

# Hálózati elemzések az üzleti életben

*Kovács Gyula – Sixtep Kft.*

INNOVATÍV BI KONFERENCIA  
*Az üzleti intelligencia élvonalán* **2011.11.22.**

# Hálózat kutatás rövid ismertetése

# Königsbergi hidak problémája

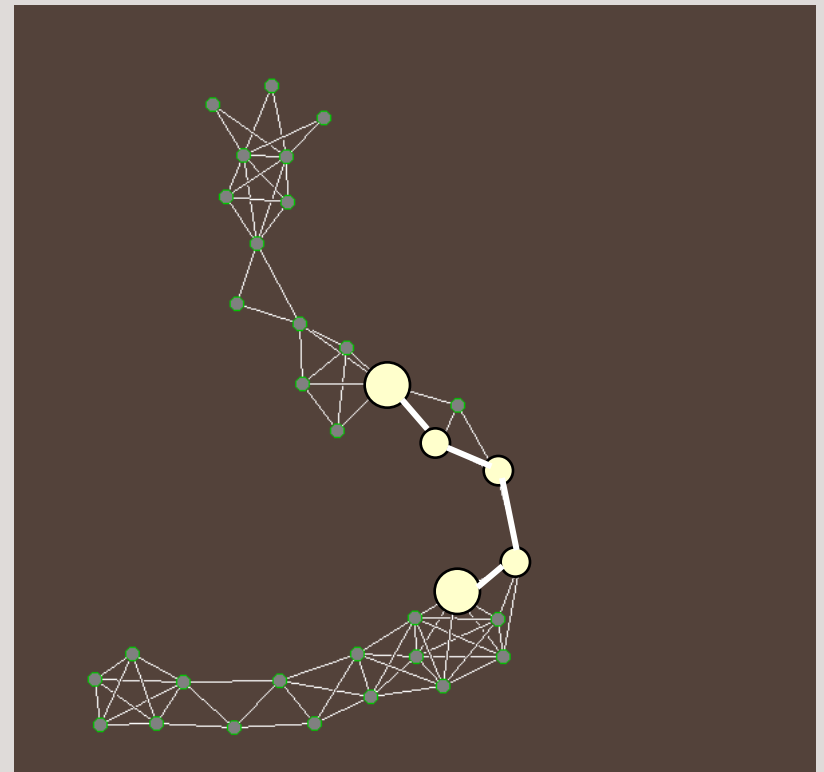
**Háttér:** A probléma története, hogy a poroszországi Königsberg (most Kalinyingrád, Oroszország) városban hét híd ívelt át a várost átszelő Prégel folyón úgy, hogy ezek a folyó két szigetét is érintették. A königsbergiek azzal a kérdéssel fordultak Eulerhez, vajon végig lehet-e menni az összes hídon úgy, hogy mindegyiken csak egyszer haladjanak át, és egyúttal visszaérjenek a kiindulópontba. 1736-ban



**Euler bizonyította, hogy ez lehetetlen. A Königsbergi hidak problémáját tekintjük az első gráfelméleti problémának (mai Königsbergben már lehetséges a bejárás).**

**Alapfogalmak:** A gráfelmélet alapfogalma a gráf, olyan struktúra, ami csúcsokból vagy szögpontokból és élekből áll, minden él két csúcs között fut. Legtöbbször egyszerű gráfokkal foglalkozunk, azaz olyanokkal, amelyekben nincs hurokél (egy csúcsot önmagával összekötő él) és nincsenek párhuzamos élek sem, tehát azonos pontok között haladó különböző élek.

