

Komplex hálózatok szerkezetének és dinamikájának feltárása és modellezése statisztikus fizikai módszerekkel

Az MTA Doktora cím elnyeréséhez készített disszertáció

Palla Gergely



MTA-ELTE Statisztikus és Biológiai Fizika Kutatócsoport

2016. március

Tartalomjegyzék

Bevezetés	1
A dolgozat tudományterülete	1
A dolgozatban tárgyalt kérdések	4
1. Topológiai fázisátalakulások	9
1.1. Hálózatok statisztikus sokasága	11
1.2. Numerikus módszerek	14
1.2.1. Egzakt leszámlálás	14
1.2.2. Monte-Carlo-módszer	16
1.3. Egyrészcskés Hamilton-függvények	16
1.3.1. Egy Ising-modellel ekvivalens energiafüggvény	18
1.3.2. Folytonos fázisátalakulás és skálafüggetlen gráfok	22
1.3.3. A nagykanonikus sokaság	26
2. k-klikkperkoláció véletlen gráfokban	27
2.1. Definíciók	29
2.2. k -klikkperkoláció Erdős-Rényi-gráfban	31
2.2.1. Generátorfüggvény formalizmus	31
2.2.2. Numerikus szimulációk	35
2.3. Irányított k -klikkperkoláció	38
2.3.1. Irányított k -klikkek	38
2.3.2. Irányított k -klikkperkoláció az Erdős-Rényi-gráfban	42
3. Hálózati csoportkeresés k-klikkperkolációval	45
3.1. A k -klikkperkoláció mint átfedő csoportkeresési módszer	49
3.1.1. A csoportkeresési algoritmus	50
3.1.2. Az irányított csoportkeresési algoritmus	53
3.2. Lokális csoportszerkezet	55
3.3. Teljes csoportszerkezet	60
3.3.1. A paraméterek optimális beállítása	60
3.3.2. Csoportstatisztikák	61
3.3.3. A csoportok hálózata	65

4. Időfejlődő csoportdinamika	67
4.1. A hálózatok és a csoportok előállítása	69
4.1.1. Súlyozott, időfejlődő hálózatok	69
4.1.2. A statikus csoportok	70
4.1.3. A csoportok ellenőrzése	72
4.1.4. Időfejlődő csoportok	75
4.2. A csoport-időfejlődés statisztikus tulajdonságai	78
4.2.1. Alapvető statisztikák	78
4.2.2. Stacionaritás és élettartam	80
4.2.3. A csoportmegszűnés jóslása	82
5. Multifraktál alapú véletlengráf-modell	85
5.1. Gráfsorozatok termodinamikai határesetek	88
5.2. Véletlengráf-generálás multifraktállal	90
5.3. Statisztikák meghatározása analitikus úton	93
5.3.1. Fokszámeloszlás	94
5.3.2. Klaszterezettségi együttható	95
5.3.3. Szomszédok átlagos fokszáma	95
5.3.4. Az izolált csúcsok problémája	96
5.4. A gráfgeneráló optimalizálása	98
Összefoglalás	103
Köszönetnyilvánítás	107
Függelék	108
A Komponensszintű Hamilton-függvények	109
B További eredmények a k-klikkperkolációval kapcsolatban	117
C A csoportkereséssel kapcsolatos további eredmények	123
D A multifraktál alapú gráfgeneráló termodinamikai határesetek	135
Saját publikációk	139
Irodalomjegyzék	140

Bevezetés

A komplex hálózatok témaköre egy hagyományosan interdiszciplináris kutatási terület, mely egyaránt szorosan kötődik a gráfelmélethez, a szociológiához, a biológiához, az informatikához, a közgazdaságtanhoz, és végül, de nem utolsósorban a statisztikus fizikához. A hálózatos megközelítés alapgondolata az, hogy egy vizsgált rendszert bizonyos alapegységek (pl. routerek, web-lapok, személyek, gének, fehérjék, vállalatok, stb.), és ezen alapegységek közti kapcsolatok, kölcsönhatások együtteseként reprezentálunk, egy gráfstruktúrát hozva létre. Ezt a leírásmódot sikeresen alkalmazták gén- és fehérje-kölcsönhatások vizsgálatától, valamint biokémiai folyamatok tanulmányozásától kezdve, telekommunikációs, e-mail és online közösségi hálózatokon keresztül az Internetig, közlekedési, elektromos és egyéb technikai hálózatokig, illetve cégek, bankok és egyéb gazdasági szereplők közti hálózatok tanulmányozásáig. A dolgozat e rendkívül szerteágazó alkalmazási területek közös elméleti háttéréből válogat, olyan témakörökre koncentrálva, melyek nagyon erősen kötődnek a statisztikus fizikához.

A dolgozat tudományterülete

A hálózatkutatás gyökerei a gráfelmélethez nyúlnak vissza, mely Leonhard Euler munkássága révén született meg 1735-ben, amikor a tudós a legendás königsbergi hidak problémájával kezdett foglalkozni [66]¹. A nevezetes feladvány egy olyan bejárását kereste Königsberg hídjainak, mely minden hídon pontosan egyszer halad át, az útvonal megszakítása nélkül. Euler frappáns módon bebizonyította, hogy ilyen bejárás nem adható, és a bizonyítás során bevezette a csúcsok (csomópontok), és a csúcsokat összekötő élek (vonalak) fogalmát, egy természetes reprezentációt kínálva Königsberg városrészei és az azokat összekötő hidak számára a fenti problémában. Magát a „gráf” kifejezést J. J. Sylvester használta először 1878-ban, a kémiában a molekulák ábrázolására használt diagramok mintájára [198].

Az Euler úttörő munkássága révén kialakuló gráfelmélet a XX. század elejére már egy fontos, különálló területévé vált a diszkrét matematikának; az első gráfelméleti kézikönyv 1936-ban jelent meg König Dénes szerzőségével [112]. Ezen terület fejlődése során a komplex hálózatok szempontjából egy különösen fontos mérföldkő volt a véletlen gráf fogalmának bevezetése 1959-ben, mely Erdős Pál és Rényi Alfréd nevéhez kötődik [63, 64]. Az általuk definiált gráf úgy jön létre, hogy N csúcs között minden lehetséges párt véletlen-

¹A probléma megoldását először 1735-ben adta elő a Szentpétervári Akadémián, a publikáció viszont csak 1741-ben jelent meg.

szerűen összekötünk egy adott p valószínűséggel (páronként függetlenül). Ezzel átlagosan $pN(N-1)/2$ élt hozunk létre a gráfban, melyek elhelyezkedése teljesen „homogén” és véletlenszerű. (Természetesen egy másik alternatíva véletlen gráf létrehozására az, hogy nem az élbekötés valószínűségét adjuk meg, hanem direkt módon a véletlenszerűen leosztott élek számát, ahogy azt Edgar N. Gilbert javasolta [80].)

Az Erdős–Rényi-gráf egyik alapvető tulajdonsága, hogy a csúcsok átlagos távolsága (extrém élbekötési valószínűségektől eltekintve) a csúcsok számának logaritmusával skálázik². Ez egy nagyon fontos eltérés például egy szabályos rácshoz képest, ahol az átlagos távolság a rácspontok számának függvényében hatványszerűen változik. Ezt a jelenséget szokás „kisvilág-effektusnak” hívni, illetve általánosabban az olyan hálózatokat, melyekben az átlagos távolság nagyon kicsi a rendszermérethez képest, „kisvilág-tulajdonságúnak” nevezzük. Az régóta sejthető volt, hogy például az emberi kapcsolatok hálózata kisvilág-tulajdonságú. Elsőként Karinthy Frigyes vetette fel az 1929-ben megjelent Láncszemek című novellájában [108] azt a feltételezést, hogy akár különböző kontinenseken véletlenszerűen választott emberek között is meglepően kevés ismerősön keresztül vezet út. Az első mérés, mely alátámasztotta ezt a tulajdonságot, Stanley Milgram amerikai szociológus nevéhez fűződik 1967-ben [134]. Az általa vezetett kísérletben véletlenszerűen választott önkénteseket kértek meg Nebraskában és Kansasban, hogy juttassanak el egy levelezőlapot egy számukra ismeretlen bostoni célszemélyhez személyes ismeretségek sorozatán keresztül. Érdekes módon a levelezőlapok egy része célba ért, és ezek átlagosan 5-6 lépést tettek meg. Ez az eredmény a „hat lépés távolság” néven vált híressé, és nagy mértékben hozzájárult ahhoz, hogy az emberi kapcsolatok hálózatának kisvilág-tulajdonsága széles körben ismertté váljon.

A kisvilág-tulajdonság, valamint a teljesen véletlenszerű élleosztás miatt az Erdős–Rényi-gráf nagyon sokáig szinte „egyeduralkodó” elméleti modellje volt általában a komplex rendszereket leíró hálózatoknak. Ebben döntő szerepe volt annak is, hogy egészen az 1990-es évek végéig nem álltak rendelkezésre könnyen hozzáférhető, részletes adatok az ilyen hálózatokról. Az természetesen nyilvánvaló volt, hogy például az emberi kapcsolatok hálózata, vagy a közlekedési hálózatok, a gazdasági hálózatok, stb., nem alkotnak olyan szabályos struktúrát, mint egy kristályrács, hiszen egyes kiragadott csúcsok lokális környezetét vizsgálva eltérések, rendezetlenség, látszólagos véletlenszerűségek voltak felfedezhetők. (Pl. az egyes emberek közvetlen ismerőseinek száma más és más, ezen ismerősök közti kapcsolatháló is egyénenként változó, stb.) E rendezetlenségek és a nagy skálájú, részletes adatok hiánya miatt kézenfekvő volt Erdős–Rényi-gráfként modellezni ezeket a hálózatokat, melyben minden kapcsolat teljesen véletlenszerűen jön létre.

Ez a kép gyökeresen változott meg az ezredforduló környékén, amikor az informatikai, információtechnológiai fejlődésnek köszönhetően hirtelen egyre több és több nagyméretű hálózatos adatbázis jött létre és vált széles körben hozzáférhetővé [7, 130]. A társszerzői [145, 146, 15], tudományos hivatkozási [167, 176], internetes és webes [69, 2, 115, 8, 163], különféle biológiai [210, 138, 218], társas kapcsolati [123], nyelvi [43], valamint filmes hálózatok [218] első elemzései hamar kimutatták, hogy az Erdős–Rényi-modell több lényeges

²Két csúcs távolsága alatt az őket összekötő legrövidebb út éleinek számát értjük.

ponton eltér a valós, komplex rendszereket reprezentáló hálózatoktól³. Az egyik fontos eltérés abban rejlik, hogy ez utóbbi hálózatok nagy mértékű tranzitivitást mutatnak a következő értelemben: egy csúcs közvetlen szomszédai jó eséllyel egymásnak is közvetlen szomszédai. Ezen tulajdonság mérésére vezették be a csúcsok klaszterezettségi együtthatóját [97, 214, 218], mely megadja, hogy az adott csúcs szomszédai közt lehetséges élek mekkora hányada van ténylegesen jelen. E mutató tükrében a fent említett eltérés egy valódi rendszert leíró hálózat és egy vele azonos számú csúcsot és élt tartalmazó Erdős–Rényi-gráf között az, hogy az átlagos klaszterezettségi együttható az Erdős–Rényi-gráfban sokkal kisebb [218]. Az első alternatív véletlen gráf modellt, mely a kisvilág-tulajdonság megtartása mellett relatíve nagy átlagos klaszterezettségi együtthatójú gráfokat képes generálni, Duncan Watts és Steven Strogatz vezették be 1998-ban [218].

Megközelítésük lényege, hogy egy szabályos rácshoz hasonló struktúrából kiindulva véletlenszerűen átkötjük az élek egy bizonyos hányadát. A kezdeti állapot kialakítása olyan, hogy biztosítja a nagy klaszterezettségi együtthatót, ám a szabályos rácshoz hasonló szerkezet miatt az átlagos távolság a rendszermérettel hatványszerűen skálázik, azaz a hálózat ezen a ponton még nem kisvilág-tulajdonságú. Ezért van szükség az élek kis hányadának véletlenszerű átkötésére: általában már néhány véletlenszerű, „hosszútávú” él is elég ahhoz, hogy az átlagos távolság drasztikusan lecsökkenjen és fellépjen a kisvilág-effektus. Ugyanakkor a magas klaszterezettségi együtthatót az élek kis hányadának átkötése még nem tudja lerontani. Ennek révén a Watts–Strogatz-modellben az átkötött élek hányadának növelése során egy viszonylag széles tartományban a kapott hálózat egyaránt rendelkezik a kisvilág-tulajdonsággal és a nagy klaszterezettségi együtthatóval.

A másik alapvető eltérés a valós rendszereket leíró hálózatok és az Erdős–Rényi-modell között az egyes csúcsokhoz kapcsolódó élek számának eloszlásában rejlik. Egy adott csúcs éleinek számát a csúcs fokszámának hívjuk, ennek eloszlása a fokszámeloszlás. Az Erdős–Rényi-gráfban ez lényegében Poisson-eloszlást követ, ezzel ellentétben a mérések nagy relatív szórású, a nagy fokszámok tartományán hatványszerűen lecsengő eloszlásokat mutattak [167, 176, 69, 115, 8, 14, 105, 210, 146, 123, 43]. Az ilyen tulajdonságú hálózatokat skálafüggetlennek hívjuk. Az első véletlengráf-modell, mely skálafüggetlen fokszámeloszlású gráfot generál, Barabási Albert-László és Albert Réka nevéhez kötődik [14].

Az általuk 1999-ben javasolt megközelítés egyik sajátossága, hogy a gráf egy növekedési folyamat révén áll elő, melyben a csúcsokat egyesével adjuk hozzá a rendszerhez. A másik nagyon lényeges eleme a modellnek az ún. preferenciális kapcsolódási szabály⁴: az új csúcsok bekötésénél a már meglévő csúcsok közül a fokszámmal arányos valószínűség szerint választunk véletlenszerűen. Ezáltal a már sok kapcsolatot begyűjtő, nagy fokszámú csúcsok-

³Bár Derek J. de Solla Price már 1965-ben felismerte, hogy a tudományos hivatkozási hálózatok fokszám-eloszlása az Erdős–Rényi-modelltől eltérően hatványszerűen cseng le [167], eredményei szélesebb körben csak a 2000-es évek elején váltak ismertté.

⁴A preferenciális kapcsolódás gondolata egyéb területeken már korábban is felvetődött, mint például a Pólya-folyamat esetén [62] a matematikában, a Yule-folyamatnál [225] a biológiában, a Gibrat-modellnél [79] a közgazdaságtanban, az ún. Máté-effektusnál a szociológiában [131], a vagyoneeloszlás lassú lecsengésének lehetséges magyarázatánál [229], a városméret, a szógyakoriság és egy adott szerző által közölt tudományos publikációk számára vonatkozó eloszlások hatványszerű lecsengésére vonatkozó Simon-modellnél [191], valamint a tudományos hivatkozottság hatványszerű lecsengésére vonatkozó Price-modellnél [168].

nak sokkal jobb esélye lesz az új csúcsokkal érkező új élek megszerzésére egy kis fokszerű csúcsra viszonyítva. Belátható, hogy ezen a „gazdag még gazdagabbá válik” effektus révén hosszútávon egy hatványszerűen lecsengő (azaz skálafüggetlen) fokszerűeloszlás alakul ki a rendszerben.

A Watts–Strogatz-modell és a Barabási–Albert-modell bevezetése robbanásszerű lökést adott a terület fejlődésének, és manapság a modern hálózat tudomány, vagy más néven a komplex hálózatok témakörének kialakulását sokan e két rendkívül nagy hatású eredmény 1998-as illetve 1999-es megjelenéséhez kapcsolják. A terület kezdeti fejlődésének fontosabb eredményeiről nagyon jó összefoglalót nyújt az Albert Réka és Barabási Albert-László által közölt [7] publikáció, valamint a José F. F. Mendes és Szergej N. Dorogovcev szerzőségével megjelent könyv [130]. Egy körülbelül 5 évvel későbből nyújt áttekintést a Mark E. J. Newman, Barabási Albert-László és Duncan J. Watts szerkesztése mellett kiadott könyv [152]. Ezt követte 2010-ben egy összefoglaló jellegű, akár tankönyvként is nagyon jól használható könyv megjelenése, melynek szerzője szintén Mark E. J. Newman volt [151]. Végül, de nem utolsósorban egy nagyon alaposan kidolgozott oktatási segédanyag egyetemista és doktorandusz hallgatók számára a 2015-ben megjelenő, online már elérhető tankönyv Barabási Albert-László szerzőségével [13], mely nagy részletességgel tárgyalja a témakör legfontosabb fejezeteit. A komplex hálózatok matematikai hátterét a véletlengráf-elmélet adja, melyhez egy nagyszerű bevezetést nyújt Bollobás Béla könyve [30]. Egy további nagyon érdekes olvasmányt nyújt Csermely Péter könyve a gyenge élek szerepéről a biokémiai és biológiai hálózatok, a társadalmi és emberi kapcsolathálók, valamint a gazdasági hálózatok stabilitásában, stabilizálásában [53].

A témakör jelenlegi népszerűségét jelzi, hogy évente több ezer tudományos publikáció jelenik meg, melyek teljesen vagy részben a komplex hálózatokhoz köthetők. Ennélfogva a terület természetes módon tovább bontható egymástól néhol jobban, néhol kevésbé jól elkülönülő részterületekre, mint például a hálózati csoportkeresés (vagy más néven hálózatklasszterezés) problémaköre, az időfejlődő hálózatok témaköre, az emberi kapcsolatok hálózatának vizsgálata és modellezése, a hálózati topológián zajló különböző dinamikai folyamatok vizsgálata, a többretegű hálózatok területe, és végül de nem utolsósorban a biológiai és bioinformatikai hálózatok nagyon tág (és ugyanakkor rendkívül fontos) témaköre. A dolgozat ezen részterületek közül többet is érint, azonban a mi esetünkben a tárgyalt kérdések közös vonása a statisztikus fizikához való szoros kötődés.

A dolgozatban tárgyalt kérdések

A statisztikus fizikai megközelítés nagyon fontos eleme volt a modern hálózatkutatásnak a kezdetektől fogva. Ennek egyik oka abban rejlik, hogy a vizsgált rendszerek általában nagyon sok alapegységet tartalmaznak, és emiatt egy „mikroszkopikus” leírásmód, melyben minden egyes egységet és annak kapcsolatait külön vizsgáljuk, kezelhetetlenné válik. Ehelyett általában, a statisztikus fizikában megszokott sokrészecskés rendszerekhez hasonlóan, eloszlásokkal, átlagokkal és egyéb statisztikai jellemzőkkel dolgozunk. Ami a fizikusok számára nagyon érdekes a komplex hálózatok statisztikus jellemzőiben az a nagy fokú univerzalitás: annak ellenére, hogy a korábban felsorolt rendszereknek (mint pl. az Internet,

a publikációs, filmes vagy biológiai adatbázisok, stb.) látszólag semmi közük sincs egymáshoz, a belőlük képzett hálózatok statisztikus tulajdonságai mégis nagyon sok hasonlóságot mutatnak. Ez azt sejteti, 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. Ennél fogva nagyon sok fizikus kapcsolódott be a hálózatkutatásba, magukkal hozva a statisztikus fizikai szemléletmódot, melynek révén számos statisztikus fizikából származó módszer, megközelítés, modell került átültetésre a komplex hálózatok témakörébe.

A természet és a társadalom különböző alrendszeit leíró hálózatok struktúrája időben gyakran változó képet mutat. Vannak olyan rendszerek, amelyeknél a hálózat mérete folyamatosan növekszik (pl. az Internet), más esetekben a rendszerméret nagyjából állandó, viszont a csúcsok közti kapcsolatok erősen időfüggőek, ami a hálózati topológia folyamatos változásával jár (pl. egy részvények közti korrelációs hálózat). Természetesen általános esetben mindkét fajta változás, a növekedés és az átrendeződés is előfordulhat egy komplex hálózat időfejlődése során. A konstans méretű, időben átrendeződő hálózatok természetes módon modellezhetők gráfsokaságokkal, melyek elemei az adott méretnél lehetséges gráftopológiák. Egy adott topológia előfordulási valószínűségét általában egy struktúrától függő energia és egy hőmérsékletnek megfeleltethető paraméter határozza meg (ez utóbbi tekinthető úgy is mint egy külső zaj). A hőmérséklet változtatásával bizonyos energiafüggvények esetén fázisátalakulásokhoz hasonló változások figyelhetők meg. Ezen topológiai fázisátalakulásokat mutatja be a dolgozat 1. fejezete.

