

Komplex hálózatok szerkezetének és dinamikájának feltárása és modellezése statisztikus fizikai módszerekkel

Az MTA Doktora cím elnyeréséhez készített
tézislevelet

Palla Gergely



MTA-ELTE Statisztikus és Biológiai Fizika Kutatócsoport

2016. március

Bevezetés, a kutatások előzménye

A komplex hálózatok témaköre egy körülbelül 15 éves múlttra visszatekintő interdiszciplináris terület, mely egyaránt kötődik a gráfelmélethez, a statisztikus fizikához, a biológiához és a biokémiához, a szociológiához, a közgazdaságtanhoz és az informatikához [2, 57, 96]. A hálózatos megközelítés alapgondolata az, hogy a vizsgált rendszert bizonyos alegységek (pl. személyek, gének, fehérjék, routerek, cégek, stb.), és ezen alegységek közti kapcsolatok, kölcsönhatások együtteseként reprezentáljuk, mely egy gráfként jelenik meg. Ennek révén a terület gyökerei a gráfelmélethez nyúlnak vissza, melynek alapjait Leonhard Euler fektette le még a XVIII. században [25], és mely a XX. század elejére már egy fontos, önálló területévé vált a diszkrét matematikának [46]. A jelenleg is folyó hálózatkutatások szempontjából a gráfelmélet egyik legfontosabb mérföldköve a véletlen gráfok bevezetése volt Erdős Pálnak és Rényi Alfrédnek köszönhetően az 1950-es évek végén [23, 24]. Az általuk definiált véletlen gráf (melyre szokás klasszikus véletlen gráfként is utalni), rögzített számú csúcsot tartalmaz, melyek között minden lehetséges párt egymástól függetlenül egy adott valószínűséggel összekötünk. Ez a modell a mai napig egy nagyon fontos referenciapontot képez a valós rendszereket reprezentáló hálózatok vizsgálatánál, hiszen megmutatja, hogy miként viselkedne egy hálózat akkor, ha a kapcsolatok teljesen homogén módon, véletlenszerűen lennének szétosztva a csúcsok között.

Érdekes módon az 1990-es évek végén, illetve a 2000-es évek elején jelentős mértékben pont az indította be a hálózatkutatás robbanásszerű fejlődését, hogy nagyon lényeges eltérésekre derült fény a valós rendszereket leíró hálózatok és az Erdős–Rényi-gráf egyes tulajdonságai között. A másik fontos tényező, ami közrejátszott a hálózatos megközelítés elterjedésében az volt, hogy ekkortájt váltak először széles körben elérhetővé nagy adatbázisok a legkülönbözőbb rendszerekről, kezdve a molekuláris biológiai adatbázisoktól [88, 60, 96], társszerzői [61, 62, 6], filmes [96] és szociológiai adatbázisokon keresztül [52] az informatikai és egyéb technológiai adatbázisokig [27, 1, 48, 3, 71]. Nagy adatrendszerek kezelését, átláthatóvá tételét sok esetben jelentősen megkönnyíti a hálózatos leírásmód mint „első közelítés”. Emellett sok esetben már pusztán a hálózatszerkezet is érdekes, nemtriviális statisztikus tulajdonságokkal rendelkezik, és fontos összefüggésekre világíthat rá.

Az egyik érdekes eltérés a valódi rendszereket leíró hálózatok és az Erdős–Rényi-

modell között abban rejlik, hogy az előbbi rendszerek annak ellenére, hogy globálisan általában ritkák (azaz az átlagos fokszám messze elmarad a rendszermérettől), lokálisan mégis sűrűek, ha egy-egy véletlenszerűen kiválasztott csúcs közvetlen szomszédai között található élek valószínűségét vizsgáljuk. Ennek jellemzésére vezették be a klaszterezettségi együtthatót [40, 92, 96], mely megadja egy csúcs közvetlen szomszédai közt található élek relatív hányadát a szomszédok közt lehetséges élek maximális számához viszonyítva. A tapasztalatok szerint a klaszterezettségi együttható általában jelentősen nagyobb egy valós rendszert reprezentáló hálózat esetén, mint egy azonos méretű és azonos átlagos élsűrűségű Erdős–Rényi-gráfban.

A másik nagyon lényeges eltérés a valódi rendszereket leíró hálózatok és a klasszikus véletlen gráfok között a csúcsok fokszámeloszlásában jelentkezik: Az Erdős–Rényi-gráf esetén ez egy binomiális eloszlás (amit a gyakorlatban Poisson-eloszlással szokás közelíteni), míg a valódi rendszerek esetén általában egy hatványszerű lecsengés tapasztalható az eloszlásban a nagy fokszámok tartományán [72, 74, 27, 48, 3, 5, 43, 88, 62, 52, 20]. A statisztikus fizikában előforduló skálázó függvények és skálázó eloszlások mintájára az ilyen fokszámeloszlással rendelkező rendszerekre a „skálafüggetlen hálózat” elnevezés terjedt el.

A valódi rendszereket leíró hálózatokban tapasztalt, az Erdős–Rényi-modelltől lényegesen eltérő univerzális tulajdonságok új véletlengráf-modellek bevezetését tették időszerűvé. Ezek közül a két legfontosabb, melyeknek rendkívül nagy hatása volt a terület fejlődésére a Watts–Strogatz-modell [96] és a Barabási–Albert-modell [5]. A Duncan Watts és Steven Strogatz által javasolt modell segítségével olyan véletlen gráfokat lehet előállítani, melyeknek nagy a klaszterezettségi együtthatója, és fellép bennük a kisvilág-effektus. Ez utóbbi azt a jelenséget takarja, hogy a csúcsok közti átlagos távolság¹ a rendszermérethez képest nagyon kicsi. Az Erdős–Rényi-gráf esetén például (extrém élbekötési valószínűségektől eltekintve), az átlagos távolság a csúcsok számának logaritmusával skálázik, illetve a valós rendszereket leíró hálózatok is kisvilág-tulajdonságúak.

A Watts–Strogatz-modell esetén egy szabályos rácshoz hasonló struktúrából indulunk ki, melynek szerkezete úgy van kialakítva, hogy biztosítsa a magas klaszterezettségi együtthatót. Azonban ez a struktúra még nem „kis világ”, ugyanis a szabályos szerkezet miatt a távolság a rendszermérettel hatványszerűen skálázik. Annak érdekében, hogy a kisvilág-effektus is fellépjen, elég az élek egy kis hányadát véletlenszerűen átkötni. Ezzel drasztikusan lecsökkenthetjük az átlagos távolságot anélkül, hogy a klaszterezettség számottevően csökkenne.

A Barabási Albert-László és Albert Réka által bevezetett modell a klaszterezettség helyett a skálafüggetlenségre koncentrál [5]. A megközelítésük lényege, hogy egy kezdeti kis maghoz lépésenként egy-egy új csúcsot kötünk hozzá (azaz a hálózat időben növekvő), és az új élek másik végéhez a fokszámmal arányos valószínűség

¹Két csúcs távolsága alatt az őket összekötő legrövidebb út éleinek számát értjük.

szerint véletlenszerűen választunk egy-egy csúcsot a már meglévő csúcsok közül; ez az ún. preferenciális kapcsolódási szabály. A preferenciális kapcsolódásnak köszönhetően a már sok kapcsolatot begyűjtő, nagy fokszerű csúcsoknak sokkal jobb esélye lesz az új csúcsokkal érkező új élek megszerzésére egy kis fokszerű csúcsához viszonyítva. Belátható, hogy ennek révén hosszútávon egy hatványszerűen lecsengő (azaz skálafüggetlen) fokszerűeloszlás alakul ki a növekvő hálózatban.

