

# KÖZÖSSÉGI HÁLÓZATOK PRIVÁTSZFÉRÁT ÉRINTŐ KÉRDÉSEI

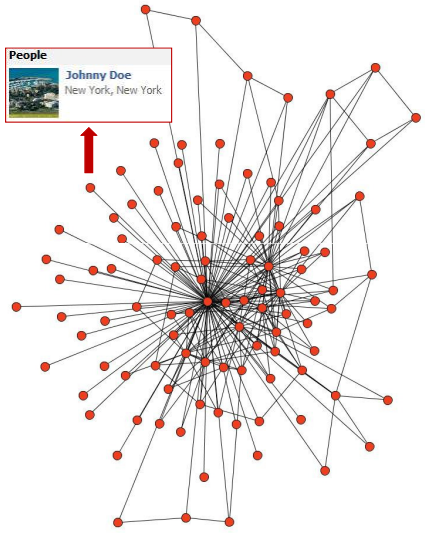
*Privátszférát Erősítő Technológiák*

Gulyás Gábor György
Óraadó
BME Híradástechnikai Tanszék
gulyasg@hit.bme.hu

2011. május 3.,
Budapest

## Bevezető

- Mi az a közösségi háló?
  - Egy szolgáltatás mögötti gráf jellegű struktúra
  - Csomópontok = felhasználók
  - Élek = kapcsolatok, interakció
- Mik jellemzik ezeket?
  - Önkéntes adatszolgáltatás
  - Érzékeny (és személyes) adatok, meta-információk megosztása






## Lehet-e általánosítani?

- Mi tartozhat még ide? Például:
  - Azonnali üzenetküldés (partnerlista, konferenciák)
  - Csevegő szolgáltatások (szobák látogatottságai)
  - Torrent hálózatok (letöltési kapcsolatok)
  - Mobil telefónia (híváslista, telefonkönyv, sms küldés)
  - E-mail kommunikáció (küldés, fogadás)
- Bármilyen szolgáltatás, ami alapjául egy kapcsolati hálózatot tudunk találni.
  - Kapcsolat: azonosítás, leíró információk (profil), interakcióra lehetőség nyílik



## Szereplők és lehetőségeik, céljaik

| | Lehetőségek | Célok |
|---|---|---|
| **Felhasználó** | (Szolgáltatásfüggő, de általában nem sok) | ▪ Adatok feletti rendelkezés<br>▪ Hozzáférés szabályozás |
| **Szolgáltató** | ▪ Mindent lát, bármire képes<br>▪ Teljes adatbázis, hálózatstruktúra | ▪Érdekeik az adatok exportálásához, eladásához vezethetnek:<br>○ Anyagi haszon előteremtése<br>○ Együttműködés harmadik féllel<br>○ Szolgáltatás üzleti értékének növelése, márkanév építése |
| **Többi felhasználó** | ▪ Nyilvános kapcsolatok listázása<br>▪ Publikus események megfigyelése<br>▪ Új regisztrációk készítése<br>▪ Új kapcsolatok létesítése | ▪ Állapot, tevékenység kompromittálása<br>▪ Profilok közötti kapcsolat megmutatása<br>▪ Egyéni adatgyűjtő akciók |
| **Harmadik fél** | ▪ Publikus adatok, mint kiegészítő források<br>▪ Anonimizált adatbázis exportok<br>▪ Kis mértékű beavatkozások (ld. többi felhasználó) | ▪ Profilozás<br>▪ Megfigyelés<br>▪ Személyre szabott szolgáltatások:<br>○ Hirdetések<br>○ Dinamikus árak |

# AZ ADATAINK MI VAGYUNK?

(Kép forrása: Szabad adatok, védett adatok 2 borító)



# Mi van egy profilon?

- Személynév vagy egyéb megnevezés
- Becenevek, felhasználónevek — Többé-kevésbé egyediek
- Elérhetőségek
  - E-mail, telefon, …
  - IM: MSN, Skype, …
  - SN: Facebook, Twitter, … — Ezek mind egyediek!
- Egyéb jellemzők
  - Születésnap
  - Munkahelyek — Önmagukban nem egyediek!
  - …

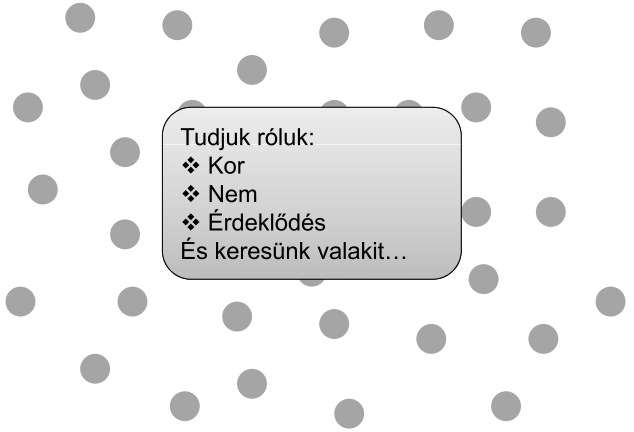


# Mi van egy profilon? (2)

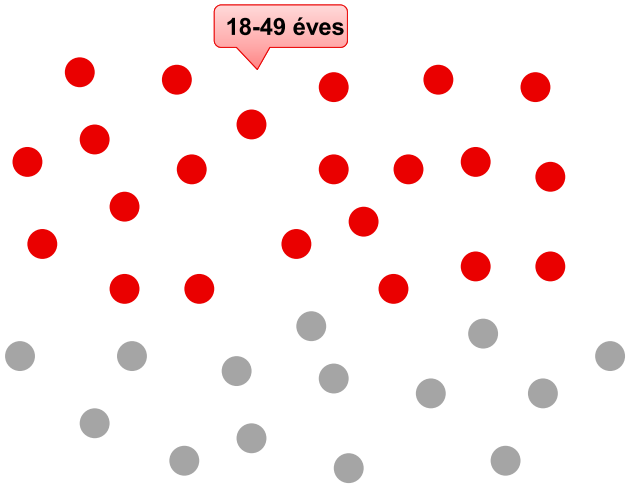
- Nem szöveges, multimédiás tartalmak
  - Képek
  - Zenék
  - Videók — Nehezen feldolgozhatóak!
- Meta-információk
  - Kapcsolatok
  - Csoport tagságok — Kicsit bonyolult a helyzet, de kezelhető
  - …