Szintén fázisátalakulással foglalkozik a 2. fejezet is, azonban ott az említett gráfsokaságokon nyugvó keretrendszert elhagyva a statisztikus fizikán belül jól ismert perkolációs átalakulás egy speciális esetét tanulmányozzuk véletlen gráfokon. A perkolációs klaszterek természetes megfelelői egy hálózat esetén az összefüggő komponensek, melyeken belül bármely csúcsból eljuthatunk bármely másik csúcsba élek sorozatán keresztül. Egy hálózat „perkolál”, ha van benne egy óriás komponens, melynek mérete összevethető a teljes rendszermérettel. A véletlen gráfokkal kapcsolatos egyik első eredmény éppen a perkoláció kritikus pontjának levezetése volt Erdős Pál és Rényi Alfréd által [64]. A 2. fejezetben a véletlen gráf élperkolációs problémáját általánosítjuk a k -klikkperkoláció fogalmának bevezetésével. A k -klikkek k darab csúcsból álló teljes részgráfoknak felelnek meg egy hálózatban (melyeken belül minden csúcs minden másik csúccsal össze van kötve). Egy természetesen adódó k -klikkszomszédsági fogalom bevezetésével bármely hálózat esetén felrajzolhatjuk a k -klikkek gráfját is, melyben a csúcsok az eredeti hálózat k -klikkjeinek felelnek meg, az élek pedig azt jelzik, hogy az adott k -klikkpárok (a 2. fejezetben részletezett definíció alapján) szomszédosak. Az így kapott k -klikkgráf perkolációs átalakulását tanulmányozzuk részletesen a 2. fejezetben az Erdős–Rényi-modell esetén.

A k -klikkperkoláció gyakorlati jelentőségét az adja, hogy a k -klikkperkolációs klaszterek nagyon természetes módon használhatók hálózati csoportkeresésre. A természet és a társadalom különböző alrendszeit leíró hálózatok nagy részénél a csoportok, vagy más néven modulok, klaszterek, csoportosulások egy nagyon érdekes szerveződési szintet képeznek az egyes csúcsok szintje fölött és a teljes hálózat szintje alatt. A hálózati csoportok

szemléltetésére kézenfekvő példát nyújtanak az emberi kapcsolatok hálózataiban fellelhető családok, baráti körök, munkahelyi közösségek, illetve a fehérje-kölcsönhatási hálózatokban jelen lévő fehérjefunkciós csoportok és fehérjekomplexek. A csoportoknak nincs egy széles körben elfogadott egységes definíciója, de általában olyan szerkezeti egységnek feleltetjük meg őket, melyen belül a kapcsolatok nagyon sűrűek, míg a csoportok között jelentősebben ritkábbak. A csoportkereséssel kapcsolatos elméleti és gyakorlati problémák vizsgálata az egyik legintenzívebben kutatott területe a komplex hálózatok témakörének. A dolgozat 3. fejezete mutatja be, hogy miként lehet a 2. fejezetben bevezetett k -klikkperkolációs klasztereket csoportkeresésre használni. Ezen megközelítés egyik nagyon nagy előnye, hogy természetes módon engedi meg a csoportok közti átfedéseket.

Ezt követően a 4. fejezet a k -klikkperkolációs csoportok időfejlődését vizsgálja nagyméretű, emberi kapcsolatokat leíró hálózatokban empirikus adatok alapján. A bemutatásra kerülő elemzés egyfelől egy több tízezer szerzót magába foglaló tudományos társszerzőségi hálózatra, másfelől egy kb. 4 millió anonimizált felhasználóból álló telefonhívási hálózatra terjedt ki. Bár a szomszédos időlépéseknél kinyert csoportok összeillesztése egy nemtriviális feladat, mégis érdemes elvégezni, hiszen ezáltal tárul fel a csoportok életútja, mely nagyon változatos képet mutat mindkét adatrendszer esetén. Egyes csoportok nagyfokú stabilitást mutatnak az időben és szinte teljesen „stacionáriusok”, míg mások dinamikusán változnak, növekednek vagy zsugorodnak, más csoportokkal összeolvadnak, vagy részekre bomlanak. A csoportok autokorrelációján alapuló statisztikai vizsgálat egy érdekes összefüggést mutatott ki a csoportok mérete, változékonysága és várható élettartama között, mely érvényes mindkét hálózatban annak ellenére, hogy a két rendszer teljesen független egymástól.

Végül az 5. fejezetben elkanyarodunk a hálózati csoportok témakörétől, hogy egy multifraktál alapú általános hálózatmodellt mutassunk be, mely ugyanazon keretrendszeren belül teszi lehetővé akár nagyon eltérő tulajdonságokkal rendelkező véletlen gráfok legyártását. Fraktálgeometria nagyon gyakran fordul elő a természetben mind az élővilágban (pl. növényi gyökök vagy lombzat, állati vagy emberi belső szervek, érhálózat stb.), mind a geológiai képződmények esetén (pl. folyók vízgyűjtő rendszere, szigetek partvonala stb.). Fraktálgeometrián játszódó fizikai folyamatok esetén a fizikai mennyiségek térbeli eloszlása sokszor válik multifraktállá. Emellett a multifraktálok fontos szerepet töltenek be a káoszelméletben is. Az 5. fejezet a multifraktálok egy új alkalmazását mutatja be, ahol egy hálózat éleit véletlenszerűen kötjük be, és az élbekötési valószínűségeket egy multifraktál segítségével adjuk meg. A multifraktál megfelelő megválasztásával az így legyártott hálózat tulajdonságai nagyon széles skálán változhatnak, például a fokszámoszlás lehet skálafüggetlen (mint a Barabási–Albert-modell esetén), vagy Poisson-eloszlású (mint az Erdős–Rényi-modellben).

A soron következő fejezetek mindegyike ad egy részletesebb bevezetőt is a fejezetben tárgyalt témakörhöz, majd ezt követi az adott területen elért eredmények ismertetése. A saját eredmények legtömörebb összefoglalását a disszertációhoz kapcsolódó tézisfüzet adja meg. A disszertációban az irodalmi hivatkozások esetén a következő jelöléseket használjuk:

- T előtagot kapnak (pl. [T1]) a tézisfüzetben felsorolt tézispontok alapjául szolgáló saját publikációk. (Az egyszerűség kedvéért ezen publikációk számozása a disszertációban és a tézisfüzetben megegyezik).

- S előtagot kapnak (pl. [S1]) a dolgozathoz kapcsolódó további saját publikációk.
- előtag nélkül szerepelnek (pl. [1]) az általános irodalomjegyzék publikációi.

1. fejezet

Topológiai fázisátalakulások

A valódi rendszereket reprezentáló hálózatok jelentős része időben változó szerkezetet mutat. Egyes hálózatoknál a növekedés a legfontosabb jelenség, azaz a rendszerbe idővel újabb és újabb csúcsok csatlakoznak be új élekkel kapcsolódva hozzá a hálózat már korábban létező részéhez. Más esetekben az élek átrendeződését figyelhetjük meg jelentős méretbeli változás nélkül. Ez utóbbi folyamatot általánosan úgy képzelhetjük el, mintha egy fajta „ágensek” próbálnának egymással kapcsolatokat létesíteni egy zajos környezetben. (Az ágensek megfelelhetnek embereknek, géneknek, vállalatoknak, stb.). Egy új kapcsolat létrejöttének, illetve egy már létező kapcsolat megszűnésének valószínűsége így függ mind a zajtól, mind azon előnytől vagy hátránytól, amit a kapcsolat létrejötte vagy megszűnése okoz. A környezeti feltételek lehetnek olyanok, hogy az ágensek „nagy felfedezőkedvvel” hozzák meg döntéseiket, és szinte válogatás nélkül kötődnek hozzá véletlenszerűen választott másik ágensekhez. De természetesen a környezeti feltételek változásával előfordulhat, hogy az ágensek számára a konzervatív, kockázatkerülő magatartás válik preferálttá, és ennek révén az őket összekötő hálózat topológiája jelentősen átalakulhat. Erre szolgáltat egy érdekes példát a David Stark és Vedres Balázs által közölt publikáció [196], melynek eredményei szerint a bizonytalanság (várható fluktuációk) szintjétől függően egyes gazdasági hálózatok szerkezete egy csillagszerű állapotból átszerveződhet egy lokálisan sűrűbben huzalozott, klikkszerű struktúrába.

A folyamatosan átrendeződő hálózatok leírására, modellezésére kínál egy általános keretrendszert a hálózatok statisztikus sokasága. Ebben a statisztikus fizikai ihletésű megközelítésben az élek átrendeződése az állapottérben való bolyongásnak felel meg, ahol az egyes állapotokhoz bizonyos szabályok szerint jól definiált valószínűségeket rendelünk, és az állapotok közti átmeneti valószínűségek kielégítik a részletes egyensúly feltételét. Mielőtt rátérnénk az általunk javasolt formalizmusra, röviden összefoglaljuk az irodalomban korábban megjelent gráfsokaságok legfontosabb tulajdonságait.

Zdzislaw Burda és munkatársai egy nulladimenziós térelmélet Feynman-diagramjai alapján vezették be a sokaság elemeit, és a különböző gráfok valószínűségét a megfelelő Feynman-diagramok súlyának feleltették meg [40, 41]. Az így kapott sokaságot egy fraktáldimenzió, illetve egy spektrális dimenzió jellemezte, és a vizsgálatok főként a kapott gráfok topológiája és e két paraméter közti összefüggésekre koncentráltak. Az eredmények

1. FEJEZET: Topológiai fázisátalakulások

szerint a kétdimenziós paramétertérben jól elkülönül két fázis, melyek közül az egyikben viszonylag homogén a fokszámoszlás, míg a másikban a hálózat összezuhan néhány nagy-méretű „csillagba”, melyek központjában egy-egy nagyon nagy foksámú csúcs található, melyet túlnyomórészt 1 foksámú csúcsok vesznek körül. A két fázis közti határvonalon skálafüggetlen fokszámoszlású gráfok jelennek meg.

Egy alternatív, a klasszikus hamiltoni rendszerekhez hasonló formalizmust javasolt a gráfsokaság partíciós függvényének bevezetésére Johannes Berg és Michael Lässig [20]. A csúcsok foksámaitól függő Hamilton-függvény mellett bevezettek egy β paramétert is, mely az inverz hőmérsékletnek felelt meg. A különböző gráftopológiák valószínűségét a Hamilton-függvény és ezen paraméter értékéből lehetett meghatározni, a klasszikus statisztikus fizikában megszokott módon. A [20] publikáció eredményei szerint ha a Hamilton-függvény magasabb rendű tagokat is tartalmaz, az a magas β tartományban foksám-korrelációk megjelenéséhez vezethet.

Egy hasonló modell vizsgálatait mutatja be a [11] publikáció, mely esetben a Hamilton-függvény a szomszédos csúcsok foksámjainak arányától függött, és ezáltal az éltrendeződési dinamika a disszortatív¹ és magas klaszterezettségű állapotok felé hajtotta a rendszert. Egy paramétertől függően a hálózat lényegében a következő három fázis valamelyikébe rendeződött be: exponenciális fokszámoszlású hálózat, skálafüggetlen hálózat, illetve a csillagszerű topológiák által dominált fázis. Azonban ebben a modellben az élek átkötésénél nem egyenletes valószínűséggel választunk az élek közül, ami miatt a részletes egyensúly feltétele nem teljesül.

Ebben a fejezetben a topológiai fázisátalakulások szemszögéből vizsgáljuk meg a hálózatok átrendeződését. Ennek megfelelően az általunk bevezetett gráfsokaságok esetén is megjelenik egy hőmérsékletnek megfelelő paraméter, mely az élkötéseket befolyásoló zajszintnek feleltethető meg. Ha nagyon magas a hőmérséklet és ezzel együtt a zajszint, akkor az élek gyakorlatilag teljesen véletlenszerűen kötődnek át egyik csúcsról a másikra, és a Hamilton-függvény konkrét formájától függetlenül a hálózat egy Erdős–Rényi-gráfra hasonlít. Azonban ha csökkentjük a hőmérsékletet (és ezzel együtt az élkötéseknél a zajszintet), akkor a hálózat előbb utóbb olyan topológiák felé kezd el átrendeződni, melyek előnyösek a megadott Hamilton-függvény szempontjából. Ilyenkor figyelhetünk meg topológiai fázisátalakulást, melynek során a hálózat topológiája valamilyen drasztikus átalakuláson megy keresztül.

Fontos leszögezni, hogy az általunk tanulmányozott jelenségek merőben eltérnek azoktól a vizsgálatoktól, melyeknél eredetileg szabályos rácson definiált statisztikus fizikai modelleket ültettek át különböző véletlen gráf topológiákra [186, 110, 82]. Továbbá nem lehet a topológiai fázisátalakulásokat egy kalap alá vonni a növekedő hálózat modellekben tapasztalt fázisátalakulásokkal [23, 201] vagy az ún. kvázi statikus hálózatok átalakulásaival [139]. A topológiai fázisátalakulásokhoz egy jól definiált rendparaméter társítható, és az átalakulás kritikus pontjában a partíciós függvényből származtatott termodinamikai függvények szingulárisává válnak. Az általunk javasolt megközelítés alapjaiban hasonlít a J. Berg és M. Lässig által bevezetett, Hamilton-függvényeken alapuló gráfsokasághoz [20]. Egy lényeges

¹Egy hálózat disszortatív ha a szomszédos csúcsok foksámjai „antikorreálnak”, azaz a nagy foksámú csúcsok előszeretettel kapcsolódnak kis foksámú csúcsokhoz és fordítva.

eltérés viszont az, hogy a mi esetünkben nem csak (az egyes csúcsok közvetlen környezetétől függő) lokális energiafüggvények jöhetnek szóba, hanem az egyes csúcsok helyett az energia függhet bármilyen, a hálózatra jellemző globális tulajdonságtól is. Egy általános Hamilton-függvény (és a hőmérsékletet, vagy zajszintet megadó paraméter) által vezérelt élatrendeződési dinamika a topológiai fázisátalakulások mellett egy természetes keretrendszer adhat hálózattopológia-optimalizálási kérdések tárgyalásához is². Ideális esetben egy megfelelő energiafüggvény választásával pont az adott feladat szempontjából optimális topológia áll elő a rendszer lehűtése során.

A fejezetben bemutatásra kerülő eredmények a *Physical Review E* folyóiratban kerültek publikálásra a disszertáció szerzőjének első szerzőségével [T1]. Ezen felül az eredményekből született egy *Physica A* cikk [T2], és egy összefoglaló jellegű *Lecture Notes in Physics* publikáció is [S1] (mindkettő a disszertáció szerzőjének társszerzőségével). A topológiai fázisátalakulások tanulmányozását Vicsek Tamás javasolta először, és az erre vonatkozó kutatások vezetése, koordinálása is neki köszönhető. A vizsgálatok során a Monte-Carlo-szimulációkat Farkas Illés végezte el, az egzakt leszámolás alapján alapuló megközelítést a disszertáció szerzője dolgozta ki. Az analitikus számításoknál a komponensmérettől függő energiafüggvényekre vonatkozó eredményeket Derényi Imre adta meg, míg az egyrészezcskés energiafüggvények esetét a disszertáció szerzője vizsgálta.

1.1. Hálózatok statisztikus sokasága

A fejezet további részében a rögzített N számú csúcsból és M darab súlyozatlan élből álló irányítatlan hálózatok statisztikus sokaságával foglalkozunk. A csúcsokról feltesszük, hogy egyedi azonosítóval rendelkeznek (mint pl. egy index 1 és N között), és ez által egymástól megkülönböztethetők; ezzel szemben az élek között csak a végpontjaik alapján tudunk különbséget tenni, és nem hordoznak magukon egy további egyedi azonosítót. Az élatrendeződési dinamika során csak egyszerű gráfokat engedünk meg, azaz egy csúcspár között maximum 1 él húzódhat, és egy él két végpontja nem tartozhat ugyanahhoz a csúcshoz. Az így kapott sokaság a csúcsok és élek rögzített száma révén a statisztikus fizikában ismert kanonikus sokaságnak felel meg: ha bevezetünk egy T hőmérsékletnek megfelelő paramétert, a partíciós függvény (a [20] publikáció formalizmusával analóg módon) a

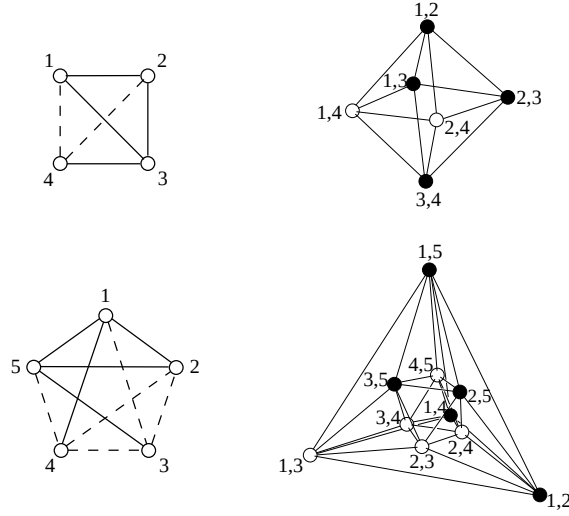
$$Z(T) = \sum_{\{\mathcal{G}_a\}} e^{-E_a/T} \quad (1.1)$$

alakban írható fel, ahol az összes lehetséges N csúcsból és M élből álló $\mathcal{G}_a$ egyszerű gráfra összegzünk, és E_a adja meg az adott gráf energiáját.

A hálózat éleinek átrendeződését a következő módon képzelhetjük el: az átrendeződés elemi lépése lehet például egy véletlenszerűen választott él egyik végének átkötése egy másik csúcshoz, vagy esetleg egy él teljes törlése és ezzel párhuzamosan egy újonnan választott csúcspár összekötése. A lépés elfogadási valószínűségét a kezdeti $\{\mathcal{G}_a\}$ és $\{\mathcal{G}_b\}$ végállapot

²Ilyen jellegű problémákat mutattak be például a [92, 206] publikációk, ahol többek közt egy lokális keresési algoritmus szempontjából keresték a legkedvezőbb hálózatszerkezetet, az élek túlterheltségének esetleges figyelembevételével.

1. FEJEZET: Topológiai fázisátalakulások



1.1. ábra. Az élatrendeződési dinamika és a Kawasaki-féle rácsgáz közötti megfeleltetés. Az ábra bal oldalán két kisméretű hálózat látható, melyekben az élek folytonos vonallal lettek behúzva, a lehetséges további élhelyeket (melyek a jelenlegi állapotban üresek), szaggatott vonalak jelzik. Az ábra jobb oldalán az adott hálózatoknak megfelelő rácsozatokat láthatjuk, melyekben a rácspontok a lehetséges éleket reprezentálják. A fekete rácspontok betöltöttek, azaz a nekik megfelelő csúcspár között van él (a rácspont két indexe mutatja, hogy melyik két csúcstról van szó). Ezzel szemben a fehér rácspontok üresek, ugyanis a nekik megfelelő csúcspárok nincsenek összekötve. Két rácspont akkor szomszédos, ha az egyik indexük közös. Egy él egyik végének átkötése ezáltal egy részecske szomszédos rácspontba való áthelyezésének felel meg a rácson. (Az ábra forrása a [T1] publikáció.)

energiájának különbsége, $\Delta E_{ab} = E_b - E_a$ határozza meg a Metropolis–Hastings-algoritmus szerint [132, 95]: amennyiben az új konfiguráció energiája alacsonyabb, a lépést automatikusan elfogadjuk. Azonban ha az új energia magasabb, az adott élatkötési lépést csak $e^{-\Delta E_{ab}/T}$ valószínűséggel fogadjuk el. Ennek révén $T \rightarrow \infty$ esetén a dinamika a teljesen véletlenszerű élatthuzalozáshoz tart (hiszen minden lépést elfogadjunk, függetlenül az energiakülönbségtől), ezért a hálózat topológiája (kellően hosszú idő után) meg fog egyezni egy Erdős–Rényi-gráffal. Ezzel szemben alacsony hőmérsékleten az alacsonyabb energiájú topológiák előfordulási valószínűsége megnő.