A Watts–Strogatz-modell és a Barabási–Albert-modell bevezetése után a 2000-es évek elején a komplex hálózatok témaköre nagyon intenzív fejlődésnek indult, és népszerűsége a hozzá kapcsolódó publikációk éves száma alapján még jelenleg is töretlenül növekszik. (Például a Web of Science kb. tízezer olyan találatot ad csak a 2015-ben megjelent publikációk között, melynél a címben, vagy kivonatban, vagy a kulcsszavak között szerepel a „complex networks” kifejezés). Ez többek között azzal is járt, hogy mára már nagyon szerteágazóvá vált a terület, így a különböző problémakörökből itt csak azokat mutatjuk be, melyeket a disszertációban tárgyalt eredmények érintenek. Ezen részterületekben az a közös, hogy mind nagyon erősen kapcsolódnak a statisztikus fizikához és a statisztikus fizikai módszerekhez.

A statisztikus fizikának a kezdetektől fogva igen jelentős hatása volt a hálózat-kutatás fejlődésére. Ez részben annak köszönhető, hogy a vizsgált valós rendszereket leíró hálózatok általában nagyon sok csúcsból állnak, és emiatt egy „mikroszkopikus” leírásmód, melyben minden egyes csúcsot és annak kapcsolatait külön vizsgáljuk, kezelhetetlenné válik. Ezzel szemben kézenfekvő a statisztikus fizikában megszokott sokrészecskés rendszerekhez hasonlóan az eloszlások, átlagok és egyéb statisztikai jellemzők használata. Ami igazán vonzóvá tette ezt a területet a fizikusok számára az a (klaszterezettség és a skálafüggetlenség kapcsán már említett) nagy fokú univerzalitás volt a feltárt hálózatok statisztikai jellemzőiben: Sok esetben olyan hálózatok is rendkívül hasonló tulajdonságokat mutattak, melyeknek egyébként semmi közük nem volt egymáshoz (mint pl. az említett publikációs, filmes, vagy biológiai adatbázisok). Ez arra utal, hogy nagyon általános törvények és mechanizmusok formálják a komplex hálózatok struktúráját, melyek feltárása, megértése egy nagy előrelépést jelent abban, hogy egy átfogó képünk legyen egyszerre nagyon sok és sokféle rendszer viselkedéséről, szerkezetéről.

Az egyik kutatási terület, mellyel a disszertáció foglalkozik (és melynek kötődését a statisztikus fizikához nem kell külön hangsúlyozni), a gráfsokaságok és a topológiai fázisátalakulások problémaköre. A valós rendszereket leíró hálózatok struktúrája általában időben változó, hiszen egyes kapcsolatok idővel megszűnhetnek, illetve új élek is létrejöhetnek, valamint természetesen a csúcsok összetétele és teljes száma is változhat. Egy hálózat átrendeződése modellezhető egy állapottérben történő bolyongással, ahol az egyes állapotokhoz bizonyos szabályok szerint valószínűségeket rendelünk, és az állapotok közti átmeneti valószínűségeket kielégítik a részletes egyensúly feltételét. Gráfsokaság természetesen többféle módon is definiálható, például a

különböző gráfok valószínűségét megfeleltethetjük egy nulladimenziós térelméletben a Feynman-diagramok súlyának [17, 18], de természetesen használhatunk klasszikus hamiltoni rendszerekhez hasonló formalizmust is [8].

A gráfsokaságok esetén is megadható egy, a hőmérsékletnek megfelelő paraméter, melynek szemléletes jelentése a hálózat átrendeződését befolyásoló zajszint: Nagyon magas hőmérsékleten az élek gyakorlatilag teljesen véletlenszerűen kötődnek át egyik csúcsról a másikra, míg alacsony hőmérsékleten egyre inkább a megadott Hamilton-függvény szempontjából optimális (alacsony energiájú) állapotok felé rendeződik át a hálózat. Topológiai fázisátalakulásról akkor beszélhetünk, ha a rendszer hűtése során a hálózat szerkezete egy drasztikus átalakuláson megy keresztül, és az átalakulás kritikus pontjában a partíciós függvényből származtatott termodinamikai függvények szingulárisá válnak. A valós rendszereket leíró hálózatok esetén is megfigyelhetők esetenként jelentős szerkezetbeli átalakulások, például bizonyos gazdasági hálózatok szerkezete a bizonytalanság (várható fluktuációk) szintjétől függően egy csillagszerű állapotból átszerveződhet egy lokálisan sűrűbben huzalozott, klikkszerű struktúrába [82]. Emellett a gráfsokaságok segítségével leírt életrendeződési dinamika egy természetes keretrendszert adhat olyan hálózattopológia-optimalizálási kérdések tárgyalásához is, mint pl. a [38, 86] publikációkban vázolt problémák. Ilyenkor ideális esetben egy megfelelő energiafüggvény választásával pont az adott feladat szempontjából optimális topológia áll elő a rendszer lehűtése során.

A disszertáció egy másik témaköre, melynél szintén nyilvánvaló a statisztikus fizikához való erős kötődés, a k -klikkperkoláció vizsgálata véletlen gráfokban. A szabályos rácsokon definiált perkolációs modellek egy alapvető eszközt kínálnak egy folyadék porózus anyagon történő átáramlásának, illetve magának a porózus anyagnak a leírására [16, 34]. Természetesen a perkoláció jelenségét véletlen gráfokon is tanulmányozták, hiszen az egyik legelső fontos eredmény az Erdős–Rényi-moddal kapcsolatban a perkoláció kritikus pontjához tartozó élbekötési valószínűség levezetése volt [24, 11]. A disszertáció ennek a problémának egy olyan általánosítását tárgyalja, melynél élek helyett k -klikkek (k csúcsból álló teljesen összekötött részgráfok) perkolációjára koncentrálnak. Ez egyfelől egy elméleti szempontból fontos fejlemény, hiszen egy történetileg híres eredményt helyez tágabb kontextusba. Másfelől a k -klikkperkoláció hálózati csoportkereső módszerként is használható, mely az évek során rendkívül sikeresnek bizonyult, emiatt gyakorlati szempontból is nagy a jelentősége.

A hálózati csoportok (más néven modulok, csoportosulások, klaszterek) sűrű, erősen összekötött részgráfoknak felelnek meg egy hálózatban [79, 80, 26, 45, 32, 63, 77], mint pl. egy család vagy baráti kör az emberi kapcsolatok hálózatában [79, 95]. Hasonló egységek előfordulnak fehérje-kölcsönhatási hálózatokban (pl. fehérjefunkciós csoportok) [73, 81], pénzügyi illetve gazdasági hálózatokban (pl. ipari szektorok) [68, 39], valamint játékelméleti modellekben is [85, 87, 84]. A csoportok feltérképe-

zése nagyban segítheti az említett hálózatok szerkezetének, működésének alaposabb megismerését, mélyebb megértését. (A területről egy átfogó leírást ad Santo Fortunato összefoglaló cikke [29].)