- **Csúcsok**
- **Élek:** két csúcsot összekötő egyenes. Vannak irányított és irányítatlan élek, illetve vannak súlyozott élek
- **Utak:** Az út élek egymáshoz csatlakozó sorozata, amely egy csúcsot legfeljebb egyszer tartalmaz
- **Összefüggőség:** bármelyik két csúcs között van út



# Milyen hálózat modellezi legjobban a valódi világot?

- Véletlen gráfok (Erdős-Rényi modell, 1959): *Az Erdős–Rényi modell a gráfelméletben két rokon, véletlen gráfok előállítására szolgáló modell neve. Az egyik változata egyenlő valószínűséggel választ az összes adott élszámú gráf közül, a másikonál minden él egymástól függetlenül egy adott valószínűséggel van behúzva.*
- Kis világok (Barabási-Albert modell, 1999): *A Barabási–Albert-modell a komplex hálózatok (gráfok) fejlődésének egy modellje, mely magyarázattal szolgál azok gyakori skálafüggetlen tulajdonságára, azaz arra, hogy a fokszám eloszlásuk gyakran negatív kitevőjű hatványfüggvény szerint cseng le..*

## A kis világ gráf kinézete

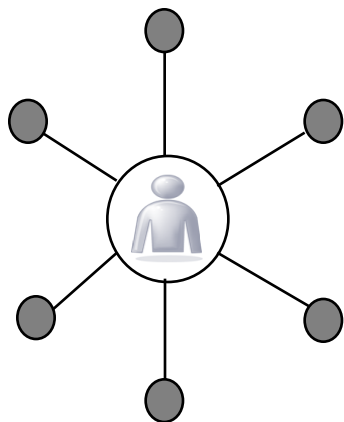
Egy  $n$  pontú  $G$  gráf kis világ, ha<sup>1</sup>

1. Ritka:  $e(G) \approx c_1 n$
2. Kicsi az átmérője:  $\text{diam}(G) \approx c_2 \log n$
3. A fokszámok eloszlása hatványtörvényt követ:  $k$ -fokú pontjainak a száma  $= c_3 k^{-\alpha}$ , valamely  $\alpha > 0$ -ra.

# Milyen szintjei vannak a közösségi hálózatok elemzésének?

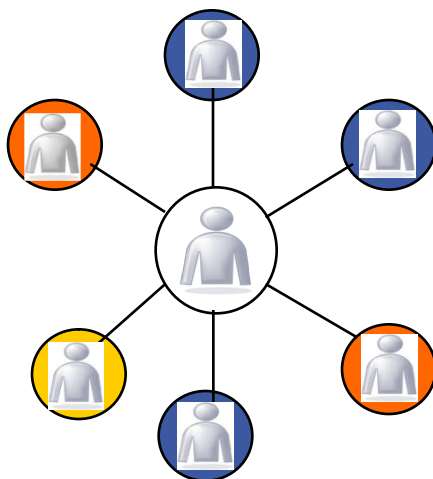
## Csúcs elemzése

- Számoljuk meg kapcsolatainak számát
- Rangsoroljuk a „legjobb” kapcsolatát



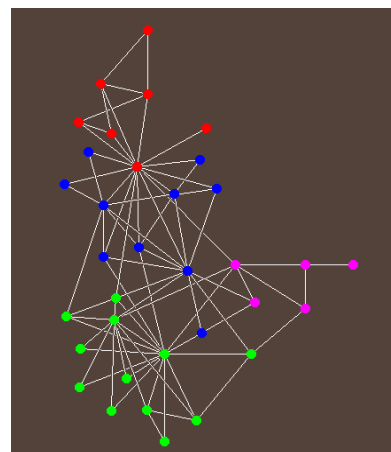
## Kapcsolat elemzése

- Jellemezd a kapcsolatokat
- Mindenkit a kapcsolatain keresztül írd le
- Keresd a szociális „burkot”



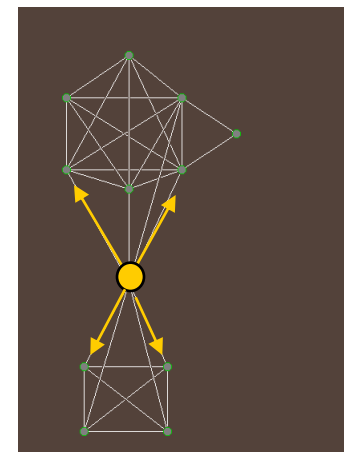
## Közösségek elemzése

- Azonosítsd be a közösségeket és klasztereket
- Mindenkire találd meg a saját közösségeit



## Közösségek szerep elemzése

- Azonosítsd be a véleményformálókat
- Elemezd a véleményformálók befolyását a környezetükre