A fenti dinamika teljesíti a részletes egyensúly feltételét és formálisan ekvivalens egy Kawasaki-féle, állandó részecskeszámú rácsgázzal, melyben a részecskék egy speciális topológiájú rácson mozognak [29, 173]. A rácspontok az összes lehetséges $N(N-1)/2$ él pozíciónak felelnek meg a hálózatban, a részecskék pedig az átrendeződő éleket reprezentálják: egy él egyik végének átkötése az adott részecske elmozdítását jelenti egy szomszédos rácsponton, ahogy azt az 1.1. ábra szemlélteti.

Az (1.1) egyenletben megadott partíciós függvény nagyon sok olyan tagot tartalmaz,

melyek azonos topológiájú, egymással izomorf³ gráfok járulékának felelnek meg. Mivel elsősorban topológiai átalakulásokra vagyunk kíváncsiak, a továbbiakban feltesszük, hogy az energiafüggvény csak a hálózat topológiájától függ, a csúcindexek konkrét leosztásától nem. Ha egy adott gráftopológiát $\mathcal{T}_\alpha$ -val jelölünk, az (1.1) kifejezést átírhatjuk a topológiákra való összegzéssé a

$$Z(T) = \sum_{\{\mathcal{T}_\alpha\}} \mathcal{N}_\alpha e^{-E_\alpha/T} \quad (1.2)$$

formában, ahol $\mathcal{N}_\alpha$ a $\mathcal{T}_\alpha$ topológiával rendelkező gráfok száma, míg E_α az ehhez tartozó energiát adja meg. Ennek a kifejezésnek egy másik ekvivalens formája a

$$Z(T) = \sum_{\{\mathcal{T}_\alpha\}} e^{-E_\alpha/T + \ln \mathcal{N}_\alpha} = \sum_{\{\mathcal{T}_\alpha\}} e^{-F_\alpha/T} \quad (1.3)$$

formula, melynek alapján bevezethetjük a $\mathcal{T}_\alpha$ topológiához tartozó F_α szabadenergiát,

$$F_\alpha = E_\alpha - T \ln \mathcal{N}_\alpha = E_\alpha - TS_\alpha, \quad (1.4)$$

illetve a $\mathcal{T}_\alpha$ topológia entrópiáját,

$$S_\alpha = \ln \mathcal{N}_\alpha. \quad (1.5)$$

Amint korábban említettük, egy topológiai fázisátalakulás kritikus pontjánál a partíciós függvényből származtatott termodinamikai függvények szingulárisá válhatnak. Egy ilyen fázisátalakulás nyomon követéséhez célszerű bevezetni egy rendparamétert, melynek értékéből megállapítható, hogy a hálózat éppen melyik fázisban tartózkodik. Amikor például az Erdős–Rényi-gráf perkolációs átalakulásához hasonló fázisátalakulást fogjuk vizsgálni, egy természetes választás a $\varphi = \varphi_s = M_1/M$ rendparaméter, mely a hálózat legnagyobb összefüggő komponensében található élek arányát adja meg a hálózat teljes élszámához viszonyítva. Amikor azonban egy olyan topológiai fázisátalakulással van dolgunk, melynél az élek egy hatalmas „csillagot” alakítanak ki azzal, hogy egyik végükkel mind ugyanahhoz a csúcshoz kapcsolódnak hozzá, célszerűbb a $\varphi = \varphi_d = d_{\max}/M$ rendparaméterre váltani, mely a hálózatban előforduló legnagyobb $d_{\max}$ fokszám és az élek számának hányadosa.

Az adott φ rendparaméterhez egy $F(\varphi, T)$ feltételes szabadenergiát is társíthatunk az

$$e^{-F(\varphi, T)/T} = Z(\varphi, T) = \sum_{\{\mathcal{G}_\alpha\}_\varphi} e^{-E_\alpha/T}, \quad (1.6)$$

összefüggés alapján, ahol $\{\mathcal{G}_\alpha\}_\varphi$ azon gráfok részhalmaza, melyeknél a rendparaméter értéke φ . Egy fázisátalakulás során a rendparaméter értéke gyorsan változik, például egy kezdeti $\varphi = 0$ értékről a kritikus pontban meredeken elkezd a nagyobb értékek felé növekedni. Ilyenkor általában a $F(\varphi, T)$ feltételes szabadenergia minimumhelye is megváltozik. A minimumhely fokozatos eltolódása általában folytonos fázisátalakulásra, vagy cross-over jelenségre utal, míg a minimumhely ugrásszerű megváltozása, vagy több lokális minimum megjelenése elsőrendű fázisátalakulásra utal.

³Két gráf izomorf, ha a csúcsok indexeinek megfelelő permutációjával az egyik gráf átvihető a másikba.

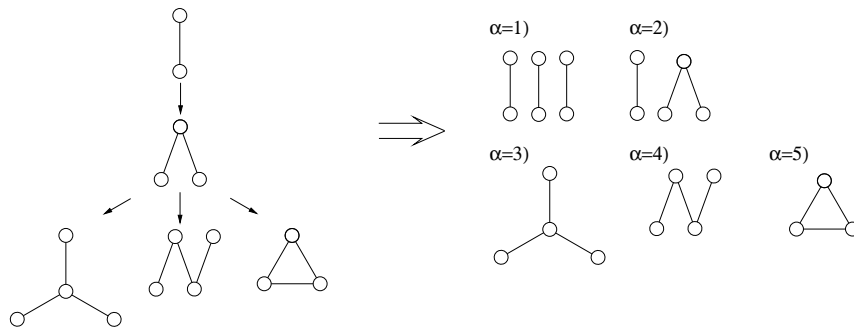
1.2. Numerikus módszerek

Ebben az alfejezetben a topológiai fázisátalakulások vizsgálatánál használt numerikus módszereket ismertetjük. Az analitikus számításokat ezek után a konkrétan vizsgált Hamilton-függvényekkel együtt mutatjuk be az 1.3. alfejezetben. Numerikus szempontból két, gyökeresen eltérő eljárást alkalmaztunk. Az egyik az egzakt leszámlálás módszere volt, melynek előnye, hogy teljesen egzakt végeredményeket ad, viszont csak kis méretű hálózatokra lehet alkalmazni. A másik a Monte–Carlo-módszeren alapuló megközelítés volt, mely ugyan csak közelítő eredményeket szolgáltat, viszont nagy rendszerek vizsgálatát is lehetővé teszi.

1.2.1. Egzakt leszámlálás

Az egzakt leszámlálás módszere egy sikeres megközelítésnek számít véletlen bolyongási problémák, perkoláció, illetve polimerek tanulmányozásában [19, 100, 18, 55]. Ez által motiválva kis hálózatok esetén mi is kiszámoltuk az egzakt partíciós függvényt úgy, hogy előbb meghatároztuk az összes lehetséges állapot valószínűségét egyesével. (A további alfejezetekben ismertetett eredményekhez kapcsolódó egzakt leszámláláson alapuló vizsgálatokat a disszertáció szerzője végezte el).

Eljárásunk lényege, hogy egy adott M számú élt tartalmazó sokaság esetén először rekurzívan előállítjuk az összes összefüggő, maximum M élt tartalmazó topológiát. Ilyenkor az $m + 1$ élből álló, összefüggő gráfokat úgy kapjuk, hogy az összes m élből álló összefüggő gráfhoz az összes lehetséges módon hozzákötünk 1 új élt. Ezek után az összesen M élt tartalmazó topológiák előállíthatók a kisebb, $m \leq M$ élből álló összefüggő gráfok „kombinációjaként”, ahogy azt az 1.2. ábra szemlélteti.



1.2. ábra. Az összesen M élt tartalmazó topológiák előállítása az egzakt leszámlálás módszerénél. Az eljárást itt $M = 3$ esetén szemléltetjük. Először rekurzívan megkonstruáljuk az összes lehetséges összefüggő, $M \leq 3$ élt tartalmazó gráfot, ahogy az ábra bal oldalán láthatjuk. Majd ezek összes lehetséges kombinációjából állnak elő az $M = 3$ élt tartalmazó topológiák, amint azt az ábra jobb oldala mutatja. (Az ábra forrása a [T1] publikáció).

Természetesen az összefüggő, $m + 1$ élből álló gráfok előállítása során egy nagyon fontos lépés, hogy egy potenciális új gráfot gondosan összehasonlítsunk a már korábban megkapott $m + 1$ élből álló gráfokkal, hiszen egy adott $m + 1$ élből álló topológia több úton is előállhat m élből álló gráfokból egy új él bekötésével. Ilyenkor két gráf összehasonlításánál

először a foksámeloszlásokat vetjük össze, hiszen ha ezek eltérnek, a két gráf nem lehet izomorf. Ha a foksámeloszlások megegyeznek, először permutáljuk a csúcs indexeket az egyik gráfban úgy, hogy a két gráfban azonos foksámú csúcsok azonos index részhalmazzal rendelkezzenek. Ezek után végig megyünk az összes csúcson és megvizsgáljuk, hogy az adott csúcs szomszédainak index halmaza egyezik-e a két gráfban. Ha nem, megpróbáljuk úgy permutálni a szomszédok indexeit (természetesen csak azonos foksámú csúcsok indexei között cserélgetve), hogy a cserék után megegyezzen a két halmaz. Ha sikerül ilyen módon minden szomszédsági viszonyt azonossá tenni a két gráfban, akkor a két gráf izomorf és nem egy új gráffal van dolgunk. Ha a csúcs indexek permutációival nem lehetséges a szomszédsági viszonyokat azonossá tenni, az adott két gráf topológiája eltérő.

Egy további fontos lépés egy adott $\mathcal{T}_\alpha$ topológiához tartozó $\mathcal{N}_\alpha$ meghatározása. Amennyiben a topológiának nincs szimmetriája, $N_\alpha = N! / (N - N_{t_\alpha})!$, ahol N_{t_α} a legalább egy éllel rendelkező csúcsok száma (ezek „vesznek részt a topológiában”), $N - N_{t_\alpha}$ pedig az izolált csúcsok száma. Az általános esetben a fenti N_α -t még el kell osztani azon csúcsindex permutációk számával, melyek nem változtatják meg a topológiát. Például ha a topológia n azonos részgráfból áll (mint az 1.2. ábrán az $\alpha = 1$ -es topológia, mely $n = 3$ azonos részgráfból tevődik össze), akkor $\mathcal{N}_\alpha$ -t el kell osztani $n!$ -sal. Továbbá ha a topológia egyik izolált részgráfján belül az indexek l permutációja önmagába viszi át a részgráfot, az is behoz egy $1/l$ -es faktort. Az 1.2. ábrán mutatott példáknál az $\alpha = 1$ és $\alpha = 2$ topológiáknál minden részgráfnál $l = 2$, míg az $\alpha = 3$ és $\alpha = 5$ esetekben $l = 3!$, végül az $\alpha = 4$ -es topológiánál megint $l = 2$. Ez alapján az 1.2. ábrán mutatott topológiákhoz az

$$\mathcal{N}_1 = \frac{N!}{(N-6)! 2^3 3!}, \quad \mathcal{N}_2 = \frac{N!}{(N-5)! 2^2}, \quad (1.7)$$

$$\mathcal{N}_3 = \frac{N!}{(N-4)! 3!}, \quad \mathcal{N}_4 = \frac{N!}{(N-4)! 2}, \quad \mathcal{N}_5 = \frac{N!}{(N-3)! 3!} \quad (1.8)$$

értékek tartoznak.

Ha meghatároztuk $\mathcal{N}_\alpha$ -t, ennek alapján az adott topológia valószínűségét a

$$p_\alpha = \frac{\mathcal{N}_\alpha e^{-E_\alpha/T}}{Z} \quad (1.9)$$

összefüggésből kapjuk. Ezek után bármely $\mathcal{Q}$ termodinamikai mennyiség várható értékét meg tudjuk határozni a

$$\langle \mathcal{Q} \rangle = \sum_{\{\mathcal{T}_\alpha\}} \mathcal{Q}(\mathcal{T}_\alpha) p_\alpha \quad (1.10)$$

formula segítségével. Az egzakt leszámítás nagyon nagy előnye az, hogy a topológiákhoz tartozó $\mathcal{N}_\alpha$ értékeket elég egyszer meghatározni, utána a fenti összefüggések alapján tetszőleges energiafüggvény választás mellett viszonylag gyorsan meg tudjuk határozni a vizsgált termodinamikai függvények viselkedését. Ezzel szemben viszont a Monte-Carlo-módszernél a szimulációkat minden egyes energiafüggvény esetén előlről kell kezdeni. Az egzakt leszámítás módszerének hátránya viszont az, hogy a lehetséges topológiák száma M -el exponenciálisan nő, ezért ez a megközelítés csak kis M -ek esetén alkalmazható.

1.2.2. Monte–Carlo-módszer

A Monte–Carlo-módszer egy széleskörűen használt megközelítés olyan problémák vizsgálatánál, melyeknél nehéz (vagy gyakorlatilag lehetetlen) a keresett eloszlásokat analitikus formában megadni. Lényege a nagyon sokszor ismételt véletlen mintavételezés, melynek segítségével numerikusan egy közelítő eredményt kapunk a vizsgált eloszlásról. (A további alfejezetekben ismertetett eredményekhez kapcsolódó Monte–Carlo-szimulációkat Farkas Illés végezte el).

A mi esetünkben az 1.1. alfejezetben említett, a Metropolis–Hastings-algoritmust követő élatrendeződés lett a szimulációk alapja. (Mint említettük, egy véletlenszerűen választott él egyik végének véletlenszerű átkötésénél az új és a kiinduló állapot közti energiakülönbség, illetve a hőmérséklet szabja meg az átkötés elfogadási valószínűségét). Azonban a legegyszerűbb formájában ez az algoritmus nagyon lassúvá válhat, ha az energiaminimumnak megfelelő topológia nehezen elérhető egyszerű véletlenszerű átkötögetések révén. Egy jó példa erre egy olyan energiafüggvény, melynek minimuma egy csillagszerű állapot, ahol minden él egyik végével egyazon csúcshoz kapcsolódik. Ha N nagy, akkor nagyon sokáig tart, mire minden él „megtalálja” a csillag központját, pusztán véletlenszerű átkötögetéssel. Az 1.1. ábrán bemutatott rácsgáz analógiából is látszik, hogy a rendszer bizonyos esetekben lassan relaxál: a rács speciális szerkezete miatt a részecskék közötti kölcsönhatások nagyon sűrűek, hiszen egy rácspontnak $2(N-2)$ első szomszédja van, ezen felül $N^2/2 - 5N/2 + 3$ másod szomszédal rendelkezik, és két lépésnél egy rácspont sincs messzebb.

A gyorsabb relaxáció érdekében számos esetben a párhuzamos lehűtésen alapuló technikát [102] alkalmaztuk a későbbi alfejezetekben bemutatásra kerülő vizsgálatok során. Ez tekinthető a szimulált hűtéses megközelítés továbbfejlesztett verziójának, melynél a vizsgált rendszer sok másolatát szimuláljuk párhuzamosan, eltérő hőfürdőkben, adott hőmérsékletű hőfürdőként is több példányban. Egy melegebb hőfürdőben szimulált másolat a fázistér nagyobb részét képes bejárni, viszont nem térképezi fel részletesen a meglátogatott részt. Ezzel szemben egy hidegebb hőfürdőben szimulált másolat csak a fázistér kis részén mozog, azt viszont részletesen, nagy mélységig feltérképezi. Két, a hőmérséklet szempontjából szomszédos hőfürdő között időről időre megpróbálunk egy-egy cserét végrehajtani. Ilyenkor mindkét hőfürdőből véletlenszerűen kiválasztunk egy-egy másolatot, majd megállapítjuk a csere elfogadási valószínűségét a Metropolis–Hastings-szabály szerint, végül az adott valószínűséggel a cserét végre is hajtjuk. (Ha a magasabb hőmérsékletű hőfürdőből vett másolat energiája alacsonyabb, a csere valószínűsége 1, ha nem, akkor a ΔE energia különbség és ΔT hőmérséklet különbség alapján a csere valószínűsége $e^{-\Delta E/\Delta T}$.) Ennek révén egy adott hőfürdőben az alacsonyabb energiájú másolatok nagyobb valószínűséggel kerülnek át a szomszédos alacsonyabb hőmérsékletű hőfürdőbe, mint a magasabb energiájú másolatok.

1.3. Egyrészecskés Hamilton-függvények

Egy gráfsokaság esetén a Hamilton-függvényt nagyon sokféle módon lehet megválasztani. Mint korábban említettük, az általunk feltételezett, Metropolis–Hastings-algoritmust kö-

vető dinamika $T \rightarrow \infty$ esetén nem igazán vesz tudomást az energiáról és a nagy entrópiájú topológiákat favorizálja. Ahhoz, hogy topológiai fázisátalakulást tapasztalhassunk a hőmérséklet csökkentésével, ennek alapján olyan energiafüggvényt kell választanunk, melynek minimuma alacsony entrópiájú állapotnak felel meg. Azonban még ilyen típusú energiafüggvényből is nagyon sokfélét lehet konstruálni. Egy természetes ötlet ún. egyrészcskés energiát feltételezni, mely az egyes csúcsok fokszámait veszi figyelembe és az

$$E = \sum_{i=1}^N f(d_i) \quad (1.11)$$

alakban áll elő, ahol d_i jelöli az i csúcs fokszámát. Egy másik lehetőség olyan energiafüggvények vizsgálata, melyek a hálózat szerkezetét egy magasabb szinten veszik figyelembe, és például csak a komponensek méretétől függenek. Ez utóbbi, a komponensektől függő energiafüggvényekhez kapcsolódó analitikus számításokat Derényi Imre végezte el, a Monte-Carlo-szimulációkat pedig Farkas Illés implementálta, ezért az ezekkel kapcsolatos eredményeket az A függelékben foglaljuk össze. A fejezet további részében az (1.11) kifejezéssel leírható energiafüggvényekre szorítkozunk. (Az egyrészcskés energiafüggvényekkel kapcsolatos analitikus számításokat és egzakt leszámlálást a disszertáció szerzője végezte el, a Monte-Carlo-szimulációkat pedig Farkas Illés).

Az élatrendeződési dinamika során az egyik alapfeltevésünk az volt, hogy az élek száma időben állandó. Ennek az (1.11) kifejezésre vonatkozó egyik egyszerű következménye az, hogy amennyiben $f(d_i)$ a fokszámnak lineáris függvénye, úgy az összenergia a hálózat struktúrájától függetlenül állandóvá válik (hiszen a csúcsok összesített fokszáma az élek számának kétszeresével egyenlő). Annak érdekében, hogy a hőmérséklet csökkentésével az Erdős-Rényi-modellben ismert perkolációs átalakuláshoz hasonló fázisátalakulásokat tapasztaljunk, melyek során az élek kondenzálódnak egy óriás komponensbe vagy egyéb, még speciálisabb struktúrákba, az $f(d_i)$ függvénynek a lineárisnál gyorsabban kell csökkennie a fokszám függvényében.