Bár a témakörnek messzire nyúló gyökerei vannak a Keningham–Lin-algoritmus [44], a spektrális biszekció [7], a hierarchikus klaszterezés, illetve a k -középklaszterezés [55] révén, mégis az első átütő módszer, mely nagyon sikeressé vált és a mai napig rendkívül széles körben ismert, a Girvan–Newman-algoritmus volt [32]. A javasolt megközelítés legfontosabb eredménye a modularitás bevezetése volt, mely lehetővé tette a csoportfelbontás minőségének kvantitatív mérését, pusztán a hálózati topológia (és természetesen a csúcsok csoportbesorolása) alapján [66]. A modularitás használata gyorsan nagyon nagy népszerűsége telt szert, és a mai napig a legelfogadottabb csoportminőség mérésére szolgáló mennyiségnek számít. Bár a modularitás egzakt maximumának megkeresése egy NP-teljes probléma [15], a széles körben használt csoportkereső eljárások egy része a modularitás maximalizálásán alapszik, valamilyen heurisztikus módszert használva mint például a szimulált hőkezelés [37], az extrém optimalizáció [21], a LOUVAIN-algoritmus [9], vagy a spektrális optimalizáció [65, 64, 90, 76, 83].

A modularitásmaximum keresésének egyik érdekessége, hogy ekvivalens egy olyan spinrendszer alapállapotának keresésével, melynél a spinek a hálózat csúcsaihoz társulnak, és az élek mentén hatnak kölcsön egymással a Potts-modellhez hasonlóan (és a szokásos ferromágneses kölcsönhatás mellett még megfelelő további tagokat vezetünk be a Hamilton-függvényben) [75]. A modularitáson alapuló csoportkeresésnek ugyanakkor hátrányai is vannak, melyek közül az egyik a felbontási határ problémája [30], mely szerint egy összefüggő hálózatban a teljes élszám gyökénél jelentősen kevesebb élből álló csoportokat nem tudunk feltárni ezzel a megközelítéssel. A másik nagy hátrány az, hogy a csoportok közti átfedések nem megengedettek (legalábbis a modularitás eredeti alakján nyugvó módszereknél).

A disszertációban bemutatott k -klikkperkolációs módszer egy természetes megoldást jelent ezekre a problémákra. Ennél a megközelítésnél nem lép fel a felbontási határ problémája, és a legfontosabb újítása az volt, hogy olyan hálózatokban is lehetővé tette a csoportok feltárását, melyeknél nagyon jelentősek a csoportátfedések. Ennek egyik hozadéka, hogy megnyílik az út olyan, korábban nem tanulmányozott statisztikák vizsgálata előtt, melyek nem értelmetlenek (vagy triviálisak) akkor, ha csak elvélve (vagy egyáltalán nem) fordulnak elő csoportátfedések.

A hálózati csoportkeresés egyik fontos alkalmazási területe az emberi kapcsolathálózatokban található kisebb-nagyobb csoportok feltárása. Az emberi kapcsolatok vizsgálata hagyományosan a szociológia témakörébe tartozik, ahol már az 1930-as években is használtak gráfokat a tanulmányozott kisebb emberi közösségek kapcsolathálózatának megjelenítésére [97, 33, 92, 28]. Az ezredforduló környékén az ilyen irányú kutatások új lendületet kaptak különféle automatikusan gyűjtött, nagymé-

retű adatbázisok elterjedésével és hozzáférhetővé válásával, melyek segítségével a korábbi, kérdőíves felméréseken alapuló vizsgálatokhoz képest több nagyságrenddel nagyobb hálózatokat lehetett tanulmányozni [69, 70, 94]. Ezen újfajta, nagy skálájú vizsgálatok egyik első, a szociológia szempontjából is fontos eredménye a Granovetter-hipotézis² igazolása volt egy több, mint 4 millió felhasználót tartalmazó mobiltelefon-hívási hálózatban [69, 70]. A disszertáció egyik fejezete ezen hálózat adataira támaszkodva a hálózati csoportok időfejlődésének statisztikai tulajdonságait vizsgálja. Az idevágó kutatásokba bevontuk emellett a Cornell Egyetem könyvtárának internetes cikkarchívumán [91] alapuló társszerzőségi hálózatot is. Pusztán az emberi kapcsolathálózatok időfejlődését (csoportelemzés nélkül) több adatbázis alapján is vizsgálták már korábban [52, 6, 41, 22, 89, 98, 67], és néhány tanulmány külön felhívta a figyelmet arra, hogy a csoportok időfejlődésének tanulmányozása nagyban segítheti a társadalom önszervező folyamatainak mélyebb megértését [35, 42, 36, 51, 47].

Végül az utolsó témakör, melyet a disszertáció érint, az a statisztikus fizikában jól ismert multifraktálok egy olyan alkalmazásával foglalkozik, melynek révén egy nagyon általános véletlengráf-generáló algoritmust lehet megalkotni. Korábban említésre került, hogy a komplex hálózatok területének kezdeti népszerűsége nagyban köszönhető a Watts–Strogatz-modell és a Barabási–Albert-modell sikerességének. Természetesen idővel számos további véletlengráf-modell merült fel az irodalomban, melyek többnyire vagy egy kiszemelt hálózatjellemzőre koncentráltak (mint például skálafüggetlen fokszámoszlás egy jól meghatározott exponenssel), vagy egy viszonylag jól behatárolt hálózattípust vizsgáltak (mint például az emberek közti kapcsolathálókat, vagy metabolikus hálózatokat, stb.). Mivel az ilyen modellek keretein belül előállítható hálózatok tulajdonságai csak viszonylag szűk tartományon változtathatók, természetes módon merült fel az igény általánosabb véletlengráf-modellek megalkotása iránt, melyek egy szélesebb skálán teszik lehetővé mesterséges hálózatok generálását a felmerülő lényeges tulajdonságok (pl. fokszámoszlás, klaszterezettség) szempontjából.

A legegyszerűbb modell, melynek alapgondolata ebbe az irányba mutat az ún. konfigurációs modell, mely tetszőleges fokszámoszlású véletlen hálózat generálását teszi lehetővé [58, 59]. Amennyiben egy valós rendszert leíró hálózatot szeretnénk ezzel a módszerrel modellezni, a gyakorlatban az eredeti hálózat éleinek véletlenszerű átkötögetésével állítjuk elő a mintákat (a fokszámoszlás megtartásával). Bár egy ilyen megközelítés könnyen célra vezethet, nem igazán segít feltárni az adott tulajdonságokkal rendelkező hálózatok esetleges rejtett összefüggéseit. Ennek révén van létjogosultsága az olyan generatív módszereknek, mint a rejtettparaméter-modell [19, 10],

²Mark Granovetter hipotézise szerint két ember kapcsolatának erőssége az egymásnak szentelt idő, anyagi ráfordítás, érzelmi intenzitás, bizalom és kölcsönös segítség/szívesség kombinációja. Továbbá a hipotézis szerint ez a kapcsolaterősség monoton növekedő függvénye a relatív átfedésnek, mely a közös ismerősök száma osztva a két személy összes ismerőseinek számával.

a dK -sorozatok módszere [56], az exponenciális véletlengráf-modell [31, 93, 78], a Parisi-mátrixokon alapuló módszer [4], vagy a Kronecker-gráfokon alapuló módszer [49, 50]. A disszertációban bemutatott, multifraktálokra alapuló véletlengráf-modell esetén a hálózat struktúráját kódoló objektum hierarchikus, mely egy rekurzív eljárás során keletkezik, és ez által önhasznós a szerkezete. Ugyanakkor a módszer nagyon erősen támaszkodik Lovász László és munkatársainak gráfsorozatok konvergenciájával kapcsolatos korábbi eredményeire [54, 14, 53], illetve közeli rokonságot mutat a Bollobás Béla és munkatársai által bevezetett véletlengráf-generáló eljárással is [12, 13].