# Mik azok az anonimitási halmazok?

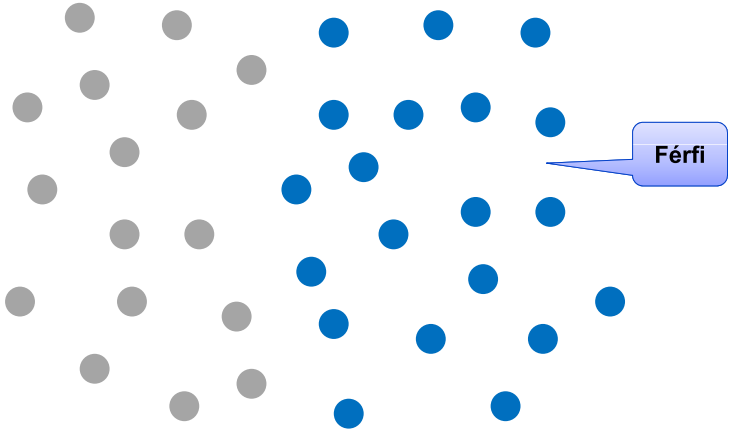


# Mik azok az anonimitási halmazok? (2)



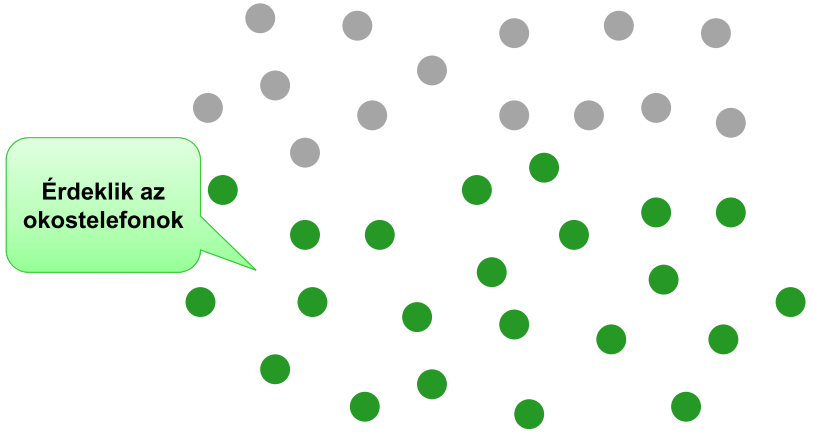



# Mik azok az anonimitási halmazok? (3)



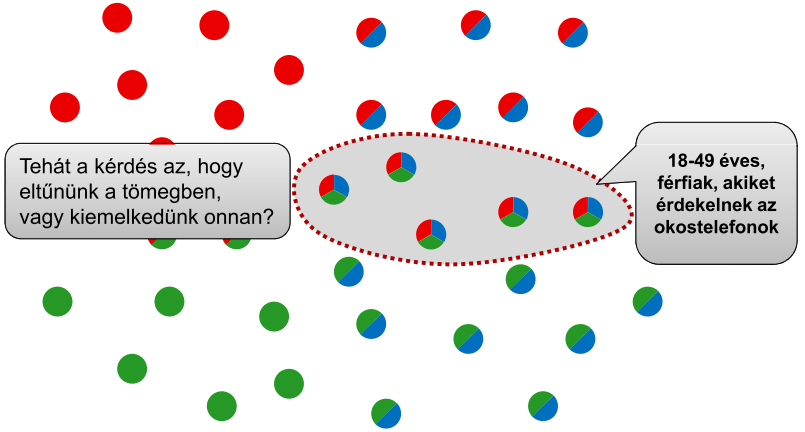



# Mik azok az anonimitási halmazok? (4)

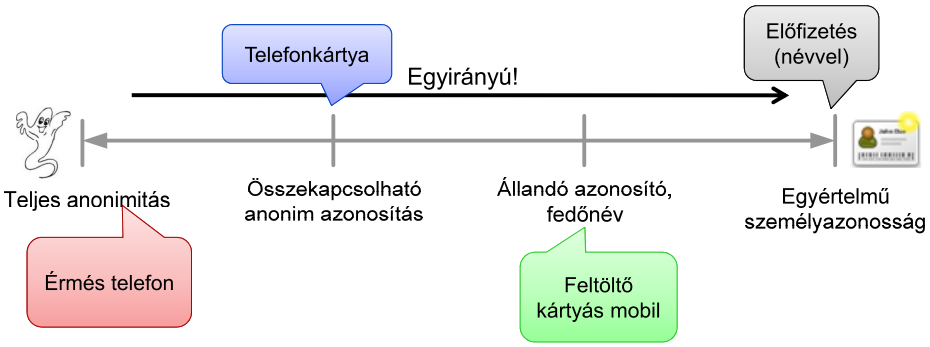





# Mik azok az anonimitási halmazok? (5)



# Azonosítási szintek



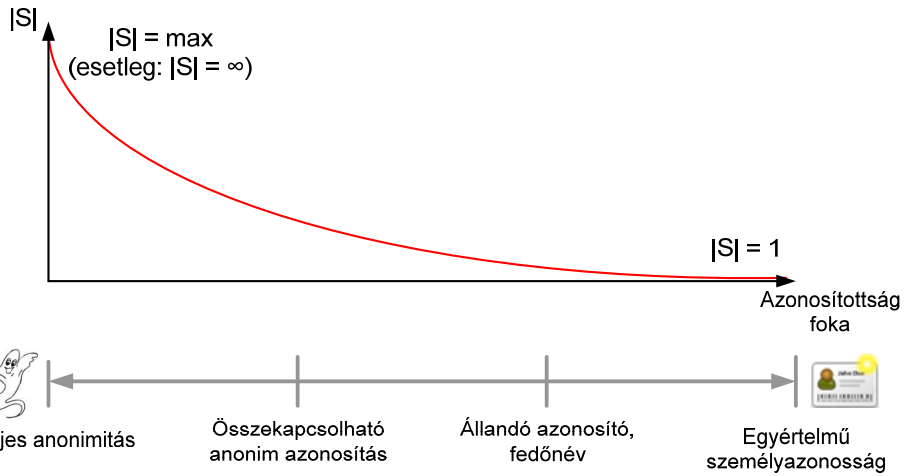

(Forrás: Goldberg, 2000)



# MIT ÁRUL EL RÓLUNK A PROFILUNK?

# Azonosíthatóság és anonimitási halmazok

- S: anonimitási halmaz






# Egyedi azonosítók

- Személynevek
  - Általában egyediek, de ha nem, a többivel könnyen kombinálható…
- Elérhetőségek
  - Használjunk eltérő elérhetőségeket
  - Második telefonszám, eldobható e-mail cím, stb.
- Más is elárulhat minket
  - Pl. fényképeink az online albumokban

# Felhasználónevek, becenevek

- Felhasználónevek, becenevek
  - Többé-kevésbé egyediek, de sokszor elérhetőek
  - Két adatbázisban a leghatékonyabb kapocs

**Meglepő statisztika:**
A felhasználónevek 30%-a független csak a személynévtől!
(10 millió Google és eBay felhasználónév alapján.)

