

Eligazodás a világ hálózataiban

Készítette:

Pluhár András

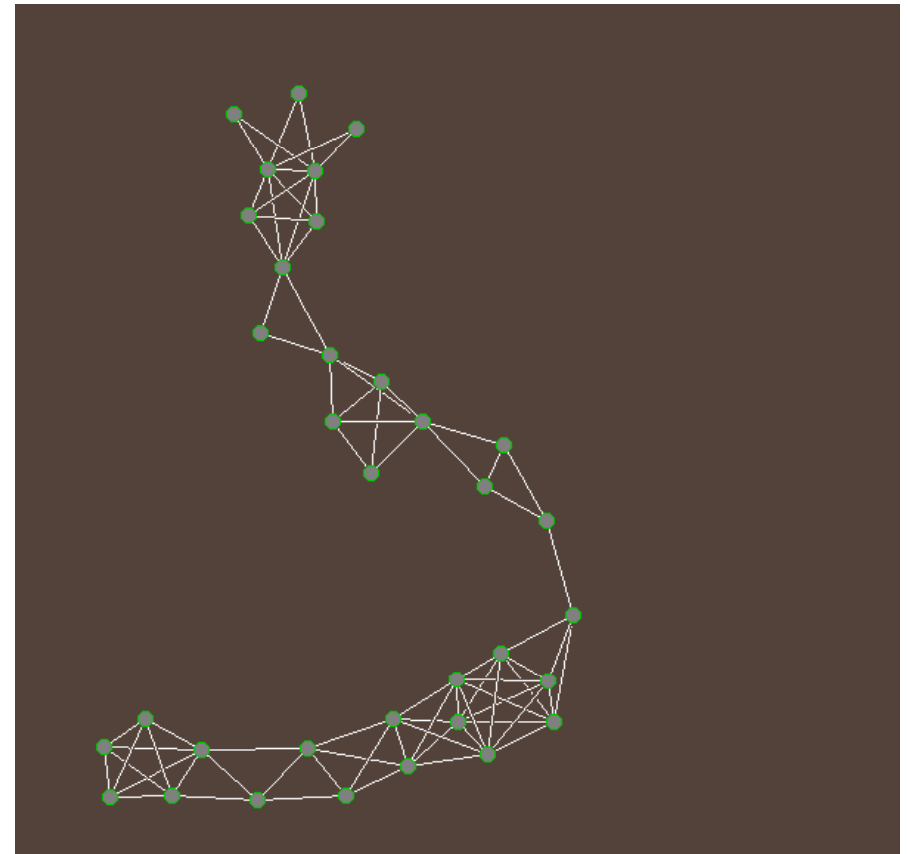
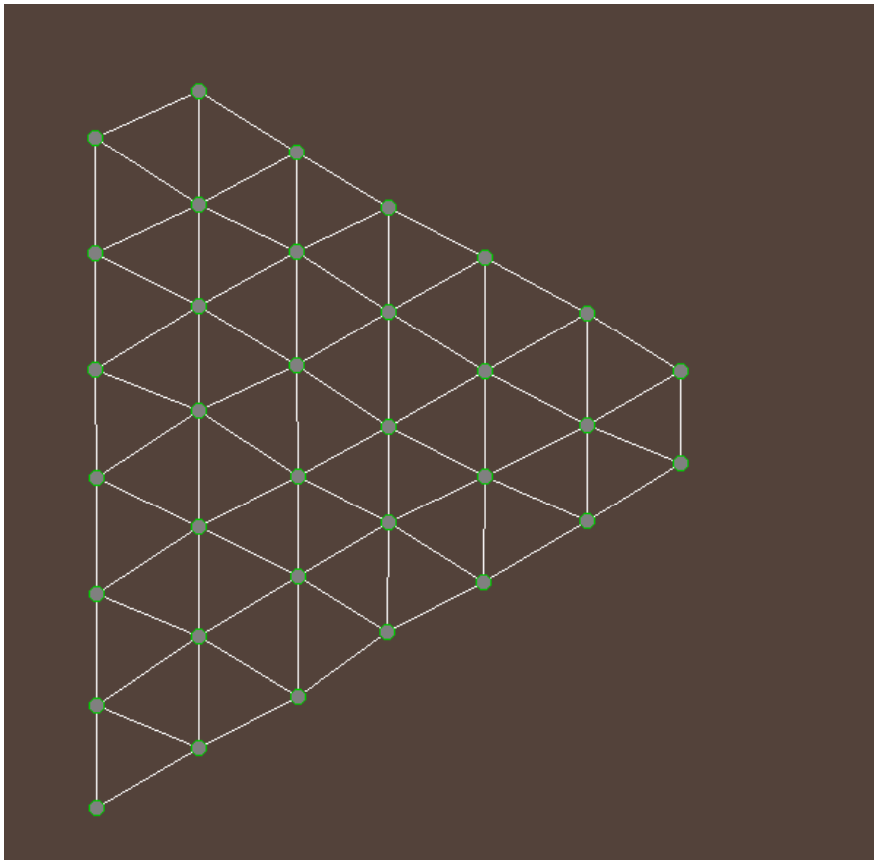
Szegedi Tudományegyetem, 2008

Adatbányászat általában

- Óriási számú adat (telefon, LHC)
- Technikai problémák
- Gyűjtés és tárolás
- Tisztítás
- Lekérdezés
- Csoportosítás
- Jogi problémák

Alapfogalmak

- Csúcsok
- Élek
- Utak
- Összefüggőség



Hálózat kutatás – modern gráfok elemzése

- World Wide Web (Barabási)
- Internet
- Szociológiai hálók (blogok, közösségi lapok,...)
- Üzleti kapcsolatok
- Biológiai hálók (gének, járványterjedés, táplálkozási)

Közös jellemzők?

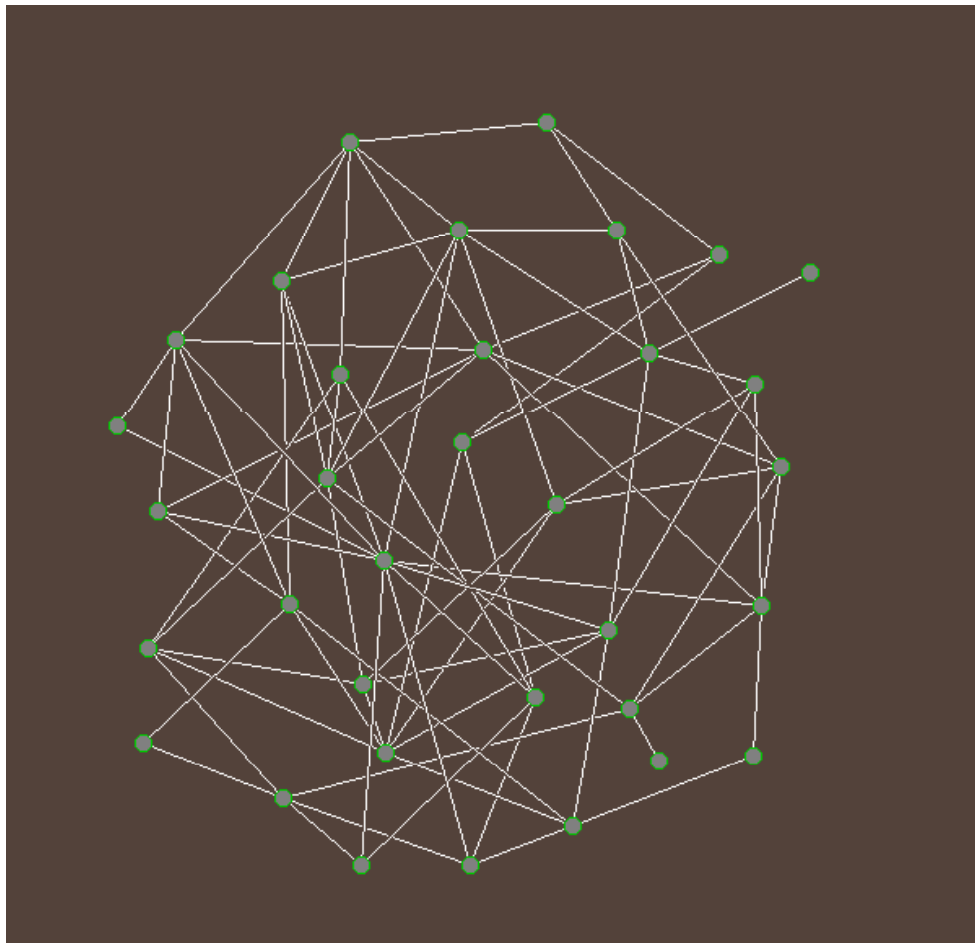
Kisvilág gráfok

Kis-világ gráfok

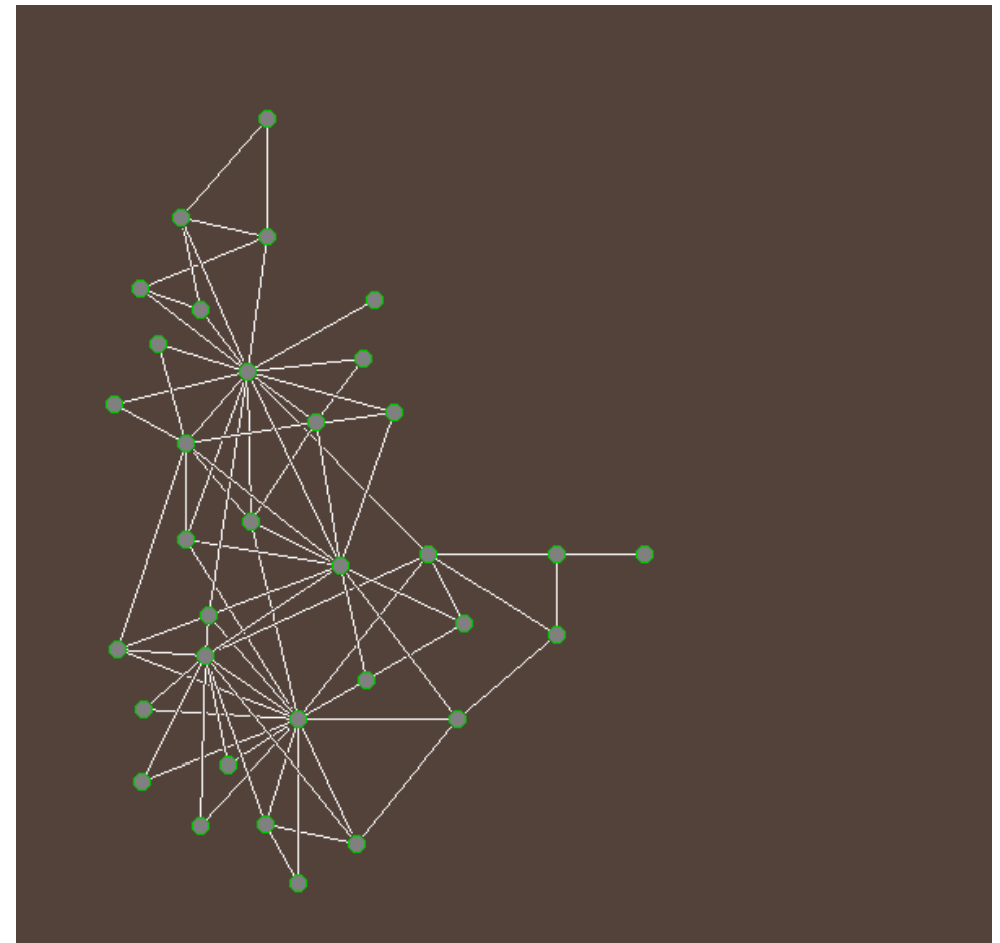
- Kevés élük van (konstans \times pontszám)
- Kicsi az átmérőjük (kézfogás 6, www 19)
- Sűrűsödések (háromszögek) vannak
- A fokszám eloszlás hatványtörvényes