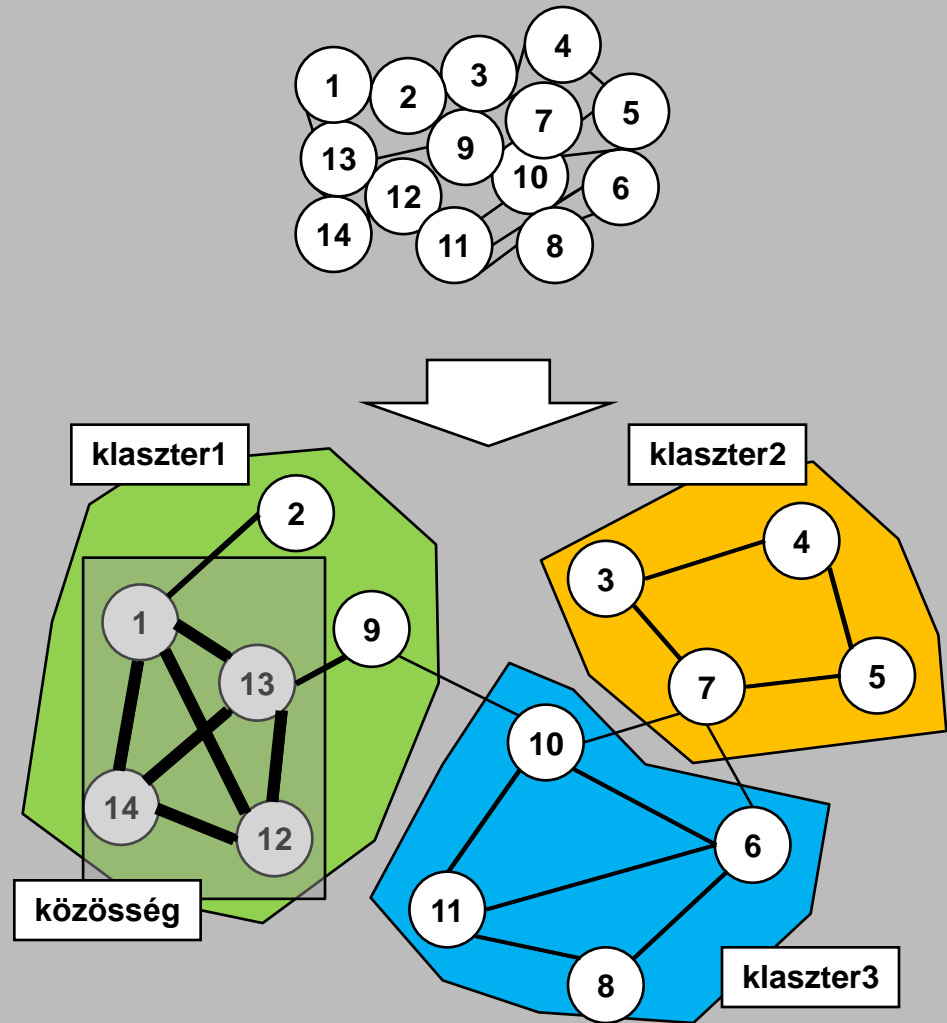


# Milyen kérdéseket kell megválaszolnia a hálózatkutatásnak az üzleti életben?

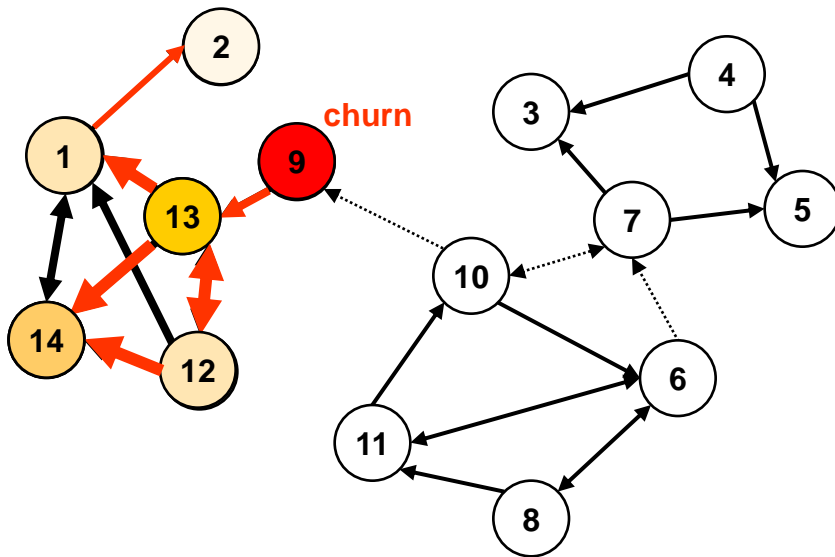
**1. Készíts térképet! –  
Hogyan lehet ábrázolni a  
gráfokat? (emeld ki a gráf  
struktúráját)**