- Mekkora valószínűséggel van két név mögött ugyanaz a személy?

(Forrás: Perito et al., 2011)

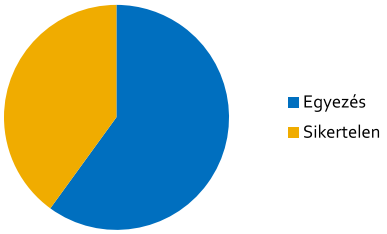


# Felhasználónevek, becenevek (2)

- Módszer
  - Valószínűségi (Markov-lánc alapú) modell
  - Nyelvi jellemzők figyelembe vételével

`gulyas.gabor` vs. `guly.asgabor`

- Ellenőrzés
  - Egy 20 000 valós felhasználónevet tartalmazó adatbázison






# Egyéb jellemzők

- Több jellemző figyelembe vételével az anonimitási halmaz mérete jelentősen csökken

`{irányítószám, nem, születési dátum}`

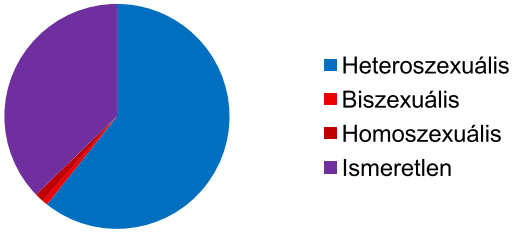
- Egy 1990-es felmérés szerint az USA állampolgárainak 87%-át egyedileg azonosította*

(* Forrás: Sweeney, 2000)



# Egyéb jellemzők (2)

- Nem kell nekünk feltenni: ismerőseinknek elég!
- A jellemzők megtippelhetőek*
  - MIT Facebook tagjait vizsgálták (kb. 2500 fő)




  - 10 főről más forrásból tudták, de nem vallotta be homoszexuális beállítottságát
  - Az ismerőseik attribútumai szerint mind sikerült megtalálni

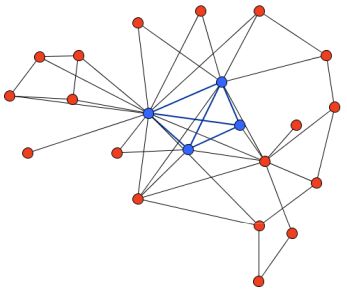
(* Forrás: Jernigan & Mistree, 2009)

# Mi a megoldás?

- Felhasználói tudatosság
  - Minden egyre inkább publikussá és valós időben kereshetővé válik
- Vannak titkosításra alkalmas programok*
  - A titkosítás értelmezhetetlen szövegbe rejti az információt (Van, ahol ezért törlés jár, pl. Facebook.)
  - Pl.


(* Forrás: Paulik et al., 2010)





**A KAPCSOLATRENDSZERÜNKBEN IS BENNE VAGYUNK?**

# Mi a megoldás? (2)

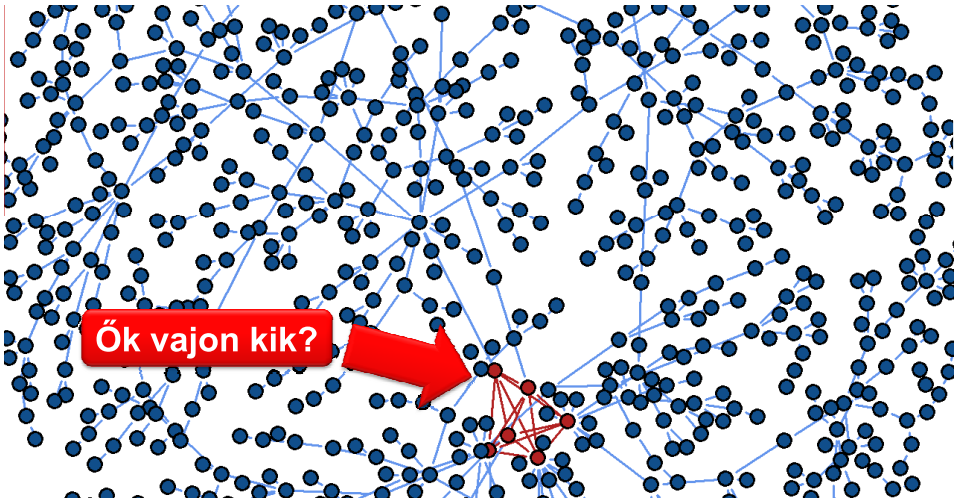
- Használhatunk szteganográfiát is*
  - Hozzunk létre egy értelmesnek tűnő álprofilt


  - Majd kezdjünk el alternatív profilokat gyártani: http://stegoweb.pet-portal.eu

(* Forrás: Besenyei et al., 2011)



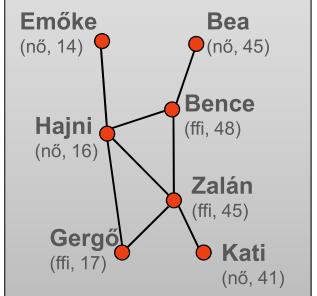
# Információ a kapcsolatrendszerben



# Az anonimizálás folyamata

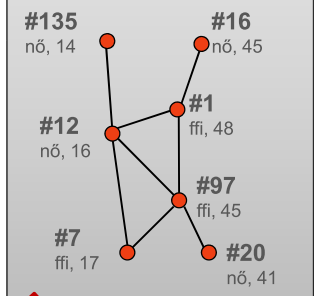
**Forrás: eredeti hálózat**

❖ Eredeti struktúra
❖ Eredeti profilok: nevek, jellemzők.



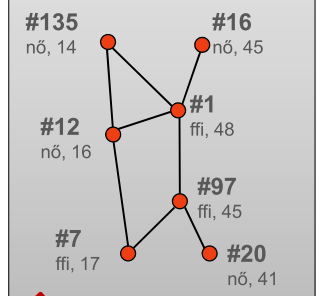

**Anonimizált hálózat**

❖ Eredeti struktúra
❖ Profilok álnéven, eredeti jellemzők.