Mivel az energiafüggvény minden egyes csúcs fokszámát figyelembe veszi, az ezzel járó élatrendeződési dinamikára tekinthetünk úgy is, mint egy versengésre a csúcsok között, ahol egy csúcs fitnessét a kapcsolatai határozzák meg. Ezt a gondolatmenetet követve bevezethetünk egy alternatív energiafüggvényt is, melynél ugyan szintén a csúcsokra összegzünk, viszont a járuléka a csúcs szomszédaitól jön az

$$E = \sum_{i=1}^N \sum_{i'} g(d_{i'}) \quad (1.12)$$

formában, ahol i' az i csúcshoz kapcsolódó csúcsokat jelöli. (Ezt az alternatív felírásmódot Derényi Imre javasolta először). Ilyenkor tehát az i csúcs egy $g(d_{i'})$ járulékot gyűjt be minden egyes szomszédjától. Cserébe ezek a szomszédok, amikor rájuk kerül sor, mind egy $g(d_i)$ járulékot fognak begyűjteni az i csúcstól. Ennek révén az i teljes hozzájárulása az energiához $d_i g(d_i)$ alakban adható meg (hiszen d_i szomszédja van). Ez alapján ha

$$f(d_i) = d_i g(d_i), \quad (1.13)$$

1. FEJEZET: Topológiai fázisátalakulások

akkor az (1.11) formában megadott és az (1.12) alternatív alakban felírt energiafüggvények megegyeznek egymással.

1.3.1. Egy Ising-modellel ekvivalens energiafüggvény

Egy természetesen adódó választás az energiafüggvény alakjára az, amikor minden élpárhoz, melyek egyik végükkel közös csúcshoz kötődnek, egy $-\mathcal{J}$ járulékot rendelünk. Az energiafüggvény ilyenkor az

$$E = -\frac{\mathcal{J}}{2} \sum_{i=1}^N d_i(d_i - 1) = -\frac{\mathcal{J}}{2} \sum_{i=1}^N d_i^2 + \frac{\mathcal{J}M}{2} \quad (1.14)$$

alakban áll elő, azaz $f(d_i) = -(\mathcal{J}/2)d_i^2$, illetve $g(d_i) = -(\mathcal{J}/2)d_i$. Mivel az élek száma időben állandó, a második, konstans tagot el lehet hagyni az (1.14) kifejezésben, hiszen semmilyen hatása nincs az élek átrendeződési dinamikájára. Talán még természetesebb ezt az energiafüggvényt az 1.1. ábrán bemutatott rácsgázképben felírni, hiszen a hálózatban egy azonos csúcshoz kapcsolódó élpár két szomszédos betöltött rácspontnak felel meg a rácsgázban. Ez alapján az (1.14) energiafüggvény a rácsgázképben az

$$E = -\mathcal{J} \sum_{\langle \alpha, \beta \rangle} b_\alpha b_\beta \quad (1.15)$$

alakot veszi fel, ahol az összegzés a szomszédos rácspontokra történik és $b_\alpha = 1$, ha az adott rácspont be van töltve, illetve $b_\alpha = 0$ egyébként. A fenti energiafüggvény megegyezik a gázokban fellépő nukleációs folyamatok tanulmányozásánál használt energiafüggvénnyel a szabályos rácson definiált rácsgázok esetén. Két szomszédos részecske között ilyenkor egy $-\mathcal{J}$ kötési energia lép fel, mely az általunk tanulmányozott hálózati életrendeződési dinamika esetén két „szomszédos” él kötési energiájának felel meg, ahol az élszomszédság alatt egy közös végpontot értünk. Természetesen az energiát mérhetjük a $\mathcal{J}$ egységeiben is, ezért az általánosság megszorítása nélkül a továbbiakban feltesszük, hogy $\mathcal{J} = 1$, és a formulákból elhagyjuk $\mathcal{J}$ -t.

A rácsgázképben felírt energiafüggvényt át lehet alakítani úgy, hogy az eredmény az Ising-modellel legyen ekvivalens, amennyiben a b_α változóról áttérünk a $z_\alpha = 2b_\alpha - 1$ spin jellegű változóra. Ennek segítségével az energiafüggvény az

$$\begin{aligned} E &= - \sum_{\langle \alpha, \beta \rangle} b_\alpha b_\beta = - \sum_{\langle \alpha, \beta \rangle} \frac{1+z_\alpha}{2} \frac{1+z_\beta}{2} = \\ &= -\frac{1}{4} \sum_{\langle \alpha, \beta \rangle} z_\alpha z_\beta - \frac{1}{2} \sum_{\alpha=1}^{N(N-1)/2} z_\alpha - \frac{1}{8} N(N-1)(N-2) \end{aligned} \quad (1.16)$$

alakban írható fel, ahol kihasználtuk, hogy a rácsgázképben a rácspontok száma $N(N-1)/2$, ami azzal jár, hogy összesen $N(N-1)(N-2)/2$ szomszédos rácspontpár fordul elő a rendszerben. Az így nyert kifejezés megfelel egy Ising-modellben felírt Hamilton-függvénynek külső mágneses tér jelenlétében.

A következőekben rátérünk a feltételes szabadenergia meghatározására. Az energiaminimumhoz tartozó topológia egy „csillagnak” felel meg ahol minden él egy közös csúcs-hoz csatlakozik az egyik végével. Ennek megfelelően a magas hőmérsékleten tapasztalható rendezetlen állapot és az alacsony hőmérsékleten bekövetkező rendeződés közti átalakulás nyomon követésére a $\varphi_d = d_{\max}/M$ rendparaméter tűnik a legalkalmasabbnak, mely a hálózatban előforduló legmagasabb foksám és a teljes élszám hányadosa.

Az ehhez tartozó feltételes szabadenergiát az alábbi módon lehet megbecsülni. Az egyszerűség kedvéért tegyük fel, hogy van már a rendszerben egy csúcs mely az élek több, mint a felét begyűjtötte, azaz $\varphi_d > 1/2$. A termodinamikai határesetben az ehhez tartozó járulék mellett a többi tagot el lehet hanyagolni az energiafüggvényben. Az entrópiával kapcsolatos tag megbecsüléséhez a következő faktorokat kell figyelembe venni: a „csillag” középpontját összesen N csúcs közül választhatjuk ki. Ha ezt fixáltuk, akkor a hozzá kapcsolódó $d_{\max}$ darab él másik végét a fennmaradó $N - 1$ csúcs között kell kiosztani, ami egy $\binom{N-1}{d_{\max}}$ faktort eredményez. A maradék éleket bárhogyan kioszthatjuk, mindössze annyi a megszorítás, hogy a „csillag” középpontjához nem köthetjük őket, aminek révén kapunk még egy $\binom{(N-1)(N-2)/2}{M-d_{\max}}$ faktort. Ezeket összeszorozva megkapjuk a $d_{\max}$ foksámú „csillagot” tartalmazó gráfok számát a sokaságban,

$$\mathcal{N}_{\text{csill}} \approx N \binom{N-1}{d_{\max}} \binom{(N-1)(N-2)/2}{M-d_{\max}} \approx N \binom{N}{d_{\max}} \binom{N^2/2}{M-d_{\max}}. \quad (1.17)$$

Ez alapján, a Stirling-formula felhasználásával valamint az $\mathcal{O}(\ln N)$ nagyságrendű tagok elhanyagolásával a feltételes szabadenergiában szerepet játszó $\ln \mathcal{N}_{\text{csill}}$ tényező az

$$\begin{aligned} \ln \mathcal{N}_{\text{csill}} \approx M \left[\frac{N}{M} \ln \frac{N}{M} + \frac{N^2}{2M} \ln \frac{N^2}{2M} - \left(\frac{N}{M} - \varphi_d \right) \ln \left(\frac{N}{M} - \varphi_d \right) \right. \\ \left. - \varphi_d \ln \varphi_d - (1 - \varphi_d) \ln(1 - \varphi_d) \right. \\ \left. - \left(\frac{N^2}{2M} - 1 + \varphi_d \right) \ln \left(\frac{N^2}{2M} - 1 + \varphi_d \right) \right] \end{aligned} \quad (1.18)$$

alakra hozható. Látható, hogy a termodinamikai határesetben a vezető rendben

$$\ln \mathcal{N}_{\text{csill}} \approx -\varphi_d M \ln N, \quad (1.19)$$

ahol a φ_d -től független tagokat elhagytuk. Ez alapján a feltételes szabadenergia az

$$F(\varphi_d, T) \approx f(\varphi_d M) + \varphi_d M T \ln N \quad (1.20)$$

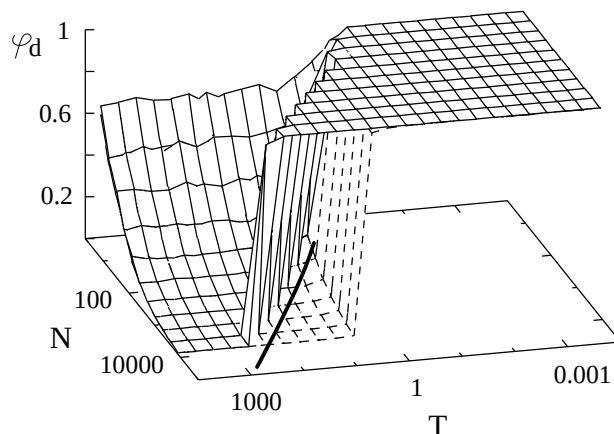
alakban írható fel. Ez a kifejezés a termodinamikai határesetben jó közelítést ad minden olyan esetben, amikor a hálózat tartalmaz egy legalább $D > M/2$ foksámú „csillagot” és az energiafüggvény alakja olyan, hogy ennek járuléka mellett a többi csúcs járuléka elhanyagolható. Konkrétan az (1.14) alakú energiafüggvény esetén az

$$F(\varphi_d, T) \approx M [-\varphi_d^2 M + \varphi_d T \ln N] \quad (1.21)$$

kifejezésre jutunk.

Az (1.21) formula által megadott parabola maximuma $\varphi_d = T/M \ln N$ -nél van. Ha $T \rightarrow 0$, akkor ez a maximumhely a nullához tart, ami azt jelenti, hogy $F(\varphi_d, T)$ egy csökkenő parabolát képez φ_d függvényében a $[0,1]$ intervallumban, azaz a szabadenergia minimuma a $\varphi_d = 1$ -nél lesz, ami a minden élt magába gyűjtő „csillag” topológiának felel meg. Ezzel szemben ha a hőmérséklet a $T_1 = M/\ln N$ spinodális pont fölé emelkedik, akkor a parabola maximuma kimozdul a $[0,1]$ intervallumon kívülre, és a feltételes szabadenergiá már emelkedő tendenciát mutat a $[0,1]$ intervallumon. Ennek megfelelően $F(\varphi_d, T)$ minimumhelye egy alacsony φ_d -nél lesz, ami egy rendezetlen állapotnak felel meg. A minimum pontos helyét nem tudjuk megbecsülni az (1.21) alapján, hiszen az alkalmazott közelítés egyik feltétele, hogy a $\varphi_d > 1/2$ tartományban vizsgálódjunk. Mindenesetre köztes hőmérsékleteken a parabola maximuma a $[0,1]$ intervallumba esik, két jelentősen eltérő topológiát választva el egymástól: az egyik a nagyon magas fokszerű „csillag”, a másik a rendezetlen állapot, ahol minden csúcsnak kicsi a fokszáma. Ilyenkor a két extrém állapot közül az egyik stabil, a másik metastabil. A két állapot stabilitását az (1.21) egyenlet alapján olyan hőmérsékleteken lehet vizsgálni, melyeknél a parabola maximuma a $[1/2,1]$ intervallumba esik.

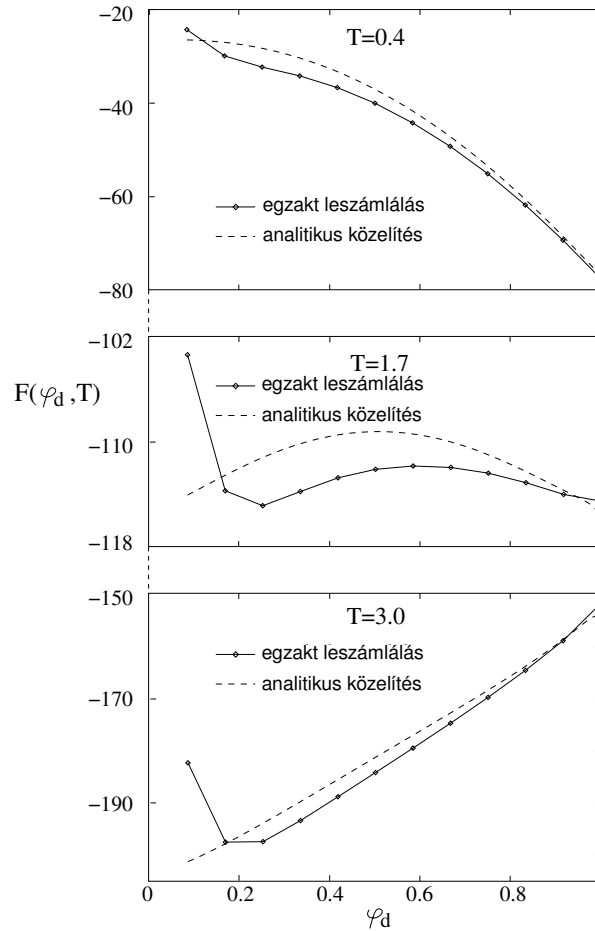
A fentiek alapján a „csillagszerű” állapot kialakulása a rendezetlen állapotból hűtés során egy elsőrendű fázisátalakulásnak felel meg, hiszen egy bizonyos hőmérséklet-tartományon a feltételes szabadenergiának egy stabil és egy metastabil minimuma is van. Ezt maximálisan alátámasztják a numerikus szimulációk eredményei is. A Monte–Carlo-



1.3. ábra. A Monte–Carlo-szimulációk eredményei a $\varphi_d = d_{\max}/M$ rendparaméterre különböző méretű, $\langle d \rangle = 0.5$ átlagos fokszerű gráfsokaságokban az $E = -\sum_i d_i^2/2$ energiafüggvény mellett. A folytonos vonalakkal jelölt eredmények esetén a szimulációk a „csillag” állapotból indultak, míg a szaggatott vonalakkal jelölt esetekben egy Erdős–Rényi-gráfból. Az adatpontok egyszeri futtatások eredményeiből származnak, a rendparaméter értéke minden esetben a $t = 100N$ és $t = 200N$ közti Monte–Carlo-lépések átlagának felel meg. A vastag folytonos vonal az analitikusan meghatározott $T_1 = M/\ln N$ spinodát mutatja. (Az ábra forrása a [T1] publikáció.)

szimulációk eredményeit az 1.3. ábra mutatja be, melynél változó méretű hálózatok esetén vizsgáltuk meg a φ_d rendparaméter hőmérsékletfüggését. Látható, hogy nagyobb N -ek esetén egy jelentős hiszterézis alakul ki a „csillagszerű” állapotból indított, valamint az Erdős–Rényi-gráfból indított szimulációk között.

Kisméretű rendszereken megvizsgáltuk a szabadenergia viselkedését az egzakt leszámolás módszerével is. Az 1.4. ábrán látható, hogy alacsony hőmérsékleten $F(\varphi_d, T)$ a φ_d



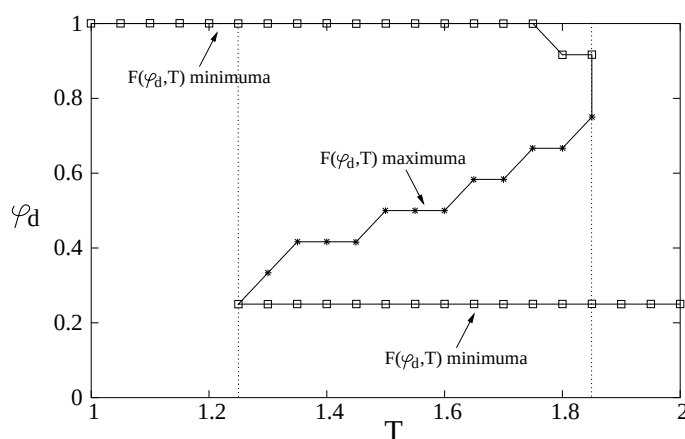
1.4. ábra. Az $F(\varphi_d, T)$ feltételes szabadenergia a φ_d függvényében három különböző hőmérsékleten az $f(d_i) = -d_i^2$ energiafüggvény esetén. A folytonos vonallal összekötött pontok az egzakt leszámolás eredményeit mutatják $N = 48$ és $M = 12$ esetén. A szaggatott vonal az (1.21) alapján kapott analitikus közelítést mutatja. (Az ábra forrása a [T1] publikáció.)

rendparaméter függvényében csökkenő tendenciát mutat az egész $[0, 1]$ intervallumon, ezért a minimum a $\varphi_d = 1$ -nél van, ami a maximális méretű „csillag” topológiának felel meg. Ezzel szemben magas hőmérsékleten $F(\varphi_d, T)$ minimuma egy viszonylag alacsony φ_d értéknél található, ami egy rendezetlen állapotnak felel meg. A köztes hőmérsékleteken viszont van egy tartomány, ahol $F(\varphi_d, T)$ -nek maximuma van φ_d szerint a $[0, 1]$ intervallum középső

1. FEJEZET: Topológiai fázisátalakulások

részen, ami azzal jár, hogy két versengő minimumhely jelenik meg egyszerre, egyik a $\varphi_d = 1$ -nél, a másik alacsony φ_d értéknél. Ez a viselkedés az elsőrendű fázisátalakulások jellemzője és egybevág a Monte-Carlo-szimulációknál látott hiszterézissel. Az 1.4. ábrán az is szembeötlő, hogy az (1.21) egyenleten alapuló analitikus közelítés nagyon jól reprodukálja az egzakt eredményeket már ilyen kis rendszerméreteknél is.

Az egzakt leszámolás módszerével ábrázoltuk a feltételes szabadenergia-függvény spinodális görbét is az 1.5. ábrán, mely az $F(\varphi_d, T)$ minimumhelyéhez tartozó φ_d -t mutatja a T függvényében. Az 1.4. ábrához hasonlóan itt is látható, hogy alacsony illetve magas hőmérsékleteken a feltételes szabadenergiának csak egy minimumhelye van, előbbi esetben a „csillag” topológiának megfelelő $\varphi_d = 1$ értéknél, utóbbi esetben pedig egy rendezetlen állapotnak megfelelő alacsony φ_d értéknél. Ezek mellett van egy kiterjedt köztes hőmérséklet-tartomány is, ahol mindkét minimumhely megtalálható. Ezen a tartományon feltüntetjük az $F(\varphi_d, T)$ maximumhelyét is, mely természetesen a két versengő minimumhely között helyezkedik el.



1.5. ábra. Az $F(\varphi_d, T)$ feltételes szabadenergia spinodális görbéje az $E = -\sum_i d_i^2$ energiafüggvény esetén, $N = 48$, $M = 12$ rendszerméret mellett az egzakt leszámolásból kapott eredmények alapján. A négyzetek az $F(\varphi_d, T)$ minimumhelyét mutatják, a csillagok pedig a maximumhelynek felelnek meg. (Az ábra forrása a [T1] publikáció.)

1.3.2. Folytonos fázisátalakulás és skálafüggetlen gráfok

Egy további energiafüggvény, mely természetesen adódik az $f(d_i) = -d_i \ln d_i$, illetve ekvivalens módon a $g(d_i) = -\ln d_i$ függvények által definiált eset. Régóta ismert, hogy a biológiai érzékelésnél az élőlények a külön fajta ingerek (pl. hang, fény, kémiai koncentráció, stb.) intenzitás változásaira egy „logaritmikus skálán” érzékenyek. Például ha vesszük a leghalkabb hangot, amit még képesek vagyunk meghallani, illetve a leghangosabb hangot, amit még képesek vagyunk elviselni, akkor ezek intenzitása között sok nagyságrend eltérés van, és mi ezen a skálán teszünk különbséget halk és hangosabb hangok között. Ennek révén az általunk érzékelt csekély mértékű hangerő változás is általában többszörös tényleges

intenzitásnövekedéssel jár. Az említett $g(d_i) = -\ln d_i$ energiafüggvény ezt a viselkedést modellezi a „fokszámérzékelés” esetén: feltesszük, hogy a szomszédok fokszámai közt lévő különbségekre a csúcsok logaritmikus skálán érzékenyek.