Klasszikus és kisvilág gráfok

Erdős-Rényi



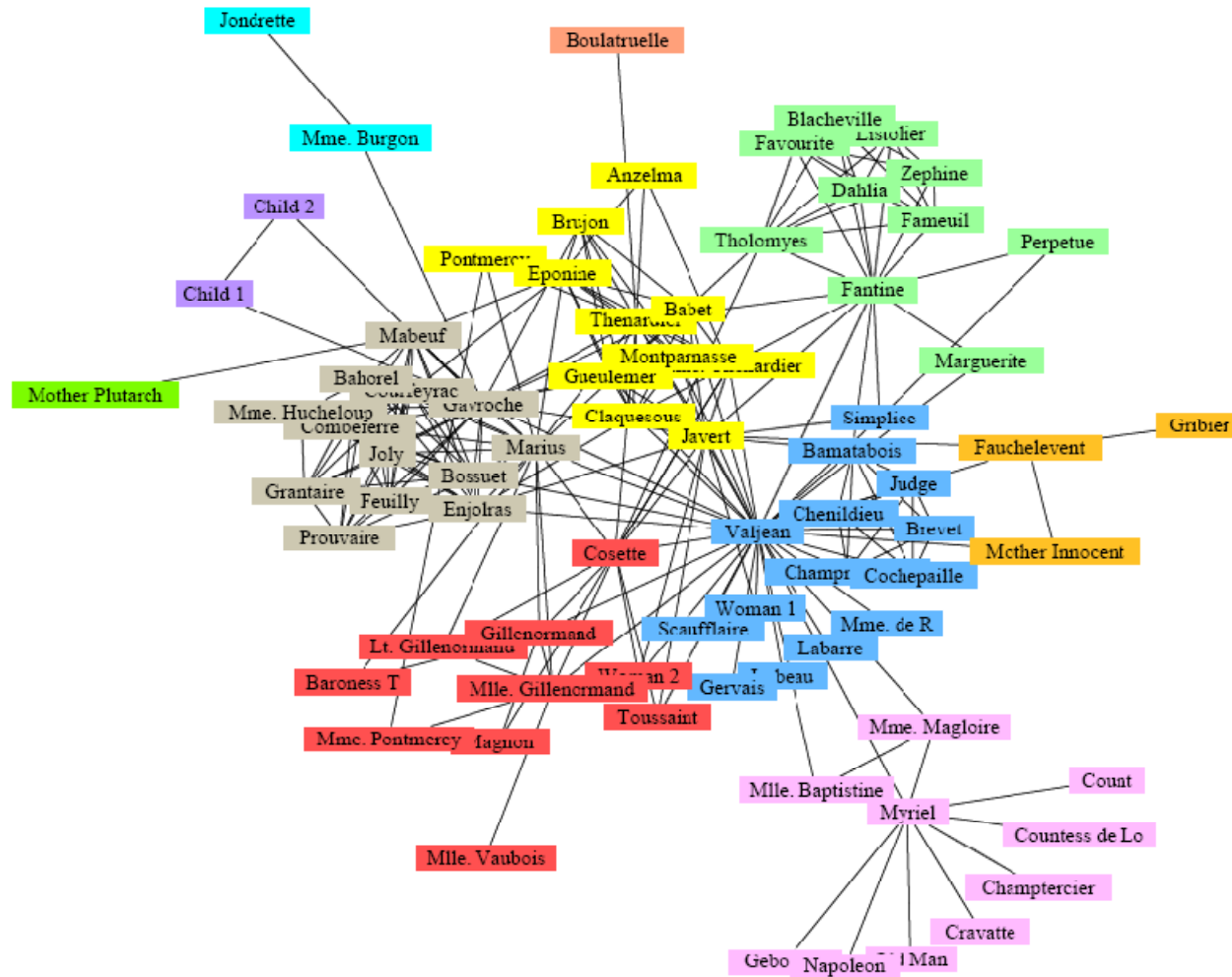
Zachary



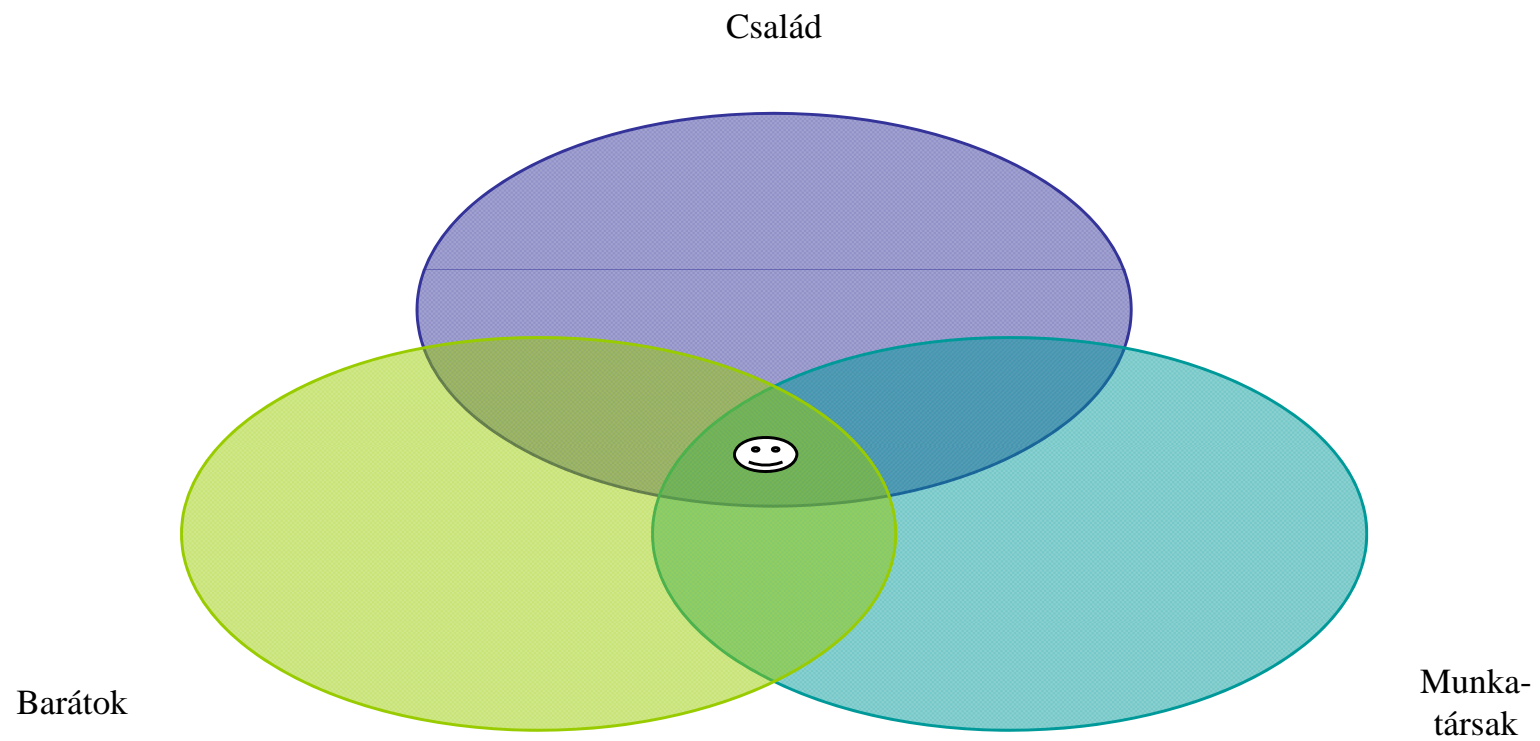
Klaszterezés

- Alapfogalom (=hasonló elemek csoportosítása)
- Elvi lehetetlenség (Kleinberg)
- A gyakorlatban jó algoritmusok
- Hálózat alapú klaszterezés
- Rajz

Példa klaszterezésre (Newman)

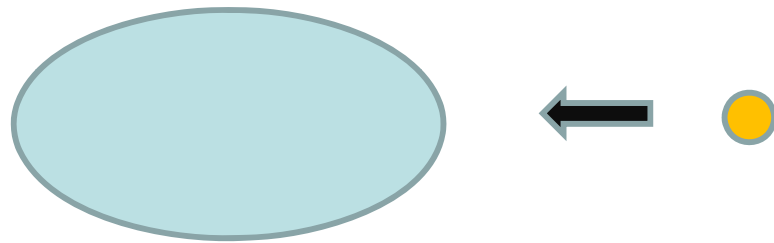


Közösségek



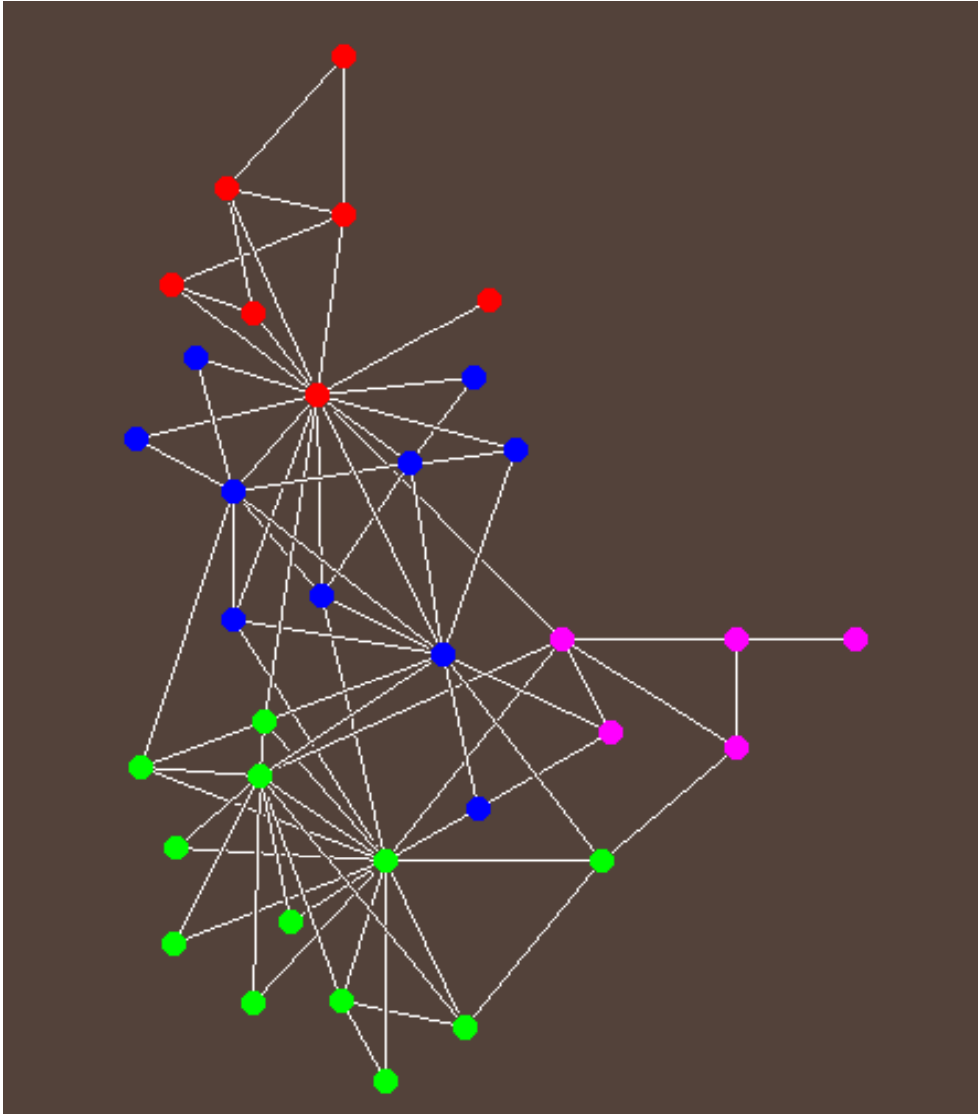
Alulról építkező eljárások

Beillesztjük, ha $f(A, x) \geq c(|A|)$



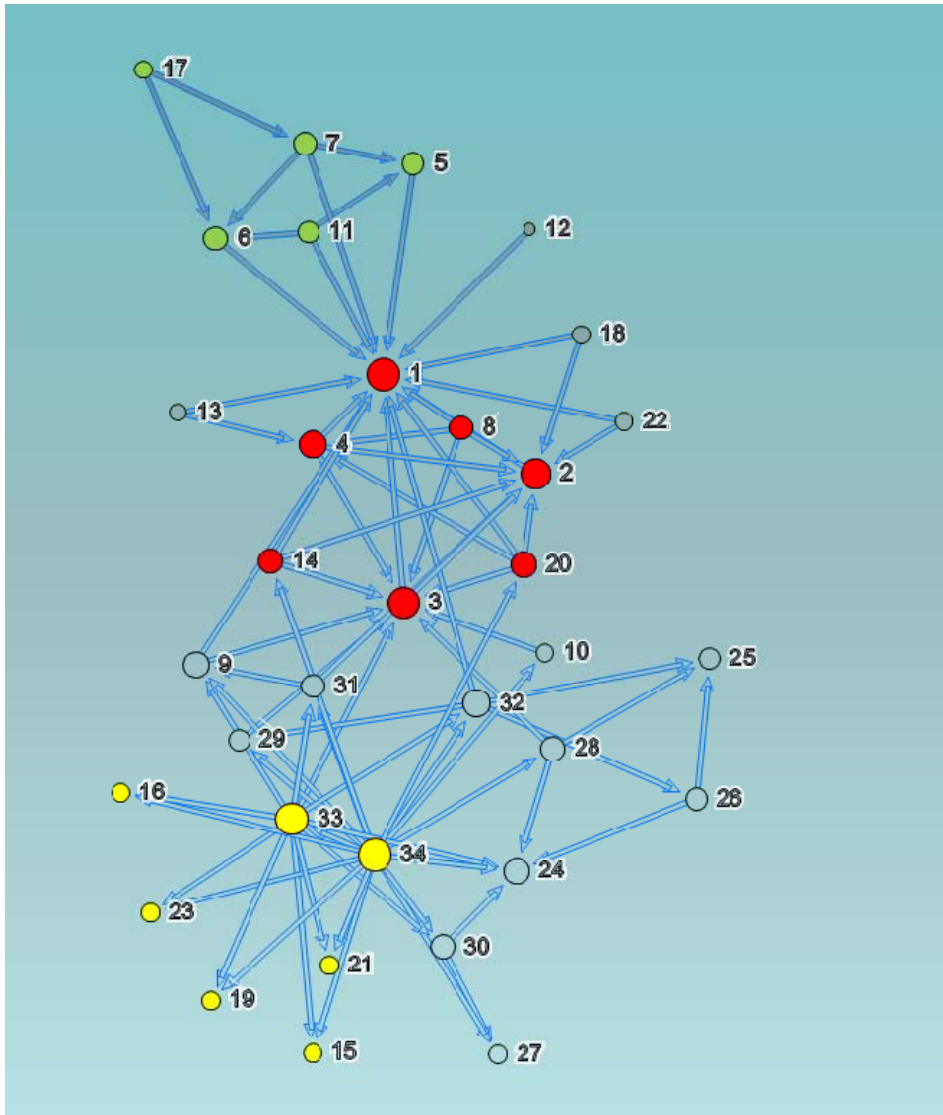
Itt f egy távolságfüggvény, c paraméter

A Zachary gráf klaszterei



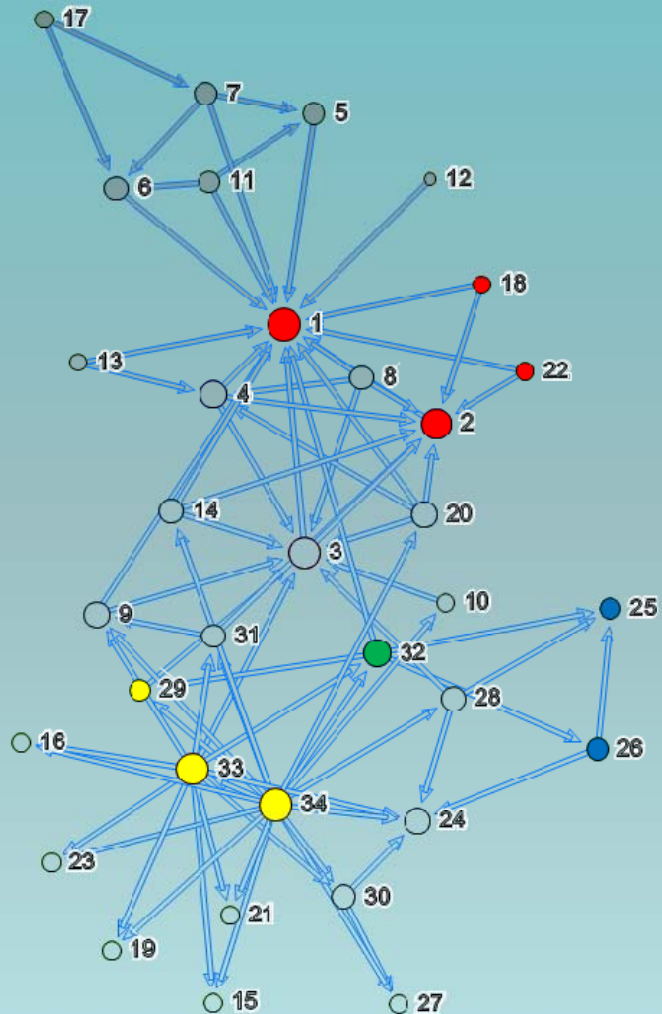
- Az 1978-as vizsgálat közben a klub kettévált.
- A piros és kék az egyik, a zöld, rózsaszín a másik.
- A stabilitás fontos kérdés egyes alkalmazásoknál.