**Perturbáció**

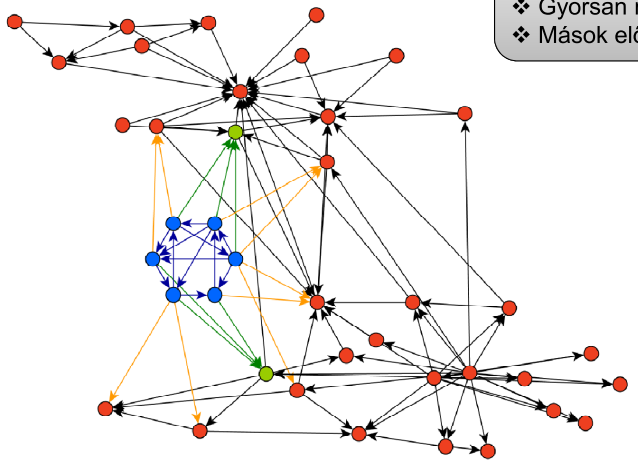
❖ Módosított struktúra
❖ Profilok álnéven, eredeti jellemzők. (néha ez is módosul)






# Aktív támadás: új struktúra beszúrása

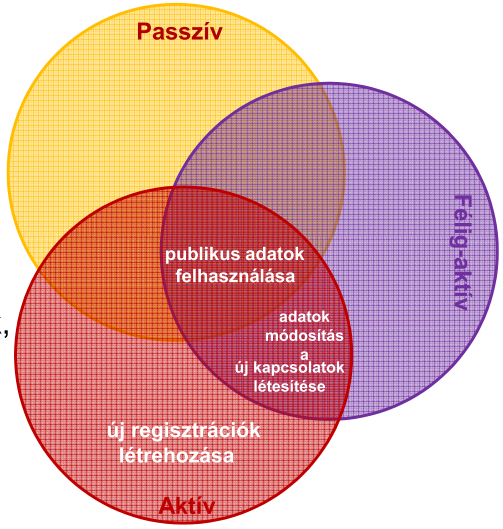
**Követelmények**:
❖ Egyedi struktúra
❖ Gyorsan megtalálható
❖ Mások előtt „rejtett”



(Forrás: Backstrom et al., 2007)

# Hogyan keressük meg őket?

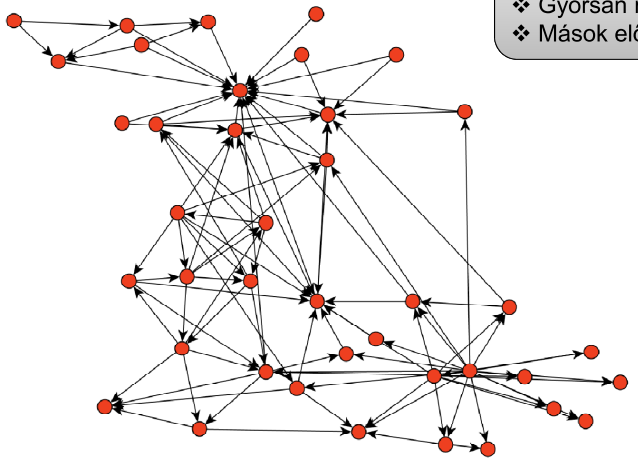
- Egy anonimizált, de többlet-információt tartalmazó hálóban keresünk
  - többlet információ: rejtett kapcsolatok, jellemzők, stb.
  - hozzáférhetnek kutatók, üzleti partnerek, stb.
- A cél 1 vagy több felhasználó újraazonosítása