**2. Színezd ki a térképet! –  
Hol vannak a gráf erősen  
összefüggő részgráfjai  
(közösségek, klaszterek)**

**3. Mérd a forgalmat! –  
Hogyan terjed az információ  
(mém) a hálózaton belül? (ki  
kit fertőz meg milyen  
mértékben)**



## Példa



*A fenti ügyfél elvándorlása szignifikánsan növeli 5 másik ügyfél elvándorlási valószínűségét*

- A (súlyozott) gráfokon belül fertőzési szimulációk végezhetők el
- Egy szimulációs futása végén minden csúcs kap egy fertőzési valószínűséget („mekkora valószínűséggel fertőződik meg?”)
- A kapott fertőzési valószínűség valójában egy prediktív score – maga a szimuláció prediktív modellezésnek tekinthető



**Milyen hálózatok  
keletkeznek az üzleti  
életben?**

- **Kapcsolati hálózatok:** *A csúcsok közötti élek valamilyen tranzakciót vagy kapcsolatot reprezentálnak. Ez lehet infokommunikációs szektorban hívás vagy más egyéb kommunikáció, míg pénzügyi szektorban banki átutalás. Céginformációs adatbázisban két cég közötti él tulajdonosi kapcsolatot reprezentálhat. A kapcsolati hálózatban az élek irányítottak!*
- **Attribútum alapú hálózatok:** *A csúcsokat azért kötjük össze, mert valamelyik attribútumuk megegyezik. Biztosítási szektorban két káreseményt összekötünk ha van közös szereplője, vagy céginformációs adatbázisban két céget összekötünk ha közös a címük. Az attribútum alapú hálózatok irányítatlan éleket tartalmaznak (egyenrangú kapcsolatok)*
- **Vegyes hálózatok:** *Azon hálózatok, amikor az élek reprezentálhatnak tranzakciót (irányított) és reprezentálhatnak közös attribútumot (irányítatlan). Egy ilyen tipikus vegyes hálózat a cégek közötti utalások és céginformációs kapcsolatok együttes reprezentációja.*

# Tranzakciós hálózat – egy adott bank ügyfelei közötti utalások

SIXTEP - Network Software v2.1

STEPS: LOADING 1 BUILDING 2 APPLICATIONS 3

Used Memory: 402MB  
Free Memory: 962MB  
Total Memory: 1366MB  
Max Memory: 1488MB

New Project  
Open Project  
Save Project  
Add graph  
Attach  
Filter  
Settings  
Exit

Added graph:  
1. C:\matroid\OFFICE\VI\SIXTEP\4\_projekte\KOMPLETEX\VI\_FULL\5\_DEMO\1\graf.csv

Added attributes:  
1. [C:\matroid\OFFICE\VI\SIXTEP\4\_projekte\KOMPLETEX\VI\_FULL\5\_DEMO\1\eginfo.atr]

Vertex sets	Vertices/Edges	Description
Graph (1)		
1970.01.01 - CSV graph	1852/1115	built from: C
Clusters (738) - Modularity		
15502003	4/3	cluster
11321561	3/2	cluster
11246615	2/1	cluster
10849844	12/11	cluster
28431569	2/1	cluster
19014812	5/4	cluster
14489765	3/2	cluster
10964033	2/1	cluster
11525150	2/1	cluster
13237912	2/1	cluster
12675746	3/2	cluster
10389395	3/2	cluster
13341194	2/1	cluster
11376190	2/1	cluster
13898610	2/1	cluster
12794939	2/1	cluster
14094510	2/1	cluster
13739744	2/1	cluster
12155169	4/3	cluster
10550180	3/2	cluster
11563408	2/1	cluster
12910607	2/1	cluster
15350727	3/2	cluster
14414183	2/1	cluster
10569016	2/1	cluster

