

Az OTKA K68669 azonosítójú,

## **Komplex hálózatok moduláris szerkezete**

című pályázat szakmai beszámolója

### **1. Bevezetés**

Az utóbbi évtizedben a hálózati megközelítés több fontos sikert hozott biológiai, technológiai, szociológiai és gazdasági problémák vizsgálatánál, melynek eredményeként mára a komplex hálózatok témaköre önálló, interdiszciplináris tudományterületté fejlődött [1-2]. Egy kiemelten fontos témakört képviselnek ezen a belül a hálózati csoportokkal (más néven modulokkal, klaszterekkel, csoportosulásokkal) kapcsolatos kutatások [3]. A pályázat résztvevői már rendelkeztek korábbi eredményekkel a hálózati csoportosulások témakörében, melyek közül legfontosabb egy klikk perkoláción alapuló csoportkeresési módszer kifejlesztése volt [4]. Ennélfogva a pályázathoz kapcsolódó kutatások egyrészt a klikk perkolációs módszer alkalmazási körének bővítését, valamint magának a módszernek a továbbfejlesztését tűzték ki célul. A hálózati csoportok szervesen kapcsolódnak a hálózati hierarchia témaköréhez is, hiszen egy köztes szerveződési szintet képviselnek az individuális csúcsok és az egész hálózat szintje között. A kutatások másik célterületét így a hálózati hierarchiához kötődő különböző problémák vizsgálata tette ki. A következőkben a kutatómunka eredményeit szűkebb témakörökre bontva mutatjuk be.

### **2. A klikk perkolációs módszer alkalmazásai**

A csoportok központi szerepe miatt az utóbbi években számos alternatív hálózat klaszterező módszert dolgoztak ki [3]. Ezek döntő többsége izolált csoportokat keres, azaz egy csúcs egyszerre csak egy csoporthoz tartozhat. Ezzel szemben a valódi hálózatokban a csoportok gyakran átfednek egymással, (pl. mindannyian egyszerre vagyunk tagjai a családuknak, a baráti körünknek és a munka helyi közösségünknek). Erre a problémára kínál megoldást a klikk perkolációs módszer [4], mely természetes módon engedi meg a csoportok közti átfedéseket. Ebben a megközelítésben a csoportok  $k$ -klikkek ( $k$  csúcsból álló, teljesen összekötött részgráfok) perkolációs klasztereinek felelnek meg, ahol két  $k$ -klikket akkor mondunk szomszédosnak, ha  $k-1$  csúcsuk közös. Egy külön érdekesség, hogy az Erdős-Rényi-féle véletlen gráfban az élbekötési valószínűség változtatásával klikk perkolációs átalakulások figyelhetők meg (minden egyes  $k$  értékhez tartozik egy kritikus élbekötési valószínűség), melyek a jól ismert él perkolációs átalakulás általánosításának felelnek meg [5].

#### **2.1 Időfejlődő csoportok vizsgálata**

A hálózati csoportokkal foglalkozó korábbi kutatások szinte kivétel nélküli „statikus”, azaz időben állandó csoportokkal foglalkoztak. Ezzel szemben számos valós hálózatban a kapcsolatok idővel változhatnak, és ennek megfelelően a csoportok is egy „dinamikus” képet mutatnak, azaz egy adott csoport összetétele időről időre változhat, (elég hosszú idő elteltével akár teljesen le is cserélődhet). A pályázathoz kapcsolódó kutatások egyik célja volt a csoport időfejlődés statisztikus tulajdonságainak vizsgálata nagyméretű társaskapcsolat hálózatokban [6-11].

A vizsgált rendszerekben az egyes időlépésekhez tartozó gráfokat a klikk perkolációs módszerrel klasztereztük, így előállt egy-egy „pillant felvétel” sorozat a hálózatokban található csoportokról. Ezek

összefűzése időfejlődő csoportokká egy nem triviális feladat, hiszen a csoportok mérete lépésről lépésre nőhet vagy csökkenhet, csoportok összeolvadhatnak vagy széteshetnek több kisebb csoportra, teljesen új csoportok jelenhetnek meg és régiak tűnhetnek el. Kidolgoztunk egy eljárást, mely lehetővé tette az időfejlődő csoportok nyomon követését, a módszer részletei megtalálhatók a [6]. hivatkozásban.