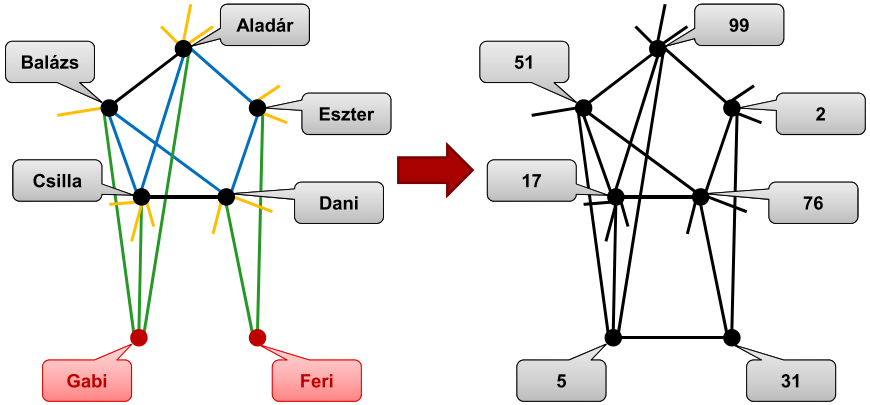



# Aktív támadás: új struktúra beszúrása (2)

**Követelmények**:
❖ Egyedi struktúra
❖ Gyorsan megtalálható
❖ Mások előtt „rejtett”

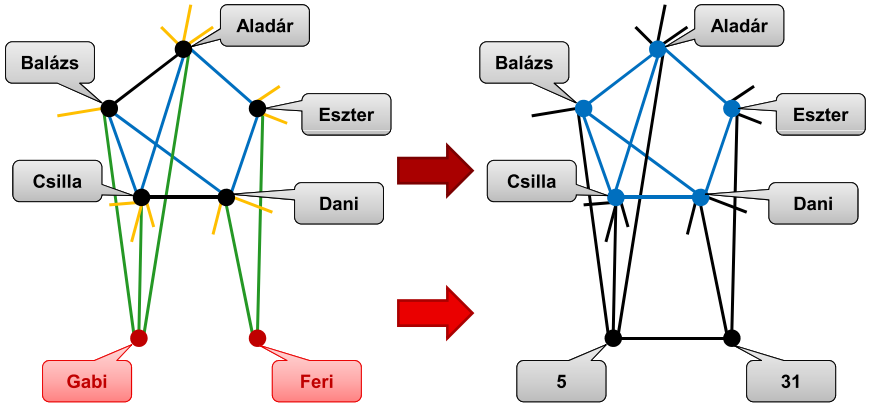


# Félig-aktív és aktív támadás






# Félig-aktív és aktív támadás (2)

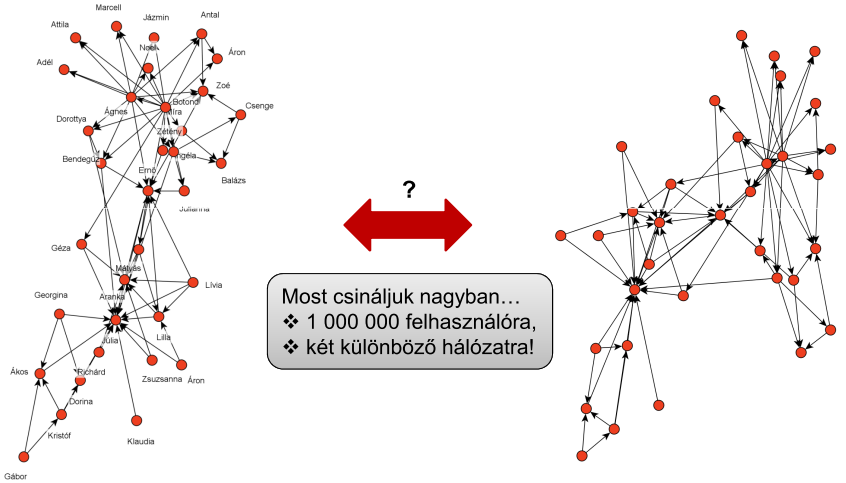





# Félig-aktív és aktív támadás (3)

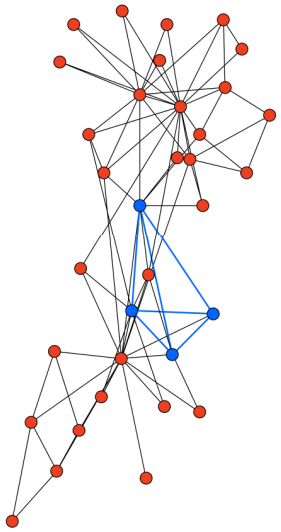



# Passzív támadás



# Passzív támadás (2)

- Alapötlet:
  - Egy másik közösségi hálózat publikus adatainak, mint kiegészítő információforrásnak alkalmazása
  - Csak a strukturális információk figyelembe vétele
    - Fokszám
    - Közös szomszédok száma
    - $f_{err}$=1±ε hibafaktor
- Az algoritmus fázisai:
  1. Kiindulási mag keresése: 4-es klikkek újra azonosítása
  2. Terjedési fázis: a beazonosított klikktől az azonosítás folytatása

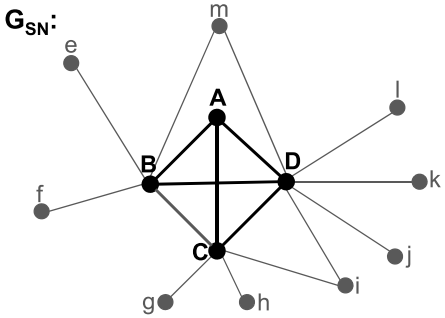


(Forrás: Narayanan & Shmatikov, 2009)

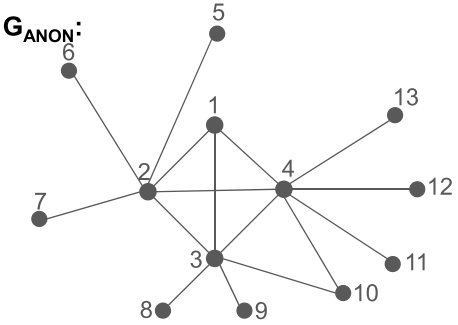


# Kiindulási mag keresése

$G_{SN}$:




$G_{ANON}$:




**Csomópontok fokszáma:**

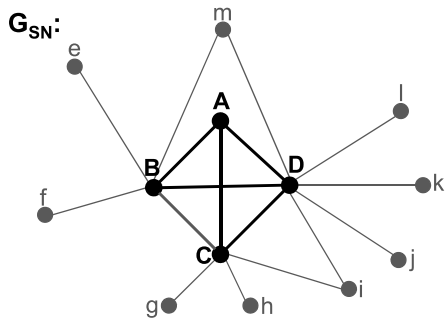
| $v_i$ | $d(v_i)$ | $v_j$ | $d(v_j)$ | $d_{err}$ |
|---|---|---|---|---|
| A | 3 | 1 | 3 | 1 |
| B | 6 | 2 | 6 | 1 |
| C | 6 | 3 | 6 | 1 |
| D | 8 | 4 | 7 | 0,875 |

$0{,}65 < d_{err}(D,4)=0{,}875 < 1{,}35$ ✓ (ε=35%)

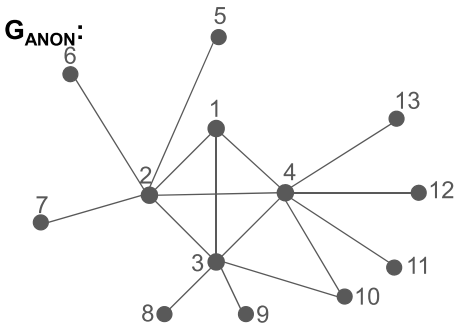


# Kiindulási mag keresése (2)

$G_{SN}$:




$G_{ANON}$:




**Közös szomszédok:**

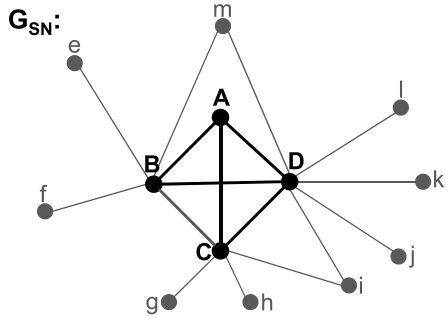
| $G_{SN}$ | $N(v_i,v_j)$ | $G_{ANON}$ | $N(v_k,v_l)$ | $d_{err}$ |
|---|---|---|---|---|
| A,B | 2 | 1,2 | 2 | 1 |
| A,C | 2 | 1,3 | 2 | 1 |
| A,D | 2 | 1,4 | 2 | 1 |
| B,C | 2 | 2,3 | 2 | 1 |
| B,D | 3 | 2,4 | 2 | 0,667 |
| C,D | 3 | 3,4 | 3 | 1 |

$0{,}65 < d_{err}(\{B,D\},\{2,4\}) = 0{,}667 < 1{,}35$ ✓ (ε=35%)

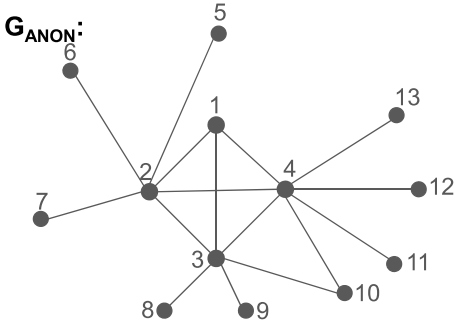


# Kiindulási mag keresése (5)

$G_{SN}$:




$G_{ANON}$:



