMX	<< HormTherapy Echogenicity	score -968.097005
TAMX	<< Parity HormTherapy Echogenicity	score -968.097005
TAMX	<< PillUse	score -968.100588
TAMX	<< PillUse Parity	score -968.100588

TAMX	<< PillUse Echogenicity	score -968.100588
TAMX	<< PillUse Parity Echogenicity	score -968.100588
TAMX	<< Meno Pathology Papillation	score -968.100729
TAMX	<< FamHist PillUse	score -968.345694
TAMX	<< FamHist PillUse Parity	score -968.345694
TAMX	<< FamHist PillUse Echogenicity	score -968.345694
TAMX	<< Pathology Papillation Locularity	score -968.38073
TAMX	<< FamHist HormTherapy	score -968.39079
TAMX	<< FamHist Parity HormTherapy	score -968.39079
TAMX	<< FamHist HormTherapy Echogenicity	score -968.39079
TAMX	<< Papillation Locularity	score -968.492236
TAMX	<< Parity Papillation Locularity	score -968.492236
TAMX	<< Papillation Locularity Echogenicity	score -968.492236
TAMX	<< PillUse HormTherapy	score -968.522304
TAMX	<< PillUse Parity HormTherapy	score -968.522304
TAMX	<< PillUse HormTherapy Echogenicity	score -968.522304
TAMX	<< Pathology	score -968.605909
TAMX	<< Parity Pathology	score -968.605909
TAMX	<< Pathology Echogenicity	score -968.605909
TAMX	<< Parity Pathology Echogenicity	score -968.605909
TAMX	<< FamHist PillUse HormTherapy	score -968.821647
TAMX	<< Age Pathology	score -969.362254
TAMX	<< Age Parity Pathology	score -969.362254
TAMX	<< Age Pathology Echogenicity	score -969.362254
TAMX	<< Age Papillation Locularity	score -969.485515
TAMX	<< Meno Pathology	score -969.496078
TAMX	<< Meno Parity Pathology	score -969.496078
TAMX	<< Meno Pathology Echogenicity	score -969.496078
TAMX	<< Pathology Locularity	score -969.73876
TAMX	<< Parity Pathology Locularity	score -969.73876
TAMX	<< Pathology Locularity Echogenicity	score -969.73876
TAMX	<< Meno Papillation Locularity	score -969.762537
TAMX	<< Age Pathology Locularity	score -970.606696
CA125	<< Pathology Ascites	score -924.470081
CA125	<< Parity Pathology Ascites	score -924.470081
CA125	<< Pathology Echogenicity Ascites	score -924.470081
CA125	<< Pathology Ascites Bilateral	score -924.470081
CA125	<< Meno Pathology Ascites	score -924.660641
CA125	<< Pathology Locularity Ascites	score -924.880268
CA125	<< Pathology ColScore Ascites	score -925.054801
CA125	<< Pathology Papillation Ascites	score -925.234923
CA125	<< Pathology TAMX Ascites	score -926.076369
CA125	<< Pathology	score -929.717835