A Zachary nagy közösségei



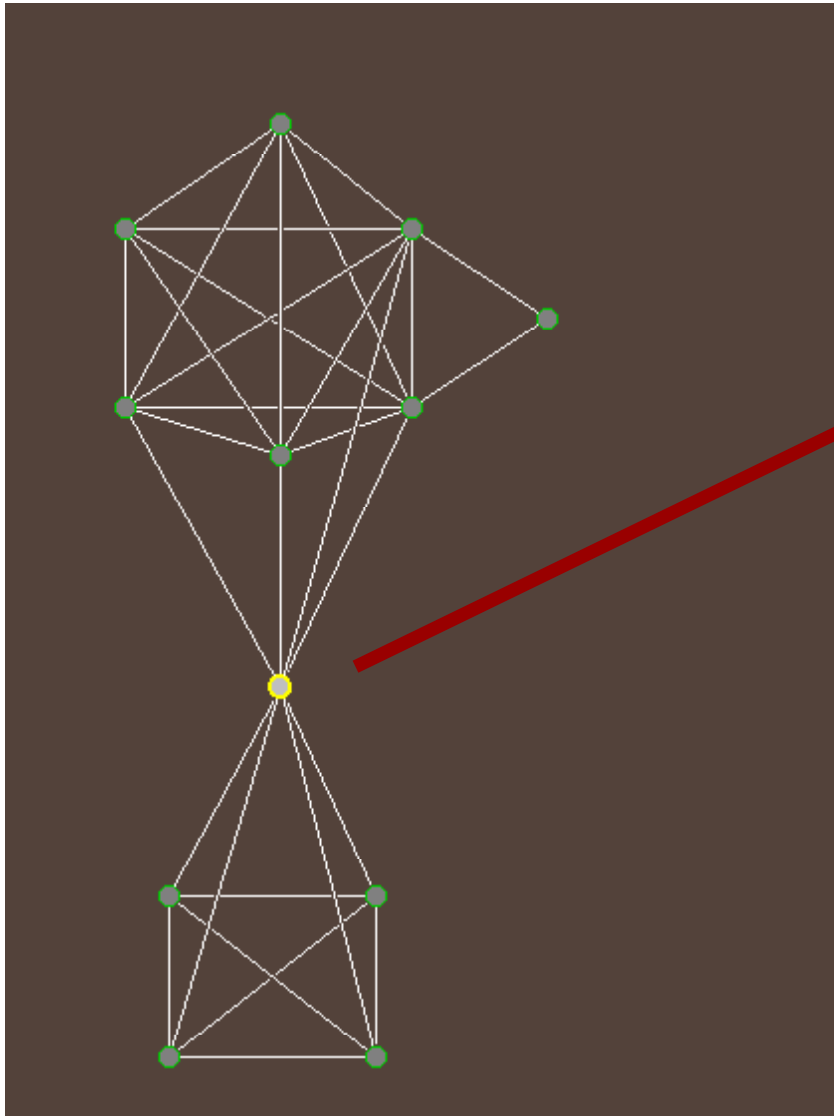
- Két nagy közösség az 1-es pont (japán mester), a másik a két segédje, 33-34 körül.
- A piros közösség miatt kerül a 3-as pont a helyére.
- Kérdés: van-e még kritikus fontosságú pont?

Kisebb közösségek



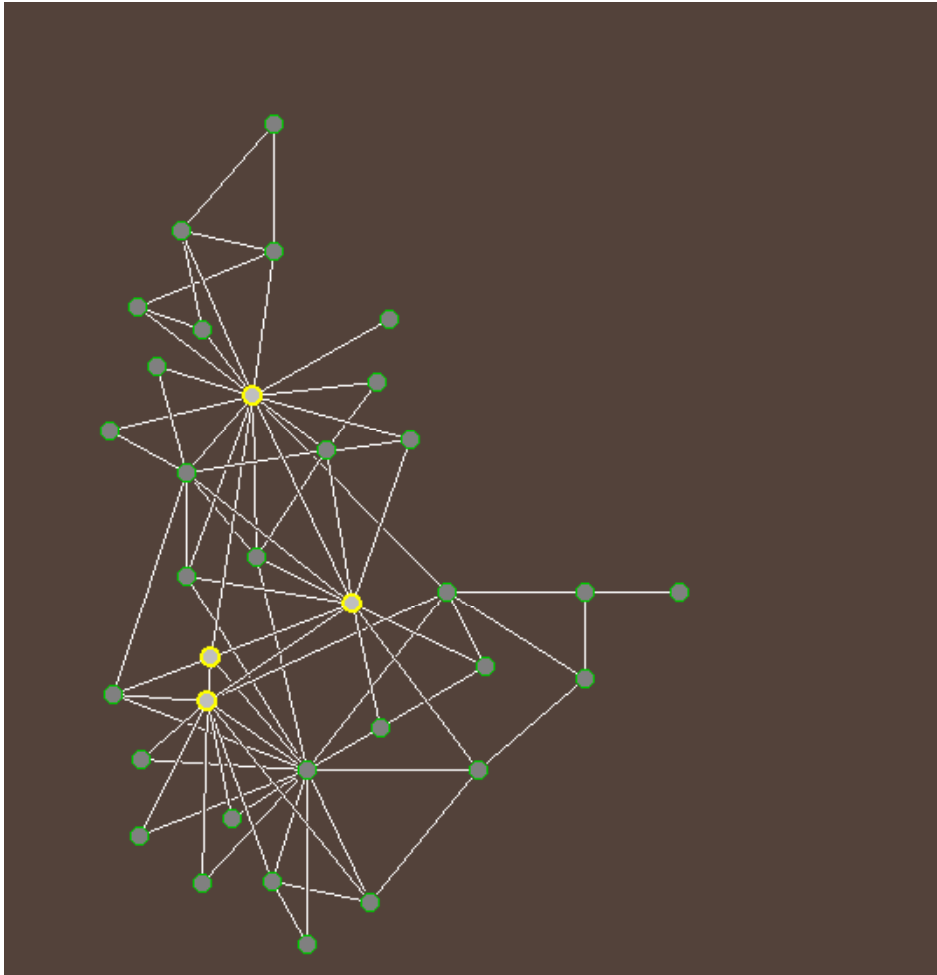
- Vannak kisebb közösségek is részben az 1-es pont köti őket össze, ill. itt a 32-es.

Közösség metszet gráf



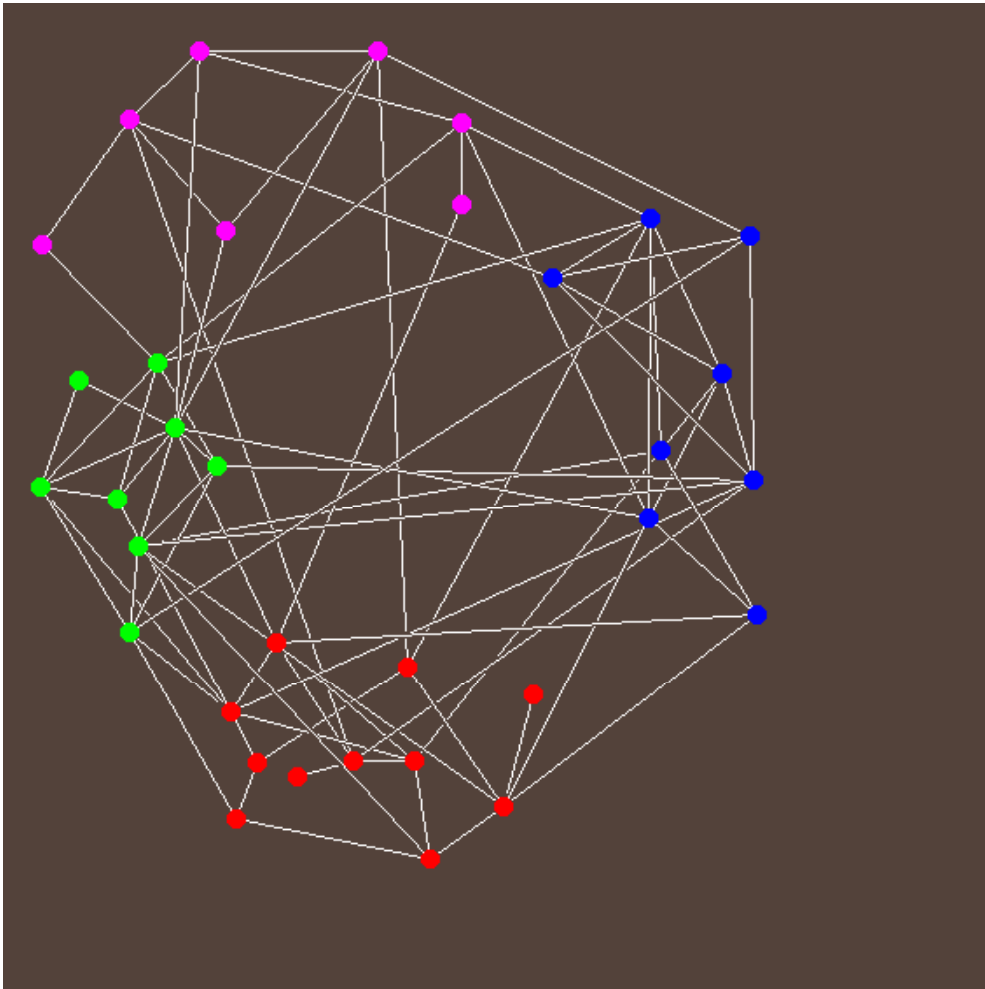
- A metszet gráfban van egy vágópont. Ez az $\{1, 3, 9, 33\}$ közösség. A többi közösség intenzív kapcsolatban áll a vele egy oldalon levőkkel.

A kritikus közösség



- A $\{1, 3, 9, 33\}$ közösség. A 3-as és a 9-es pontnak kellett súlyos döntést hozni. A közösségek vizsgálata nélkül a 9-es pont szerepe nem tárható fel.

Erdős-Rényi gráf klaszterei

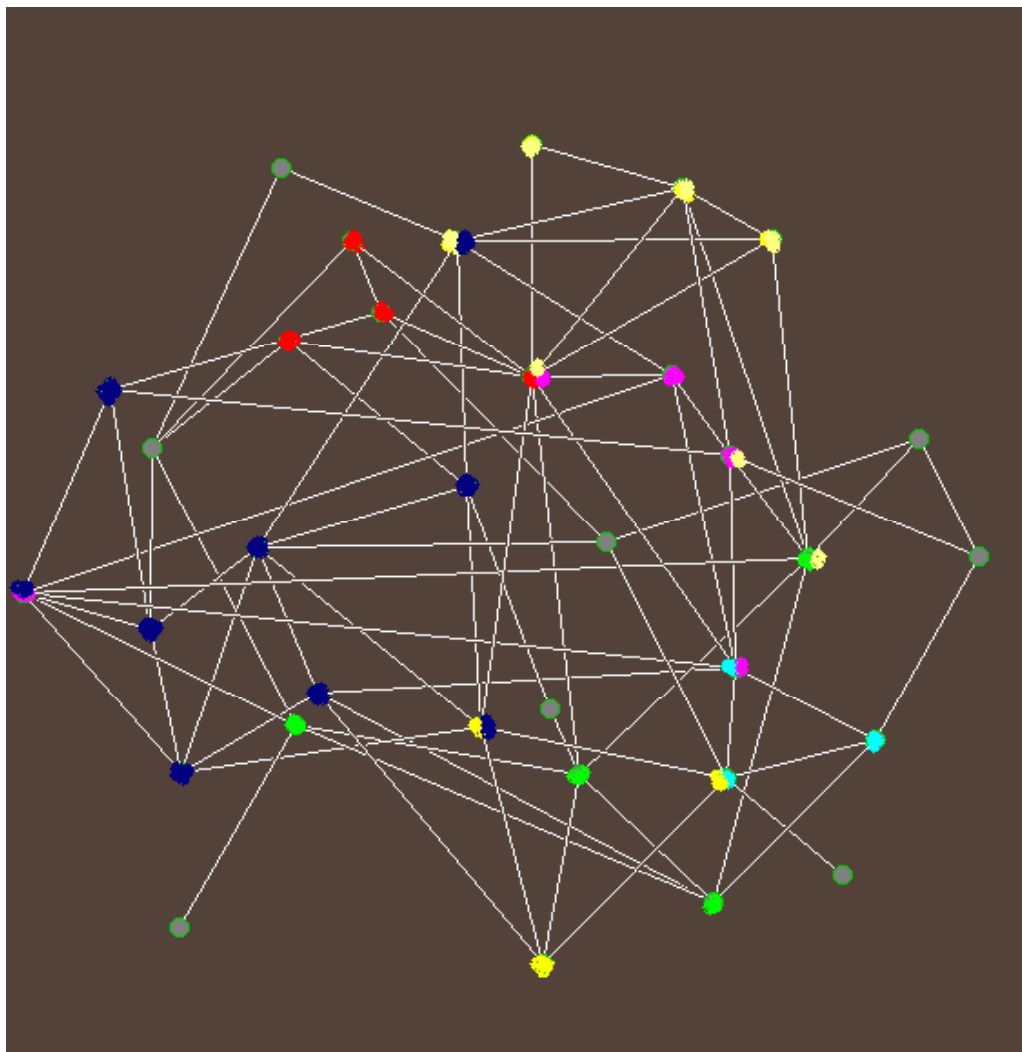


Jellemező:

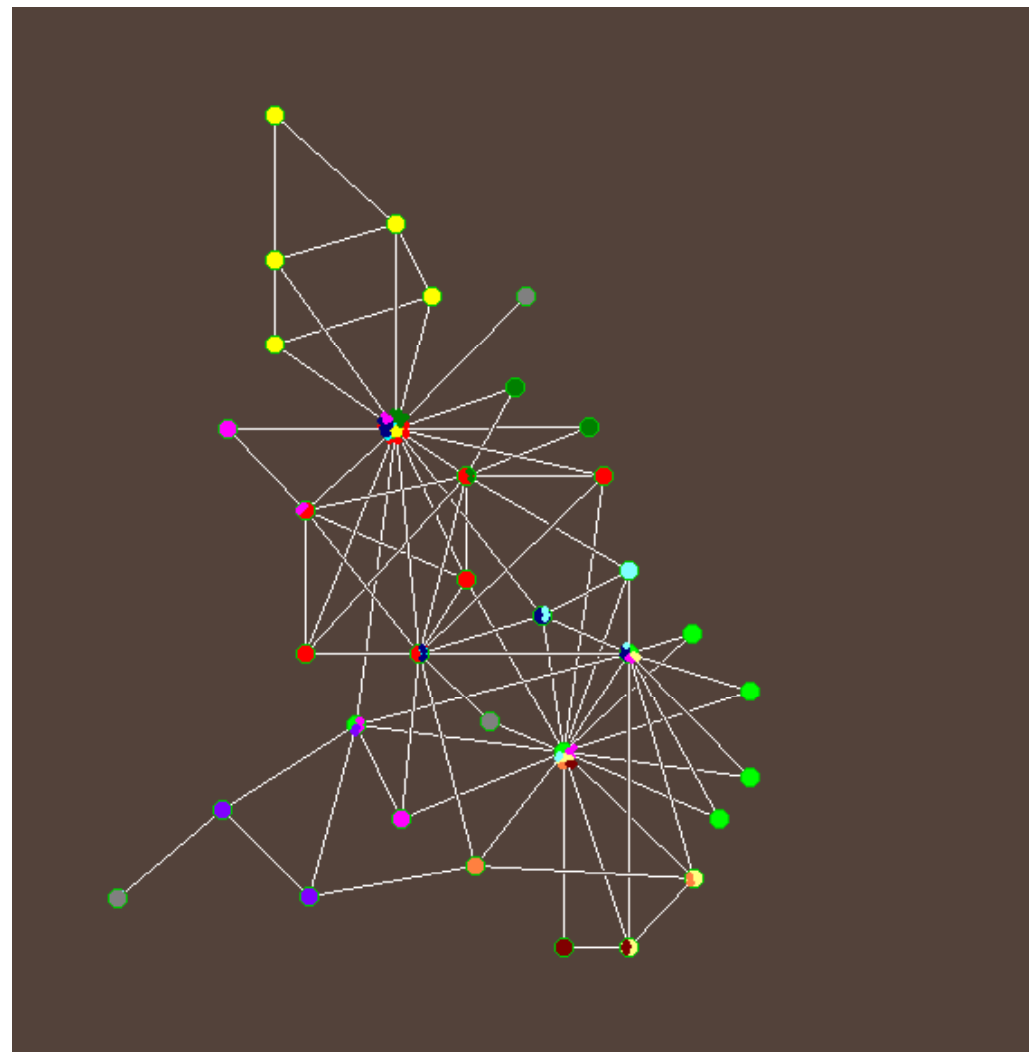
- Nincsenek „éles” klaszterhatárok
- A gráf lényegében homogén.