Graph - subgraph

Mouse mode Rendering

Adott bank ügyfele

Utalás iránya

Másik bank ügyfele

Clusters created (Modularity maximization heuristic). Calculation time: 00:02:574(mm:ss:SSS)

# Attribútum alapú hálózat – káreseteket akkor kötünk össze, amikor van közös szereplő/autó/bankszámlaszám

SIXTEP - Network Software v2.1

STEPS

LOADING 1 BUILDING 2 APPLICATIONS 3

Used Memory: 78MB  
Free Memory: 1175MB  
Total Memory: 1254MB  
Max Memory: 1488MB

New Project  
Open Project  
Save Project  
Add graph  
Attach  
Filter  
Settings  
Exit

Added graph: 1. 0  
Added attributes: 1.

Vertex sets	Vertices/Edges	Descri
569721	10/13	compc
449810	10/19	compc
424105	10/33	compc
435230	10/45	compc
516522	11/55	compc
522109	11/15	compc
428297	11/55	compc
442692	11/55	compc
432134	12/66	compc
547633	12/64	compc
441602	13/78	compc
451829	13/78	compc
495495	15/105	compc
521331	15/44	compc
462104	15/105	compc
434561	15/87	compc
448504	16/96	compc
432404	16/68	compc
424108	16/105	compc
450677	17/121	compc
428853	17/96	compc
436947	18/32	compc
453699	18/153	compc
508068	19/154	compc
449968	19/135	compc
436990	20/190	compc
434913	21/175	compc
486426	21/90	compc

Graph

Graph - subgraph

Mouse mode Rendering Insurer

Közös szereplő

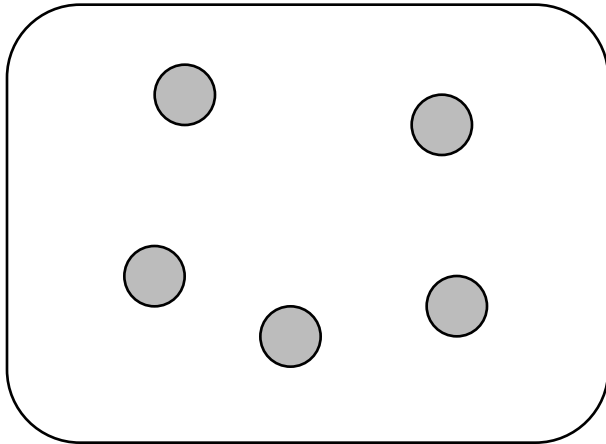
Minden csúcs egy káreset

Szinezés fraud score alapján

Graph loaded

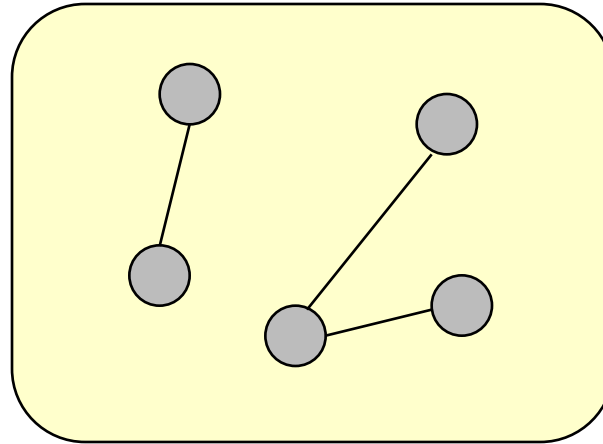
# Üzleti életben az igazi kihívás a hálózatok építése

**Reprezentativitás:** A hálózatkutatás eredményei jól alkalmazhatók az üzleti életben keletkező hálózatokon. Az igazi kihívás az adott üzleti probléma minél jobb reprezentációja hálózatok segítségével: (i) mikor húzunk be élt két csúcstól?, (ii) a behúzott él mennyire erős kapcsolatot reprezentál?



