

Az OTKA F47203 azonosítójú,
Komplex hálózatok vizsgálata
című pályázat szakmai zárójelentése

1. Bevezetés

Az élővilág (beleértve az emberek világát is) számos komplex alrendszere sikeresen reprezentálható a rendszert alkotó elemek közti kölcsönhatások hálózatának segítségével. Az ilyen komplex hálózatoknak megfelelő gráfok általában már a topológia szintjén is számos nem triviális jelenséget produkálnak, úgy mint skála független fokszámeloszlás megjelenése, fürtösödés, anomális átmérő, érdekes foksám korrelációk, stb. [1, 2, 3, 4, 5].

Ezen tágabb témakörön belül a pályázathoz kapcsolódó kutatások két területre koncentráltak. A pályázat résztvevőinek korábbi eredményei a hálózatok statisztikus fizikájához kötődnek, így 2004-ben ezen korábbi kutatások folytatásaként hálózati Hamilton-függvények visszafejtésével foglalkoztunk. 2005-től viszont egy másik nagyon dinamikus fejlődő részterület felé fordultunk, mely a hálózatok moduláris szerkezetét, a hálózatokban található csoportosulásokat, klasztereket, modulokat vizsgálja.

2. Hálózati Hamilton-függvények visszafejtése

A komplex hálózatok témakörének egyik fontos kutatási területét a hálózatok topológiai változásainak statisztikus fizikai tárgyalása teszi ki. Elsőként Burda és munkatársai írtak fel a statisztikus fizikában megszokott állapotösszeget egy hálózat lehetséges állapotainak segítségével [6, 7], nem sokkal később Berg és Lässig ezen formalizmus segítségével azt vizsgálták, hogy különböző Hamilton-függvények milyen korrelációkra vezetnek a hálózatban [8]. Az említett, a gráf szerkezetétől függő Hamilton-függvény megválasztására több lehetőség kínálkozik [8, 9], a pályázat résztvevőinek korábbi kutatásai ezen különböző Hamilton-függvényeknek a gráf struktúrájára gyakorolt hatásait vizsgálták [10, 11, 12]. Ebben a megközelítésben a hőmérséklet az átrendeződések zajszintjének felel meg, a Hamilton-függvény pedig egy (bizonyos szempontból) optimális konfigurációt kódol. Az eredmények szerint több olyan egyszerű Hamilton-függvény is adható, melynél a hőmérséklet (zajszint) csökkentésével a hálózat olyan strukturális átrendeződésen megy át, mely analóg a statisztikus fizikában megszokott fázisátalakulásokkal.

A pályázathoz kapcsolódó kutatások során elsőként (a topologikus fázisátalakulásokkal foglalkozó vizsgálatok folytatásaként) kidolgoztunk egy módszert, mely lehetővé teszi a valós hálózatok átrendeződéseinek elemzéséből a hálózat időfejlődését (dinamikáját) meghatározó Hamilton-függvény visszakövetgetését [13]. Eljárásunkkal olyan hálózatok vizsgálhatóak, ahol az átrendeződési folyamatok kielégítik a részletes egyensúly feltételeit, valamint az időegységre jutó él-átrendeződések száma mellett eltörpül a megjelenő új csúcsok és élek száma. Módszerünk továbbá feltételezi, hogy a visszafejtendő energiafüggvény előállítható az egyes csúcsok fokszámától függő tagok összegeként. Megvizsgálva több nagyméretű valós hálózatot arra az eredményre jutottunk, hogy az átrendeződéseket irányító empirikus Hamilton-függvényeket egy olyan univerzális (az egyes csúcsok fokszámától függő) függvénnyel lehet megadni, melynél az egyes csúcsok járuléka a teljes energiához a csúcs fokszámának és a fokszám logaritmusának szorzatával arányos [13]. Megmutattuk, hogy ez az energia alak konzisztens a természetben található növekvő hálózatoknál tapasztalt preferenciális kapcsolódási szabállyal. (A preferenciális csatolási szabálynak engedelmessé váló hálózatok esetén egy új csúcs bekötésénél a már létező csúcsok a fokszámukkal arányos valószínűséggel válnak az új csúcs szomszédává, ezáltal nyerve egy új élt.)