Közösségek

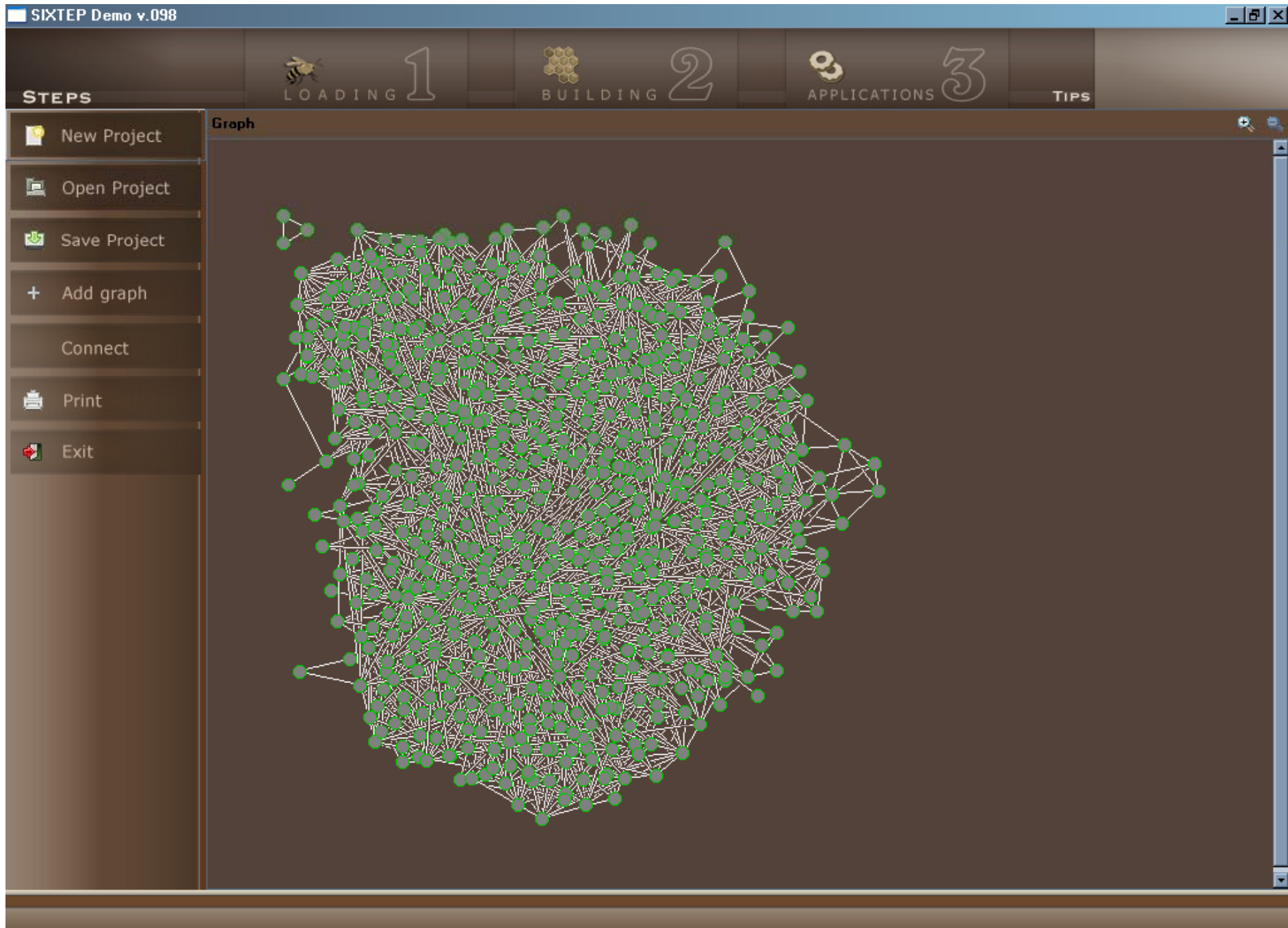
Erdős-Rényi



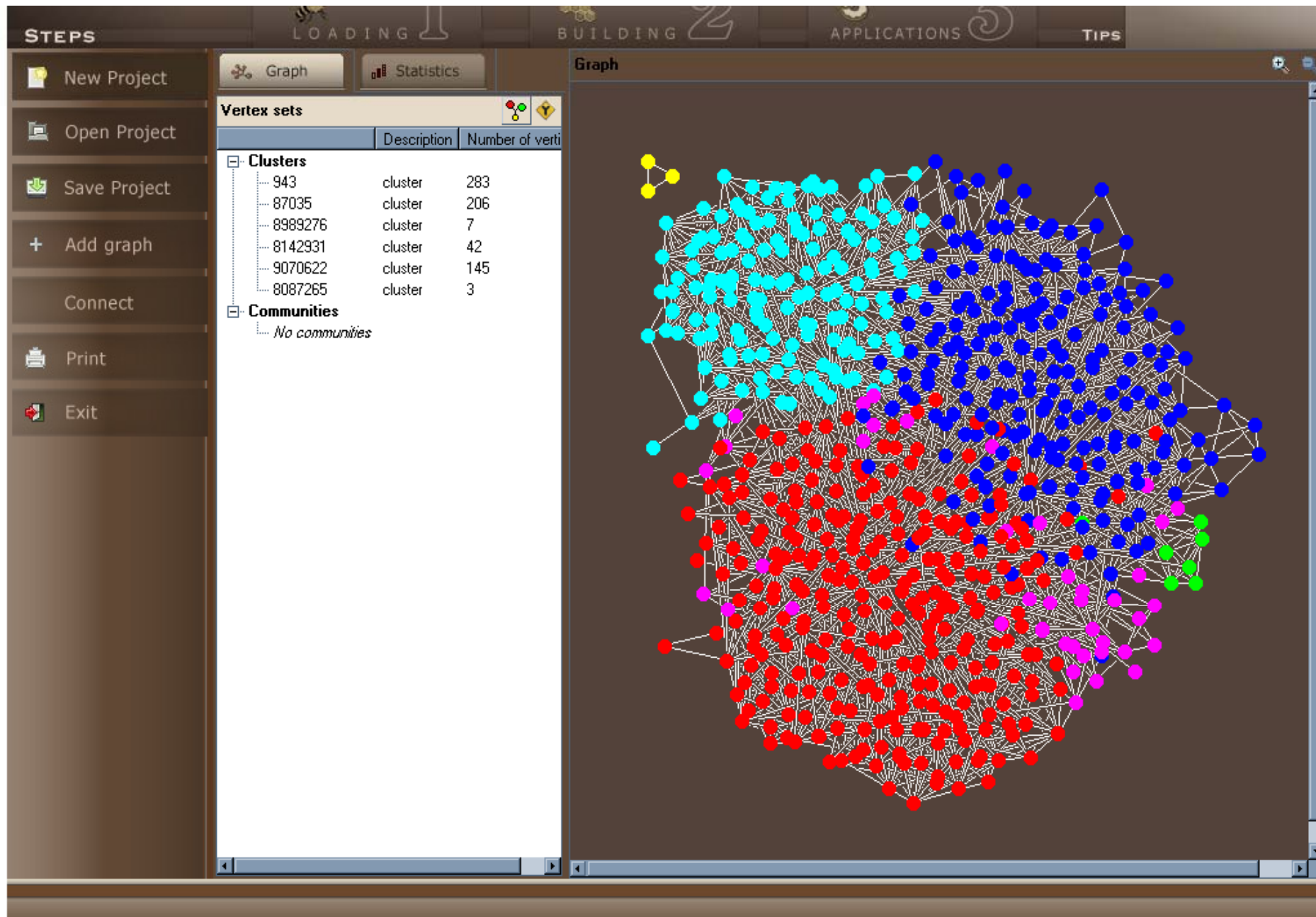
Zachary



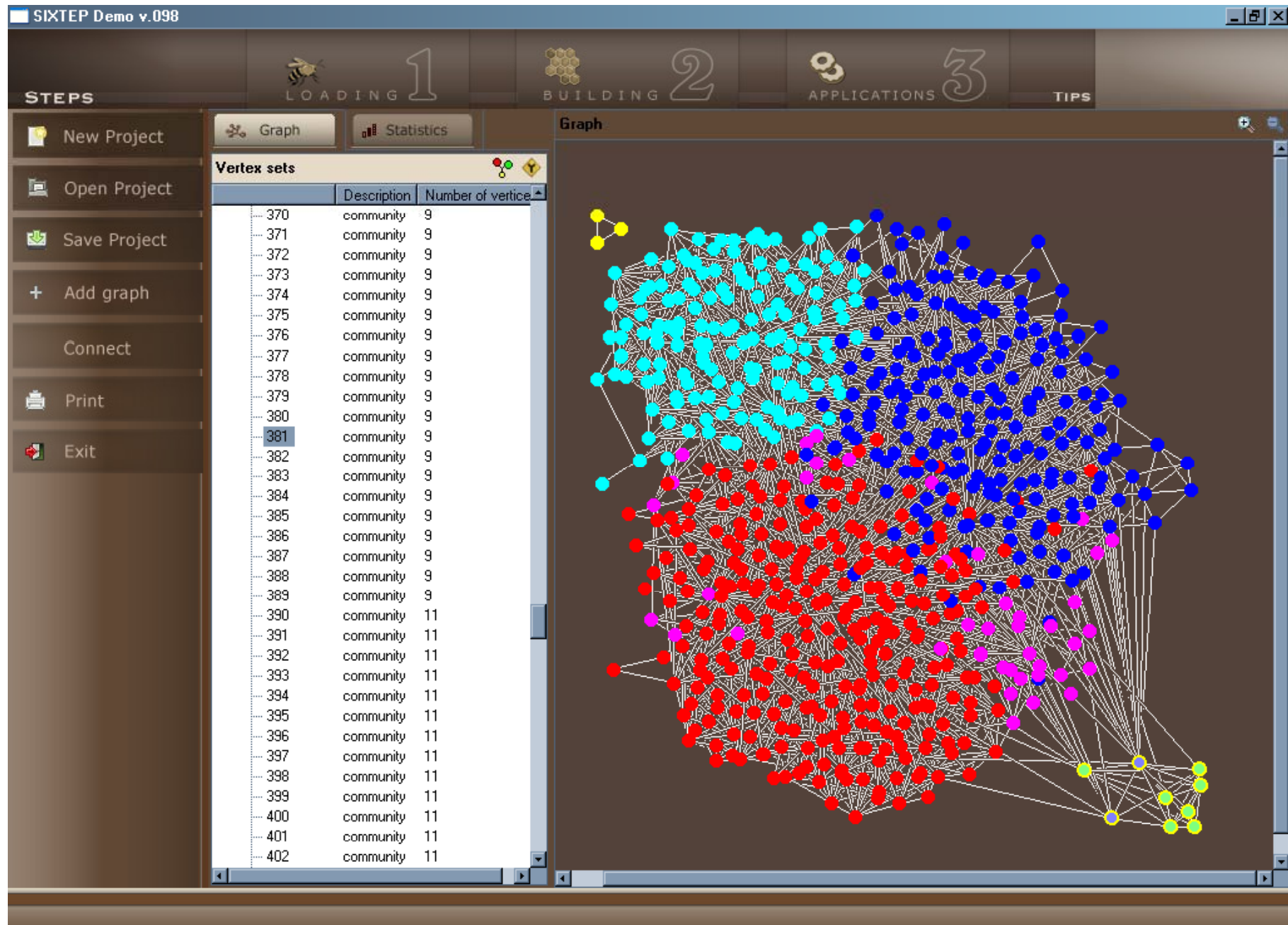
Esettanulmány – közösségi háló



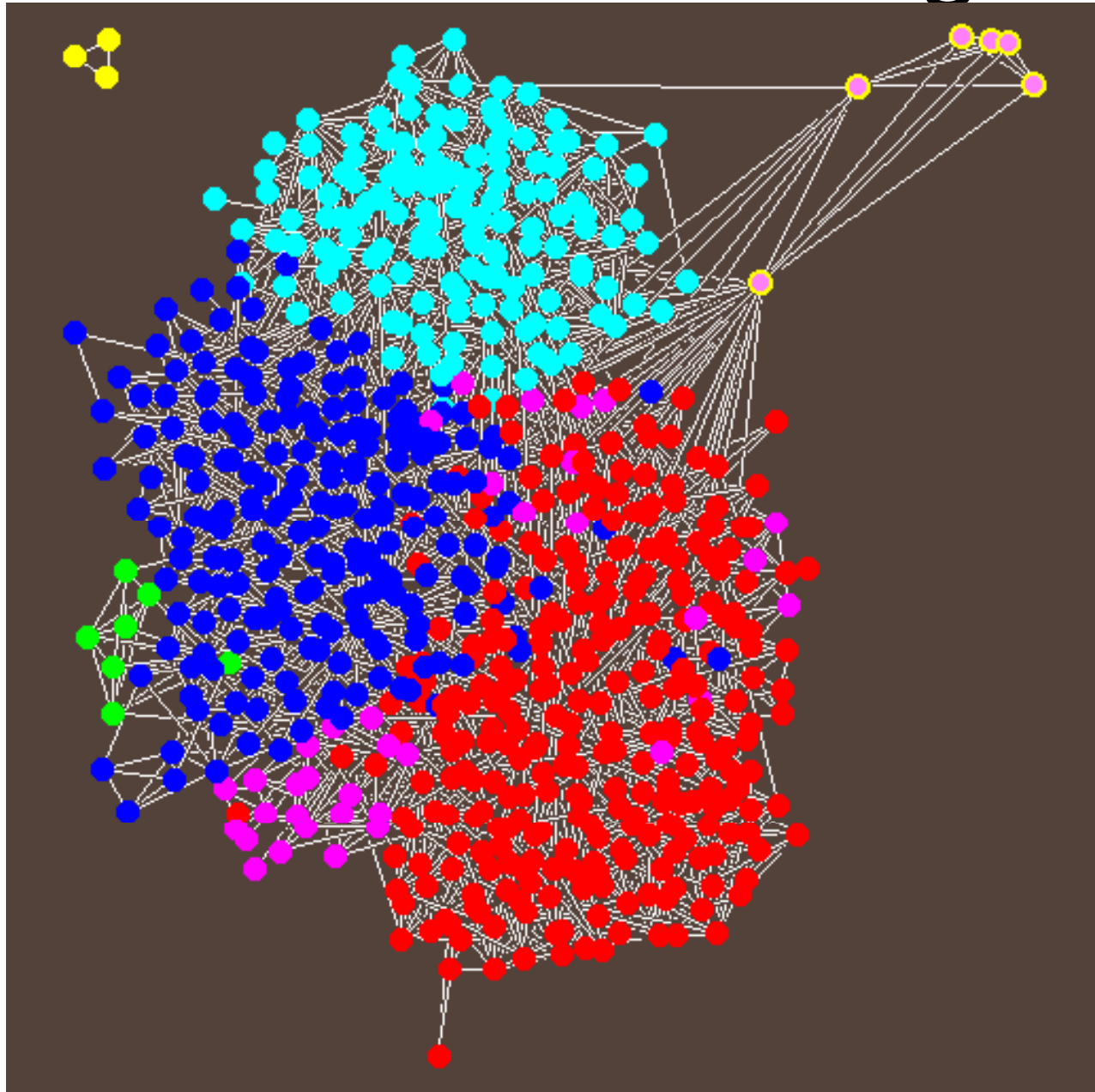
Klaszterek



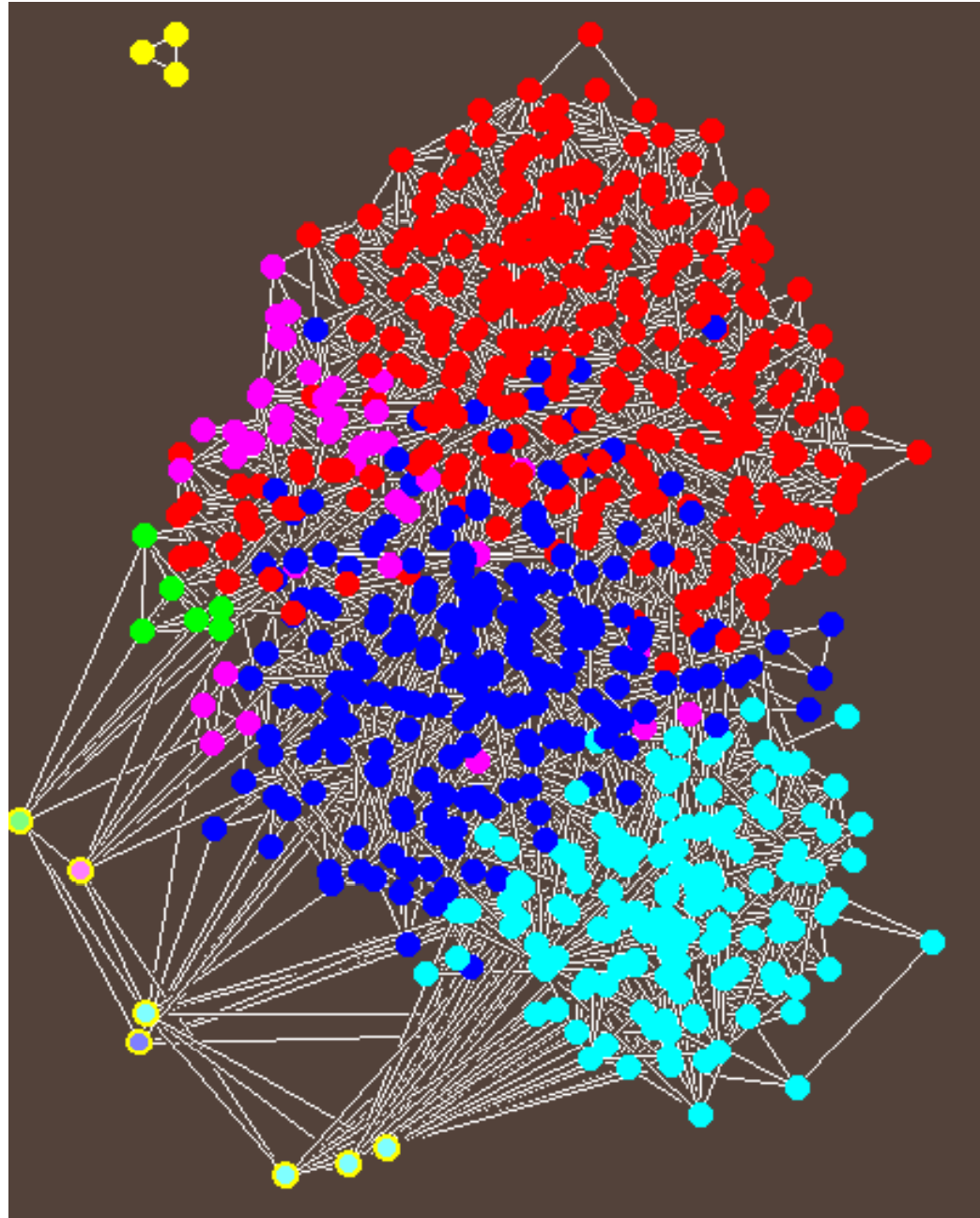
Közösségek-klaszter kiterjesztése



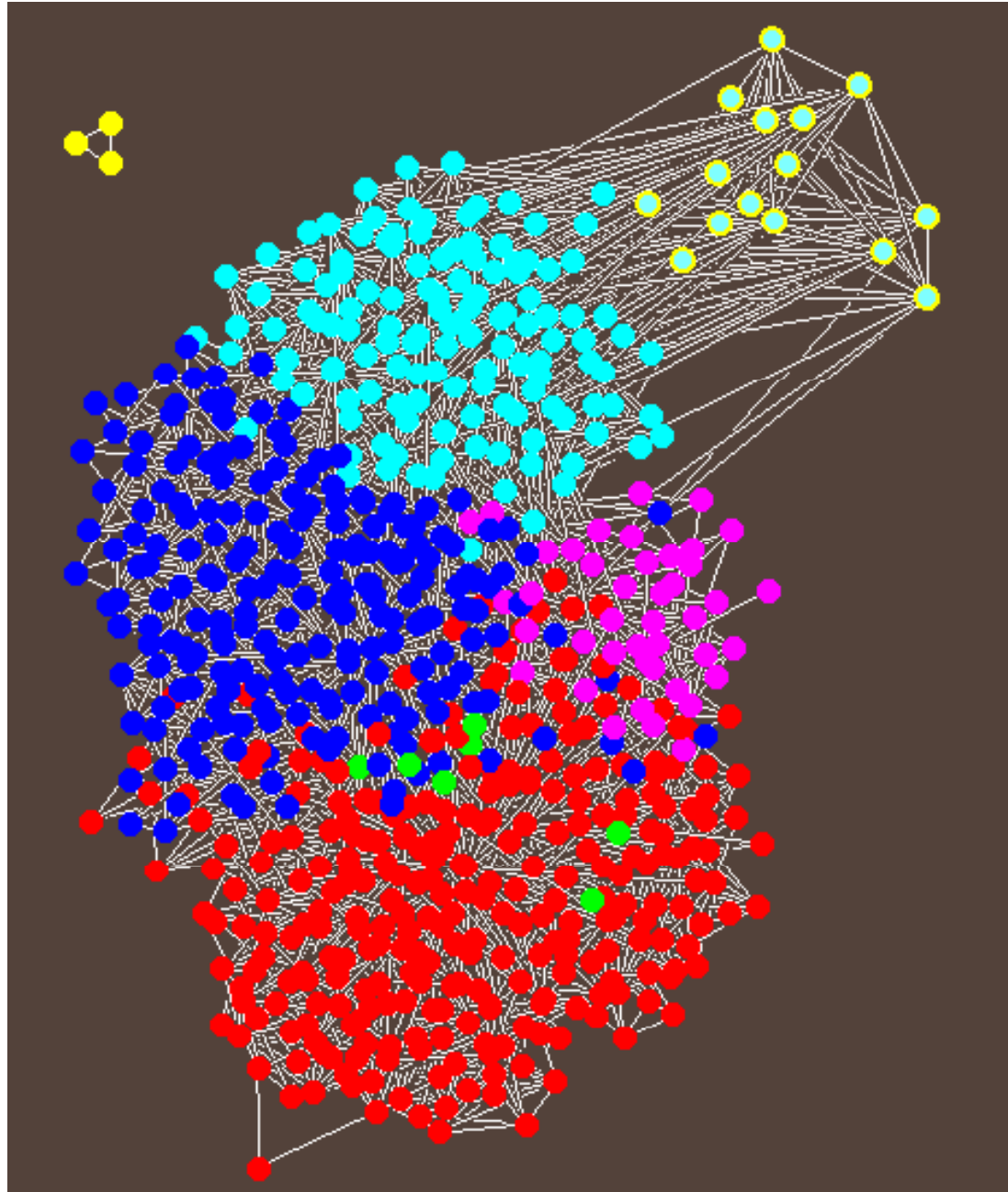
Közösségek - klaszteren belüli szoros közösség