Célkitűzések

A kutatások célkitűzése nagy vonalakban a komplex hálózatok szerkezetének feltárása és modellezése volt a statisztikus fizika eszköztárának segítségével. Ahogy már a bevezetésben is említésre került, a disszertáció több témakört is érintett, melyek egymáshoz lazán kapcsolódnak.

A topológiai fázisátalakulások vizsgálata során az egyik cél az volt, hogy megadjunk egy statisztikus fizikai alapokon nyugvó általános keretrendszert állandó méretű, időben változó szerkezetű hálózatok modellezéséhez. A kutatások során kidolgozott, a kanonikus sokaságnak megfelelő gráfsokaságban olyan energiafüggvényeket kerestünk, melyek esetén a rendszer hűtése során olyan drasztikus változások következnek be a hálózatszerkezetben, melyek megfelelnek egy fázisátalakulásnak.

A k -klikkperkoláció tanulmányozása során az egyik célkitűzés az volt, hogy az Erdős–Rényi-modellben korábban jól ismert élperkolációs átalakulást általánosítsuk a k -klikkek esetére. Egy másik, ugyan olyan fontos cél volt, hogy a k -klikkperkolációs klaszterek segítségével adjunk egy olyan csoportdefiníciót, mely megengedi a csoportok közti átfedéseket és a gyakorlatban is használható valós rendszereket reprezentáló hálózatok klaszterezésére. Továbbá az általunk bevezetett csoportkereső módszer lehetővé tette több olyan csoportátfedésekkel kapcsolatos statisztika empirikus vizsgálatát, melyeket a korábbi módszerek által feltárt csoportokon nem lehetett tanulmányozni.

A hálózati csoportkereséssel kapcsolatos kutatások folytatásaként a k -klikkperkolációs csoportok időfejlődését tanulmányoztuk nagyméretű, emberi kapcsolatokat reprezentáló hálózatokban. Ezeknél a vizsgálatoknál az egyik cél a nagy skálájú csoportidőfejlődés-elemzés módszertani kidolgozása volt. A technikai részletek kidolgozása után tudtunk nekilátni az idevonatkozó kutatások lényegi céljának megvalósításához, mely a csoportösszetétel időbeli változásának statisztikai elemzése volt.

Végül a multifraktál alapú véletlengráf-moddellel kapcsolatos kutatások célja egy általános hálózatgeneráló eljárás kifejlesztése volt, mely lehetővé teszi egymástól jelentősen különböző tulajdonságokkal rendelkező véletlen gráfok előállítását ugyan

abban a keretrendszerben. A modell kidolgozásánál (a természetükénél fogva hierarchikus és önhasznó) multifraktálokat kombináltuk egy korábbi véletlengráf-generáló eljárással, mely a gráfsorozatok konvergenciájának vizsgálatánál került bevezetésre.

Új tudományos eredmények (tézispontok)

1. Analitikus megfontolások alapján meghatároztam a szabadenergia függését vezető rendben a $\varphi_d = d_{\max}/M$ rendparamétertől és az effektív hőmérséklettől az egyrészcskés energiafüggvények által meghatározott kanonikus gráfsokaságokban [T1, T2]. (A rendparaméter képletében $d_{\max}$ a hálózatban előforduló legnagyobb foksámot, M pedig az élek számát jelöli). A szabadenergiára kapott kifejezés a termodinamikai határesetben jó közelítést ad minden olyan esetben, ahol a hálózat egyik csúcsához az élek több, mint fele hozzákapcsolódik, és az energiafüggvény alakja olyan, hogy a többi csúcs járuléka emellett elhanyagolható. Ezen feltételeknek eleget tesz kellően alacsony effektív hőmérsékleten az $E = -\sum_{i=1}^N d_i^2$ energiafüggvény (ahol az összegzés a csúcsokra történik, és d_i az i csúcs foksámát jelöli). Az ezen energiafüggvény által „vezérelt” gráf dinamikája ekvivalens az Ising-moddal egy olyan rácson, melynek rácspontjai a hálózat N csúcsa közt lehetséges $N(N-1)/2$ él pozíciójának felelnek meg.

Az analitikus számítások megmutatták, hogy ha az effektív hőmérséklet nullához tart, akkor a szabadenergia minimuma $\varphi_d = 1$ -nél egy „csillagszerű” állapotban van, melynél a hálózat egyik csúcsához az összes él hozzákapcsolódik. Ezzel szemben ha az effektív hőmérséklet magasabb, mint $M/\ln N$, akkor a szabadenergia minimuma $\varphi_d = 0$ -nál egy rendezetlen állapotban van, melynél a legnagyobb foksámú csúcsához kapcsolódó élek száma M -hez képest elhanyagolható a termodinamikai határesetben. A köztes hőmérsékleteken a szabadenergiának két minimuma van φ_d függvényében a $[0,1]$ tartományon, melyek az imént leírt két extrém állapotnak felelnek meg, és melyek közül az egyik stabil, a másik metastabil. Az átalakulás a „csillagszerű” és a rendezetlen állapot között ezek alapján elsőrendű [T1, T2].

Az analitikus számításokat az egzakt leszámolás módszerével numerikus úton is ellenőriztem kisméretű gráfokon. Az egzakt leszámolás eredményei teljes mértékben alátámasztották az analitikus megközelítés eredményeit, mely szerint a szabadenergiának két minimuma van ϕ_d függvényében egy szélesebb hőmérséklet-tartományon. Ezen minimumok közül a „csillagszerű” állapot (ahol $\phi_d = 1$) válik stabillá, ha az effektív hőmérséklet nullához tart, és ellenkezőleg a rendezetlen állapot (ahol $\phi_d = 0$) lesz stabil magas hőmérsékleten [T1].

2. Generátorfüggvények segítségével meghatároztam a k -klikkperkoláció $p_c(k)$

kritikus élbekötési valószínűségét az Erdős–Rényi-gráfban [T3]. Amennyiben az élbekötési valószínűség kisebb, mint $p_c(k)$, a legnagyobb k -klikkperkolációs klaszter mérete a termodinamikai határesetben elhanyagolható a rendszermérethez képest. Ezzel szemben ha az élek valószínűsége nagyobb, mint $p_c(k)$, a legnagyobb k -klikkperkolációs klaszter relatív mérete a termodinamikai határesetben nullánál nagyobb értékhez tart, azaz a gráf tartalmaz egy „óriás” k -klikkperkolációs klasztert. A két állapot között az átalakulás folytonos, és a $k = 2$ esetben $p_c(k)$ visszaadja az Erdős–Rényi-gráf perkolációs átalakulásának kritikus élbekötési valószínűségét.