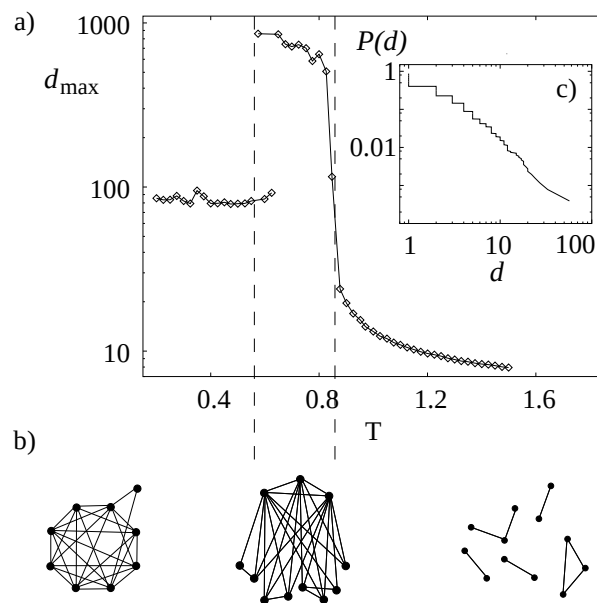
Ilyen esetben az energiaminimumhoz tartozó állapot egy klikknek felel meg, melyben minden csúcs minden másik csúccsal össze van kapcsolva. Természetesen általános M esetén nem feltétlen lehet a gyakorlatban megvalósítani, hogy az élek egyetlen hatalmas klikkbe rendeződjenek be, ilyenkor az alapállapot topológia egy, az adott élszámból kialakítható maximális méretű klikkhez nagyon közeli konfigurációnak felel meg. Az egyetlen hatalmas „csillagnak” megfelelő topológia szintén nagyon kedvező energetikailag, hiszen mind a maximális klikk, mind a „csillag” esetén az energia vezető rendben $-M \ln M$ szerint skálázik az élek számával. A két topológia közti energiakülönbség domináns tagja ehhez képest csupán $\sqrt{M} \ln \sqrt{M}$ nagyságrendű.

Az 1.3.1. alfejezethez hasonlóan itt is a $\varphi_d = d_{\max}/M$ rendparamétert használhatjuk a fázisátalakulások nyomon követésére, ahol $d_{\max}$ a hálózatban előforduló legnagyobb fokszám. Az említett két alacsony energiájú topológia esetén φ_d két jól megkülönböztethető értéket vesz fel, hiszen $d_{\max} \approx \sqrt{2M}$ a maximális klikk esetén, míg $d_{\max} \approx M$ a „csillag” esetén. Az 1.6. ábrán mutatjuk be az idevonatkozó Monte–Carlo-szimulációk eredményeit. Látható, hogy ha magas hőmérsékletről kezdjük el hűteni a rendszert, akkor a rendezetlen fázisból először néhány nagyon nagy fokszámú csúcs emelkedik ki, melyek szinte az összes élt begyűjtik, és nagyon hasonlítanak a fent említett „csillagszerű” topológiára, majd a legalacsonyabb energiának megfelelő klikkszerű állapot csak a legalacsonyabb hőmérsékleten áll be. Ennek magyarázata az, hogy a klikknek nem csak az energiája kisebb, mint a csillagé, hanem az entrópiája is lényegesen alacsonyabb, ezért csak olyan hőmérsékleten válhat uralkodóvá, ahol a szabadenergiában az entrópiával kapcsolatos tagokat elnyomják az energiához kapcsolódó tagok. A legvalószínűbb topológiákat kis rendszerméreteket mellett az egzakt leszámlálás módszerével is vizsgáltuk, ennek eredményeit foglalja össze az 1.7. ábra. A $T = 0.65$ -nél, illetve a $T = 0.3$ -nál kapott topológiák teljes összhangban vannak a Monte–Carlo-szimulációknál látott eredményekkel.

Az 1.6. ábrán hiszterézist figyelhetünk meg a $d_{\max}$ görbében a klikkszerű topológia és a „csillagszerű” topológia közti átalakulásnál, ami elsőrendű fázisátalakulásra utal. Ezzel szemben a másik átalakulásnál, mely a magas hőmérsékleti rendezetlen állapot és a „csillagszerű” állapot között lép fel, nincs hiszterézis, és az átalakulást egy divergencia kíséri a hőkapacitásban is, ami azt mutatja, hogy ez egy folytonos átalakulás. Amennyiben $\varphi_d > 1/2$, használhatjuk az (1.20) analitikus közelítést a feltételes szabadenergiára ebben az esetben is, ami az $f(-\varphi_d M) = -(\varphi_d M) \ln(\varphi_d M)$ helyettesítéssel az

$$F(\varphi_d, T) \approx M(T-1) \ln(N) \varphi_d \quad (1.22)$$

egyenletre vezet. A fentiekkel összhangban ez alapján azt kapjuk, hogy a „csillagszerű” topológia stabil $T < 1$ esetén (az $F(\varphi_d, T)$ minimuma a $\varphi_d = 1$ -nél van), míg $T > 1$ esetén instabillá válik. Az (1.22) egyenlet jóslata alapján továbbá a $T = T_c = 1$ -nél bekövetkező átalakulásnál φ_d lépcsőszerűen esik le nullára, hiszterézis nélkül, ami egy folytonos fázisátalakulásnak felel meg, végtelen nagy kritikus exponenssel. (Az 1.6. ábrán látható eltérés a kritikus hőmérsékletben a Monte–Carlo-szimulációk esetén nagy valószínűség szerint a



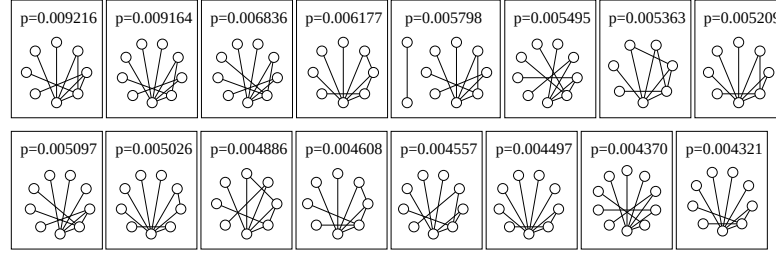
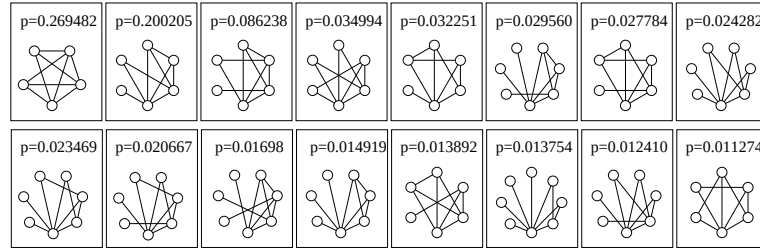
1.6. ábra. A Monte–Carlo-szimulációk eredményei az $E = -\sum_i d_i \ln d_i$ energiafüggvény mellett egy $N = 10224$ csúcsból és $M = 2556$ élből álló gráfsokaság esetén. a) A hálózatban előforduló legnagyobb fokszaám a hőmérséklet függvényében. Az adatpontok a $t = 5000N$ és $t = 20000N$ Monte–Carlo-lépések között lettek kiátlagolva egyetlen futtatás során. b) Tipikus topológiák a három jól elkülöníthető hőmérséklet-tartományban: egy rendezetlen, Erdős–Rényi-gráfhoz hasonló állapot magas hőmérsékleteken, ahol $d_{\max} = \mathcal{O}(1)$, néhány nagyon nagy fokszaámú csúcs, sok közös szomszéd a köztes hőmérsékleteken, ahol $d_{\max} = \mathcal{O}(M)$ és egy klikk az alacsony hőmérsékleteken, ahol $d_{\max} = \mathcal{O}(\sqrt{M})$. c) A fokszaámok kiegészítő eloszlásfüggvénye, $P(d)$ a $T = 0.8$ hőmérsékleten, a $t = 600N$ Monte–Carlo-lépésnél. (Az ábra forrása a [T1] publikáció.)

véges méret effektusoknak köszönhető.)

Egy külön érdekessége az $E = -\sum_k d_k \ln d_k$ energiafüggvény által definiált gráfsokaságoknak, hogy a rendezetlen állapotból a „csillagszerű” állapotba való átalakulás során skálafüggetlen hálózatok jelennek meg a kritikus pont közelében. Ezt demonstrálja az 1.6c ábra, melyen a fokszaámok kiegészítő eloszlását⁴, $P(d)$ -t ábrázoltuk a fokszaám függvényében. Látható, hogy ez hatványszerűen csökken a magas fokszaámok tartományában, nagyjából egy -2 -es exponenssel. A fokszaám sűrűségfüggvény ezek alapján szintén hatványszerűen cseng le, $\rho(d) \sim d^{-\gamma}$, ahol az exponens $\gamma \approx 3$. Ez az érték megegyezik az Barabási–Albert-modell által jósolt exponenssel, ami mint látni fogjuk, nem pusztán a véletlen műve.

A Monte–Carlo-dinamika során ha egy élt az i csúcshoz próbálunk kötni, akkor az energiaváltozást az $f(d_i) = -d_i \ln d_i$ deriváltjával, $\Delta E = 1 - \ln d_i$ -vel becsülhetjük. A Metropolis–

⁴A kiegészítő eloszlásfüggvény definíciója szerint $P(x)$ azt a valószínűséget adja meg, hogy az adott változó (mennyiség) nagyobb, mint x , azaz ha a hagyományos eloszlásfüggvény $F(x)$, akkor a hozzá tartozó kiegészítő eloszlás $P(x) = 1 - F(x)$.

a) $T=0.65$ b) $T=0.3$ 

1.7. ábra. A legvalószínűbb topológiák az $E = -\sum_i d_i \ln d_i$ energiafüggvény által definiált gráfsokaságban $N = 48$ csúcs és $M = 12$ él esetén az egzakt leszámlálás alapján. (Min-den esetben csak a legnagyobb összefüggő komponenst mutatjuk) a) Ha $T = 0.65$, akkor a legvalószínűbb topológiák tartalmaznak egy, a többihez képest kimagasló fokszerű csúcsot, melynek sokszor vannak közös szomszédai a fokszerű szerint utána következő egy-két csúccsal. b) $T = 0.3$ -nál már egy klikkszerű állapot a legvalószínűbb topológia, és a sorban utána következő néhány konfiguráció is lényegesen kiegyensúlyozottabb, mint az a) ábrán. (Az ábra forrása a [T1] publikáció.)

Hastings-szabály szerint az átkötés elfogadási valószínűsége az $e^{-\Delta E/T}$ faktorról arányos, ami $T = T_c = 1$ -nél egy d_i -vel arányos valószínűséget eredményez. Ez alapján a kritikus pontnál az éltrendeződési dinamika a preferenciális csatolási szabályhoz hasonló módon rendezi át a hálózat struktúráját, ami megmagyarázza, hogy miért alakul ki olyan fokszámeloszlás, mint a Barabási–Albert-modell esetén.

Egy további érdekesség, hogy a maximális klikknek illetve a „csillagnak” megfelelő topológiák megegyeznek R. Guimerà és munkatársai által a [92] publikációban bemutatott két alapvető topológiával, melyeket lokális keresésre optimalizált hálózatoknál találtak az esetlegesen fellépő csomagtörölődések figyelembevételével. A köztes hőmérséklet-tartományon látott néhány nagy „csillag”, melyeknek sok közös szomszédjuk van, nagyon hasonlít a kis-számú párhuzamos keresések esetén optimális topológiára, míg az alacsony hőmérsékleten felbukkanó klikk a nagyszámú párhuzamos keresések esetén optimális homogén topológiának felel meg. Mivel a mi esetünkben e két „közel optimális” topológia teljesen természetes módon, az éltrendeződési dinamikának köszönhetően jelenik meg, felvetődik az ilyen típusú modellek alkalmazásának ötlete egyéb hálózatoptimalizálási problémáknál is. A megközelítés lényege, hogy az optimális topológiát szimulált hőkezeléssel keressük, egy

1. FEJEZET: Topológiai fázisátalakulások

megfelelően választott energiafüggvény segítségével.

1.3.3. A nagykanonikus sokaság

Természetes módon vetődik fel az a kérdés, hogy vajon a nagykanonikus sokaságban is lehet-e vizsgálni a fejezetben bemutatott topológiai fázisátalakulásokat. Ennek kapcsán érdemes megjegyezni, hogy mindkét vizsgált egyrézecskes energiafüggvénynél alacsony hőmérsékleteken olyan állapotok jelennek meg, melyeknél a csúcsok egy része „makroszkopikus” számú élt gyűjt be. Emiatt a teljes rendszer energiája máshogyan skálázik a rendszermérettel alacsony és magas hőmérsékleten. Magas hőmérsékleten a hálózat rendezetlen és nagyon sok pici izolált komponensből áll, melyek nagyjából egyforma méretűek. Ennek révén a rendszerméret növelése itt csupán ezen kis izolált komponensek számának növekedésével jár, és a teljes rendszer összenergiája N -nel arányos lesz. Ezzel szemben az alapállapot energiája például az $f(d_i) = -\sum_i d_i^2$ esetén M^2 -tel skálázik, hiszen az alapállapot egy minden élt begyűjtő „csillag”, míg az $f(d_i) = -d_i \ln d_i$ esetén az alapállapot energiája arányos $M \ln M$ -mel. Ez alapján nem tudunk olyan csatolási állandót megadni, mely minden hőmérsékleten extenzívvé tenné az energiát.

A fentiek alapján a rendezetlen fázis, melynél a teljes hálózat energiája extenzív, tanulmányozható a nagykanonikus sokaságban is. Ilyenkor J. Berg és M. Lässig eredményei alapján a foksám eloszlás a

$$p(d) = c \frac{e^{-f(d)/T - \mu d}}{d!} \quad (1.23)$$

alakban adható meg [20], ahol c egy normálási konstans és μ felel meg a kémiai potenciálnak, melynek segítségével például be lehet állítani a kívánt átlagos foksámot. Az általunk javasolt $f(d) = -d \ln d$ energiafüggvény esetén a fenti kifejezés a Stirling-formula segítségével a

$$p(d) = c \frac{e^{-(\mu-1)d}}{\sqrt{2\pi d}} d^{(1/T-1)d} \quad (1.24)$$

azonosságra vezet. Ha $T > 1$, akkor a fenti eloszlás az exponenciálisnál gyorsabban cseng le, emiatt várhatóan minden csúcsnak viszonylag alacsony foksáma lesz. Ezzel szemben $T < 1$ esetén az eloszlás divergenssé válik a magas foksámok tartományán, ami azt jelzi hogy $T = T_c = 1$ -nél egy fázisátalakulás következik be. A hirtelen megnövekedő valószínűséggel megjelenő nagy foksámú csúcsok miatt azonban a $T < 1$ tartományon többé már nem extenzív az energia, ezért a nagykanonikus sokaság is érvényét veszti. Mindenesetre ha a magas hőmérsékletek felől közelítjük meg a kritikus pontot, akkor a kritikus pontra tekinthetünk úgy mint a nagykanonikus sokaság érvényességének határára. Ha egy fokkal általánosabb energiafüggvényt választunk az $f(d) = -(d - \gamma) \ln d$ formában és $\mu = 1$ értéket állítunk be a kémiai potenciálra, akkor ebben a $T = 1$ pontban elérhető, hogy a foksám sűrűségfüggvény a $\rho(d) \sim d^{-(\gamma+1/2)}$ alakot vegye fel, és a hálózat skálafüggetlenné váljon. Azonban ez a viselkedés egyáltalán nem robusztus, $\mu > 1$ esetén a foksámeloszlás exponenciálisan lecseng, $\mu < 1$ esetén viszont divergál a nagy foksámoknál.

2. fejezet

k -klikkperkoláció véletlen gráfokban

A perkoláció jelensége széleskörűen tanulmányozott a statisztikus fizikában, ahol folyadék porózus anyagon történő átáramlását, illetve magát a porózus anyagot a következő egyszerű modellel írjuk le [37, 86]: Egy szabályos rács csúcsait (vagy éleit) egy adott p valószínűséggel véletlenszerűen betöltjük, a betöltött rácshelyeket feleltetvén meg az anyagban véletlenszerűen elhelyezkedő üregeknek. Kis p esetén természetesen a betöltött rácshelyek ritkán állnak össze nagyobb klaszterekké, helyett kisméretű, izolált klasztereket tapasztalunk. Azonban egy kritikus p_c betöltési valószínűségnél egy folytonos fázisátalakuláshoz hasonló módon megjelenik a rendszerben egy óriási klaszter (perkoláló klaszter), melynek mérete már nem elenyésző a teljes rendszermérethez viszonyítva, ezáltal lehetővé téve például a folyadék átáramlását az anyagban.

Természetesen a perkoláció jelenségét szabályos rács helyett véletlen gráfokon is lehet vizsgálni. Ilyenkor a gráf összefüggő komponenseit feleltetjük meg a fent említett, szomszédos betöltött rácshelyekből összeálló klasztereknek. Ahogy azt már korábban említettük, az egyik legelső fontos eredmény az Erdős–Rényi-moddal kapcsolatban a kritikus élbekötési valószínűség levezetése volt, mely szerint $p = p_c = 1/N$ esetén megjelenik a hálózatban egy óriás komponens, melynek relatív mérete még $N \rightarrow \infty$ esetén sem tart nullához [64, 30]. Ezen „klasszikus” élperkolációs jelenség több módosítását illetve általánosítását is vizsgálták a komplexhálózat-kutatás területén az elmúlt 15 évben.

Az egyik nagy érdeklődést kiváltó problémakör a robbanásszerű perkoláció (angolul *explosive percolation*) jelenségéhez kapcsolódik [1, 48, 171, 228, 172, 57, 9], melynél egy üres (éleket nem tartalmazó) gráfból indulva iterációnként mindig két lehetséges új élt választunk véletlenszerűen, melyek közül azt kötjük be ténylegesen is a hálózatba, melynek végpontjaihoz kapcsolódó komponensek méretének szorzata kisebb. (Ezt a „szorzatszabályt” Bollobás Béla javasolta először az átalakulás bekövetkezésének késleltetésére). Egy ilyen folyamatnál az óriás komponens egy lényegesen nagyobb kritikus bekötött élhányadnál jelenik meg a teljesen véletlenszerű élbekötési folyamathoz viszonyítva, ugyanakkor az átalakulás sokkal drasztikusabb is, mint a „hagyományos” esetben. Bár a numerikus szimulációk kezdetben azt sejtették, hogy a rendparaméternek szakadása van a kritikus pontban, Rui A. da Costa és munkatársainak egy későbbi munkája [54], illetve Oliver Riordan és Lutz Warnke eredményei [180, 181] még általánosabb kereteken belül megmutatták, hogy

2. FEJEZET: k -klikkperkoláció véletlen gráfokban

az átalakulás folytonos.

Egy másik fontos kapcsolódó témakör az egymástól kölcsönösen függő hálózatok perkolációs átalakulásának vizsgálata [39, 162, 75, 76, 77]. Számos olyan hálózattal találkozhatunk a mindennapokban, melynek működését jelentősen befolyásolhatja egy másik hálózat: 2003-ban például Olaszországban néhány erőmű leállása kihatott a helyi internetszolgáltatásra, melynek révén további erőművek álltak le, és a folyamat végeredménye egy országos szintű, jelentős áramszünet lett [184]. Az ehhez hasonló jelenségek miatt az egymástól függő hálózatok robusztusságának illetve sérülékenységének vizsgálata gyakorlati szempontból különösen fontos [39]. A véletlen meghibásodások hatását első közelítésben egy véletlenszerű csúcstávoltítási folyamattal modellezhetjük, mely a rendszer fragmentációjához vezet, ha elérjük a kritikus pontot [39, 162, 75].

Ebben a fejezetben az Erdős–Rényi-gráf „klasszikus” élperkolációs átalakulásának egy olyan általánosításáról lesz szó, melynél élek helyett k -klikkek (k csúcsból álló teljesen összekötött részgráfok) perkolációját vizsgáljuk. 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. Ez utóbbi aspektus annyira fontos, hogy ezt külön fejtjük ki a 3. fejezetben, míg a jelen fejezetben elsősorban a k -klikkperkoláció kritikus pontjával kapcsolatos elméleti eredményekre koncentrálnunk.