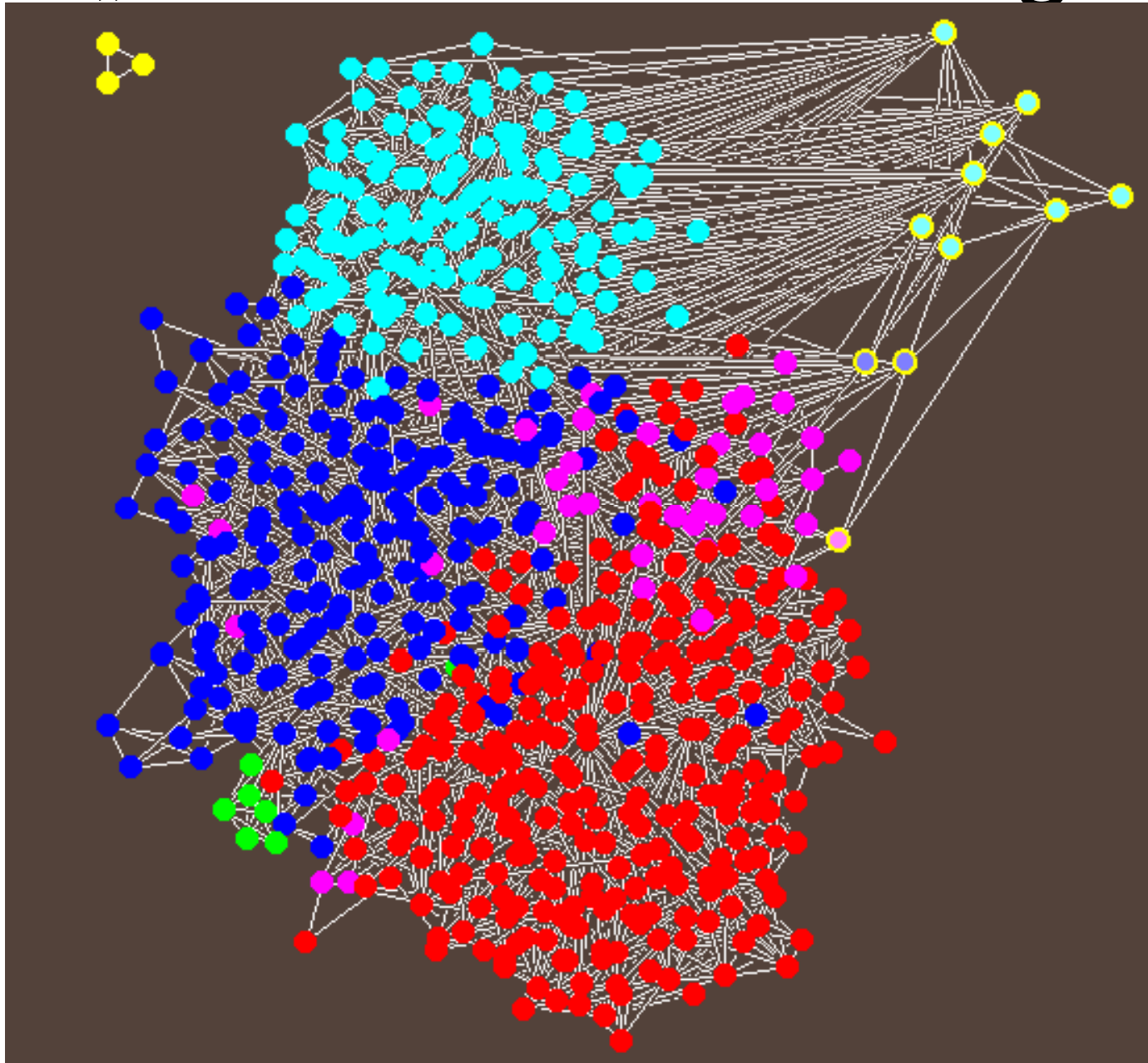
Közösségek – Nem „hagyományos” közösség



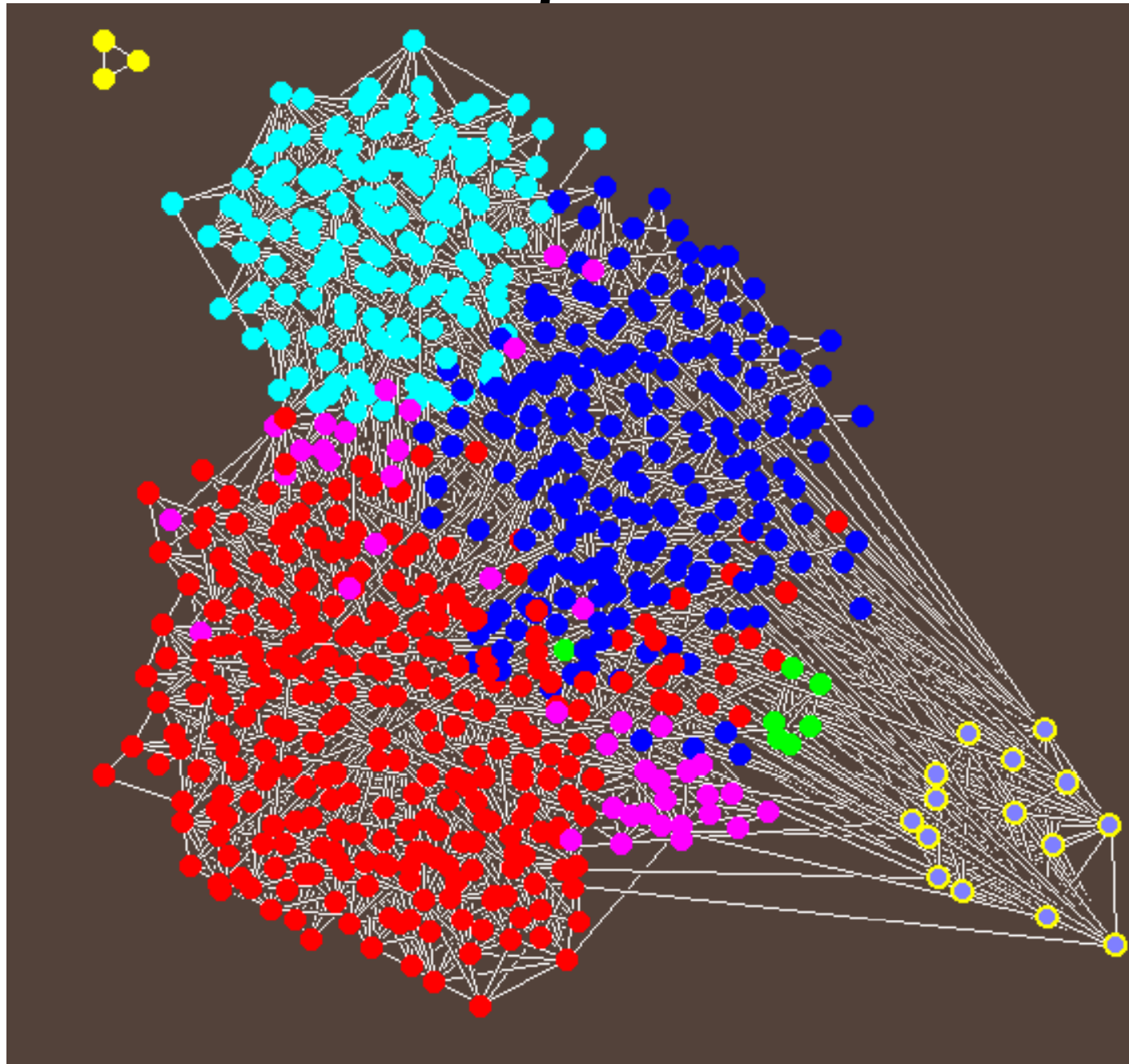
Közösségek – egy „nyilvánvaló” közösség



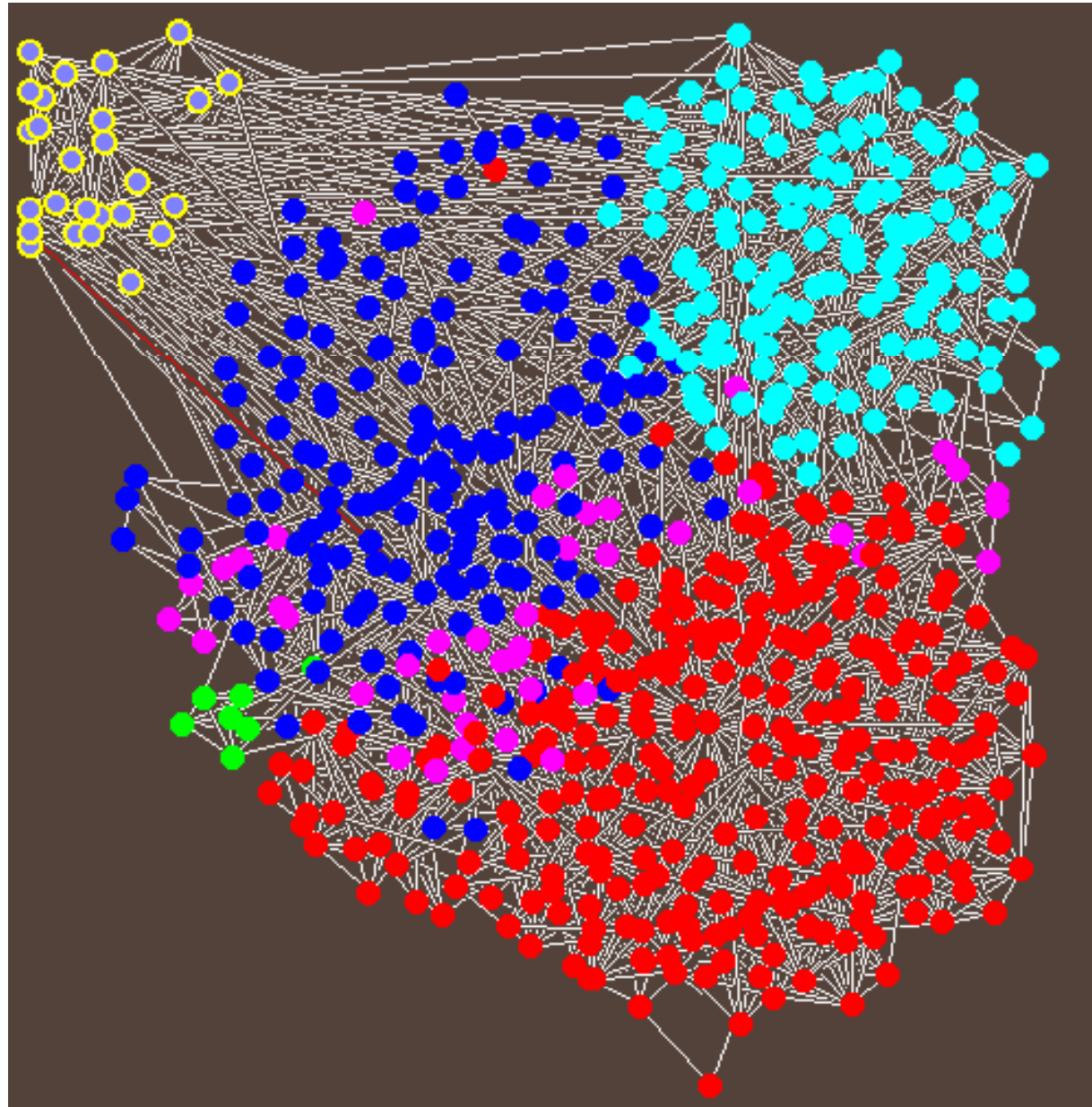
Közösségek – egy látszólag „sokszínű” közösség