Az időfejlődő csoportok számos statisztikáját vizsgáltuk, az egyik legérdekesebb jelenséget a csoportok tag-autokorrelációs függvénye mutatta, mely az időben átlagosan gyorsabban csengett le nagyobb méretű csoportok esetén a kisebbekhez viszonyítva [6-11]. Ez azt jelenti, hogy a nagyobb csoportok dinamikusabban változnak, bennük gyorsabban cserélődnek a tagok. Ennek jellemzésére bevezettük a csoport stacionaritás fogalmát is, mely a szomszédos időlépések közti tagkorreláció átlaga a csoport életútja során. Az általunk vizsgált hálózatok egy nagyon érdekes összefüggést mutattak a stacionaritás, a csoportméret és a csoport várható élettartama között. (Egy csoport élettartama azon időlépések száma, amíg a csoport jelen van a rendszerben, ami tekinthető egy fajta „fitness” értéknek is, hiszen a „fitt” csoportok sokáig élnek, míg az alacsony fitnessszel rendelkezők hamar eltűnnek). Kiderült, hogy ha a csoportméret és a stacionaritás függvényében ábrázoljuk a várható élettartamot, akkor annak maximuma nagyobb stacionaritás értékeknél található kis csoportok esetén, míg kisebb stacionaritásnál nagy csoportok esetén [6-11]. Ez azt jelenti, hogy a kis csoportok esetén az optimális, ha tagösszetételük az időben nem, vagy csak keveset változik. Ezzel szemben a nagyobb csoportoknak állandóan meg kell újulniuk a fennmaradáshoz, a tagoknak folyamatosan cserélődniük kell ahhoz, hogy a csoport tovább tudjon élni. Ezen felül megmutattuk azt is, hogy tagokhoz illetve a csoportokhoz tartozó kapcsolaterősségek segítségével meg lehet jósolni a tagok kiválási, valamint a csoportok szétesési valószínűségét is [6-11].

Ezekhez az eredményekhez szorosan kapcsolódnak a csoport-hálózat időfejlődésével kapcsolatos korábbi kutatások is [11-13]. (Ebben a hálózatban a csúcsok egy-egy csoportnak felelnek meg, melyek akkor vannak összekötve, ha van köztük átfedés). Az eredmények szerint a csoport-hálózat dinamikáját is a preferenciális csatlakozás mechanizmusa vezérli, a csúcsok hálózatához hasonlóan.

## 2.2 Bioinformatikai hálózatok vizsgálata

A klikk perkolációs módszert alkalmaztuk bizonyos mikroRNS-ek és az általuk gátolt mRNS-ek hálózatának tanulmányozására: azonosítottuk a mikroRNS-ek funkció csoportjait és javaslatot tettünk kísérletekre [14]. Vizsgáltunk továbbá sejten belüli jelátviteli hálózatokat. A speciálisan erre a célra indított <http://Signalink.org> weboldalon bemutatjuk a [15]-es cikk három fő eredményét: (i) egységes kézi gyűjtési adatok, jól dokumentált gyűjtési szabályokkal, (ii) a jelátviteli útvonalakban résztvevő fehérjék és gyógyszer célpont fehérjék bioinformatikai előrejelzése, és (iii) az előrejelzett útvonal fehérjék közül hat Notch fehérje kísérletes ellenőrzése *C. elegans*-ban.

## 3. A klikk perkolációs módszer tovább fejlesztése

### 3.1 A klikk perkolációs módszer kiterjesztése irányított hálózatokra

A pályázathoz kapcsolódó kutatások egyik célja a klikk perkolációs módszer kiterjesztése volt irányított hálózatokra [16-18]. Ezekben a rendszerekben az élek aszimmetrikus kapcsolatot jelölnek, (pl. a hálózatban lezajló terjedési folyamatok irányát, vagy emberek közti kapcsolatok hálózatában alá-fölé rendeltséget, stb.). A módszer irányított esetre történő átültetésénél az eredeti irányítatlan *k*-klikkeket lecseréltük olyan irányított *k*-klikkekre, melyekben az élek iránya csak néhány megengedett konfigurációt vehet fel [16]. Az irányított klikk perkolációs módszer segítségével nagyméretű biológiai, szóasszociációs, e-mail és web hálózatokat vizsgáltunk. A kapott csoportokat ahol lehetett, a csúcsokhoz társítható címkék segítségével ellenőriztük. Az eredmények azt mutatták, hogy a vizsgált hálózatok két nagy osztályba sorolhatók az irányított csoportosulások viselkedése alapján [16].

Az irányított csoportosulások átfedéseiben található csúcsok kiemelt központi szerepét külön is

tanulmányoztuk e-maileken, illetve kérdőíveken alapuló társaskapcsolat hálózatokban [17]. Eredményeink szerint a külön fajta centralitást jellemző mennyiségek (betweenness centrality, closeness) érdekes eltérő viselkedést mutatnak a csoportok átfedéseiben található csúcsokon a két hálózat esetén. Ezen felül megvizsgáltuk, hogy miként módosul az Erdős-Rényi-féle véletlen gráfban a  $k$ -klick perkolációs átalakulás kritikus pontja az irányított esetre történő áttéréskor [16,18].