Összeségben a hálózatok statisztikus fizikájával kapcsolatos eredményeink egy új nézőpontot kínálnak a szerkezetileg különböző hálózatok összehasonlításához, rendszerezéséhez és tovább mélyítik a kapcsolatot a hagyományos statisztikus fizika és a komplex hálózatokelmélete között.

3. Hálózatok moduláris szerkezete

3.1. Hálózati csoportok, modulok, klaszterek

A komplex hálózatok témakörének egy másik nagyon fontos új területe az ún. csoportosulásokkal (más néven hálózati klaszterekkel, modulokkal, vagy kohézív csoportokkal) foglalkozik. Ezek a szerkezeti egységek az átlagosnál sűrűbben összekapcsolt részgráfoknak felelnek meg, és egyelőre nincs általánosan elfogadott, egységes definíciójuk [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]. Az ilyen alapegységek (pl. hasonló funkciójú proteinek [27, 28], gazdasági szektorok [29], szociális csoportok [22, 30], kooperatív játékosok [31, 32], stb.) nagyon fontos szerepet játszhatnak a hálózat szerkezeti és funkcionális felépítésében, működésében.

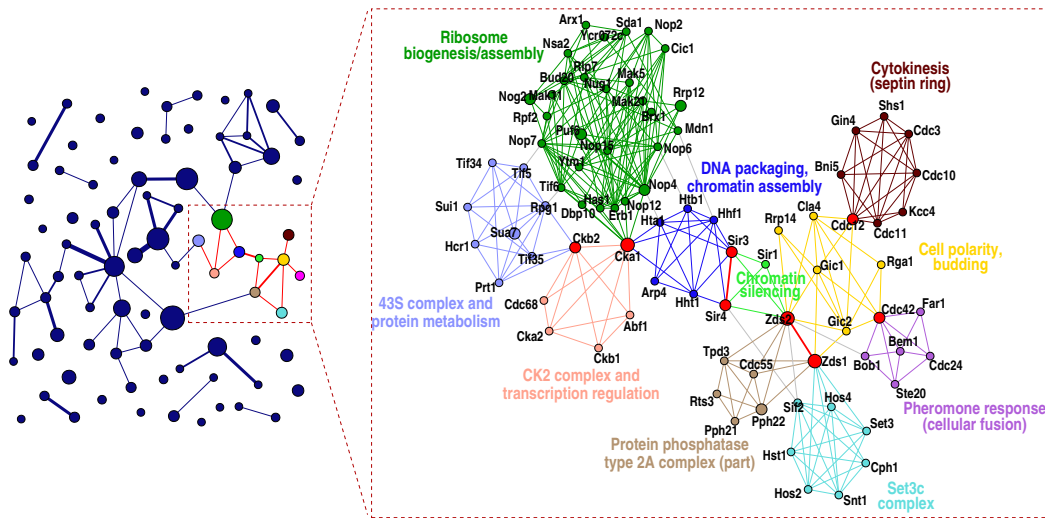
A legtöbb hálózati csoportosulás kereső módszer a hálózat egyre kisebb részekre történő darabolásán alapszik. Ennek legnagyobb hátránya az, hogy az így kapott csoportosulások nem fedhetnek át egymással. Ezzel ellentétben számos komplex rendszer esetén a csoportosulások nem izoláltak, egy-egy csúcs

egyszerre akár több csoporthoz is tartozhat, pl. egy protein egyszerre több funkciót is elláthat [33], vagy egyes emberek egyszerre több szociális csoportnak is tagjai lehetnek [34].