Az itt bemutatásra kerülő eredmények több publikációból lettek összeválogatva. A 2.2.1. alfejezetben bemutatásra kerülő, generátorfüggvényeken alapuló eredmények a *Journal of Statistical Physics*-ben jelentek meg a disszertáció szerzőjének első szerzőségével [T3], míg a k -klikkperkoláció kritikus pontjával kapcsolatos numerikus eredmények a *Physical Review Letters*-ben kerültek közlésre a disszertáció szerzőjének társszerzőségével [T4]. Ezek mellett fontos megemlíteni a disszertáció szerzőjének első szerzőségével publikált *Nature*-cikket is [T5], mely ugyan elsősorban a csoportkeresésre koncentrált, viszont ugyanakkor döntő szerepe volt abban, hogy a k -klikkperkoláció széles körben ismertté vált. A 2.3.1. alfejezet irányított k -klikkperkolációval kapcsolatos eredményei a *New Journal of Physics*-ben jelentek meg [T6] (szintén a disszertáció szerzőjének első szerzőségével). Egy áttekintő jellegű, összefoglaló írás a k -klikkperkolációról a Bollobás Béla, Kozma Róbert és Miklós Dezső által szerkesztett könyv egyik fejezeteként jelent meg [S2] (ahol a fejezet első szerzője megint csak a disszertáció szerzője volt).

A k -klikkperkoláció ötlete a [T5] publikáció szerzői közt zajló diszkussziók során merült fel úgy, mint egy ígéretes módszer hálózati csoportkeresésre. A hálózati csoportkeresés problémájának vizsgálatát Vicsek Tamás javasolta. A k -klikkperkolációról azonban hamar kiderült, hogy a csoportkeresés mellett elméleti szempontból is nagyon érdekes. A k -klikkperkoláció definíciója a [T5] publikáció szerzőinek közös eredménye. A k -klikkperkolációs klaszterek kinyerésére szolgáló algoritmust Derényi Imre és a disszertáció szerzője közösen dolgozták ki, ez a 3.1.1. alfejezetben kerül ismertetésre, a csoportkereséssel kapcsolatos eredményeknél. A módszert a disszertáció szerzője implementálta, melynek révén megnyílt az út a k -klikkperkoláció numerikus tanulmányozása előtt az Erdős–Rényi-modellben (a vonatkozó numerikus szimulációkat szintén a disszertáció szerzője végezte el).

A kritikus pontot analitikus úton elsőként Derényi Imre vezette le egy heurisztikus megközelítés segítségével, később a disszertáció szerzője is megadott egy alternatív levezetést generátorfüggvények használatával. Az irányított k -klikkperkoláció esetén a kritikus pontot a disszertáció szerzője határozta meg analitikus úton, a kapcsolódó numerikus szimulációkat Farkas Illés végezte el a disszertáció szerzője által készített programok segítségével.

2.1. Definíciók

Először felsorolunk néhány definíciót, melyek megalapozzák a k -klikkperkoláció fogalmát. Hangsúlyozzuk, hogy a k -klikk fogalma régóta ismert a hálózatkutatásban [30, 67, 25, 24], viszont a k -klikkszomszédságon alapuló k -klikkperkolációs klaszterek vizsgálata először a [T5, T4] és [T3] cikkekben jelent meg. A k -klikkperkoláció tehát a következő definíciókon alapszik:

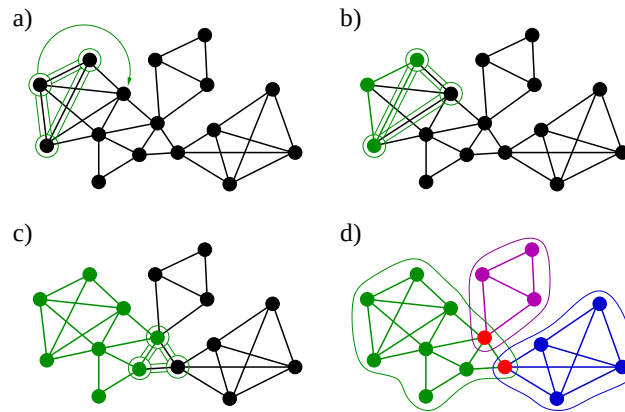
- k -klikk: egy k csúcsból álló teljes részgráf (melyben minden csúcs össze van kötve minden másik csúccsal).
- k -klikkszomszédság: két k -klikk szomszédos, ha $k-1$ csúcsuk közös (azaz csak egy-egy csúcsban különböznek egymástól).
- k -klikklánc: k -klikkek sorozata, melyben minden k -klikk szomszédos az előtte- és utána következő k -klikkel (természetesen az első és utolsó k -klikk kivételével).
- k -klikkösszekötöttség: két k -klikk összekötött, ha tagjai ugyanannak a k -klikk láncnak.
- k -klikkperkolációs klaszter: k -klikkek olyan maximális halmaza, melyben minden k -klikkpár összekötött (azaz ezen belül bármelyik k -klikkből eljuthatunk bármelyik más k -klikkbe k -klikkszomszédságok sorozatán keresztül).

A k -klikkperkolációt szemléletesen egy „ k -klikksablon” görgetésével is elképzelhetjük: A k -klikksablon izomorf egy k -klikkel, és ráhelyezhetjük a hálózat bármely k -klikkjére. Ezután megpróbálhatjuk görgetni a sablont a hálózaton a következő egyszerű szabály alapján: mindig csak szomszédos k -klikkre gördíthetjük tovább. A hálózat k -klikkperkolációs klaszterei azon részgráfoknak felelnek meg, melyeket be tudunk járni a k -klikksablon görgetésével. Általános esetben több ilyen tartományt is találhatunk a hálózatban, melyek között nem lehetséges a sablon átgörgetése; ilyenkor természetesen több k -klikkperkolációs klaszterrel állunk szemben. A sablon gördítés szemléltetése látható a 2.1. ábrán.

A k -klikkperkolációs klaszterek egyik érdekes tulajdonsága, hogy egy csúcs akár több k -klikkperkolációs klaszternek is tagja lehet. (Erre legegyszerűbb példa $k=3$ esetén két olyan háromszög, melyeknek egy csúcsuk közös.) Ennek révén a k -klikkperkoláció egy természetes megközelítést nyújt átfedő hálózati csoportok kereséséhez, amint azt bővebben is kifejtjük a 3. fejezetben.

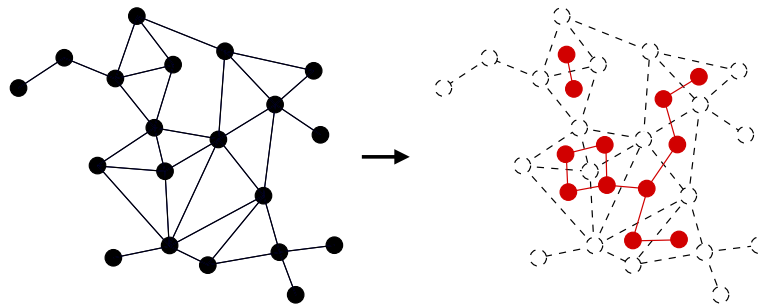
Megjegyezzük, hogy a k -klikkperkoláció megfelel a hagyományos élperkolációnak a hálózathoz rendelhető k -klikkszomszédsági gráfban: Ebben a gráfban minden egyes csúcspont

2. FEJEZET: k -klikkperkoláció véletlen gráfokban



2.1. ábra. A „ k -klikksablon” görgetése egy hálózaton $k = 3$ esetén. a) A sablon (zöld színnel jelölve) izomorf egy k -klikkkel, ezért ráhelyezhetjük a hálózat bármelyik k -klikkjére. Ezután mindig egy szomszédos k -klikkbe gördítjük tova a hálózat b) A sablon helyzete az első gördítés után. A már bejárt tartományt zöldre színezzük a hálózaton. c) Ezen a ponton már nem tudjuk még meg nem látogatott k -klikkre görgetni a sablont, ezért a zölddel jelölt k -klikkperkolációs klaszter feltérképezése véget ér. d) A hálózat k -klikkperkolációs klaszterei $k = 3$ esetén, zöld, kék és lila színekkel jelölve, a klaszterek közötti átfedésekben lévő csúcsokat pirossal jelöltük.

a hálózat egy k -klikkjét reprezentálja, és két csúcspont akkor van összekötve, ha a megfelelő k -klikkek szomszédosak. A fent definiált k -klikkperkolációs klaszterek ezen hálózat összefüggő komponenseinek felelnek meg. A k -klikkszomszédsági gráf szemléltetését a 2.2 ábrán mutatjuk be.



2.2. ábra. A k -klikkszomszédsági gráf szemléltetése $k = 3$ esetén. Minden hálózathoz hozzárendelhetünk egy k -klikkszomszédsági gráfot, melyben a csúcsok az eredeti hálózat k -klikkjeinek felelnek meg, és két csúcs akkor van összekötve, ha az adott k -klikkek szomszédosak. A jobb oldalon látható hálózathoz tartozó k -klikkszomszédsági gráfot a bal oldalon piros színnel ábrázoltuk $k = 3$ esetén (szaggatott vonallal még egyszer feltüntetve az eredeti hálózatot, a jobb átláthatóság kedvéért).

2.2. k -klikkperkoláció Erdős–Rényi-gráfban

Elsőként az Erdős–Rényi-gráf esetén vizsgáljuk meg a k -klikkperkoláció jelenségét. Ha a modell extrém eseteit tekintjük, a $p = 0$ élbekötési valószínűség esetén egyáltalán nincs k -klikkperkolációs klaszter a rendszerben (hiszen élek sincsenek), míg $p = 1$ esetén bármilyen $k \leq N$ -re az egész gráf egyetlen k -klikkperkolációs klaszternek felel meg (hiszen minden lehetséges él be van kötve). A kérdés az, hogy ha elkezdjük növelni az élbekötési valószínűséget, akkor milyen p esetén jelenik meg a rendszerben egy óriás k -klikkperkolációs klaszter. Előrevetítjük, hogy hasonlóan az óriás komponens megjelenéséhez $p_c = \frac{1}{N}$ -nél, az óriás k -klikkperkolációs klaszter megjelenése egy k -tól függő $p_c(k)$ kritikus értéknél szintén nagyon hasonló egy fázisátalakuláshoz. A $p_c(k)$ értékére először Derényi Imre adott egy heurisztikus levezetést, amit a B.1. függelékben vázolunk. Itt a továbbiakban a disszertáció szerzője által javasolt, generátorfüggvényeken alapuló levezetést ismertetjük a 2.2.1. alfejezetben. Megjegyezzük, hogy a k -klikkperkoláció fogalmát bevezető, Vicsek Tamás által vezetett csoporttól teljesen független forrásokból is születtek analitikus eredmények az átalakulás kritikus pontjával kapcsolatban: Ráth Balázs és Tóth Bálint egy parciális differenciálegyenleteken alapuló megközelítést javasoltak [174], melyet a B.2. függelékben vázolunk röviden, míg Bollobás Béla és Oliver Riordan egzakt eredményeket vezettek le [32], melyeket a B.3. függelékben sorolunk fel.

2.2.1. Generátorfüggvény formalizmus

Ebben az alfejezetben a k -klikkperkoláció kritikus pontját határozzuk meg egy generátorfüggvényeken alapuló megközelítés segítségével. Az itt bemutatásra kerülő analitikus számításokat a disszertáció szerzője végezte el, az eredmények a [T3] cikkben kerültek publikálásra a disszertáció szerzőjének első szerzőségével.

Mielőtt rátérnénk a k -klikkperkoláció kritikus pontjának vizsgálatára, összefoglaljuk a generátorfüggvények legfontosabb tulajdonságait. Ha egy nemnegatív egész értékeket felvevő ξ valószínűségi változó eloszlása $\rho(n)$ (azaz minden nemnegatív n esetén annak valószínűsége, hogy $\xi = n$ megegyezik $\rho(n)$ -nel), akkor az ehhez tartozó generátorfüggvény

$$G_\rho(x) \equiv \langle x^\xi \rangle = \sum_{n=0}^{\infty} \rho(n) x^n. \quad (2.1)$$

Egy megfelelően normált valószínűségi változó generátorfüggvénye abszolút konvergens $|x| \leq 1$ esetén (ezért szingularitásai sincsenek ezen a tartományon). Az $x = 1$ esetén egyszerűen

$$G_\rho(1) = \sum_{n=0}^{\infty} \rho(n) = 1. \quad (2.2)$$

A valószínűségi változó eloszlását és momentumait megadhatjuk a generátorfüggvény se-

2. FEJEZET: k -klikkperkoláció véletlen gráfokban

gítségével a következő összefüggések alapján:

$$\rho(n) = \frac{1}{n!} \left. \frac{d^n G_\rho(x)}{dx^n} \right|_{x=0}, \quad (2.3)$$

$$\langle \xi^l \rangle = \sum_{n=0}^{\infty} n^l \rho(n) = \left[\left(x \frac{d}{dx} \right)^l G_\rho(x) \right]_{x=1}. \quad (2.4)$$

Végezetül, ha $\eta = \xi_1 + \xi_2 + \dots + \xi_l$ ahol $\xi_1, \xi_2, \dots, \xi_l$ független valószínűségi változók (melyek nemnegatív egész értékeket vehetnek fel), akkor az η változóhoz és annak $\sigma(n)$ eloszlásához tartozó generátorfüggvény a

$$\begin{aligned} G_\sigma(x) &= \langle x^\eta \rangle = \langle x^{\xi_1} x^{\xi_2} \dots x^{\xi_l} \rangle = \langle x^{\xi_1} \rangle \langle x^{\xi_2} \rangle \dots \langle x^{\xi_l} \rangle = \\ &= G_{\rho_1}(x) G_{\rho_2}(x) \dots G_{\rho_l}(x). \end{aligned} \quad (2.5)$$

formában adható meg. A generátorfüggvényeken alapuló megközelítést általános hálózateleméleti problémákra M. E. J. Newman és munkatársai javasolták a [155] publikációban.

A k -klikkperkoláció kritikus élbekötési valószínűségét az $N \rightarrow \infty$ határesetben fogjuk levezetni. Ehhez először megadjuk egy véletlenszerűen választott k -klikkel szomszédos k -klikkek számának $u(n)$ eloszlását. Az, hogy találunk egy szomszédos k -klikket a véletlenszerűen választott k -klikkünkhöz, ekvivalens azzal, hogy találunk egy csúcsot a k -klikken kívül, melyből legalább $k-1$ él mutat az adott k -klikkbe. Egy konkrét csúcs esetén annak valószínűsége, hogy k éle mutat az adott k -klikkbe, az élek függetlensége miatt egyszerűen p^k , míg annak valószínűsége, hogy csak $k-1$ éllel kapcsolódik a k -klikkhez $k(1-p)p^{k-1}$ (hiszen ilyenkor egy „nemlétező” élt is figyelembe kell vennünk $1-p$ valószínűséggel). Összesen $N-k$ csúcs esik az adott k -klikken kívülre a hálózatban, ezért N -ben vezető rendig a szomszédos k -klikkek várható száma:

$$\langle u \rangle = (N-k) \left[k(1-p)p^{k-1} + p^k \right] \simeq Nkp^{k-1}. \quad (2.6)$$

Az élek függetlenségéből az is következik, hogy az $u(n)$ Poisson-eloszlást fog követni, ami ez alapján az

$$u(n) = \exp(-Nkp^{k-1}) \frac{(Nkp^{k-1})^n}{n!} \quad (2.7)$$

alakot ölti.

Tegyük fel, hogy a perkoláció kritikus pontja alatt vagyunk, azaz a k -klikkek ritkák, szomszédos k -klikkek pedig még ritkábbak, és a hurkok a k -klikkszomszédsági gráfban elhanyagolhatóak, azaz a k -klikkszomszédsági gráf lokálisan faszerű. Ebben az esetben egy k -klikkperkolációs klaszter méretét az alábbi elágazási folyamat alapján lehet meghatározni. Az első lépésben kiválasztjuk a klaszter egy tetszőleges k -klikkjét, majd betöltjük az első szomszédait a k -klikkszomszédsági gráfban. Ezután bármely betöltött k -klikk esetén az ő szomszédainak betöltésével haladunk tovább, kivéve azt a szomszédot, ahonnan az adott k -klikk betöltése történt. Az eredeti hálózatban ez ekvivalens a k -klikksablon továbbgördítésével, ahol elágazásonként meg többszörözzük a sablont, és nem görgetünk „visszafelé”.

Legyen $I(n)$ annak valószínűsége, hogy egy véletlenszerűen választott k -klikkből kiindulva, a fenti folyamat alapján egy n darab k -klikkből álló perkolációs klasztert tudunk felfedezni, beleszámolva magát a kiinduló k -klikket is. Továbbá jelölje $H(n)$ annak valószínűségét, hogy a fenti elágazási folyamat során egy később sorra kerülő k -klikkből továbbmenve egy n méretű ágat fedezzük fel az adott k -klikkperkolációs klaszternek. Ez ekvivalens annak valószínűségével, hogy egy véletlenszerűen választott k -klikkből az elágazási folyamatot az egyik szomszéd irányában letiltva a maradék szomszédok irányában összesen egy n darab k -klikkből álló klasztert fedezzünk fel. Végezetül jelölje $H_m(n)$ annak valószínűségét, hogy ha m darab k -klikket választunk véletlenszerűen, akkor ugyanezen korlátozott elágazási folyamat segítségével felfedezett ágak összesen n darab k -klikket tartalmaznak. Mivel a kritikus pont alatt vagyunk, a k -klikkszomszédossági gráf sok kis szétcsúszó komponensből áll, ezért $N \rightarrow \infty$ esetén annak valószínűsége, hogy két (vagy több) k -klikk az m -ből ugyanahhoz a klaszterhez tartozik, elhanyagolható. Ennek alapján a (2.5) egyenlet szerint a $H(n)$ illetve $H_m(n)$ eloszlásokhoz tartozó $G_H(x)$ és $G_{H_m}(x)$ generátorfüggvények között a következő összefüggést írhatjuk fel:

$$G_{H_m}(x) = [G_H(x)]^m. \quad (2.8)$$

A fenti generátorfüggvények meghatározásához először bevezetjük $v(n)$ -et, ami annak valószínűségét adja meg, hogy egy véletlenszerűen választott k -klikkből n szomszédos k -klikkbe tudunk átgördülni azzal a megkötéssel, hogy az egyik „irányt” nem használhatjuk. A $v(n)$ formuláját teljesen hasonló gondolatmenettel lehet levezetni, mint az $u(n)$ -et megadó (2.6-2.7) összefüggéseket, azzal a különbséggel, hogy most csak $k-1$ -féle módon választhatjuk ki, hogy az adott k -klikk melyik részhalmazához csatlakozik a külső csúcs. Ez alapján

$$v(n) = \exp\left(-N(k-1)p^{k-1}\right) \frac{(N(k-1)p^{k-1})^n}{n!}. \quad (2.9)$$

Mivel a k -klikkszomszédossági gráfot faszerűnek tekintjük, a $H(n)$ -et felírhatjuk $v(n)$ és $H_m(n)$ segítségével a

$$H(n) = v(0)H_0(n-1) + v(1)H_1(n-1) + v(2)H_2(n-1) + \dots \quad (2.10)$$

formában. Mindkét oldal generátorfüggvényét képezvén, felhasználva a (2.3) és (2.8) összefüggéseket az alábbi egyenletet írhatjuk fel:

$$\begin{aligned} G_H(x) &= \sum_{n=0}^{\infty} \left[\sum_{m=0}^{\infty} v(m)H_m(n-1) \right] x^n = \\ &= \sum_{n=0}^{\infty} \left[\sum_{m=0}^{\infty} v(m) \frac{1}{(n-1)!} \frac{d^{n-1}}{dx^{n-1}} [G_H(x)]^m \Big|_{x=0} \right] x^n = \\ &= \sum_{m=0}^{\infty} v(m) [G_H(x)]^m x = xG_v\left(G_H(x)\right), \end{aligned} \quad (2.11)$$

ahol $G_v(x)$ természetesen a $v(n)$ eloszlás generátorfüggvényét jelöli.