Az Erdős–Rényi-gráf legnagyobb k -klikkperkolációs klaszterének várható méretét numerikus szimulációk segítségével is meghatároztam az élbekötési valószínűség függvényében, az eredmények maximálisan alátámasztották a kritikus élbekötési valószínűségre vonatkozó analitikus eredményeket [T4]. A szimulációk során két fajta rendparamétert vizsgáltam: Φ a legnagyobb k -klikkperkolációs klaszter csúcsainak relatív hányadának felelt meg a rendszermérethez képest, Ψ pedig a legnagyobb k -klikkperkolációs klaszter k -klikkjeinek relatív hányadát adta meg a gráf összes k -klikkjeihez képest. A szimulációs eredmények elemzése során sikerült megmutatni, hogy rögzített k esetén megfelelő skálázással a különböző gráfméretek és különböző élbekötési valószínűségek mellett kapott Φ értékek egy univerzális görbére ejthetők össze [T4]. A Ψ rendparaméter kritikus pontban felvett értéke a numerikus eredményeim szerint hatványszerűen cseng le a rendszerméret függvényében, k -tól függő exponenssel [T4].

3. Részt vettem a k -klikkperkolációs csoportkereső algoritmus kidolgozásában. Az eljárást beprogramoztam és az elkészült program segítségével nagyméretű tudományos társszerzőségi, fehérje-kölcsönhatási és szóasszociációs hálózatok csoportszerkezetét tártam fel [T5]. A tanulmányozott hálózatokban kapott csoportok lokális elemzése megmutatta, hogy a módszer képes az emberi intuícióval jól egyező, egymással átfedő csoportokat kinyerni a vizsgált hálózatokból (pl. a társszerzőségi hálózatban kiválasztott kutató csoportjaiba a környezetében található többi kutató a különböző kutatási területek szerint lett besorolva, a szóasszociációs hálózatban kiválasztott szó csoportjai a szó különböző jelentéseinek feleltek meg, stb.).

A globális csoportszerkezet jellemzésére három, korábban nem vizsgált eloszlást vezettem be, melyek mind a csoportok közti átfedésekkel kapcsolatosak [T5]. Ezek közül a legfontosabb egy véletlenszerűen választott csoporttal átfedő többi csoport számának eloszlása volt. Ez megegyezik azon hálózat fokszámeloszlásával, melyben a csúcsok az eredeti hálózat csoportjainak felelnek meg, és két ilyen csúcs akkor van összekötve, ha az adott csoportpár át-

fed egymással. Az így kapott fokszámeloszlás egy viszonylag összetett alakot vett fel a vizsgált rendszerekben: A kis és közepes csoportfokszámok tartományán exponenciálisan csökkent egy bizonyos határig, míg ezzel szemben a nagy fokszámok tartományán egy hatványfüggvényszerű lecsengés volt tapasztalható azokban a hálózatokban, melyek mérete elég nagy volt ahhoz, hogy viszonylag magas csoportfokszámok is előfordulhassanak. Ez utóbbi, a nagy csoportfokszámok tartományán hatványszerű csökkenés nagyon hasonlít a csúcsok közti „alaphálózat” fokszámeloszlásának viselkedésére. Azonban a kisebb csoportfokszámoknál tapasztalt, exponenciális függvényhez hasonló alak a csoportfokszám-eloszlás egyedi sajátossága, mely jelentősen eltér a csúcsok közti hálózat fokszámeloszlásától. Ez a viselkedés jól illeszkedik a komplex rendszereket leíró olyan általános képbe, mely szerint egy nagyméretű összetett rendszer különböző szerveződési szintjei alapvetően hasonló struktúrákat mutatnak, megengedve ugyanakkor egyes szintspecifikus eltéréseket is.

4. Analitikus úton meghatároztam az irányított k -klikkperkoláció kritikus élbekötési valószínűségét az irányított Erdős–Rényi-gráfban [T6]. Hasonlóan az irányítatlan esethez, ha az élbekötési valószínűség alacsonyabb, mint a kritikus érték, a legnagyobb irányított k -klikkperkolációs klaszter mérete elhanyagolható a rendszermérethez képest a termodinamikai határesetben, míg ha az élbekötési valószínűség meghaladja a kritikus értéket, a gráfban felbukkan egy óriási (a rendszermérethez képest sem elhanyagolható méretű) irányított k -klikkperkolációs klaszter.

Az irányított k -klikkperkolációs klaszterek feltárására kidolgoztam és beprogramoztam egy algoritmust, mely az irányított maximális klikkek megkeresésén alapszik [T6]. Az irányított k -klikkperkolációs klaszterek egy olyan hálózati csoportdefiníciót kínálnak, mely egyfelől figyelembe veszi az élek irányítottságát, másfelől megengedi a csoportok közti átfedéseket. A módszer segítségével feltártam több nagyméretű, irányított hálózat csoportszerkezetét. A kapott irányított csoportok esetén előfordult, hogy bizonyos csoporttagok egyfajta „forrásként”, míg mások „nyelőként” viselkedtek a csoporton belül a többi csoporttaghoz kapcsolódó éleik iránya alapján. Ennek mérésére bevezettem a csoporttagok (adott csoportra vonatkozó) relatív bejövő és relatív kimenő fokszámát [T6]. Ezek a mutatók segítettek olyan csoportok közt átvezető „hidak” beazonosításában, melyeknél az adott csúcs az egyik csoportban túlnyomórészt bejövő, míg a másikban már főként kimenő élekkel rendelkezett. Az ilyen csúcsok különösen fontosak lehetnek olyan hálózatokban, melyeknél az élek iránya valamilyen áramlás, terjedés irányát jelöli ki (mint pl. információáramlás, metabolizmus, stb.), hiszen ilyenkor ezen csúcsok az egyik csoportból „begyűjtene” a másikba meg „továbbítanak”.

-
5. Kidolgoztam egy eljárást, melynek segítségével időben változó rendszerek időlépésként rögzített eseményeiből kiindulva egy olyan hálózatot állíthatunk elő, melynek élsúlyai az időben folytonosan változnak [T7]. Ez a megközelítés különösen alkalmas emberi kapcsolathálók tanulmányozására olyan egyéb célokra létrehozott adatbázisok alapján, melyekben rendszeres időközönként gyűlnek különféle felhasználói adatok; ilyenek például az interneten fellelhető tudományos cikkarchívumok, az online közösségi hálók vagy a különböző telekommunikációs rendszerek. Az említett módszer az élsúlyok előállításánál figyelembe veszi, hogy az adatbázisban rögzített események hatása a kapcsolatok erősségére időben kiterjedt, hiszen egy esemény (pl. egy közös publikáció, egy telefonhívás, stb.) által érintett személyek között a kapcsolat sok esetben feltehetően már az esemény előtt is megvolt, illetve az esemény után is fennmarad. Az élek erőssége ebben a megközelítésben egyfelől a hozzájuk társuló események súlyától és gyakoriságától függ, és időben lecseng ha megszűnik a felek között az interakció.