3.2. A klikk perkolációs módszer (CPM)

2005-ben jelentős eredményeket értünk el a hálózati csoportosulás keresés fent említett hiányosságának kiküszöbölésében [35, 36]. Az általunk javasolt új csoportosulás kereső módszer a k -klikk perkoláció jelenségén alapszik, és megengedi a csoportok közti átfedéseket. Maguk a k -klikkek olyan k csúcsból álló részgráfoknak felelnek meg, melyekben minden csúcs minden másik csúccsal össze van kötve. Két k -klikket szomszédosnak mondunk, ha $k - 1$ csúcsuk közös, és a k -klikk perkolációs klaszter pedig olyan k -klikkek uniójaként áll elő, melyben bármely két k -klikkből eljuthatunk a másikba szomszédos k -klikkek sorozatán keresztül. A mi megközelítésünkben a csoportosulások mind egy-egy k -klikk perkolációs klaszternek felelnek meg [35, 36]. Ezek átfedhetnek egymással, hiszen a rendszerben található olyan k -klikk párokat, melyek ugyan különböző k -klikk perkolációs klaszterhez tartoznak, viszont átfednek egymással néhány ($k - 1$ -nél kevesebb) csúcsban. Az ilyen átfedésekben található csúcsok mindkét csoportosulásnak tagjai lesznek. Az átfedések révén természetes módon származtathatjuk a csoportosulások hálózatát is: benne a csúcsok az egyes csoportosulásoknak felelnek meg, és két csúcs akkor van összekötve, ha az adott két csoportosulás közt van átfedés. Az így bevezetett hálózat segítségével a rendszer szerveződését, hierarchiáját egy magasabb szinten tanulmányozhatjuk (1.ábra).

A k -klikk perkoláció jelensége az általa kínált csoportosulás definíció mellett már pusztán matematikai szempontból is igen érdekes. Régóta ismert, hogy az Erdős-Rényi-féle véletlen gráfban a csúcsok kapcsolódási valószínűségének változtatásával, egy (a perkolációval analóg) fázisátalakulást figyelhetünk meg [37, 38]. Amennyiben az összekötési valószínűség olyan kicsi, hogy a csúcsok átlagos fokszáma nem éri el az egyet, a hálózat sok kisméretű izolált komponensre esik szét. Ellenben ha az összekötési valószínűség növelésével a csúcsok átlagos fokszáma meghaladja az egyet, a hálózatban megjelenik egy óriás (összefüggő) komponens, melynek mérete összemérhető a rendszermérettel. A k -klikk perkoláció tekinthető ezen folyamat általánosításának, hiszen $k = 2$ -re visszkapjuk az imént ismerttetett él-perkolációt. Az Erdős-Rényi-féle véletlen gráf esetén sikerült általánosan meghatározni a k -klikk perkoláció kritikus pontját egy egyszerű elméleti becsléssel [36], illetve egy közelítő analitikus módszerrel [39], az eredmény helyességét numerikus simulációkkal támasztottuk alá [36].



1. ábra. Az élesztő baktérium fehérje kölcsönhatási hálózatából nyert csoportosuláshálózat. A csúcsok mérete illetve az élek vastagsága a csoportosulások méretével illetve az átfedésekben lévő csúcsok számával arányos. A kép jobb oldalán néhány csoportosulás belső szerkezetét mutatjuk kinagyítva, itt a csoportosulásokat a színek különböztetik meg egymástól, az átfedéseket pirossal jelöltük. Az egyes csoportosulások jól megfeleltethetők egy-egy fehérje funkciónak.

3.3. Optimális csoportfelosztás

Az imént említett kritikus pont, (ahol először megjelenik az óriás komponens), fontos szerepet játszik a valós hálózatok csoportanalízise során is. Az ilyen vizsgálatok során általános esetben összesen egy szabad paraméterünk van, a k -klikk méret. Ezen felül súlyozott hálózatok esetén az él-súlyokat oly módon vehetjük figyelembe, hogy bevezetvén egy w súlyküszöböt a w -nél gyengébb éleket elhanyagoljuk. A súlyküszöb növelésével a csoportosulások mérete csökken és csak a legerősebben összekapcsolt részek maradnak meg. Hasonló effektust okoz k növelése is, a nagyobb k -hoz tartozó csoportosulások kisebbek, de ugyanakkor kohézívebbek is. A w és k paraméterek változtatása hasonlít egy mikroszkóp felbontásának beállításához. Ha egy konkrét csúcshoz tartozó csoportosulások érdekelnek minket, akkor azokat érdemes egy szélesebb w és k tartományban megvizsgálni. Ilyenkor csúcsról csúcsra más és más paraméter értékeknél fogjuk a legérdekesebb képet látni. Ugyanakkor a globális csoportosulás szerkezet vizsgálatához valamilyen kritérium szerint fixálni kell a súlyküszöböt és k -t. Az általunk használt kritériumot az imént tárgyalt perkoláció ihlette és azon alapszik, hogy lehetőleg a legtöbb információt hordozó