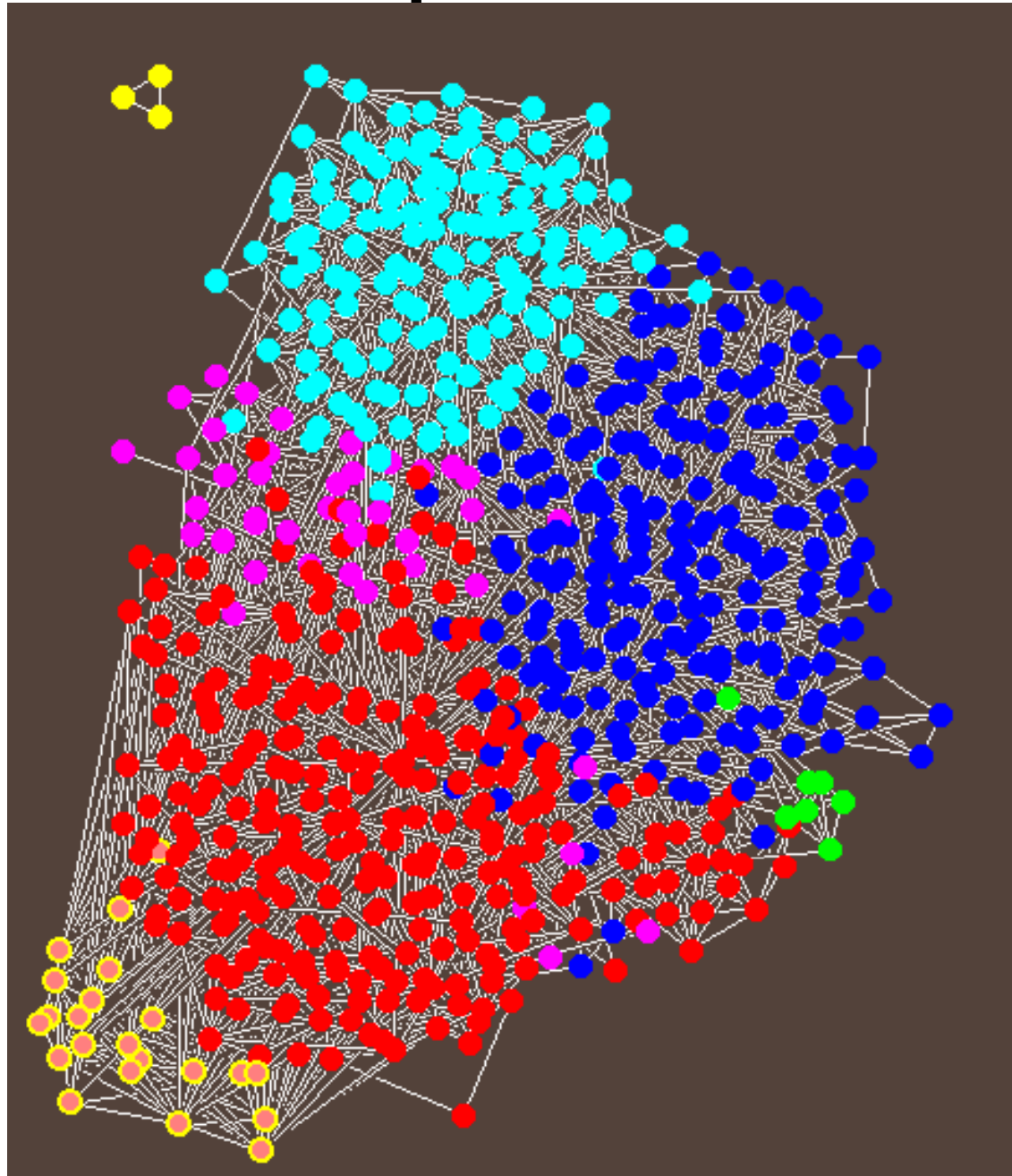
Közösségek – a közös hobby ereje



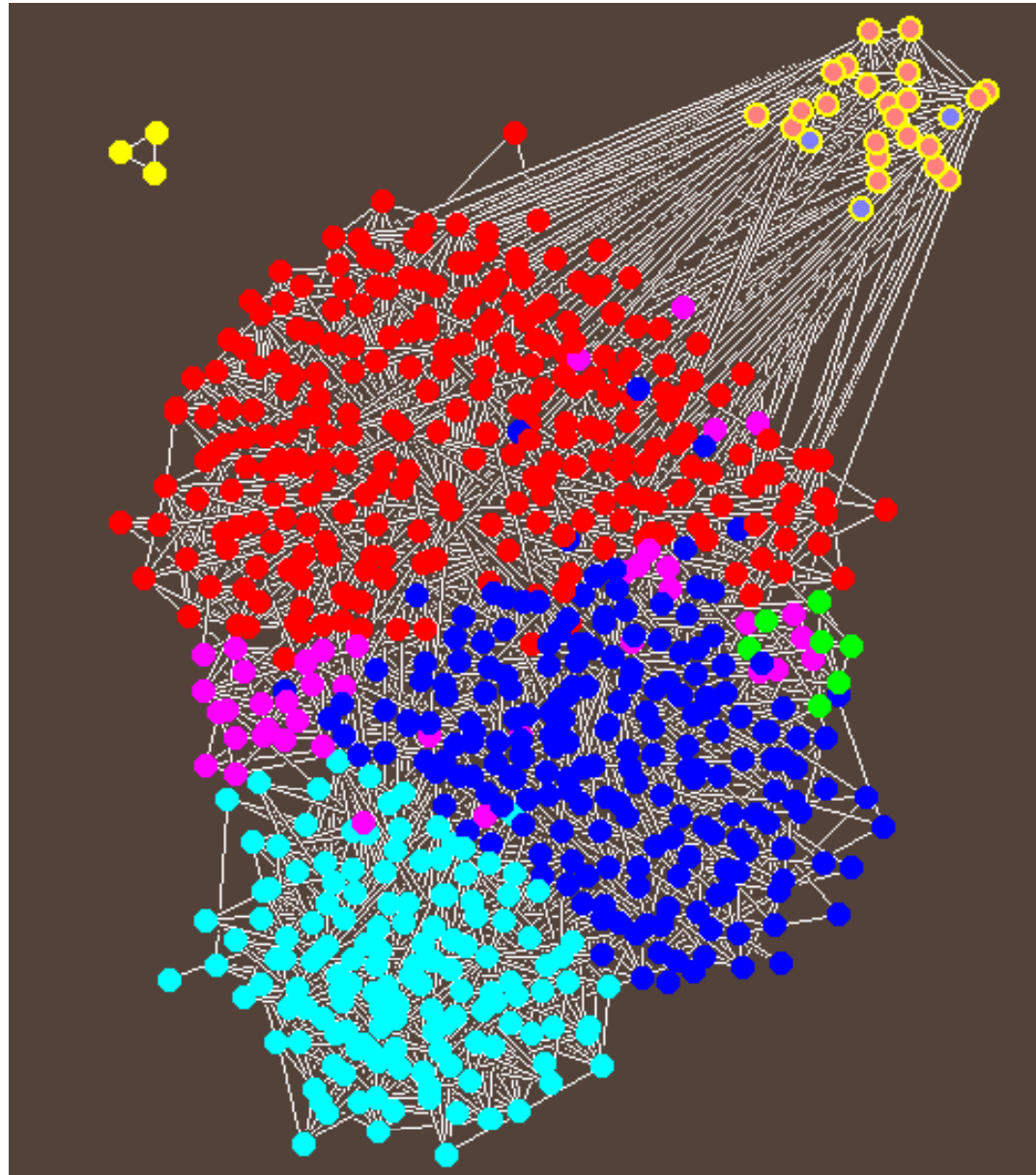
Közösségek – a családi „kiterjesztés”



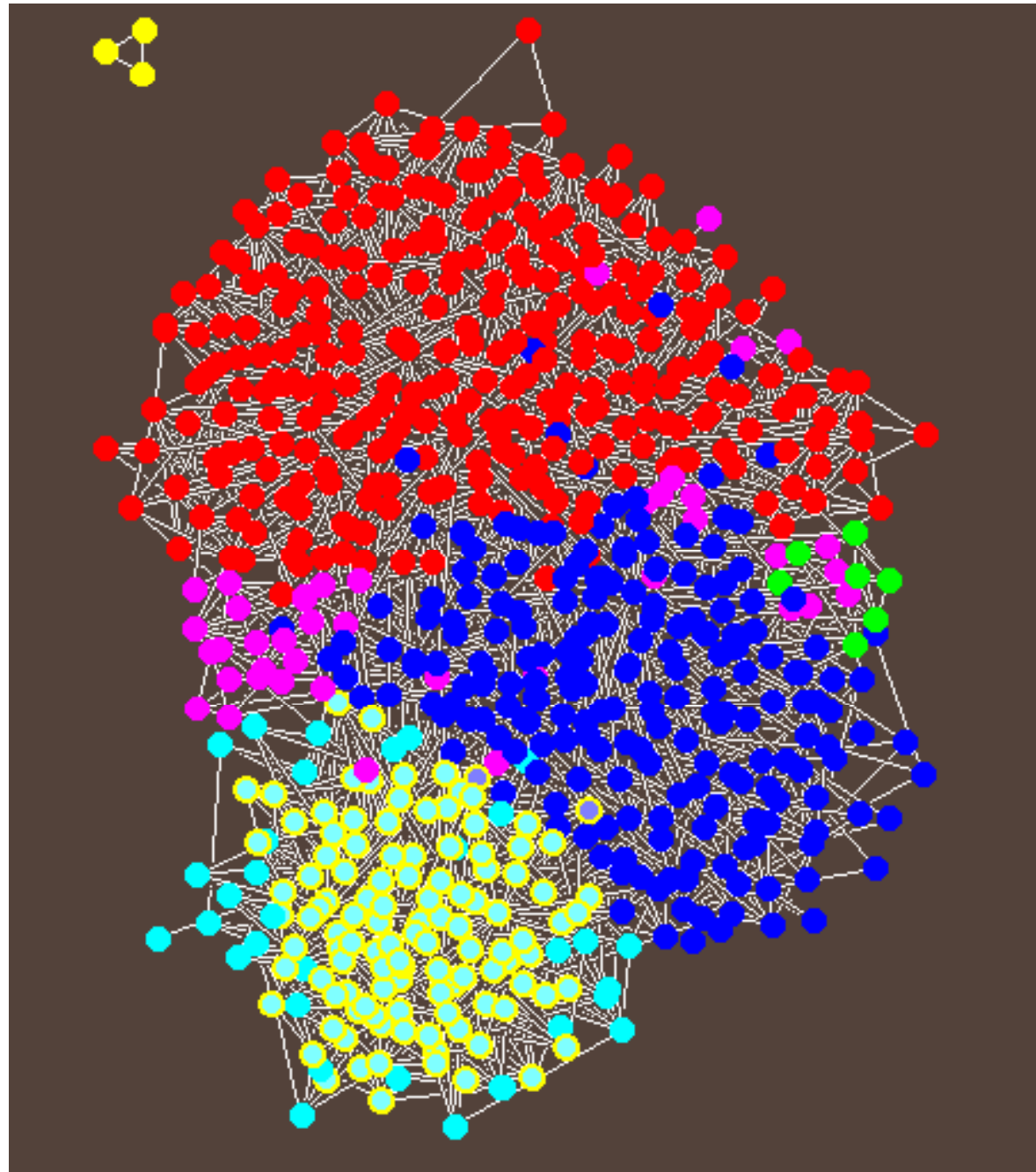
Közösségek – hallgatói-oktatói kapcsolat