A kidolgozott módszer segítségével két nagyméretű, emberi kapcsolatokat ábrázoló hálózat időfejlődő k -klikkperkolációs csoportszerkezetét tártam fel [T7]. Ezek egyike egy tudományos cikkarchívumban havi bontásban megjelenő publikációk alapján feltérképezett társszerzőségi hálózat volt, a másik pedig egy több millió felhasználóval rendelkező mobiltelefon-szolgáltató által kéthetes periódusokra aggregált anonimizált híváslisták alapján előálló hálózat. Az alapadatokból előállított, időfejlődő élsúlyok által definiált hálózatokban először minden időlépésnél megkerestem az aktuális állapotnak megfelelő, „pillanatfelvételszerű” statikus csoportokat, majd ezekből fűztem össze a dinamikus, időfejlődő csoportokat [T7]. A mobiltelefon-hívási hálózat esetén a felhasználókról rendelkezésre álló egyéb adatok csoportszintű elemzése azt mutatta, hogy a statikus csoportok egy-egy tulajdonság szerinti homogenitása jelentősen nagyobb, mint egy ugyanakkora, véletlenszerűen összeválogatott csoporté. Ez arra utalt, hogy a k -klikkperkolációs módszer vélhetően ténylegesen létező, egymáshoz közel álló emberekből álló csoportokat tárt fel [T7].

A dinamikus, időfejlődő csoportok esetén a tagösszetétel időbeli változékonyságának jellemzésére bevezettem a stacionaritás mennyiségét, mely megegyezik a szomszédos időlépésekben tapasztalt csoportösszetételek relatív átfedésének átlagával a csoport élete során [T7]. A vizsgált két nagyméretű hálózat időfejlődő csoportjainak elemzése alapján egy érdekes összefüggést tártam fel a stacionaritás, a csoportméret és a csoport várható élettartama között. (Ez utóbbi mennyiségre tekinthetünk úgy, mint egy „fitness” értékre: a fittebb csoportok tovább maradnak fenn, míg a kevésbé fitt csoportok hamar szétesnek, vagy beolvadnak egy másik csoportba). Az eredmények szerint a maximális átlagos élettartam alacsonyabb stacionaritás értékeknél volt a nagyméretű csoportok

esetén, mint a kisebb csoportoknál [T7]. Ez úgy interpretálható, hogy a kis csoportok várhatóan akkor maradnak fenn sokáig, ha időben nem igazán változik az összetételük, míg ezzel szemben a nagyméretű csoportoknak állandóan meg kell újulniuk a fennmaradásért, ezért optimális esetben az összetételük gyorsan változik.

6. Kidolgoztam a multifraktál alapú véletlengráf-modellben bevezetett általános véletlengráf-generáló eljárás részleteit [T8]. A módszer lényege, hogy egy viszonylag egyszerű, az egységnyezeten definiált, cellánként konstans generáló mértékből kiindulva egy rekurzív eljárás segítségével néhány iteráción keresztül jutunk el egy önhasználó struktúrával rendelkező élbekötési mértékhez, mely megadja a véletlen gráf éleinek valószínűségét. Az ezen eljárás segítségével előállított hálózatok legfontosabb jellemzőit (mint pl. fokszámoszlás, klaszterezettség) analitikus úton határoztam meg a paraméterek függvényében, és az eredményeket numerikus szimulációk segítségével ellenőriztem [T8]. A multifraktál alapú gráfgenerálás nagy előnye, hogy az ilyen úton előállított hálózatok tulajdonságai nagyon széles skálán változhatnak, a fokszámoszlás lehet például skálafüggetlen, vagy Poisson-eloszlású.

Mivel elég nagy szabadságunk van a generáló mérték megválasztásában, ha egy előre megadott tulajdonságokkal rendelkező véletlen gráfot szeretnénk előállítani ezzel a módszerrel, akkor az adott célfeladathoz optimális generáló mérték megadása egy nemtriviális feladat. Erre a problémára kidolgoztam egy szimulált hőkezelésen alapuló megoldást [T8], melynél egy véletlenszerű cellákból álló generáló mértékből kiindulva az eljárás során a paramétereket hozzáillesztjük a célul megjelölt tulajdonságokhoz (mint pl. kívánt fokszámoszlás, klaszterezettség, stb.). A numerikus szimulációk megmutatták, hogy ennek révén egy igen flexibilis általános véletlen gráfgeneráló módszert kapunk, mely képes egymástól jelentősen eltérő tulajdonságokkal rendelkező véletlen gráfok előállítására ugyanazon keretrendszeren belül [T8].

A tézispontokhoz kapcsolódó saját publikációk

- [T1] G. Palla – I. Derényi – I. Farkas – T. Vicsek: Statistical mechanics of topological phase transitions in networks. *Phys. Rev. E*, 69. évf. (2004), 046117. p.
- [T2] I. Derényi – I. Farkas – G. Palla – T. Vicsek: Topological phase transitions of random networks. *Physica A*, 334. évf. (2004), 583–590. p.
- [T3] G. Palla – I. Derényi – T. Vicsek: The critical point of k-clique percolation in the Erdős-Rényi graph. *J. Stat. Phys.*, 128. évf. (2007), 219–227. p.
- [T4] I. Derényi – G. Palla – T. Vicsek: Clique percolation in random networks. *Phys. Rev. Lett.*, 94. évf. (2005), 160202. p.
- [T5] G. Palla – I. Derényi – I. Farkas – T. Vicsek: Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435. évf. (2005), 814–818. p.
- [T6] G. Palla – I. J. Farkas – P. Pollner – I. Derényi – T. Vicsek: Directed network modules. *New J. Phys.*, 9. évf. (2007), 186. p.
- [T7] G. Palla – A.-L. Barabási – T. Vicsek: Quantifying social group evolution. *Nature*, 446. évf. (2007), 664–667. p.
- [T8] G. Palla – L. Lovász – T. Vicsek: Multifractal network generator. *Proc. Natl. Acad. Sci. USA*, 107. évf. (2010), 7640–7645. p.

Irodalomjegyzék

- [1] L. A. Adamic – B. A. Huberman: Internet: growth dynamics of the world-wide web. *Nature*, 401. évf. (1999), 131. p.
- [2] R. Albert – A.-L. Barabási: Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74. évf. (2002), 47–97. p.
- [3] R. Albert – H. Jeong – A.-L. Barabási: Diameter of the world wide web. *Nature*, 401. évf. (1999), 130–131. p.
- [4] V. A. Avetisov – A. V. Chertovich – S. K. Nechaev – O. A. Vasilyev: On scale-free and poly-scale behaviors of random hierarchical network. *J. Stat. Mech.*, 2009., P07008. p.
- [5] A.-L. Barabási – R. Albert: Emergence of scaling in random networks. *Science*, 286. évf. (1999), 509–512. p.
- [6] A.-L. Barabási – H. Jeong – Z. Neda – E. Ravasz – A. Schubert – T. Vicsek: Evolution of the social network of scientific collaborations. *Physica A*, 311. évf. (2002), 590–614. p.
- [7] E. R. Barnes: An algorithm for partitioning the nodes of a graph. *SIAM J. Algebr. Discrete Methods*, 3. évf. (1982), 541–550. p.
- [8] J. Berg – M. Lässig: Correlated random networks. *Phys. Rev. Lett.*, 89. évf. (2002), 228701. p.
- [9] V. D. Blondel – J.-L. Guillaume – R. Lambiotte – E. Lefebvre: Fast unfolding of communities in large networks. *J. Stat. Mech.*, 2008., P10008. p.
- [10] M. Boguñá – R. Pastor-Satorras: Class of correlated random networks with hidden variables. *Phys. Rev. E*, 68. évf. (2003), 036112. p.
- [11] B. Bollobás: *Random graphs*. 2. kiad. Cambridge, 2001, Cambridge University Press.