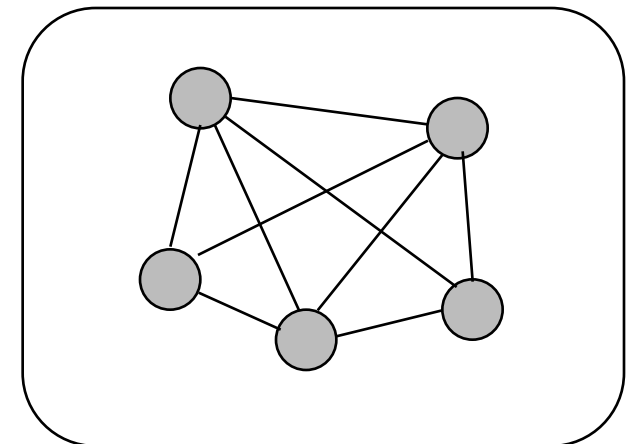
**GRÁF1**

**A csúcsok között nincs kapcsolat – nincs információ tartalom**



**GRÁF2**

**A csúcsok között vannak élek – jelentős információ tartalom**



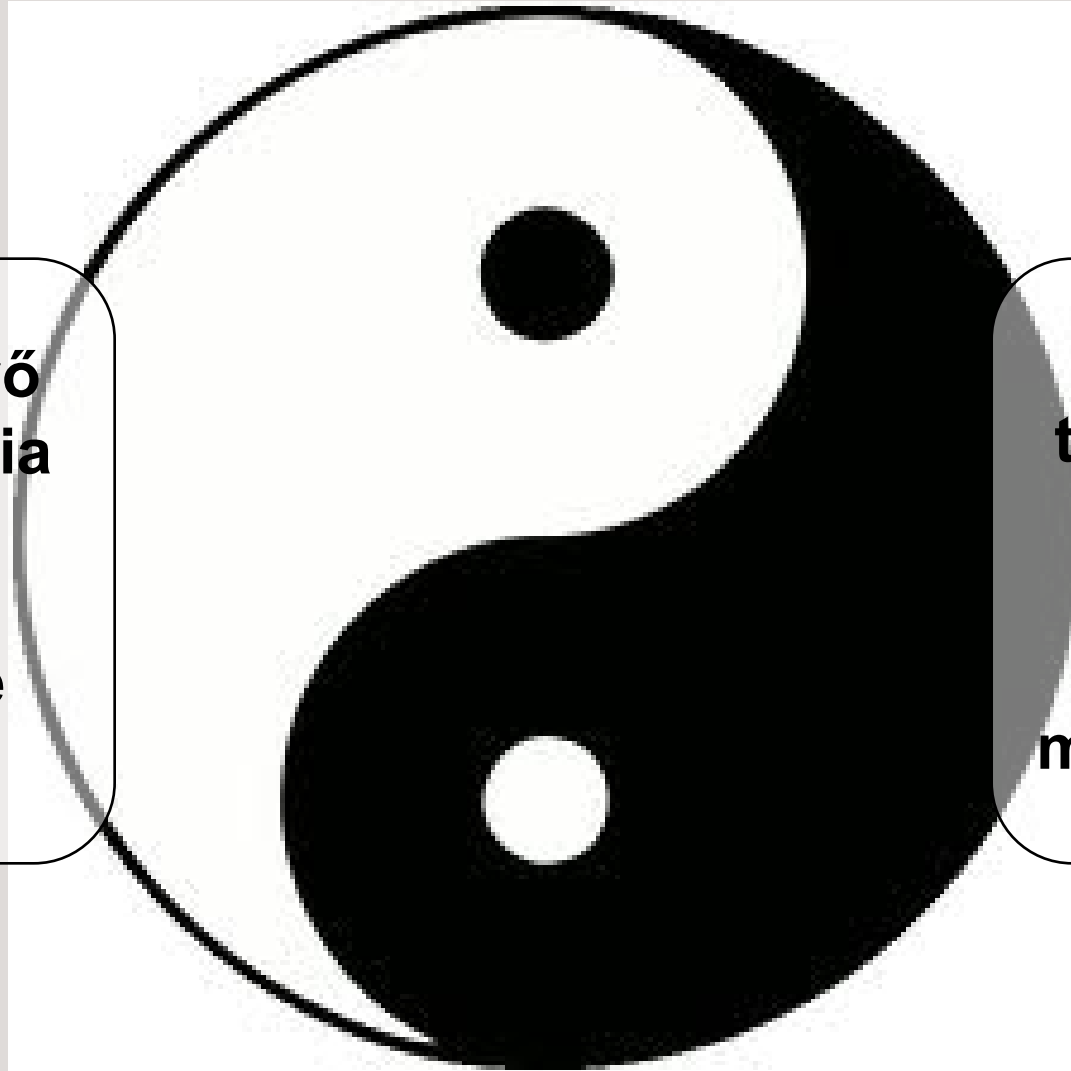
**GRÁF3**

**Minden csúcstól van él – nincs információ tartalom**

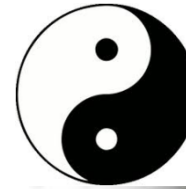
**Miért most?**

# Akkor miért is van szükség a hálózatkutatásra?

**A meglévő  
technológia  
nem ad  
minden  
kérdésre  
választ**



**Új  
technológia  
alkalmas a  
felmerült  
problémák  
megoldására**



- **Megjelent az ügyfélközpontú gondolkodás a vállalatokon belül (CRM)**
- **Adattárházak építésénél analitikai célok érvényesülnek**
- **Modellek integrációja üzleti folyamatokba**
  - ***CRM, kampány menedzsment***
  - ***Ajánló rendszerek***
- **Adatbányászati szoftverek egyre szélesebb körű használata**



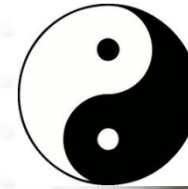


## Milyen kihívásokkal néz szembe az adatbányászat területe?

- **Performancia:** *Bizonyos üzleti problémák esetében a hagyományos adatbányászati modellek teljesítménye elmaradt attól a szinttől, hogy rentábilis legyen*
- **Rugalmasság:** *Az adatbányászati modellek integrációja nem megoldott, de integráció esetén is nehézkes a modellek frissítése. Ez többnyire a CRISP/SEMMA örökség része*