### **3.2 A klikk perkolációs módszer kiterjesztése súlyozott hálózatokra**

A komplex hálózatok jelentős részénél a csúcsokat összekötő élekhez egyfajta súly vagy intenzitás is társítható, mely megadja, hogy mennyire szoros vagy gyenge a kapcsolat az él két vége között. Ezen élsúlyok figyelembe vétele a hálózat klaszterezésben egy további lehetőség az eredmények finomítására, ezért céljaink közt szerepelt a klikk perkolációs módszer kiterjesztése súlyozott hálózatokra is [18-19]. Az alapötletet a hálózati motívumok (néhány csúcsból álló, fix konfigurációjú irányított részgráfok) vizsgálatánál bevezetett motívum intenzitás fogalma adta, mely a motívum élsúlyainak mértani közepének felel meg. A súlyozott klikk perkolációs módszernél ezt a mennyiséget ültettük át a  $k$ -klikkekre: egy adott súlyküszöb érték mellett a figyelembe vett  $k$ -klikkek intenzitása nem lehet kisebb mint a küszöbérték. A súlyozott klikk perkolációs módszerrel társszerzőségi és tőzsdei hálózatokat vizsgáltunk [18-19]. Az eredmények megmutatták, hogy az élsúlyok közti korrelációk jelentősen eltérnek a vizsgált valós hálózatokban egy (azonos méretű és sűrűségű) véletlenszerű hálózathoz képest.

## **4. Címkezett hálózatok statisztikus tulajdonságainak vizsgálata**

Számos valós hálózat esetén a csúcsokhoz címkék (tulajdonságok, attribútumok, annotációk) társíthatók, melyek nagyon sokféle információt tartalmazhatnak a csúcsokra vonatkozóan. Mint korábban említettük, ezek a címkék egy alapvető eszközt nyújtanak a csoportfelbontás ellenőrzésekor. Emellett szorosan kapcsolódnak a hálózati hierarchia problémaköréhez is, ugyanis sok esetben maguk is hierarchikusan szerveződnek, (egyes speciális címkék rész kategóriái lehetnek más, általánosabb jelentésű címkéknek). Ennélfogva a pályázathoz kapcsolódó kutatások során olyan címkezett hálózatok alapvető statisztikai tulajdonságait vizsgáltuk melyeknél a címkék egy (hierarchikus) irányított aciklikus gráfba rendezhetők alá-fölé rendeltség alapján [20-21]. A címkék és a hálózati hierarchia között több érdekes összefüggést is találtunk, melyek közül legfontosabb a címke indukált részgráfokra történő leszűkítés esetén tapasztalt ön hasonlóság volt. (Egy címke indukált részgráf az adott címkét hordozó csúcsok által kifizített részgráfnak felel meg.) A címke indukált részgráfok éleinek átlagos száma hatványoszerűen változott több nagyságrenden keresztül a részgráf méretének függvényében [20]. Ezen viselkedés jellemzésére vezettük be a címke asszortativitási exponenst, melynek nagyságából pl. meg lehet állapítani, hogy az adott hálózatban teljesül-e az, hogy a hasonló (az aciklikus gráfban egymáshoz közel található) címkékkel rendelkező csúcsok nagyobb valószínűséggel kötődnek egymáshoz (azaz a hálózat címke asszortatív). A címke indukált részgráfok átlagos klaszterezettségi együtthatója a részgráf méretének függvényében csökkenő tendenciát mutatott, viszont a csökkenés mértéke sokkal kisebb volt egy véletlen gráfhoz viszonyítva [21].

## **5. Multifraktál véletlen gráf generátor kidolgozása**

A hálózati hierarchia tanulmányozásához kapcsolódóan kifejlesztettünk egy multifraktál élbekötési mértéken alapuló általános véletlen gráf generáló módszert [22-23]. A nevezett mértéket egy viszonylag egyszerű szerkezetű, cellákra osztott generáló mértékből állítjuk elő, rekurzívan szorozgatva a cellákat magával a generáló mértékkal. Az iteráció előrehaladtával egy egyre összetettebb, hierarchikus struktúra áll elő, és a legyártani kívánt hálózat éleit ebből a multifraktálból származó valószínűségekkel kötjük be. Megmutattuk, hogy ezzel a módszerrel (megfelelő generáló mérték

választás mellett) nagyon sokféle eltérő véletlen hálózat állítható elő, és egy adott tulajdonsághoz (pl. egy adott fokszámeloszláshoz) tartozó generáló mérték egy szimulációs hőkezelésen alapuló eljárás keretében könnyen visszakereshető [22].