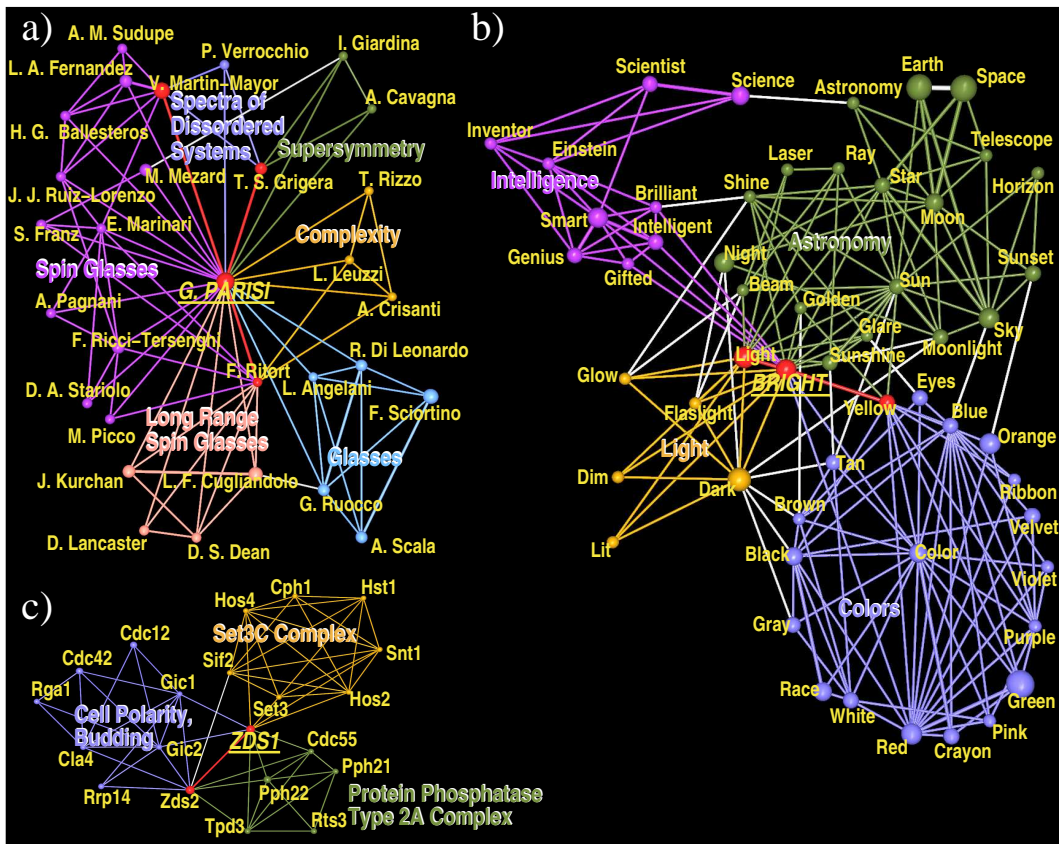
csoportosulás szerkezetet nyerjük ki [35]. Amennyiben túl alacsony w és k paramétereket választunk, a rendszer "perkolál", azaz megjelenik egy óriás-csoportosulás, mely magába foglalja a hálózat túlnyomó részét, elfedvén a csoportosulás szerkezet lokális részleteit. Ezzel szemben túl magas paraméter értékeknél csak elszórtan találunk néhány kisméretű csoportosulást, hiszen csak a legerősebben összekapcsolt, legkohézívebb részek maradnak meg. Az ideális paraméterválasztás a két véglet között található: adott k értékhez w -t úgy kell beállítani, hogy még éppen ne jelenjen meg egy óriás-csoportosulás.