2. FEJEZET: k -klikkperkoláció véletlen gráfokban

A (2.10) egyenlethez nagyon hasonló kifejezést írhatunk $I(n)$ -re is, az $u(n)$ eloszlás és $H_m(n)$ segítségével:

$$I(n) = u(0)H_0(n-1) + u(1)H_1(n-1) + u(2)H_2(n-1) + \dots \quad (2.12)$$

Ismét mindkét oldal generátorfüggvényét képezve a

$$G_I(x) = xG_u(G_H(x)), \quad (2.13)$$

egyenletre jutunk, ahol $G_u(x)$ az $u(n)$ generátorfüggvényét adja meg. A (2.4) és (2.13) összefüggések alapján egy véletlenszerűen választott k -klikkből induló elágazási folyamat révén feltérképezhető k -klikkperkolációs klaszter várható mérete

$$\langle I \rangle = G'_I(1) = G_u(G_H(1)) + G'_u(G_H(1))G'_H(1) = 1 + G'_u(1)G'_H(1). \quad (2.14)$$

A (2.11) segítségével felírhatjuk, hogy

$$G'_H(1) = G_v(G_H(1)) + G'_v(G_H(1))G'_H(1) = 1 + G'_v(1)G'_H(1), \quad (2.15)$$

amiből $G'_H(1)$ kifejezhető a következő alakban:

$$G'_H(1) = \frac{1}{1 - G'_v(1)}. \quad (2.16)$$

Ezt visszahelyettesítve a (2.14) egyenletbe az

$$\langle I \rangle = 1 + \frac{G'_u(1)}{1 - G'_v(1)} = 1 + \frac{\langle u \rangle}{1 - \langle v \rangle} \quad (2.17)$$

eredményre jutunk.

A perkoláció kritikus pontjában egy véletlenszerűen választott k -klikkhez kapcsolódó k -klikkperkolációs klaszter várható mérete, $\langle I \rangle$ divergál. A (2.17) kifejezés alapján ez akkor következik be, ha

$$\langle v \rangle = N(k-1)p^{k-1} = 1. \quad (2.18)$$

Ezt $p_c(k)$ -ra átrendezve a

$$p_c(k) = \frac{1}{[(k-1)N]^{\frac{1}{k-1}}} \quad (2.19)$$

végeredményre jutunk. A (2.19) kifejezés $k=2$ esetén visszaadja a klasszikus, $p_c = \frac{1}{N}$ élperkolációs kritikus pontot, ami megnyugtató, hiszen a k -klikkek $k=2$ esetén pusztán az éleknek felelnek meg.

2.2.2. Numerikus szimulációk

Az Erdős–Rényi-gráfon végzett numerikus szimulációink teljes egyezést mutattak a k -klikkperkolációra vonatkozó, az előző alfejezetben bemutatott analitikus eredményekkel. A vonatkozó numerikus szimulációkat a disszertáció szerzője végezte el és értékelte ki, az eredmények a [T4] publikációban jelentek meg.

Az analitikus megfontolásokkal összhangban a numerikus tapasztalatok azt mutatták, hogy az átalakulás folytonos, melyet nemuniverzális exponensek jellemeznek; ezek függenek mind k -tól, mind attól, hogy miként mérjük az óriás k -klikkperkolációs klaszter méretét. Erre több, nagyjából egyformán természetes választási lehetőség is adódik, az egyik legegyszerűbb az, amikor pusztán a bennfoglalt csúcsok számát feleltetjük meg a k -klikkperkolációs klaszter méretének. Ilyen esetben a rendparaméter az óriás klaszter relatív mérete,

$$\Phi = N_1(k)/N, \quad (2.20)$$

ahol $N_1(k)$ az óriás k -klikkperkolációs klaszter csúcsainak száma, N pedig a gráf csúcsainak száma.

Egy másik természetes választás a tartalmazott k -klikkek $\mathcal{N}_1(k)$ számában mérni az óriás k -klikkperkolációs klaszter méretét, hiszen ez felel meg a klaszterméretnek a k -klikkszomszédsági gráfban. Az ennek megfelelő rendparaméter

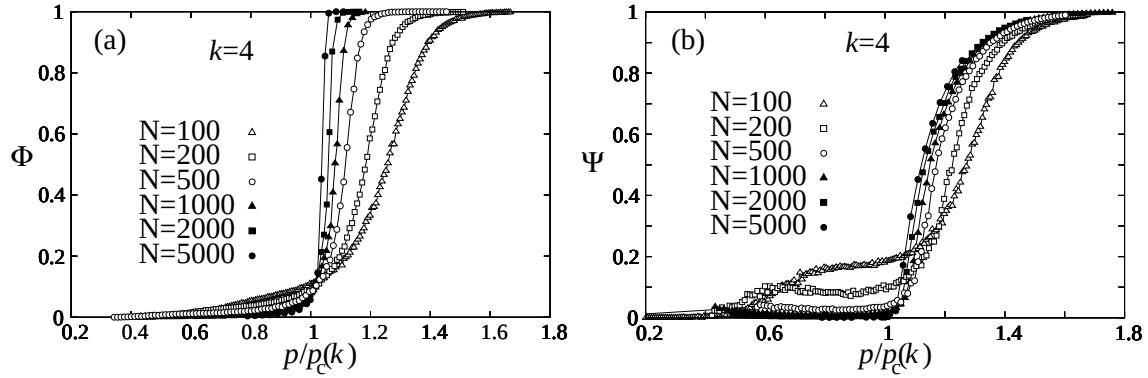
$$\Psi = \mathcal{N}_1(k)/\mathcal{N}(k), \quad (2.21)$$

ahol $\mathcal{N}(k)$ a gráfban található összes k -klikkek száma (ami a k -klikkszomszédsági gráf méretének felel meg). Az $\mathcal{N}(k)$ várható értéke

$$\langle \mathcal{N}(k) \rangle = \binom{N}{k} p^{k(k-1)/2} \approx \frac{N^k}{k!} p^{k(k-1)/2}, \quad (2.22)$$

hiszen k különböző csúcs összesen $\binom{N}{k}$ módon választható, és ezek akkor formálnak egy k -klikket ha a köztük lévő $k(k-1)/2$ él mind létezik (egymástól függetlenül p valószínűséggel). Az Erdős–Rényi-gráf klasszikus élperkolációs átmenete a $k=2$ esetnek felel meg, ilyenkor a (2.21) egyenletben definiált Ψ rendparaméter az óriás komponensben található élek relatív hányadát adja meg. Természetesen általánosan a tartalmazott l -klikkek $\mathcal{N}_1^{(l)}(k)$ számában is mérhetnénk a k -klikkperkolációs klaszterméretet $1 \leq l \leq k$ esetén. Az egyszerűség kedvéért itt a két szélsőértékre ($N_1(k) \equiv \mathcal{N}_1^{(1)}(k)$ -ra és $\mathcal{N}_1(k) \equiv \mathcal{N}_1^{(k)}(k)$ -ra) korlátozzuk vizsgálatainkat.

A számítógépes szimulációk eredményei azt mutatták, hogy az általunk vizsgált két rendparaméter másként viselkedik a kritikus pont környékén. Ezt illusztrálja a 2.3. ábra, melyen Φ -t és Ψ -t rajzoltuk fel a $p_c(k)$ -val átskálázott élbekötési valószínűség, $p/p_c(k)$ függvényében $k=4$ esetén, többféle rendszerméret (N) mellett. A Φ rendparaméter egy lépcsőfüggvényhez tart $k \geq 3$ esetén az $N \rightarrow \infty$ határesetben. Az, hogy a lépcső a $p/p_c(k)=1$ helyre esik, maximálisan alátámasztja a (2.19) egyenlet eredményét a kritikus élbekötési valószínűségekre. A Ψ rendparaméter más viselkedést mutat $k \geq 2$ esetén, mely nagyon hasonló a klasszikus élperkolációnál tapasztaltakhoz: egy olyan függvényhez tart, mely $p/p_c(k) < 1$ esetén nulla, és folytonosan növekszik fel egyhez ha $p/p_c(k)$ -t növeljük egytől végtelenig.

2. FEJEZET: k -klikkperkoláció véletlen gráfokban

2.3. ábra. A k -klikkperkoláció numerikus vizsgálata Erdős-Rényi-gráfban. a) A Φ rendparaméter az átskálázott $p/p_c(k)$ függvényében (több futtatásra átlagolva, a statisztikai hibák kisebbek a szimbólumméreteknél). A Φ egy lépcsőfüggvényhez tart az $N \rightarrow \infty$ határesetben. b) A Ψ rendparaméter a $p/p_c(k)$ függvényében, ugyanazon szimulációk esetén mint az a) panelen. A Ψ egy olyan függvényhez tart az $N \rightarrow \infty$ határesetben, mely nulla ha $p/p_c(k) < 1$ és folytonosan nő fel egyhez a $p/p_c(k)$ növelésével. (Az ábra forrása a [T4] publikáció).

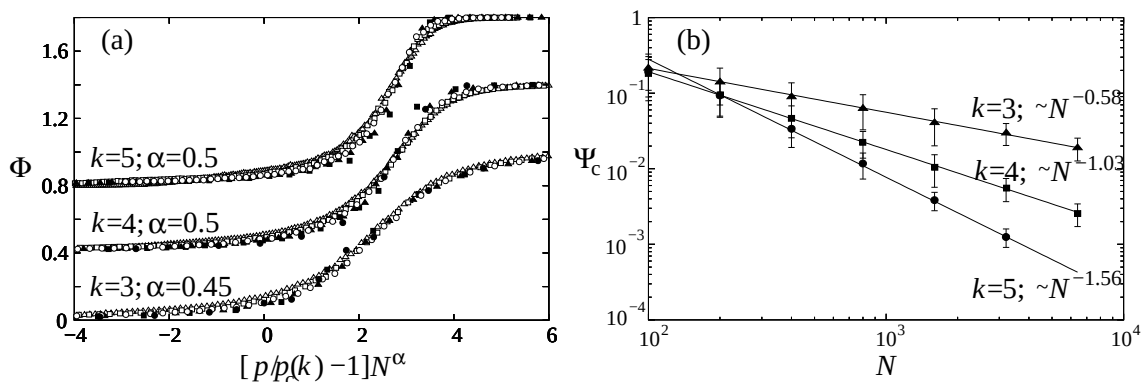
A 2.3b ábrához hasonló, folytonos átalakulások esetén a rendparaméter vizsgálata mellett a kritikus pont numerikus meghatározásának egyik legelterjedtebb módja a szuszceptibilitás ábrázolása a kontrollparaméter függvényében. Esetünkben a szuszceptibilitás a legnagyobb k -klikkperkolációs klaszter várható méretnövekedésének felel meg, amennyiben egy véletlenszerűen választott másik k -klikkperkolációs klaszterrel olvasztjuk össze, a klaszterek között a mérettel arányos valószínűséggel választva. Mivel a méretet itt a tartalmazott k -klikkek számában mérjük, a J indexű k -klikkperkolációs klaszter kiválasztásának valószínűsége $\mathcal{N}_J(k) / \sum_{J' \neq 1} \mathcal{N}_{J'}(k)$, és a szuszceptibilitás a

$$\chi = \sum_{J \neq 1} \frac{[\mathcal{N}_J(k)]^2}{\sum_{J' \neq 1} \mathcal{N}_{J'}(k)} \quad (2.23)$$

alakban írható fel, ahol az összegzésből a $J=1$ indexű, legnagyobb k -klikkperkolációs klasztert természetesen kihagytuk. Folytonos fázisátalakulások esetén véges rendszerméret mellett a szuszceptibilitásnak éles maximuma van a kritikus pontnál, mely a termodinamikai határesetben (végtelen rendszerméret esetén) divergál. Ennél fogva ha numerikusan szeretnénk meghatározni egy átalakulás kritikus pontját, célszerű a szuszceptibilitást is megvizsgálni, mely növekvő rendszerméret mellett általában egy egyre élesedő maximumot mutat a kritikus pont közelében. Magáról a szuszceptibilitásról az irányított k -klikkperkoláció kritikus pontjának vizsgálatánál mutatunk ábrát a 2.3. alfejezetben.

A 2.3. ábrán látható lépcsők szélessége hatványszerűen cseng le a rendszerméret függvényében, egy bizonyos α exponenssel. Ha Φ -t a $[p/p_c(k) - 1]N^\alpha$ függvényében ábrázoljuk, azaz megnyújtjuk a vízszintes tengelyt egy N^α faktorial, akkor a különböző N -ekhez tar-

tozó eredmények összeesnek egyetlen görbére, ahogy azt a 2.4a ábrán láthatjuk $k = 3, 4$, és 5 esetén. Az α exponensre $1/2$ -hez közeli érték adódott a tanulmányozott esetekben.



2.4. ábra. A rendparaméterek skálázása a kritikus pontnál. a) A 2.3a ábra esetén a lépcsők szélessége N függvényében hatványszerűen csökken, α kitevővel. Ez alapján a különböző rendszerméretekhez tartozó görbék összeejthetők, amennyiben a vízszintes tengelyt átskálázzuk egy N^α faktorial. Az így kapott eredményeket mutatjuk $k = 3, 4$ és 5 esetén. A jobb láthatóság kedvéért a $k = 4$ -hez és $k = 5$ -höz tartozó görbéket függőlegesen eltoltuk 0.4 -el illetve 0.8 -al. b) A Ψ rendparaméter kritikus pontban felvett értéke hatványszerűen cseng le N függvényében, jó egyezést mutatva a (2.25) kifejezéssel. (Az ábra forrása a [T4] publikáció.)

A klasszikus véletlen gráf élperkolációs átalakulása kapcsán Erdős Pál és Rényi Alfréd bebizonyították, hogy az $N \rightarrow \infty$ határesetben a legnagyobb komponens csúcsainak száma, N_1 , a kritikus $p = p_c \equiv 1/N$ esetén $N^{2/3}$ szerint divergál, azaz a Φ rendparaméter $N^{-1/3}$ -al skálázik [64]. Egy hasonló jelenség figyelhető meg a k -klikkperkoláció kritikus pontjánál is, amennyiben a Ψ rendparaméter viselkedését vizsgáljuk. Ha feltesszük, hogy a k -klikkszomszédsági gráf úgy viselkedik mint az Erdős–Rényi-gráf, akkor a legnagyobb komponensében található k -klikkek száma $p = p_c(k)$ esetén $[\mathcal{N}_c(k)]^{2/3}$ szerint kell skálázódjon. Behelyettesítve a (2.19) egyenletből kapott $p_c(k)$ -t a (2.22) kifejezésbe és elhagyván az N -től független tagokat az összes k -klikkek számára az

$$\mathcal{N}_c(k) \sim N^{k/2} \quad (2.24)$$

kifejezést kapjuk a kritikus pontban. Ez alapján a legnagyobb k -klikkperkolációs klaszter k -klikkjeinek száma, $\mathcal{N}_1(k)$ vezető rendben $[\mathcal{N}_c(k)]^{2/3} \sim N^{k/3}$, és ennél fogva a kritikus pontban azt várjuk, hogy a Ψ_c rendparaméter $[\mathcal{N}_c(k)]^{2/3} / \mathcal{N}_c(k) \sim N^{-k/6}$ szerint skálázik.

Ez a fajta skálázás azonban csak $k \leq 3$ esetén valósul meg. A $k > 3$ esetén tapasztalható eltérés oka a következő: A fenti eredmény szerint a legnagyobb k -klikkperkolációs klaszterben a k -klikkek száma $[\mathcal{N}_c(k)]^{2/3} \sim N^{k/3}$, ami $k > 3$ esetén gyorsabban nő, mint N . Másfelől viszont a k -klikkperkoláció kritikus pontjában a k -klikkszomszédsági gráf legnagyobb komponense (ami egyben a legnagyobb k -klikkperkolációs klaszternek felel meg), faszzerű, nagyon kevés hurokkal. Ez a faszzerű szerkezet a gyakorlatban azt jelenti, hogy a

2. FEJEZET: k -klikkperkoláció véletlen gráfokban

k -klikkszomszédsági gráf legnagyobb komponensében majdnem minden csúcs (k -klikk) egy külön csúcsnak felel meg az eredeti gráfban. Ennél fogva a k -klikkszomszédsági gráf legnagyobb komponensének mérete nem nőhet N -nél gyorsabban, csak maximum úgy, mint N . Így $k > 3$ esetén $\mathcal{N}_1(k) \sim N$, ami alapján a (2.24) egyenlet szerint $\Psi_c = \mathcal{N}_1(k) / \mathcal{N}_c(k) \sim N^{1-k/2}$. Összegezve az elméleti eredményünk a Ψ rendparaméter skálázására a k -klikkperkoláció kritikus pontjánál a következő

$$\Psi_c \sim \begin{cases} N^{-k/6} & \text{ha } k \leq 3 \\ N^{1-k/2} & \text{ha } k \geq 3 \end{cases} \quad (2.25)$$

A 2.4b ábrán a Ψ kritikus pontban felvett értékére vonatkozó numerikus eredményeket mutatjuk be, melyek jó egyezést mutatnak a fenti kifejezésekkel.

2.3. Irányított k -klikkperkoláció

A valódi rendszereket leíró hálózatok számos esetben irányítottak, azaz a csúcsok közt a kapcsolatok nem szimmetrikusak. Az adott rendszertől függően az élek iránya jelentheti valamilyen áramlás irányát (információ, energia, stb.), vagy jelölhet alá-fölé rendeltségi relációt (pl. főnök-beosztott viszonyt) a csúcsok között. Az ilyen rendszerek esetén természetes módon merül fel az a kérdés, hogy a hálózat csoportszerkezetének vizsgálatakor érdemes-e olyan csoportkeresési módszert választani, mely figyelembe veszi az élek irányát is a szokásos, irányítatlan hálózatokon definiált módszerek helyett. E gyakorlati szempont által motiválva ebben az alfejezetben definiáljuk az irányított k -klikkperkoláció fogalmát, mely lehetővé teszi irányított hálózati csoportok keresését. A módszer gyakorlati vonatkozásai a 3. fejezetben kerülnek kifejtésre, itt első sorban az elméleti oldalra koncentrálnunk, azon belül is az irányított k -klikkek perkolációjára Erdős–Rényi-gráfban.

Az alfejezetben bemutatásra kerülő eredmények a [T6] publikációban jelentek meg, a disszertáció szerzőjének első szerzőségével. Az irányított k -klikkek definíciója a cikk szerzői közötti hosszas diszkussziók során alakult ki. Az irányított k -klikkperkolációs klaszterek megkeresésére vonatkozó algoritmus kidolgozása és implementálása a disszertáció szerzőjének eredménye (ez a 3.1.2. alfejezetben kerül ismertetésre a csoportkeresésre vonatkozó eredményeknél). Az irányított k -klikkperkoláció kritikus pontjához tartozó élbekötési valószínűséget az Erdős–Rényi-gráfban szintén disszertáció szerzője vezette le analitikusan, a kapcsolódó numerikus szimulációkat Farkas Illés végezte el és értékelte ki.

2.3.1. Irányított k -klikkek

Egy irányítatlan hálózatban két csúcs (i és j) vagy össze van kötve, vagy nem, míg ezzel szemben egy irányított hálózatban többféle módon is kapcsolódhatnak egymáshoz: egyszeres éllel, úgy mint $i \rightarrow j$, vagy épp ellenkező irányban, úgy mint $i \leftarrow j$, illetve „dupla” éllel, úgy mint $i \rightleftharpoons j$. Az egyszerűség kedvéért kettőnél több élt, illetve egyetlen csúcshoz kapcsolódó hurkot nem engedünk meg a hálózatban.

Egy k csúcsból álló teljes gráf esetén a $k(k-1)/2$ kapcsolat összesen $3^{k(k-1)/2}$ módon lehet irányított. Mivel irányítatlan esetben ezeket a lehetőségeket mind egyként kezeltük,

az él irányok figyelembe vétele elvileg nagyon sokféle lehetőséget nyit meg az irányított k -klikkdefiníciók számára. Mindazonáltal egy természetes követelmény, hogy az irányított k -klikkek megőrizzenek egyfajta irányultságot egy magasabb szinten is ahelyett, hogy pusztán irányított élek halmazának feleljenek meg k csúcs között. Ennélfogva az irányított k -klikkdefiníciónk lényege, hogy egy irányított k -klikk csúcsai sorba rendezhetők úgy, hogy minden csúcsból mutat egy irányított él minden hozzá képest hátrább lévő csúcsba, illetve fordítva, minden csúcsba mutat egy él minden hozzá képest előrébb lévő csúcsból.