A multifraktál gráf generátor segítségével előállított hálózatokban egy érdekes csúcs izolációs effektust tapasztaltunk extrém nagy gráf méretek esetén, (a valós hálózatok mérettartományában ez általában elhanyagolható). Megmutattuk, hogy ennek hátterében az áll, hogy a multifraktál fokszámeloszlást meghatározó vetülete egy kitüntetett irányba esik [23]. Ez alapján bevezettünk egy apró módosítást a konstrukcióban, mely a multifraktál elforgatásával kiküszöböli a túlzott mértékű csúcs izolációt.

## Hivatkozások

- [1] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47-97, (2002).
- [2] J. F. F. Mendes and S. N. Dorogovtsev, *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford, (2003).
- [3] S. Fortunato, Community detection in graphs. *Physics Reports* **486**, 75-174, (2010).
- [4] G. Palla, I. Derényi, I. Farkas and T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814-818 (2005).
- [5] I. Derényi, G. Palla and T. Vicsek, Clique percolation in random networks. *Phys. Rev. Lett.* **94**, 160202 (2005).
- [6] G. Palla, A.-L. Barabási and T. Vicsek, Quantifying social group evolution. *Nature* **446**, 664-667 (2007).
- [7] G. Palla, T. Vicsek and A.-L. Barabási, Community dynamics in social networks. *Fluctuation and Noise Letters* **7**, L273-L387 (2007).
- [8] G. Palla, T. Vicsek and A.-L. Barabási: Statistical Properties of Community Dynamics in Large Social Networks. *International Journal of Agent Technologies & Systems* **1(4)**, 1-16, (2009).
- [9] Palla G. és Kertész J., Szociofizika: humán kapcsolatok hálózata nagy skálán. *Fizikai Szemle* **2008/6**, 217 (2008).
- [10] Palla G. Barabási A.-L. és Vicsek T., Társas kapcsolatok hálózata, *Természet Világa* **139 (3)**, 108-110 (2008).
- [11] G. Palla, P. Pollner, A.-L. Barabási and T. Vicsek, Social group dynamics in networks. *Adaptive Networks*, (ed.: T. Gross, H. Sayama, Springer Berlin/Heidelberg), Chap.2 p.11-38 (2009).
- [12] P. Pollner, G. Palla and T. Vicsek: Preferential attachment of communities: The same principle but at a higher level. *Europhys. Lett.* **73**, 478-484 (2006).
- [13] Palla G., Derényi I., Farkas I., Pollner P. és Vicsek T., A természet és a társadalom komplex hálózataiban található átfedő csoportosulások feltárása. *Műszaki Szemle (Az Erdélyi Magyar Műszaki Tudományos Társaság Kiadványa)* **42/2008**, 9-18 (2008).
- [14] G. Boross, K. Orosz and I. J. Farkas, Human microRNAs co-silence in well-separated groups and have different predicted essentialities. *Bioinformatics* **25**, 1063 (2008).
- [15] T. Korcsmáros, I. J. Farkas, M. S. Szalay, P. Rovó, D. Fazekas, Z. Spiró, C. Böde, K. Lenti, T. Vellai, P. Csermely, Uniformly curated signalling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* **26**, 2042 (2010).
- [16] G. Palla, I. J. Farkas, P. Pollner, I. Derényi and T. Vicsek, Directed network modules. *New Journal of Physics* **9**, 186 (2007).
- [17] P. Pollner, G. Palla, D. Ábel, A. Vicsek, I. J. Farkas, I. Derényi and T. Vicsek: Centrality properties of directed module members in social networks. *Physica A* **387**, 4959-4966 (2008).

- [18] G. Palla, D. Ábel, I. J. Farkas, P. Pollner, I. Derényi and T. Vicsek:  $k$ -clique percolation and clustering. *Handbook of Large-scale Random Networks* (ed.: B. Bollobás, R. Kozma, D. Miklós, Springer) Chap. 9 p.369-408 (2009).
- [19] I. J. Farkas, D. Ábel, G. Palla and T. Vicsek, Weighted network modules. *New Journal of Physics* **9**, 180 (2007).
- [20] G. Palla, I. J. Farkas, P. Pollner, I. Derényi and T. Vicsek, Fundamental statistical features and self-similar properties of tagged networks. *New Journal of Physics* **10**, 123026 (2008).
- [21] P. Pollner, G. Palla and T. Vicsek, Clustering of tag-induced subgraphs in complex networks. *Physica A* **389**, 5887-5894 (2010).
- [22] G. Palla, L. Lovász and T. Vicsek, Multifractal network generator. *Proc. Natl. Acad. Sci. USA* **107**, 7640-7645 (2010).
- [23] G. Palla, P. Pollner and T. Vicsek, Rotated multifractal network generator. Accepted for publication in JSTAT, arXiv:1101.2338 (2010).