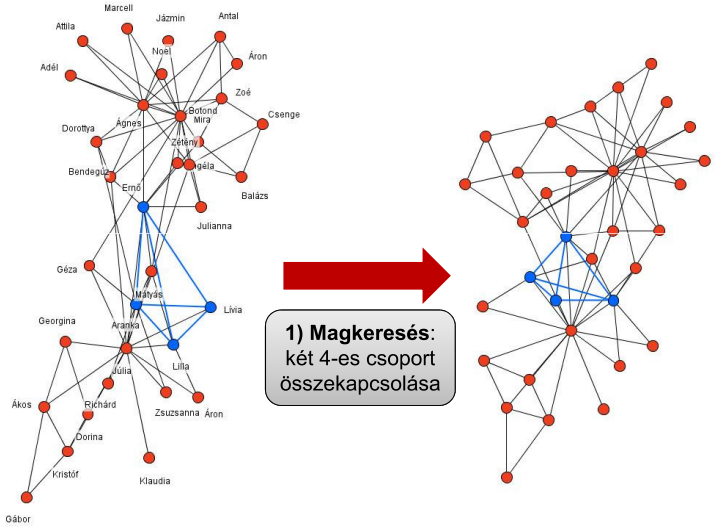

- A magot ε=35% mellett sikerült megtalálni ($f_{err}$=1±ε)
- A gyakorlatban
  - A gráfok és így az algoritmus is robosztusabbak (nagyobb fokszám, több közös szomszéd)
  - Meg is jelenhetnek élek (anonimizálás, más kapcsolatrendszer)
  - Sokkal kisebb ε is elég
  - A túl nagy ε miatt sok téves találat lehet

# Terjedési fázis

- Rekurzív algoritmus
  - Az újracímkézett halmazt bővíti
  - Eleinte ez a mag
- Lépései a $G_{SN}=(V_{SN}, E_{SN})$, $G_{ANON}=(V_{ANON}, E_{ANON})$ gráfokon
  1. Válasszunk egy $v_i \in V_{SN}$ csomópontot, aki nincs újracímkézve, de vannak már beazonosított kapcsolatai: $\{v_1,\ldots,v_k\} \in V_{SN}$
     a. Válasszuk ki a párjaikat: $\{v_{1'},\ldots,v_{k'}\} \in V_{ANON}$
     b. Vegyük az azonosítatlan szomszédjaikat: $\{v_{u1},\ldots,v_{um}\} \in V_{ANON}$
     c. Minden csomópont kapjon annyi pontot, ahány azonosított szomszédja van: $\{s_{u1},\ldots,s_{um}\}$
  2. Ha van szignifikáns eredmény $\{s_{u1},\ldots,s_{um}\}$-ben, akkor válasszuk a legmagasabb pontszámút $v_i$ párjának

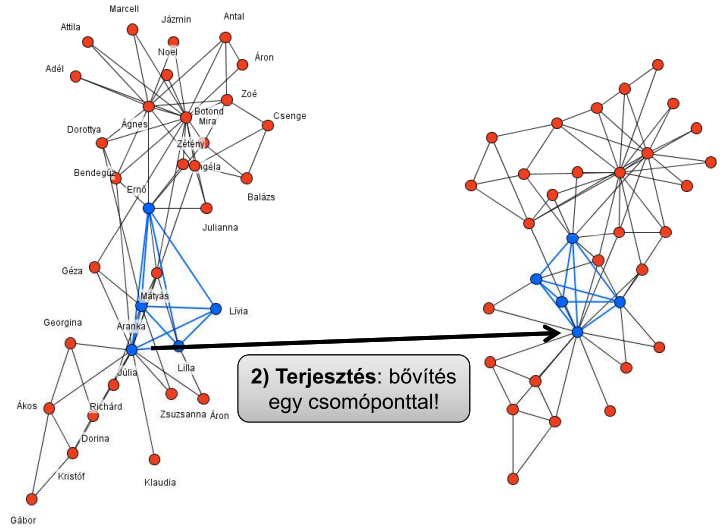


# Passzív támadás



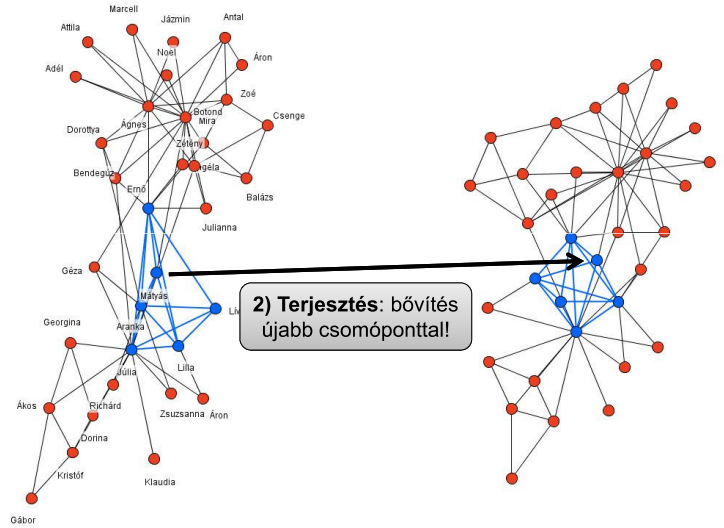



# Passzív támadás (2)



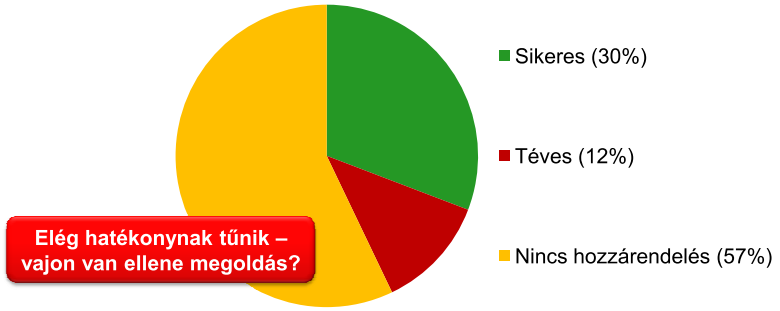



# Passzív támadás (3)



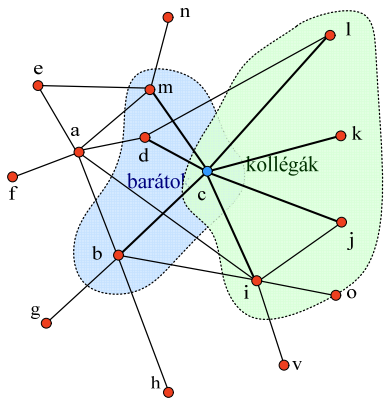
# Passzív támadás hatékonysága

- Két valódi hálózat között kerestek átfedést
  - Forrás: Flickr (3,3m felh., 53m kapcsolat)
  - Cél: Twitter (224e felh., 8,5m kapcsolat)
- Eredmények:

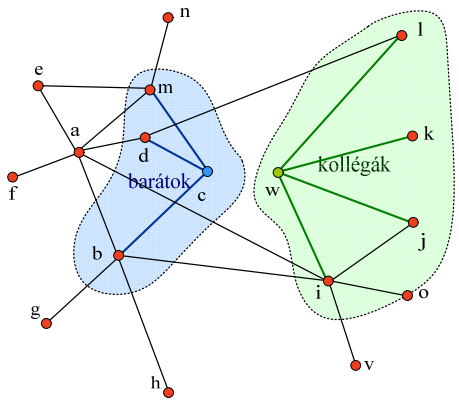





# Mi az identitás szeparáció?




Közösségi hálózat




Közösségi hálózat identitás szeparáció támogatással



# IDENTITÁSMENEDZSMENT: A JÖVŐ MEGOLDÁSA?



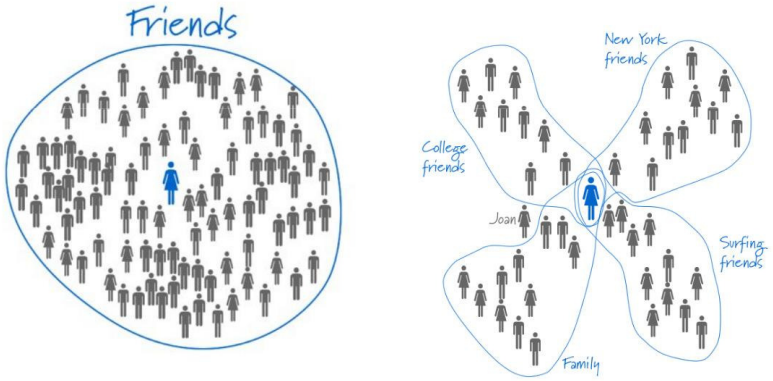
# Alapkoncepció

- Szerep alapú identitásmenedzsment
  - Más helyzetekben más információ számít
  - Köthetőek: partnerhez, szerepkörhöz
  - Időtartam: hosszú távú, tranzakcióhoz kötött
- A privátszféra oldaláról
  - Akár ellentmondó információk megosztása
  - Látszólag összeköthetetlen, független identitások
  - Teljes anonimitás lehetősége

➔Közösség/csoport alapú modellezés!

# Az élet igazolta: egyébként is szeparálunk!

**ONLINE** **OFFLINE**

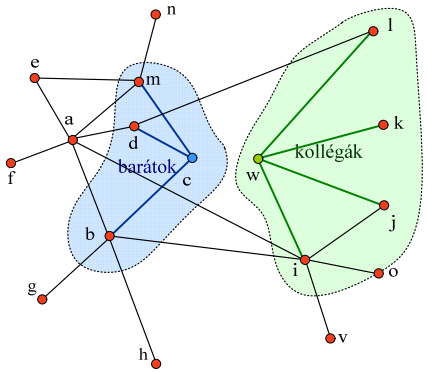



(Forrás: http://www.slideshare.net/padday/the-real-life-social-network-v2)



# Védi a kapcsolatrendszerünket?

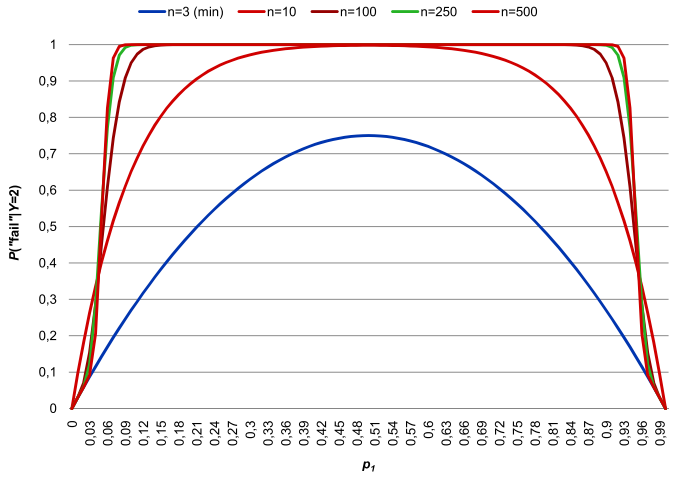
- Ismeretlen felhasználói viselkedés
  - Modellezzük valószínűségi alapokon
  - Létező identitások „darabolása”
  - Ezen belül is többféle modell lehetséges
- Modellek megadása az identitás szeparáció szerint
  - Az identitások kapcsolatai között lehet átfedés?
  - Elveszhetnek az élek?
  - Az élek besorolása független?



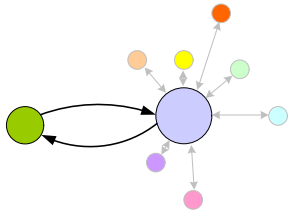



# Védi a kapcsolatrendszerünket? (2)

**Alap modell, hibavalószínűségek a passzív támadás esetén egy felhasználóra 2 új identitás esetén:**

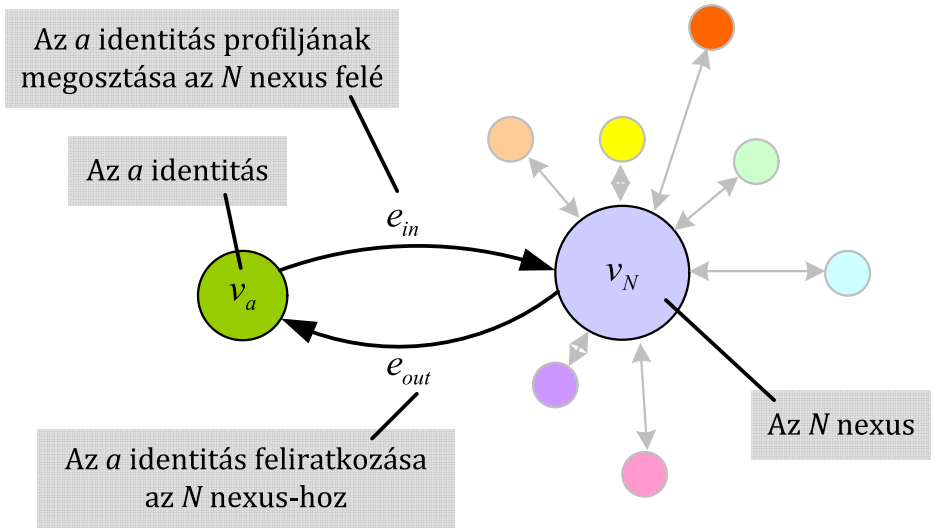




## IDENTITÁS SZEPARÁCIÓ KÖZÖSSÉGI HÁLÓKBAN: NEXUS-IDENTITY NETWORKS

# Struktúra






# Anonimitási szintek

1. Teljes anonimitás
   - Azonosító nélkül
2. Összekapcsolható anonim azonosítás
   - Rövid távon használt azonosító
   - Lokális, adott névtérben egyedi azonosító
   - Tetszőlegesen lecserélhető
   - Párhuzamosan több használható
3. Állandó azonosító
   - Hosszú távon használt azonosító
   - Globális névtérben egyedi azonosító