Az optimális csoportfelosztás problémájához kapcsolódik a hálózatok spektrális tulajdonságait vizsgáló munkánk is [40]. A jelenleg igen népszerű, az élek rekurzív kivágásán alapuló (a hálózatot diszjunkt részekre osztó) Girvan-Newman-féle klaszterező módszert vizsgálva egy tőlünk független kutatócsoport arra következtetésre jutott, hogy az optimális csoportfelbontás elérésekor a gráf spektrumának statisztikai jellemzői egy érdekes átalakuláson mennek keresztül [41]. Az Erdős-Rényi-féle véletlen gráfon és több valós hálózaton végzett vizsgálataink ezzel szemben azt mutatták ki, hogy az említett átalakulás nem az optimális csoportfelosztás elérését jelzi, hanem általánosan egy összefüggő óriás komponens megjelenésével/eltűnésével van szoros kapcsolatban [40].

3.4. Valós hálózatok vizsgálata a CPM segítségével

A bevezetett új csoportosulás kereső módszerünkkel nagyméretű (egy esetben több mint százezer élt tartalmazó) fehérje-fehérje kölcsönhatási, szóasszociációs és tudományos társszerzőségi hálózatok csoportosulásait tanulmányoztuk [35, 42, 43]. Eredményeink szerint a talált csoportosulások jelentős mértékben átfedtek egymással. Módszerünk jól adta vissza, hogy pl. azonos alakú szavak esetén az adott szó több, a különböző jelentéseknek megfelelő csoportosulásoknak mind tagja kell legyen, hasonlóan ahhoz, hogy egy egyszerre több témán is aktívan dolgozó kutatónak, vagy egy több funkcióval bíró fehérjének is több csoportosulása kell legyen (2.ábra). A fehérje-fehérje kölcsönhatási hálózatát esetén a feltárt csoportosulások összetételét fehérje annotációs adatbázisok segítségével vizsgálva azt az eredményt kaptuk, hogy a csoportok jelentős része egy-egy jól beazonosítható fehérje funkciónak feleltethető meg [42, 43]. Ezen tulajdonság alapján módszerünk lehetővé teszi a sejtműködésben eddig ismeretlen szerepű fehérjék funkciójának jóslását a feltárt csoportokhoz való tartozás alapján.

A csoportosulások lokális vizsgálata mellett a tanulmányozott rendszerek globális statisztikai jellemzőit is vizsgáltuk. Eredményeink szerint a csoportosulások méreteloszlása valamint a csoport-hálózat fokszámeloszlása is hatványszerűen cseng le, a csoportokat alkotó csúcsok hálózatához hasonlóan.



2. ábra. Példák lokális csoportosulás szerkezetre. A csoportosulásokat a színek különböztetik meg, az átfedések színe piros, a csúcsok mérete (az élék vastagsága) a csúcs (él) csoportjainak számával arányos. a) G. Parisi csoportosulásai egy társszerzőségi hálózatban, a csoportosulások a szerző különböző érdeklődési területeinek felelnek meg. b) a "bright" szó csoportosulásai egy szóasszociációs hálózatban, a csoportosulások a "bright" különböző jelentéseivel kapcsolatosak c) a ZDS1 fehérje csoportosulásai egy fehérje kölcsönhatási hálózatban, a csoportosulások a ZDS1 különböző funkcióinak felelnek meg.

Egy érdekes eltérés viszont a csúcsok hálózata és a csoportok hálózata közt az, hogy utóbbiban a foksámeloszlás exponenciálisan indul alacsony foksám értékekre [35].