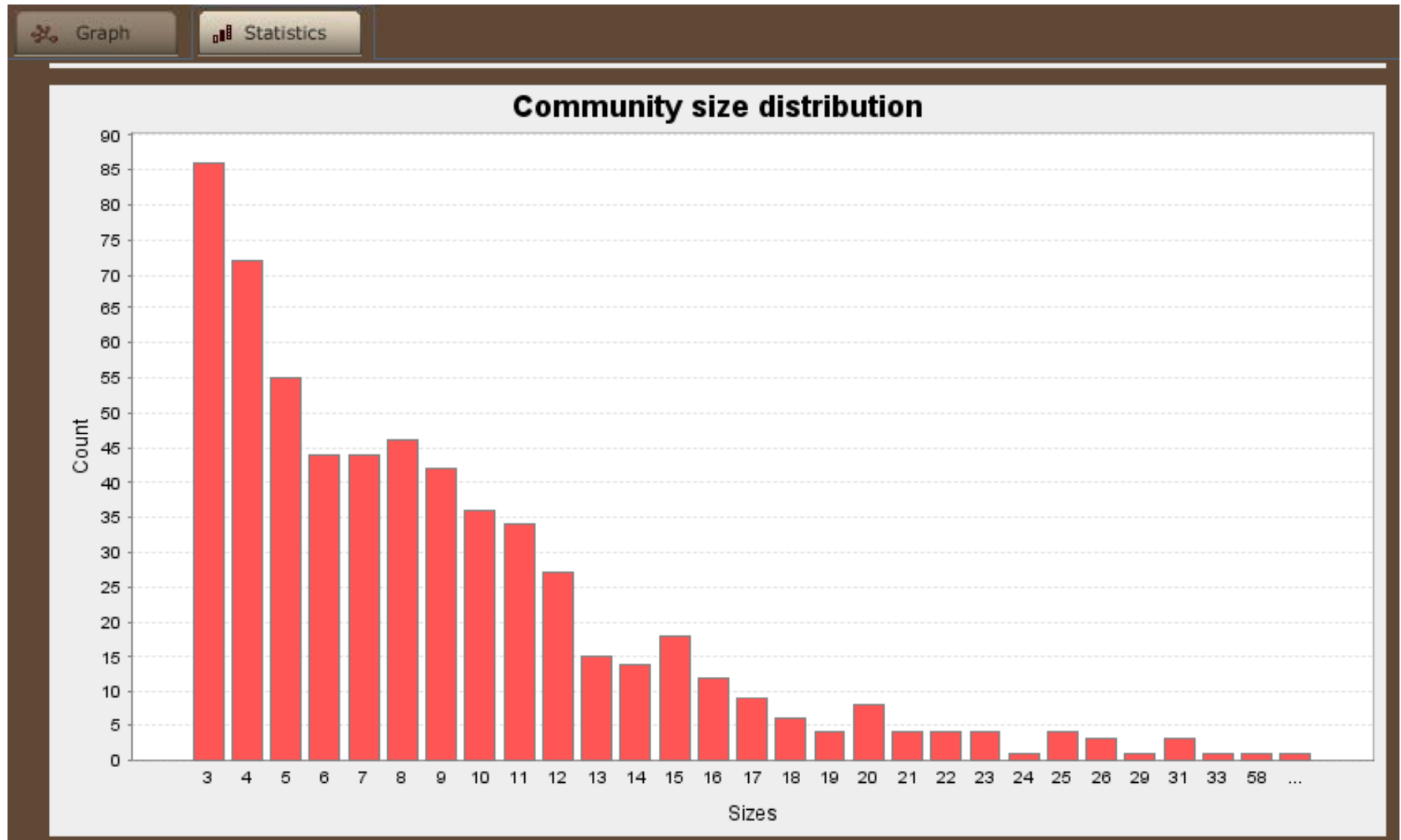
Közösségek – dolgozói kapcsolat



Közösségek – „óriás” közösség

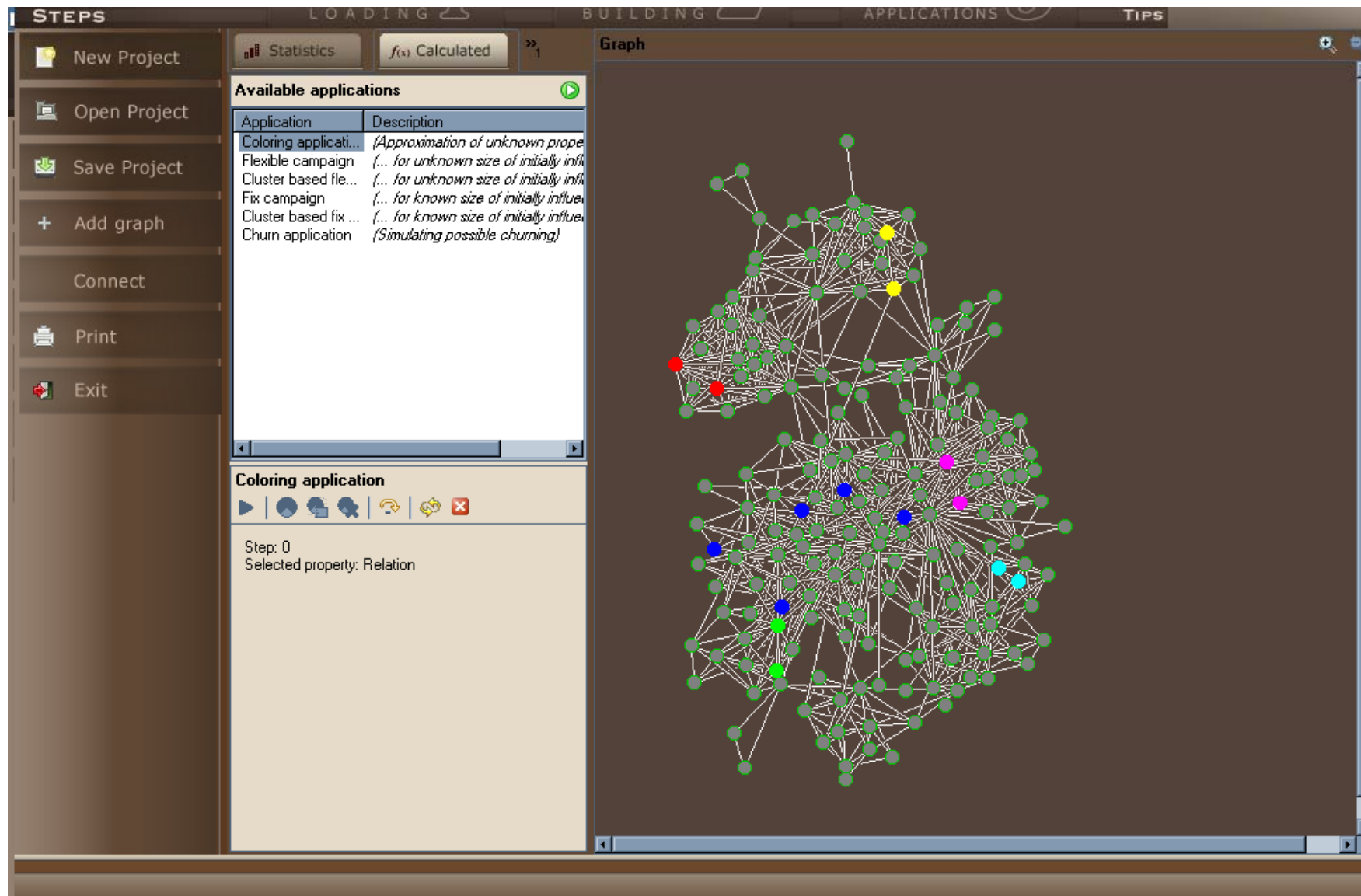


Közösségi statisztikák



Alkalmazás-attributum predikció

Inicializálás



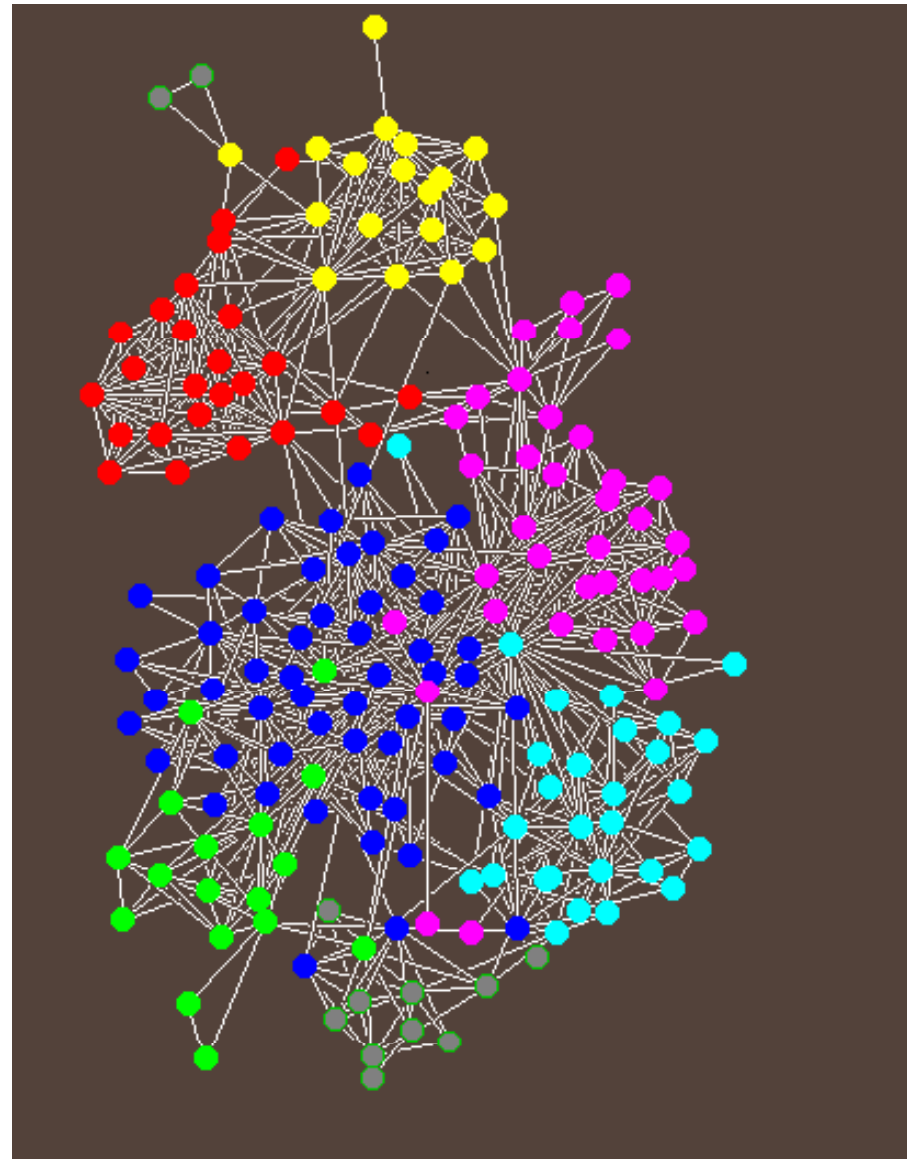
The screenshot displays a software interface for graph analysis. The interface is divided into several sections:

- STEPS:** A vertical sidebar on the left containing buttons for "New Project", "Open Project", "Save Project", "Add graph", "Connect", "Print", and "Exit".
- LOADING:** A section at the top with a "Statistics" button and a "Calculated" indicator.
- BUILDING:** A section at the top with a "Graph" label.
- APPLICATIONS:** A section at the top with a "TIPS" label.
- Available applications:** A table listing various applications and their descriptions.
- Coloring application:** A section below the table with a play button, a toolbar with icons for graph operations, and a status area showing "Step: 0" and "Selected property: Relation".
- Graph:** A large central area displaying a complex network graph with nodes and edges. Some nodes are highlighted in different colors (red, yellow, blue, green, pink).