# Reguláris vs. több identitásos közösségi háló

**Reguláris**

- Kapcsolat orientált struktúra
- Egy felhasználó egy identitással van jelen (egy felhasználó egy csomópont)
- Pontos beazonosíthatóság

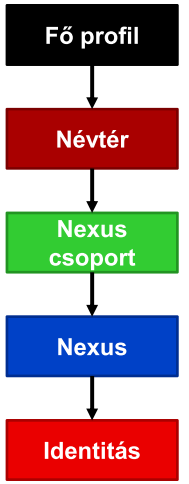
**Több identitásos**

- Közösség orientált struktúra
- Egy felhasználó több identitás által reprezentált (egy felhasználó „több csomópontból áll”)
- Pontos beazonosíthatóság
- Anonimitás
  - Teljes anonimitás
  - Összeköthetetlen identitások

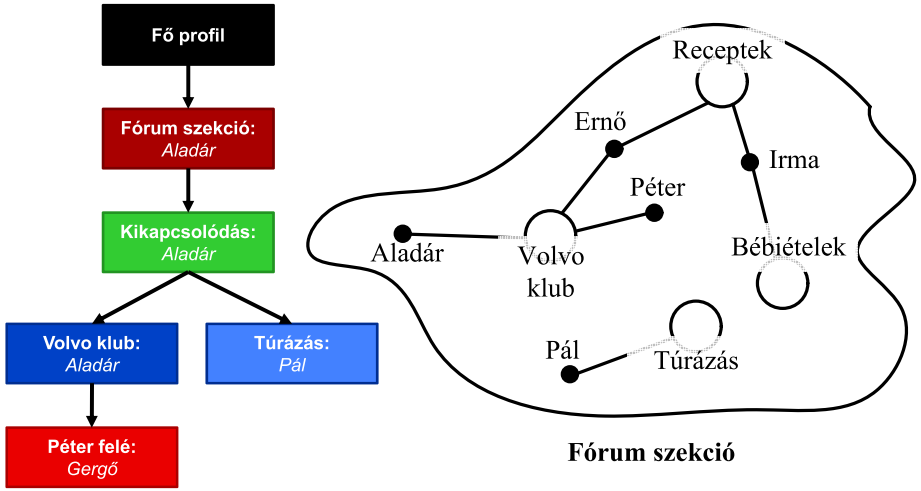


# Profil hierarchia

- Fa struktúrájú
  - Globálistól a lokális felé halad
  - Öröklődés
  - Egyre kevesebb információ kerül megosztásra
  - Ellentmondó információk lehetnek az egyes ágakon
- Csak a tulajdonos ismeri
  - (és a szolgáltató)
  - Mások számára összeköthetetlen ágak is lehetnek benne

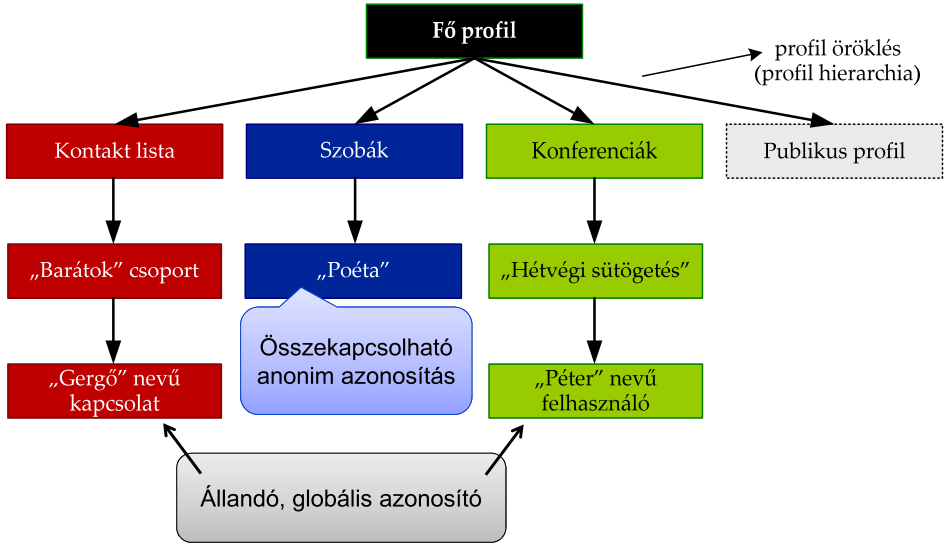


# Hálózat struktúra



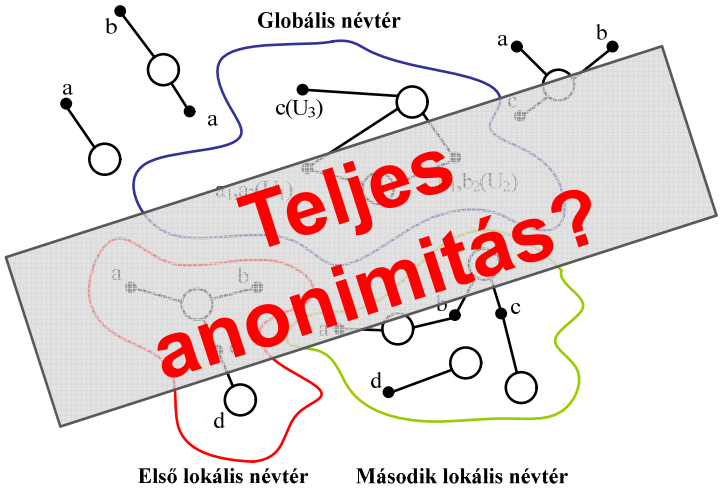



# Anonim csevegő szolgáltatások






# Hálózat struktúra (2)






# ÖSSZEFOGLALÁS

# Összefoglalás: mit tehetünk?

- Tudatosság! Tényleg szükségünk van rá?
  - Amit publikáltunk, az már visszavonhatatlanul kint van.
- Számoljunk a trendekkel! Minden egyre inkább...
  - … publikus
  - … kereshető
  - … valós idejű
- Használjunk saját vezérlésű hozzáférés szabályozási alkalmazásokat.
- Identitásszeparáció: ha más néven futunk két helyen: kerüljük az azonos feliratkozást, tevékenységeket, barátokat



# KÖSZÖNÖM A FIGYELMET!

**Gulyás Gábor György**
gulyasg@hit.bme.hu

Híradástechnikai Tanszék

http://www.hit.bme.hu

international PET portal and blog

http://pet-portal.eu