3.5. Csoportosulások hálózata

A csoportosuláshálózat foksámeloszlásának imént ismerttetett hatványszerű lecsengése egy igen érdekes kérdést vet fel. Ismert, hogy a természetben

található hálózatok döntő többségénél szintén ilyen a fokszámoszlás lecsengése, és számos esetben ez az effektus a hálózat növekedését vezérlő preferenciális csatolási szabály miatt lép fel. Ezek alapján joggal kérdezhethetjük, hogy vajon a csoportosuláshálózat fejlődésénél is fellép-e a preferenciális csatolási szabály?

A csoportosuláshálózat általában sokkal kevesebb számú csúcsot tartalmaz mint az eredeti hálózat (hiszen benne egy csúcs az eredeti hálózatban csúcsok csoportjának felel meg). Ennélfogva egy csoportosuláshálózat esetén kevesebb számú esemény (él bekötés) alapján kell eldönteni, hogy az új csúcsok preferenciálisan csatolódnak-e vagy sem, mint pl. a neki megfelelő eredeti hálózat esetén. Ezért először egy olyan módszert dolgoztunk ki, mellyel megbízhatóan lehet kimutatni a preferenciális csatolás jelenlétét vagy hiányát a csoportosulás hálózatok mérettartományába eső rendszerek esetén is [44]. Ezzel a módszerrel egy társszerzőségi hálózat csoportosulásainak időfejlődését vizsgáltuk, és eredményeink igazolták azt a sejtést, hogy preferenciális csatolási mechanizmusok formálják a rendszert a csoportosulások szintjén is. Egyrészt a csoportosuláshálózat növekedése során egy, még kapcsolatok nélküli csúcs (csoportosulás) a csoportosulásmérettel és csoportosulás fokszámmal preferenciálisan kapcsolódik be a csoportosuláshálózatba. Ezen felül az eredeti hálózatban egy csúcs, mely még egyetlen egy csoportosulásnak sem tagja, a csoportosulásmérettel és csoportosulás fokszámmal preferenciálisan fog egy csoportosuláshoz csatlakozni [44].

Eredményeink szerint tehát a társszerzőségi hálózat időfejlődését hasonló mechanizmusok vezérlik mind a csúcsok, mind a csoportosulások szintjén. A csoportosulások hálózatának növekedése a preferenciális csatolási szabály szerint történik, teljesen analóg módon az alapul szolgáló hálózat növekedésével. Ez a jelenség egy további megerősítése a rendszer különböző szerveződési szintjei közt tapasztalható hasonlóságnak.

3.6. Kapcsolódó programcsomag

A fenti eredmények mellett kifejlesztettünk egy programcsomagot is, mely a CPM használatával csoportosulásokat keres és - több más elemzéssel együtt - a talált csoportosulások hálózatát bemutatja [43, 45]. Ez a szoftver CFinder (Clique and Community Finder) néven a ingyenesen letölthető, és tudományterülettől függetlenül alkalmazható minden olyan adatrendszer elemzésére, amely hálózatként ábrázolható. A program Windows, Linux és Macintosh számítógépeken egyaránt használható.

Irodalomjegyzék

Hivatkozások

- [1] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [2] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [3] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [4] J. F. F. Mendes and S. N. Dorogovtsev, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
- [5] A. Barrat, M. Barthelemy, and A. Vespignani, *Phys. Rev. Lett.* **92**, 228701 (2004).
- [6] Z. Burda, J. D. Correia, and A. Krzywicki, *Phys. Rev. E* **64**, 046118 (2001).
- [7] Z. Burda and A. Krzywicki, *Phys. Rev. E* **67**, 046118 (2003).
- [8] J. Berg and M. Lässig, *Phys. Rev. Lett.* **89**, 228701 (2002).
- [9] M. Baiesi and S. S. Manna, *Phys. Rev. E* **68**, 047103 (2003).
- [10] I. Derényi, I. Farkas, G. Palla, and T. Vicsek, *Physica A* **334**, 583 (2004).
- [11] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, *Phys. Rev. E* **69**, 046117 (2004).
- [12] I. Farkas, I. Derényi, G. Palla, and T. Vicsek, *Lect. Notes Phys.* **650**, 163 (Springer-Verlag Berlin Heidelberg, 2004).
- [13] G. Palla, I. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek, *Phys. Rev. E* **70**, 046115 (2004).
- [14] M. Blatt, S. Wiseman, and E. Domany, *Phys. Rev. Lett.* **76**, 3251 (1996).
- [15] M. Girvan and M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **99**, 7821 (2002).
- [16] H. Zhou, *Phys. Rev. E* **67**, 061901 (2003).
- [17] M. E. J. Newman, *Phys. Rev. E* **69**, 066133 (2004).
- [18] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, *Proc. Natl. Acad. Sci. USA* **101**, 2658 (2004).