Application	Description
Coloring applicati...	(Approximation of unknown prop...
Flexible campaign	(... for unknown size of initially infl...
Cluster based fle...	(... for unknown size of initially infl...
Fix campaign	(... for known size of initially influe...
Cluster based fix ...	(... for known size of initially influe...
Churn application	(Simulating possible churning)

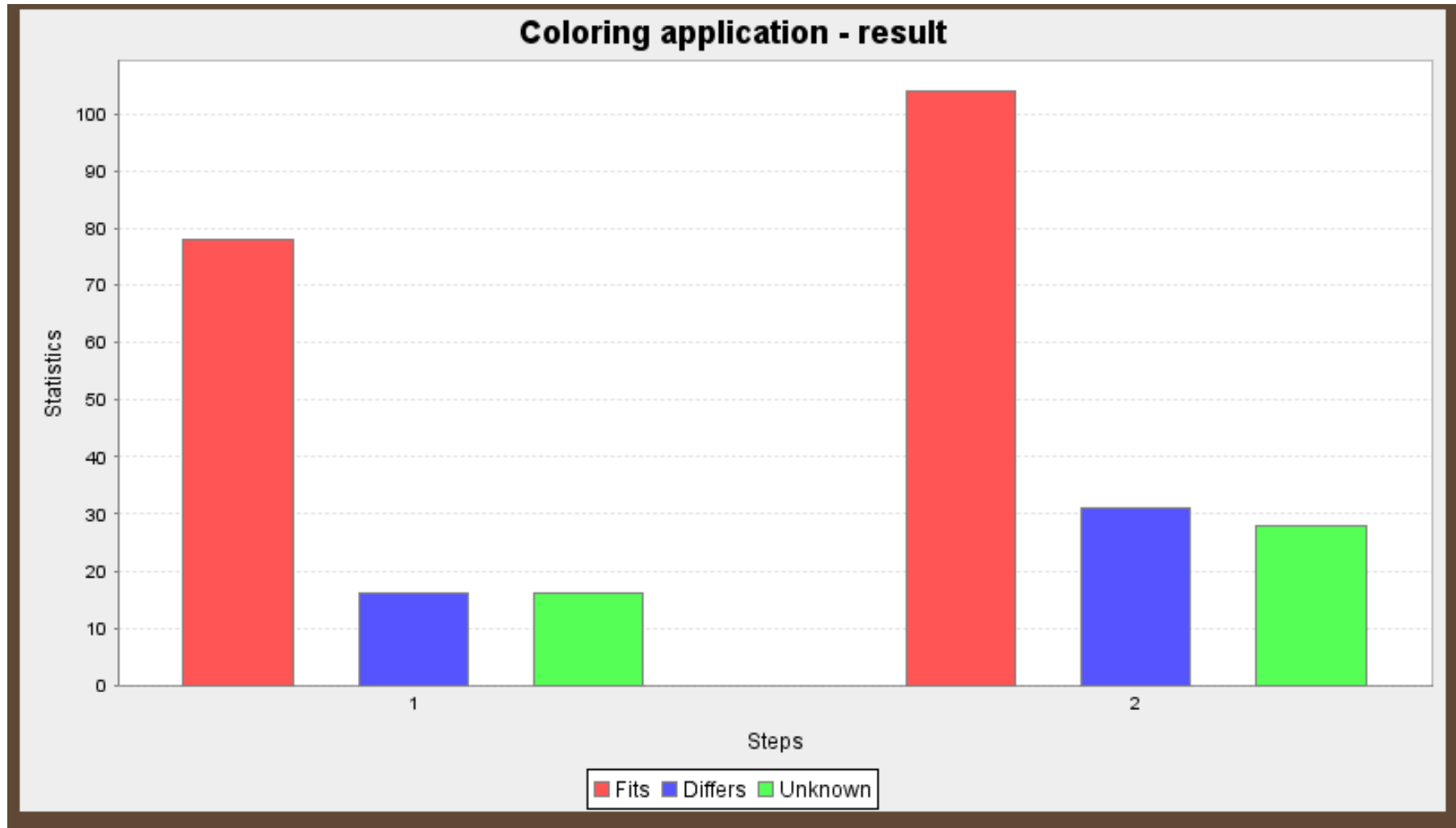
Attributum predikció

Eredmény



Attributum predikció

Statisztika



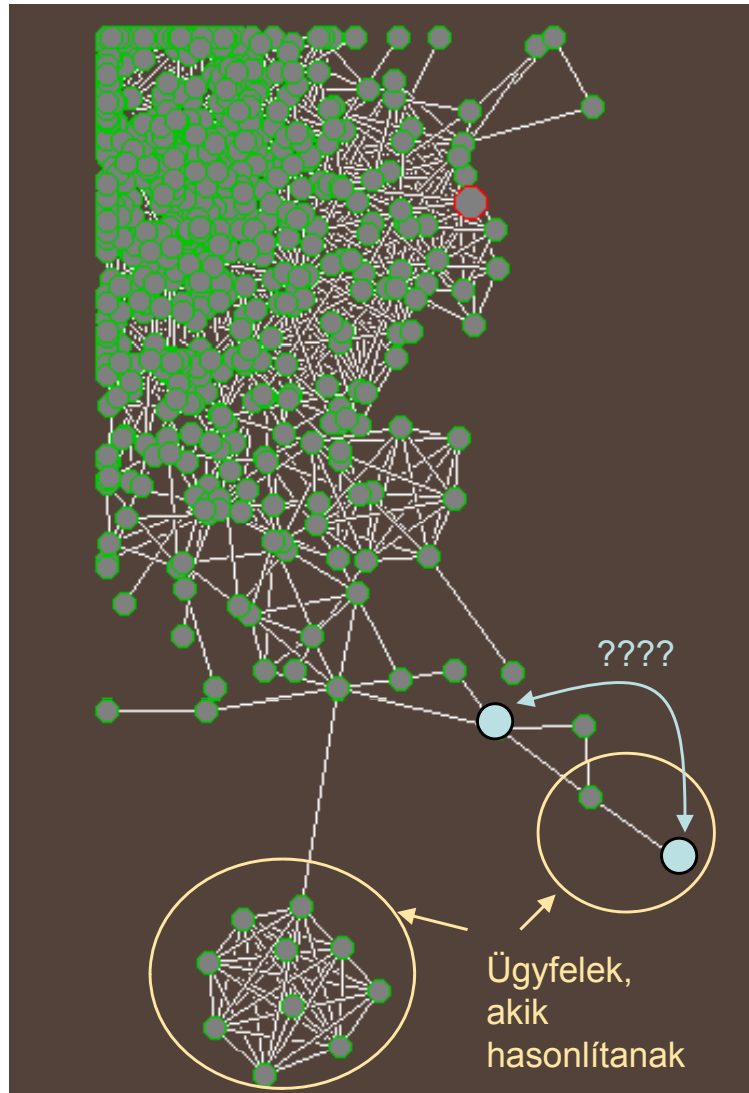
Üzleti célú alkalmazások

- Adattisztítás
- Keresztértékesítés
- Ügyfél szegmentáció

Adattisztítás – de-duplikálás

- **Alapfeladat:** Vállalati adatvagyonban ügyfél-redundancia kiszűrése (adatbázisba többszörösen felvitt ügyfelek beazonosítása, egyedi azonosító hozzárendelése)
- **Jellemző megoldás:** egy olyan szabályrendszer (távolság-függvény) kidolgozása, mely két ügyfelet akkor rendel össze, ha a közöttük lévő távolság kisebb egy előre meghatározott értéknél.

Mi itt a gráf?



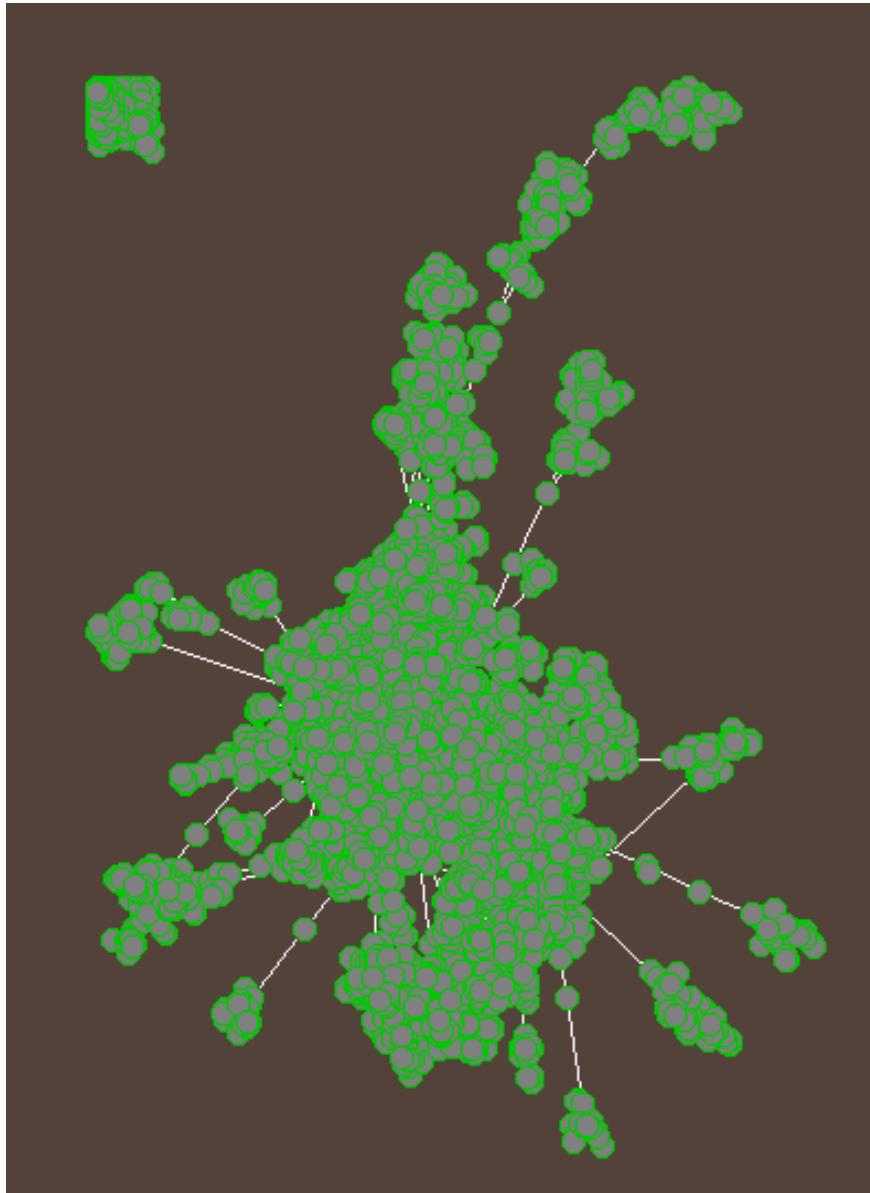
- Csúcsok: az összes meglévő „ügyfél”
- Élek: két ügyfelet akkor kötünk össze, ha a szabály rendszer szerint „kicsi” a távolság közöttük.

- Az élekhez súlyok is rendelhetők („távolság” két ügyfél között) – így az élek száma ritkítható, amennyiben csak azokat az éleket húzzuk be, ahol két ügyfél nagyon hasonlít

Netflix – „DM világbajnokság”

- **Alapadatbázis:** egy videó kölcsönző adatai, benne 480.000 ügyfél, akik 17.000 film közül választhatnak. Ami rendelkezésre áll:
 - *Egy adott ügyfél milyen filmeket kölcsönzött ki?*
 - *Mikor kölcsönözte ki?*
 - *Milyennek értékelte a filmet?*
- **Szűrés:** csak azokat elemeztük, akik legfeljebb 50 filmet kölcsönöztek ki

Mi itt a gráf?

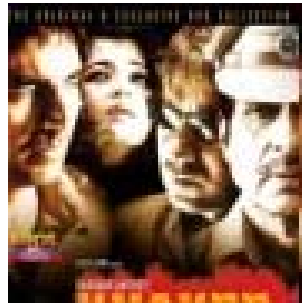


- Csúcsok: a kikölcsönözhető filmek
- Élek: két film között akkor van él, amikor legalább 4 ügyfél kivette mind a két filmet
- Közösség: azon filmek kerülnek egy közösségbe, amiket gyakran vesz ki egyazon személy

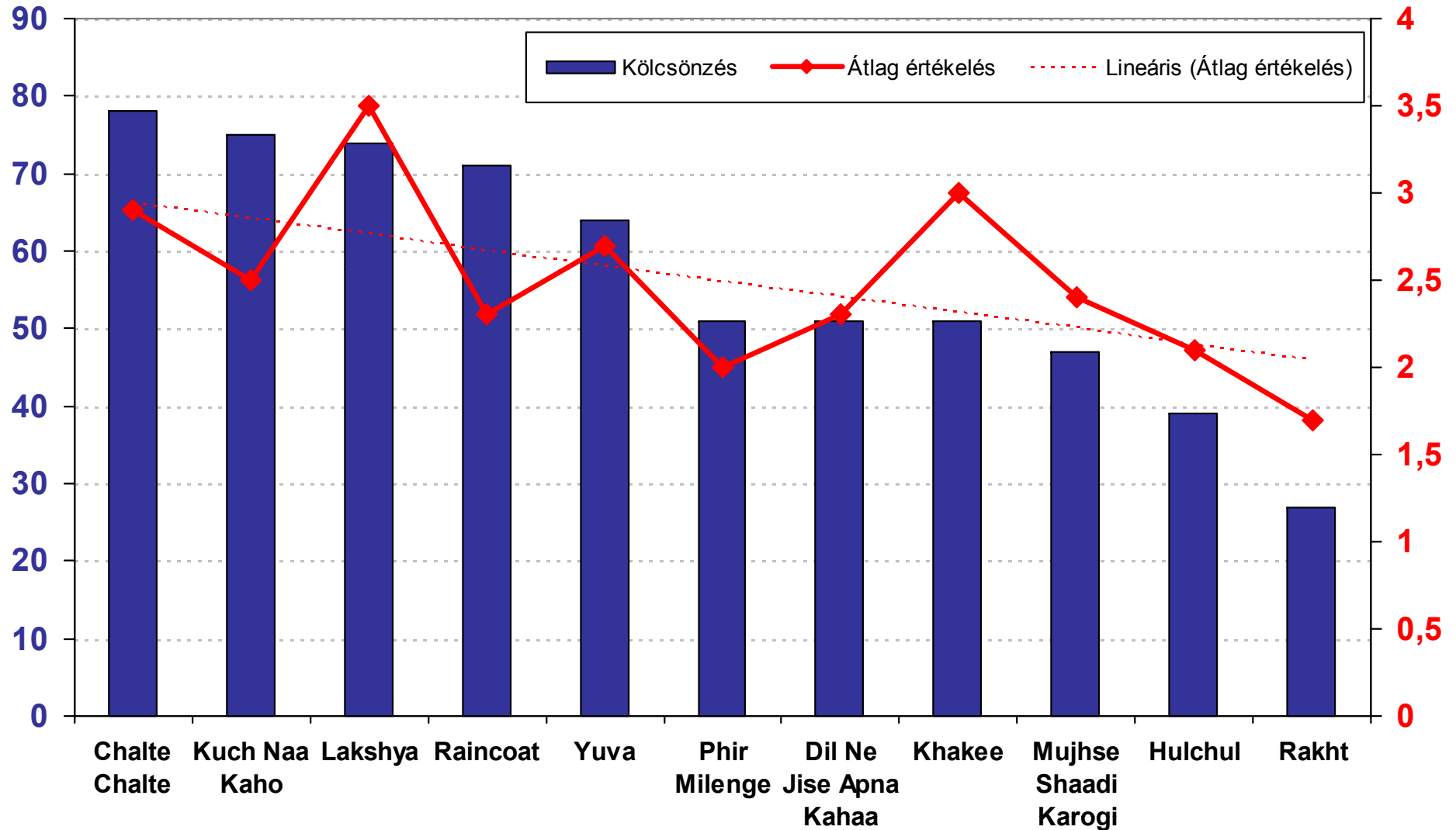
Közösség: indiai filmek (11 db)

A vizsgált közösség:

- Kuch Naa Kaho (2003)
- Chalte Chalte (2003)
- Phir Milenge (2004)
- Mujhse Shaadi Karogi (2004)
- Rakht (2004)
- Khakee (2004)
- Hulchul (2004)
- Lakshya (2004)
- Dil Ne Jise Apna Kahaa (2004)
- Yuva (2004)
- Raincoat (2005)



Regresszió - indiai filmek



Köszönöm a figyelmet!