-
- [12] B. Bollobás–S. Janson–O. Riordan: The phase transition in inhomogeneous random graphs. *Random Struct. Algor.*, 31. évf. (2007), 3–122. p.
- [13] B. Bollobás–O. Riordan: Random graphs and branching process. In B. Bollobás–R. Kozma–D. Miklós (szerk.): *Handbook of Large-scale Random Networks*. Berlin, 2009, Springer, 15–115. p.
- [14] C. Borgs–J. Chayes–L. Lovász–V. T. Sós–K. Vesztegombi: Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219. évf. (2008), 1801–1851. p.
- [15] U. Brandes–D. Dellling–M. Gaertler–R. Görke–M. Hoefer–Z. Nikoloski–D. Wagner: On modularity clustering. *IEEE T. Knowl. Data En.*, 20. évf. (2008), 172–188. p.
- [16] S. R. Broadbent–J. M. Hammersley: Percolation processes i. crystals and mazes. *Proc. Camb. Phil. Soc.*, 53. évf. (1957), 629–641. p.
- [17] Z. Burda–J. D. Correia–A. Krzywicki: Statistical ensemble of scale-free random graphs. *Phys. Rev. E*, 64. évf. (2001), 046118. p.
- [18] Z. Burda–A. Krzywicki: Uncorrelated random networks. *Phys. Rev. E*, 67. évf. (2003), 046118. p.
- [19] G. Caldarelli–A. Capocci–P. De Los Rios–M. A. Muñoz: Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.*, 89. évf. (2002), 258702. p.
- [20] R. F. Cancho–R. V. Solé: The small world of human language. *Proc. R. Soc. Lond. B*, 268. évf. (2001), 2261–2265. p.
- [21] J. Duch–A. Arenas: Community detection in complex networks using extremal optimization. *Phys. Rev. E*, 72. évf. (2005), 027104. p.
- [22] H. Ebel–J. Davidsen–S. Bornholdt: Dynamics of social networks. *Complexity*, 8. évf. (2002), 24–27. p.
- [23] P. Erdős–A. Rényi: On random graphs I. *Publ. Math. Debrecen*, 6. évf. (1959), 290–297. p.
- [24] P. Erdős–A. Rényi: On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5. évf. (1960), 17–61. p.
- [25] L. Euler: Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8. évf. (1741), 128–140. p.

-
- [26] B. S. Everitt: *Cluster Analysis*. 3. kiad. London, 1993, Edward Arnold.
- [27] M. Faloutsos–P. Faloutsos–C. Faloutsos: On power-law relationships of the internet topology. *Comput. Commun. Rev.*, 29. évf. (1999), 251. p.
- [28] Mérei Ferenc: *Közösségek rejtett hálózata*. Budapest, 1971, Közgazdasági és Jogi Könyvkiadó.
- [29] S. Fortunato: Community detection in graphs. *Phys. Rep.*, 486. évf. (2010), 75–174. p.
- [30] S. Fortunato–M. Barthélemy: Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA*, 104. évf. (2007), 36–41. p.
- [31] O. Frank–D. Strauss: Markov graphs. *J. Am. Stat. Assoc.*, 81. évf. (1986), 832–842. p.
- [32] M. Girvan–M. E. J. Newman: Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99. évf. (2002), 7821–7826. p.
- [33] M. Granovetter: *Decision Making: Alternatives to Rational Choice Models Economic Action and Social Structure: The Problem of Embeddedness*. Newbury Park, 1992, CA:SAGE.
- [34] G. R. Grimmett: *Percolation*. 2. kiad. Berlin Heidelberg, 1999, Springer-Verlag.
- [35] R. Guimerà–L. Danon–A. Diaz-Guilera–F. Giralt–A. Arenas: Self-similar community structure in organisations. *Phys. Rev. E*, 68. évf. (2003), 065103. p.
- [36] R. Guimerà–B. Uzzi–J. Spiro–L. A. N. Amaral: Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science*, 308. évf. (2005), 697–702. p.
- [37] R. Guimerà–L. A. N. Amaral: Functional cartography of complex metabolic networks. *Nature*, 433. évf. (2005), 895–900. p.
- [38] R. Guimerà–A. Díaz-Guilera–F. Vega-Redondo–A. Cabrales–A. Arenas: Optimal network topologies for local search with congestion. *Phys. Rev. Lett.*, 89. évf. (2002), 248701. p.
- [39] T. Heimo–J. Saramäki–J.-P. Onnela–K. Kaski: Spectral and network methods in the analysis of correlation matrices of stock returns. *Physica A*, 383. évf. (2007), 147–151. p.
- [40] P. W. Holland–S. Leinhardt: Transitivity in structural models of small groups. *Comp. Group Stud.*, 2. évf. (1971), 107–124. p.

-
- [41] P. Holme–Ch. R. Edling–F. Liljeros: Structure and time-evolution of an internet dating community. *Soc. Networks*, 26. évf. (2004), 155–174. p.
- [42] J. Hopcroft–O. Khan–B. Kulis–B. Selman: Tracking evolving communities in large linked networks. *Proc. Natl. Acad. Sci. USA*, 101. évf. (2004), 5249–5253. p.
- [43] H. Jeong–B. Tombor–R. Albert–Z. N. Oltvai–A.-L. Barabási: The large-scale organization of metabolic networks. *Nature*, 407. évf. (2000), 651–654. p.
- [44] B. W. Kernighan–S. Lin: An efficient heuristic procedure for partitioning graphs. *Bell Syst. Tech. J.*, 49. évf. (1970), 291–307. p.
- [45] S. Knudsen: *A Guide to Analysis of DNA Microarray Data*. 2. kiad. 2004, Wiley-Liss.
- [46] D. König: *Theorie der endlichen und unendlichen Graphen*. Providence, Rhode Island, 1936, AMS Chelsea Publishing.
- [47] G. Kossinets–D. J. Watts: Empirical analysis of an evolving social network. *Science*, 311. évf. (2006), 88–90. p.
- [48] R. Kumar–P. Raghavan–S. Rajalopagan–A. Tomkins: Extracting large-scale knowledge bases from the web. In *Proceedings of the 9th ACM Symposium on Principles of Database Systems* (konferenciaanyag). 1999, 1. p.
- [49] J. Leskovec–D. Chakrabarti–J. Kleinberg–C. Faloutsos: Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. In *European Conference on Principles and Practice of Knowledge Discovery in Databases* (konferenciaanyag). 2005, 133–145. p.
- [50] J. Leskovec–C. Faloutsos: Scalable modeling of real graphs using kronecker multiplication. In *Proceedings of the 24th International Conference on Machine Learning* (konferenciaanyag). 2007, 497–504. p.
- [51] Ch. Li–Ph. K. Maini: An evolving network model with community structure. *J. Phys. A*, 38. évf. (2005), 9741–9749. p.
- [52] F. Liljeros–Ch. R. Edling–L. A. N. Amaral–H. E. Stanley–Y. Aberg: The Web of Human Sexual Contacts. *Nature*, 411. évf. (2001), 907–908. p.
- [53] L. Lovász: *Large Networks and Graph Limits*. American Mathematical Society Colloquium Publications sorozat, 60. köt. Providence, RI, 2012, American Mathematical Society.