- [19] D. M. Wilkinson and B. A. Huberman, *Proc. Natl. Acad. Sci. USA* **101**, 5241 (2004).
- [20] L. Donetti and M. A. Muñoz, *cond-mat/0404652*.
- [21] J. Reichardt and S. Bornholdt, *Phys. Rev. Lett.* **93**, 218701 (2004).
- [22] J. Scott, *Social Network Analysis: A Handbook*, 2nd ed. (Sage Publications, London, 2000).
- [23] R. M. Shiffrin & K. Börner, Mapping knowledge domains. *Proc. Natl. Acad. Sci. USA* **101** 5183–5185 Suppl. 1 (2004).
- [24] B. S. Everitt, *Cluster Analysis*, 3th ed. (Edward Arnold, London, 1993).
- [25] S. Knudsen, *A Guide to Analysis of DNA Microarray Data*, 2nd ed. (Wiley-Liss, 2004).
- [26] M. E. J. Newman, Detecting community structure in networks. *Eur. Phys. J. B*, **38**, 321–330 (2004).
- [27] E. Ravasz, A. L. Somera, D. A. Mongru, Z. Oltvai and A.-L. Barabási *Science* **297**, 1551 (2002).
- [28] V. Spirin and L. A. Mirny *Proc. Natl. Acad. Sci. USA* **100** 12123 (2003).
- [29] J. P. Onnela, A. Chakraborti, K. Kaski, J. Kertész and A. Kanto *Phys. Rev. E* **68**, 056110 (2003).
- [30] D. J. Watts, P. S. Dodds and M. E. J. Newman *Science* **296**, 1302 (2002).
- [31] J. Vukov and Gy. Szabó *Phys. Rev. E* **71**, 036133 (2005).
- [32] Gy. Szabó, J. Vukov and A. Szolnoki *cond-mat/0506433* (2005).
- [33] A. C. Gavin *et al.* *Nature* **415**, 141 (2002).
- [34] *P. Eds Carrington and J. Scott and S. Wasserman Models and Methods in Social Network Analysis*, Ch. 7, Edited by K. Faust, Cambridge University Press, New York (2005).
- [35] G. Palla, I. Derényi, I. Farkas and T. Vicsek *Nature* **435**, 814 (2005).
- [36] I. Derényi, G. Palla and T. Vicsek *Phys. Rev. Lett.* **94**, 160202 (2005).
- [37] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).

- [38] B. Bollobás, *Random graphs*, 2nd ed. (Cambridge University Press, Cambridge, 2001).
- [39] G. Palla, I. Derényi, and T. Vicsek: *J. Stat. Phys.* **128**, 219 (2007).
- [40] G. Palla and G.Vattay: *New Journal of Physics* **8**, 307 (2006) .
- [41] F. Luo, J. Zhong, Y. Yang, R. H. Scheuermann and J. Zhou 2006 *Physics Letters A* **357**, 420 (2006).
- [42] G. Palla, I. Derényi, I. Farkas and T. Vicsek, *FEBS J.* **272**, Issue s1, 434 (2005).
- [43] B. Adamcsek, G. Palla, I. J. Farkas, I. Derényi and T. Vicsek, *Bionformatics* **22**, 1021 (2006).
- [44] P. Pollner, G. Palla, and T. Vicsek *Europhys. Lett.* **73**, 478 (2006).
- [45] <http://www.cfinder.org>