- **Nem ad választ a miértekre:** *Az adatbányászat egyik fő ígérete hogy választ kapunk a „miértekre”. A tapasztalatok azt mutatják, hogy az adatbányászati modellek fekete dobozként funkcionálnak, nem alkalmasak az ügyfél motivációk feltárására,.*



- **Adatbővítés:** *Nem strukturált adatokat feldolgozó eljárások kifejlesztése (szövegbányászat, hangbányászat, képfeldolozás), meglévő adatkörök bővítése*
- **Új összefüggések feltárása:** *ügyfél központú szemlélet „csoport szintű szemléletre” történő felváltása (közösségi hálózatok megjelenése ebben nagy segítséget nyújt). Közösségek, vélemény-formálók vizsgálata és információ terjedés mérése.*
- **Real-time technológia:** *Egyre fejlettebb informatikai háttér biztosítja az adatok minél gyorsabb elemezhetőségét. Egyre fontosabb az ügyfél reakciók beépítése a modellekbe.*



- **Adatok:** *a hálózati elemzéshez szükséges adatok rendelkezésre állnak, az elmúlt években kidolgozásra kerültek azon módszerek, melyek ezekből képesek hálózatok előállítására*
- **Hálózati szoftverek:** *egyre szélesebb kínálat hálózati szoftverekből (vizualizációs szoftverek -> komoly analitikai megoldások)*
- **Fogadókészség:** *a pénzügyi és infokommunikációs szektorban az üzleti felhasználói oldal megismerkedett a technológiával, igény van a technológia adaptációjára*
- **Nyomás:** *az elhúzódó üzleti válságban csak a rentábilis megoldásokra van üzleti igény*

**Köszönöm a figyelmet!**