-
- [54] L. Lovász–B. Szegedy: Limits of dense graph sequences. *J. Comb. Theory B*, 96. évf. (2006), 933–957. p.
- [55] J. B. MacQueen: Some methods for classification and analysis of multivariate observations. In L. M. L. Cam–J. Neyman (szerk.): *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (konferenciaanyag), 1. köt. Berkeley, USA, 1967, University of California Press, 281–297. p.
- [56] P. Mahadevan–D. Krioukov–K. Fall–A. Vahdat: Systematic topology analysis and generation using degree correlations. *ACM SIGCOMM Comp. Comm. Rev.*, 36. évf. (2006), 135–146. p.
- [57] J. F. F. Mendes–S. N. Dorogovtsev: *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford, 2003, Oxford University Press.
- [58] M. Molloy–B. Reed: A critical point for random graphs with a given degree sequence. *Random Struct. Algor.*, 6. évf. (1995), 161–179. p.
- [59] M. Molloy–B. Reed: The size of the giant component of a random graph with a given degree sequence. *Comb. Probab. Comput.*, 7. évf. (1998), 295–305. p.
- [60] J. M. Montoya–R. V. Solé: Small world patterns in food webs. *J. Theor. Biol.*, 214. évf. (2002), 405–412. p.
- [61] M. E. J. Newman: Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Phys. Rev. E*, 2001., 016132. p.
- [62] M. E. J. Newman: The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98. évf. (2001), 404–409. p.
- [63] M. E. J. Newman: Detecting community structure in networks. *Eur. Phys. J. B*, 38. évf. (2004), 321–330. p.
- [64] M. E. J. Newman: Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74. évf. (2006), 036104. p.
- [65] M. E. J. Newman: From the cover: Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, 103. évf. (2006), 8577–8582. p.
- [66] M. E. J. Newman–M. Girvan: Finding and evaluating community structure in networks. *Phys. Rev. E*, 69. évf. (2004), 026113. p.
- [67] M. E. J. Newman–J. Park: Why social networks are different from other types of networks. *Phys. Rev. E*, 68. évf. (2003), 036122. p.

-
- [68] J.-P. Onnela – A. Chakraborti – K. Kaski – J. Kertész – A. Kanto: Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E*, 68. évf. (2003), 056110. p.
- [69] J.-P. Onnela – J. Saramäki – J. Hyvönen – G. Szabó – M. A. de Menezes – K. Kaski – A. L. Barabási – J. Kertész: Analysis of a large-scale weighted network of one-to-one human communication. *New J. Phys.*, 9. évf. (2007), 179. p.
- [70] J.-P. Onnela – J. Saramäki – J. Hyvönen – G. Szabó – D. Lazer – K. Kaski – J. Kertész – A.-L. Barabási: Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. USA*, 104. évf. (2007), 7332–7336. p.
- [71] R. Pastor-Satorras – A. Vázquez – A. Vespignani: Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87. évf. (2001), 258701. p.
- [72] D. J. De Sola Price: Networks of scientific papers. *Science*, 14. évf. (1965), 510–515. p.
- [73] E. Ravasz – A. L. Somera – D. A. Mongru – Z. N. Oltvai – A.-L. Barabási: Hierarchical organization of modularity in metabolic networks. *Science*, 297. évf. (2002), 1551–1555. p.
- [74] S. Redner: How popular is your paper? an empirical study of the citation distribution. *Eur. Phys. J. B*, 4. évf. (1998), 131–134. p.
- [75] J. Reichardt – S. Bornholdt: Statistical mechanics of community detection. *Phys. Rev. E*, 74. évf. (2006), 016110. p.
- [76] T. Richardson – P. J. Mucha – M. A. Porter: Spectral tripartitioning of networks. *Phys. Rev. E*, 80. évf. (2009), 036111. p.
- [77] A. W. Rives – T. Galitski: Modular organization of cellular networks. *Proc. Natl. Acad. Sci. USA*, 100. évf. (2003), 1128–1133. p.
- [78] G. Robins – T. Snijders P. Wang M. Handcock – P. Pattison: Recent developments in exponential random graph (p^*) models for social networks. *Soc. Networks*, 29. évf. (2007), 192–215. p.
- [79] J. Scott: *Social Network Analysis: A Handbook*. 2. kiad. London, 2000, Sage Publications.
- [80] R. M. Shiffrin – K. Börner: Mapping knowledge domains. *Proc. Natl. Acad. Sci. USA*, 101. évf. (2004), 5183–5185. p.

-
- [81] V. Spirin – K. A. Mirny: Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. USA*, 100. évf. (2003), 12123–12128. p.
- [82] D. Stark – B. Vedres: Pathways of property transformation: Enterprise network careers in hungary, 1988-2000. 2001-12-081. Jelentés, 2001, Santa Fe Institute Working Paper.
- [83] Y. Sun – B. Danila – K. Josic – K. E. Bassler: Improved community structure detection using a modified fine-tuning strategy. *Europhys. Lett.*, 86. évf. (2009), 28004. p.
- [84] G. Szabó – G. Fáth: Evolutionary games on graphs. *Phys. Rep.*, 446. évf. (2007), 97–216. p.
- [85] G. Szabó – J. Vukov – A. Szolnoki: Phase diagrams for an evolutionary prisoner’s dilemma game on two-dimensional lattices. *Phys. Rev. E*, 72. évf. (2005), 047107. p.
- [86] S. Valverde – R. Ferrer Cancho – R. V. Solé: Scale-free networks from optimal design. *Europhys. Lett.*, 60. évf. (2002), 512. p.
- [87] J. Vukov – G. Szabó – A. Szolnoki: Cooperation in the noisy case: Prisoner’s dilemma game on two types of regular random graphs. *Phys. Rev. E*, 73. évf. (2006), 067103. p.
- [88] A. Wagner – D. A. Fell: The small world inside large metabolic networks. 2000. Tech. Rep. 00-07-041, Santa Fe Institute.
- [89] C. S. Wagner – L. Leydesdorff: Network structure, self-organization, and the growth of international collaboration in science. *Res. Policy*, 34. évf. (2005), 1608–1618. p.
- [90] G. Wang – Y. Shen – M. Ouyang: A vector partitioning approach to detecting community structure in complex networks. *Comput. Math. Appl.*, 55. évf. (2008), 2746–2752. p.
- [91] S. Warner: E-prints and the Open Archives Initiative. *Library Hi Tech*, 21. évf. (2003), 151–158. p.
- [92] S. Wasserman – K. Faust: *Social network analysis: methods and applications Structural Analysis in the Social Sciences*. Cambridge, 1994, Cambridge University Press.

-
- [93] S. Wasserman–P. E. Pattison: Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p^* . *Psychometrika*, 61. évf. (1996), 401–425. p.
- [94] D. J. Watts: A twenty-first century science. *Nature*, 445. évf. (2007), 489. p.
- [95] D. J. Watts–P. S. Dodds–M. E. J. Newman: Identity and search in social networks. *Science*, 296. évf. (2002), 1302–1305. p.
- [96] D. J. Watts–S. H. Strogatz: Collective dynamics of 'small-world' networks. *Nature*, 393. évf. (1998), 440–442. p.
- [97] H. C. White–S. A. Boorman–R. R. Breiger: Social structure from multiple networks. I. Blockmodels of roles and positions. *Am. J. Sociol.*, 81. évf. (1976), 730–780. p.
- [98] Y.-Y. Yeung–T. C.-Y. Liu–P.-H. Ng: A social network analysis of research collaboration in physics education. *Am. J. Phys.*, 73. évf. (2005), 145–150. p.