Mivel a dupla élek jelenléte többféle sorrendet is lehetővé tesz, ami a fenti kitételnek eleget tesz, először koncentráljunk arra az esetre, amikor csak egyszeres élekből áll a k -klikk. Ilyenkor minél előrébb van az adott csúcs a sorrendben, annál több kimenő éllel kell rendelkezzen a k -klikken belül, és ezzel párhuzamosan minél hátrább van a sorrendben, annál több bejövő éllel kell csatlakozzon a k -klikk többi tagjához. Ez alapján bevezethetjük a leszűkített ki-fokszámot, mely megadja, hogy hány él mutat az adott csúcsból a k -klikk többi tagjához, vagyis a leszűkített ki-fokszám értéke mindig nulla és $k - 1$ között mozog. Hasonló módon megadhatjuk a leszűkített be-fokszámot is, mely a k -klikk többi tagjából az adott csúcsba mutató élek számának felel meg. A csúcsok sorrendje megegyezik a leszűkített fokszámok sorrendjével: a legkisebb leszűkített be-fokszámú (és legnagyobb leszűkített ki-fokszámú) csúcs az első, a legnagyobb be-fokszámú (és legkisebb leszűkített ki-fokszámú) csúcs az utolsó az irányított k -klikken belül.

A fentiek alapján könnyű belátni, hogy egy pusztán egyszeres élekből álló k -klikk esetén az imént bevezetett irányított k -klikkdefiníciót más formában is meg lehet fogalmazni, ugyanis a következő három feltétel ekvivalens egymással [T6]:

- Bármely él a k -klikken belül egy sorrendben előrébb álló (magasabb leszűkített ki-fokszámú) csúcsból mutat egy sorrendben hátrább lévő csúcsba.
- A k -klikkben nincs irányított hurok.
- A k -klikkben minden csúcsnak más a leszűkített ki-fokszáma.

Először belátjuk, hogy a második állításból következik a harmadik, azaz ha nincs a k -klikkben irányított hurok, akkor a k -klikkben minden csúcsnak más a leszűkített ki-fokszáma. Amennyiben nincs hurok, kell lennie egy csúcsnak, amibe csak bejövő élek csatlakoznak a k -klikken belülről, hiszen ellenkező esetben végtelen sokáig lépegethetnénk a csúcsról csúcsra a k -klikken belül az élek irányát követve. Ha egyszerre megfordítjuk az összes él irányát a k -klikken belül, akkor nem hozunk létre hurkot, viszont lesz egy csúcsunk, aminek csak kimenő éle van a k -klikken belül. Ez alapján egy hurokmentes k -klikkben kell lennie egy csúcsnak, melynek csak kimenő élei vannak a k -klikken belül, azaz a leszűkített ki-fokszáma $k - 1$. Ha ezt a csúcsot kivesszük a k -klikkből, akkor létrehozunk egy hurokmentes $k - 1$ -klikket. Teljesen hasonló gondolatmenettel belátható, hogy ebben viszont kell lennie egy csúcsnak, melynek csak kimenő élei vannak a $k - 1$ -klikken belül, azaz a leszűkített ki-fokszáma $k - 2$. Mivel az eredeti k -klikkben az elvett csúcsból egy bejövő él csatlakozott ehhez a csúcsához, a leszűkített ki-fokszáma az eredeti k -klikkben is $k - 2$. Ha ezt a csúcsot is elvesszük, akkor marad egy hurokmentes $k - 2$ -klikkünk, melyre megint alkalmazhatjuk a fenti gondolatmenetet. Így szépen lépésről lépésre beláthatjuk, hogy ha nincs irányított

2. FEJEZET: k -klikkperkoláció véletlen gráfokban

hurok a k -klikkben, akkor minden leszűkített ki-fokszám előfordul $k - 1$ és nulla között, azaz minden csúcsnak más a ki-fokszáma.

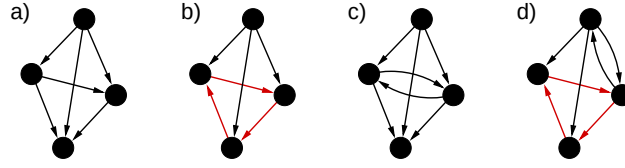
Ez a gondolatmenet azt is megmutatta, hogy a második állításból az első is következik, azaz ha nincs irányított hurok a k -klikkben, akkor minden él egy magasabb leszűkített ki-fokszámú csúcsból mutat egy alacsonyabb leszűkített ki-fokszámú csúcsba. A csúcsok képzeletbeli eltávolítását a legnagyobb leszűkített ki-fokszámú csúccsal kezdtük, e csúcs összes élére teljesül az állítás. Ezután a második legnagyobb ki-fokszámú csúccsal folytattuk, az ezen a csúcson maradó összes élre is igaz. És így tovább, ahogy megyünk végig a csúcsokon a nagyobb leszűkített ki-fokszámoktól az alacsonyabbak felé, minden élről sorban beláthatjuk, hogy igaz rá az állítás.

Ezután belátjuk, hogy a harmadik állításból következik a második, azaz ha minden csúcsnak más a leszűkített ki-fokszáma, akkor nem lehet irányított hurok a k -klikkben. Ha minden csúcsnak más a leszűkített ki-fokszáma, akkor minden lehetséges értéknek nulla és $k - 1$ között elő kell fordulnia, hiszen összesen k darab csúcs van a k -klikkben. Ez azt jelenti, hogy van egy csúcs, melynek $k - 1$ kimenő éle van a többi csúcsához, azaz már nem lehet bejövő éle egy sem, hiszen dupla éleket egyelőre nem engedünk meg. Ezen csúcs biztos nem lehet része egy irányított huroknak a k -klikken belül. Hasonló módon, a $k - 2$ a leszűkített ki-fokszámmal rendelkező csúcsnak csak egy bejövő éle van, ez viszont az imént tárgyalt, $k - 1$ leszűkített ki-fokszámú csúctól jön, ezért ez a csúcs sem lehet része egy irányított huroknak a k -klikken belül. És így tovább, ahogy végig haladunk a csúcsokon csökkenő leszűkített ki-fokszám szerint, mindegyikről belátható, hogy a bejövő élei a nagyobb leszűkített ki-fokszámú csúcsoktól jönnek, melyek nem lehetnek benne egy irányított hurokban, amiből az következik, hogy az adott csúcs sem. Így csúcsról csúcsra haladva a k -klikk összes csúcsánál kizárhatjuk, hogy része lenne egy irányított huroknak a k -klikken belül, azaz a k -klikk nem tartalmazhat irányított k -klikket.

Végül nagyon egyszerűen azt is beláthatjuk, hogy az első állításból következik a harmadik, azaz ha minden él egy magasabb leszűkített ki-fokszámú csúcsból mutat egy alacsonyabb leszűkített ki-fokszámú csúcsba, akkor minden csúcsnak más a leszűkített ki-fokszáma. Ez egyenesen következik abból, hogy ha találnánk két egyforma leszűkített ki-fokszámú csúcsot a k -klikken belül, akkor a köztük lévő élre nem lenne igaz, hogy magasabb leszűkített ki-fokszámú csúcsból mutat alacsonyabb leszűkített ki-fokszámú csúcsba.

A fenti definícióknak eleget tevő irányított k -klikkek „magasabb szintű” irányultsága természetes módon követi a csúcsok sorrendjét a csökkenő leszűkített ki-fokszám szerint. Az a csúcs, melynek csak kimenő élei vannak a k -klikk többi csúcsai felé egy fajta „forrásnak”, vagy „felső” csúcsnak tekinthető, míg a többi csúctól csak bejövő éllel rendelkező csúcs egy fajta „nyelő”, vagy „alsó” csúcs. A többi csúcs e két határeset között helyezkedik el, és ahogy haladunk az első csúctól az utolsó felé, egyre kevésbé tekinthetők a csúcsok forrásnak, viszont egyre inkább nyelővé válnak. Az irányított k -klikkdefiníciónk szemléltetése a 2.5. ábrán látható.

Fontos még kitérnünk a dupla élt is tartalmazó k -klikkek esetére is. Ilyenkor a fenti három feltétel egyike sem teljesül: A hurok jelenléte a k -klikkben triviális, hiszen maga a dupla él egy két elemű hurok. Továbbá a dupla éllel összekötött csúcspárok között az egyik élre biztosan nem fog teljesülni, hogy nagyobb leszűkített ki-fokszámú csúcsból mu-



2.5. ábra. Az irányított k -klikk definíciója. a) Egy pusztán egyszeres élekből álló irányított k -klikk $k=4$ esetén. b) Mivel az általunk definiált irányított k -klikkek nem tartalmazhatnak irányított hurkot, ezt a k -klikket nem fogadjuk el irányított k -klikknek, hiszen a pirossal jelölt élek egy hurkot képeznek. c) Egy dupla élt is tartalmazó irányított k -klikk $k=4$ esetén. (A dupla él egyik tagját elvéve egy irányított hurkoltól mentes, egyszeres élekből álló k -klikket kapunk.) d) Ebben a k -klikkben nem tudjuk úgy egyirányúvá tenni a dupla élt, hogy ne maradjon irányított hurok, ezért ez nem egy irányított k -klikk.

tat kisebb leszűkített ki-fokszámú csúcsba. Ezen felül azt is beláthatjuk, hogy dupla élt tartalmazó k -klikk esetén nem lesz minden csúcs leszűkített ki-fokszáma különböző. Egy k -klikk összes éleinek m számát felírhatjuk úgy mint

$$m = \sum_{\ell=0}^{k-1} \ell n_{\ell}, \quad (2.26)$$

ahol ℓ a lehetséges leszűkített ki-fokszámokon fut végig, az n_{ℓ} pedig megadja, hogy hány csúcs rendelkezik az adott ℓ leszűkített fokszámmal a k -klikken belül. Ha minden csúcsnak különböző a leszűkített ki-fokszáma, akkor $n_{\ell} = 1$ minden ℓ -re, és megkapjuk a csupán egyszeres éleket tartalmazó k -klikkekre vonatkozó $m = k(k-1)/2$ értéket. Viszont dupla élt is tartalmazó k -klikkek esetén $m > k(k-1)/2$, aminél fogva legalább az egyik n_{ℓ} értékének nagyobbak kell lennie, mint 1 a (2.26) egyenletben, azaz vannak olyan csúcsok, melyeknek egyforma a leszűkített ki-fokszámuk.

Ezek alapján a kérdés az, hogy miként módosítsuk a csupán egyszeres élekből álló k -klikkeknel megadott irányított k -klikkdefiníciót úgy, hogy az alkalmazható legyen dupla élt is tartalmazó esetekre is. Az általunk választott megoldás teljesen természetes és egyszerű: Irányított k -klikknek fogadjunk el minden olyan dupla éllal rendelkező k -klikket, melyből valahány él eltávolításával létre lehet hozni egy pusztán egyszeres élekből álló k -klikket, melyre már teljesülnek a korábban lefektetett feltételek. (Ezt szemlélteti a 2.5c-d ábra).

A k -klikkszomszédságot teljesen hasonlóan definiáljuk az irányítatlan esethez: két irányított k -klikk szomszédos, ha $k-1$ csúcsuk közös. Egy irányított k -klikkperkolációs klaszter ez alapján olyan irányított k -klikkek maximális uniójának felel meg, melyben bármely irányított k -klikkből eljuthatunk bármely más irányított k -klikkbe k -klikkszomszédságok sorozatán keresztül. A k -klikksablon görgetés illusztrációként teljesen ugyanúgy alkalmazható ebben az esetben is, mint azt ahogy a 2.1. alfejezetben bemutattuk az irányítatlan k -klikkperkolációs klasztereken. Az irányított k -klikkperkolációs klaszterek egy kézenfekvő irányított hálózati csoportdefiníciót nyújtanak, megközelítésünk ezen vonatkozásait bővebben a 3.1.2. alfejezetben fejtjük ki.

2.3.2. Irányított k -klikkperkoláció az Erdős–Rényi-gráfban

Az Erdős–Rényi-gráf irányított verziójában az N csúcs között összesen $N(N-1)$ különböző él lehetséges, és ezek egymástól függetlenül egységesen p valószínűséggel jönnek ténylegesen létre, átlagosan $M \simeq N(N-1)p$ élt eredményezve a rendszerben. (Az eredeti irányítatlan Erdős–Rényi-gráfban csak $N(N-1)/2$ különböző él lehetséges, ezért ugyanolyan élbekötési valószínűség esetén ott átlagosan csak $M \simeq N(N-1)p/2$ él jön létre). Az irányított k -klikkperkoláció kritikus élbekötési valószínűségét $p_c^{\text{ir}}(k)$ -val jelöljük. A $p_c^{\text{ir}}(k)$ értéke növekvő N esetén csökken, és az $N \rightarrow \infty$ határesetben nullához tart, hasonlóan mint a klasszikus élperkolációra vonatkozó $p_c = 1/N$ vagy az irányítatlan k -klikkperkoláció esetén levezetett (2.19) kifejezés.

Az egyszerűség kedvéért csak vezető rendben határozzuk meg $p_c^{\text{ir}}(k)$ -t, a nagy N határesetre koncentrálva. Tegyük fel, hogy a kritikus élbekötési valószínűség alatt vagyunk, azaz az irányított k -klikkek még nem állnak össze egy óriás perkolációs klaszterbe, ehelyett csak kisméretű, izolált klaszterekkel találkozhatunk. A k -klikksablon gördítési képben ez annak felel meg, hogy ha megpróbáljuk bejárni a hálózatot a sablon szomszédos irányított k -klikkek közötti görgetésével, akkor általában néhány gördítés után meg kell állnunk, mert nem tudunk a még fel nem tárt részek felé továbbgördülni.

Az irányított k -klikkperkoláció kritikus pontjánál azonban ebben egy változás áll be: a még fel nem fedezett szomszédos irányított k -klikkek átlagos száma eggyel válik egyenlővé, ami lehetővé teszi, hogy sok lépésen keresztül gördítsük tova a k -klikksablont. Mivel $p_c^{\text{ir}}(k)$ -t csak vezető rendben kívánjuk meghatározni, elhanyagolhatjuk azon eseteket, mikor a továbbgördülés dupla élek használatával történik: Minimum $k-1$ él jelenléte szükséges ahhoz, hogy az adott irányított k -klikkből tovább tudjunk lépni a következő (szomszédos) irányított k -klikkbe. Mivel az élek egymástól függetlenek, ennek valószínűsége arányos p^{k-1} -nel. Bár alapvetően nem tiltott, hogy ezen élek között legyenek dupla élek is, minden egyes további él jelenléte egy további p szorzó faktort jelent. Ez alapján annak valószínűsége, hogy egy dupla élt is felhasználunk a továbbgördülés során p -szer kisebb, mint annak, hogy pusztán egyszeres élek segítségével gördülünk tovább. Hasonló módon, két dupla él felhasználásának valószínűsége már p^2 -szer kisebb, stb.

Továbbgördülés esetén a sablon $k-1$ csúcsát választhatjuk ki arra, hogy áthelyezzük, és durván N csúcs jöhet számításba mint az áthelyezés célpontja. Ha nem lenne semmilyen megkötés az élek irányára vonatkozóan, akkor ezen új csúcs és az aktuális irányított k -klikk között húzódó $k-1$ él mindegyike tetszőleges irányba mutathatna, összesen 2^{k-1} különféle módon. Azonban az irányított k -klikk, melybe átgördülünk, szintén eleget kell tegyen a korábban lefektetett kritériumoknak, ami jelentősen csökkenti a szóba jövő konfigurációk számát. Az új csúcs (melybe a sablon kiválasztott csúcsa átgördül), k különféle helyet foglalhat el az új irányított k -klikk csúcsainak sorrendjében (a leszűkített ki- vagy befokszám szerint). Ezzel szemben a két szomszédos irányított k -klikk közt közös $k-1$ csúcs sorrendjét már rögzítette az előző irányított k -klikk (és ebbe a sorrendbe kell „beilleszteni” az új csúcsot). Az új csúcs sorrendben elfoglalt helyének rögzítése egyben rögzíti az élek irányát is, vagyis azt kaptuk, hogy továbbgördülés esetén csupán csak k -féle konfigurációja lehet az élírányoknak. Az eddig számba vett faktorokat összegyűjtve a következő egyenletre

juthatunk [T6]:

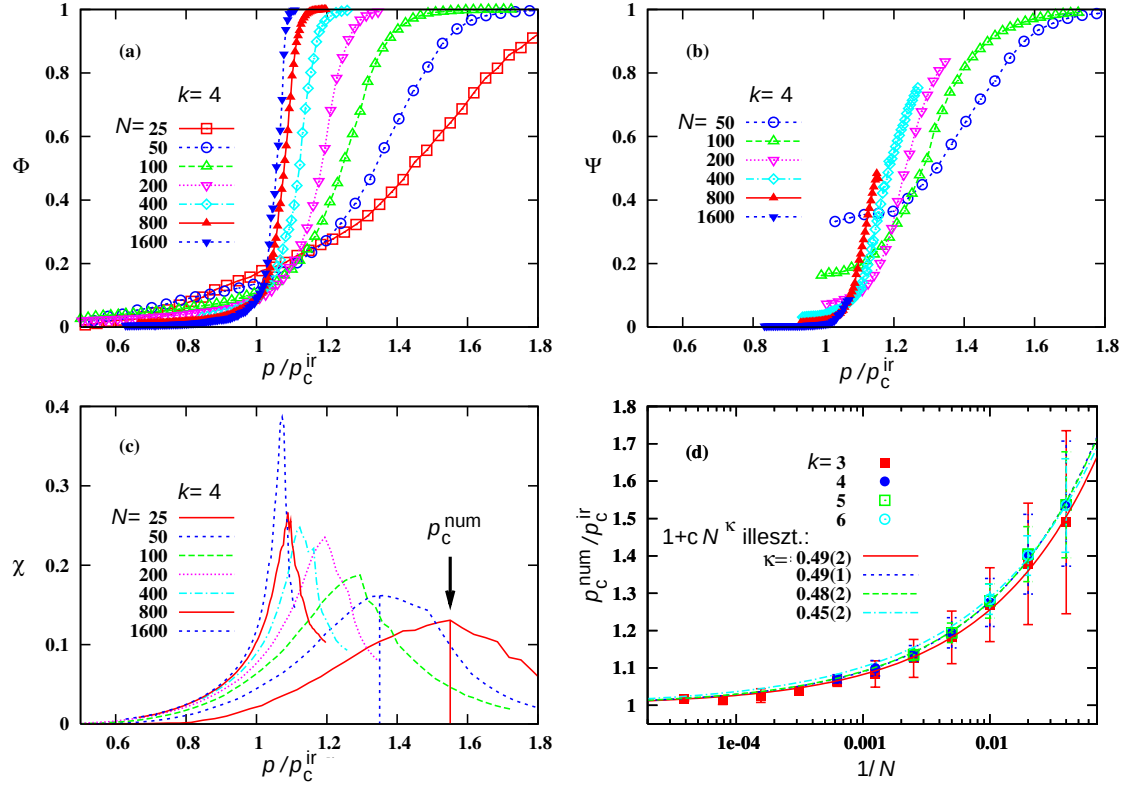
$$[p_c^{\text{ir}}(k)]^{k-1} N(k-1)k = 1, \quad (2.27)$$

ami alapján az irányított k -klikkperkoláció kritikus élbekötési valószínűsége vezető rendben

$$p_c^{\text{ir}}(k) = [Nk(k-1)]^{-1/(k-1)} = p_c(k)/k^{1/(k-1)}. \quad (2.28)$$

A $k=2$ határesetben $p_c^{\text{ir}}(k=2) = p_c/2$ (ahol itt p_c a jobb oldalon a klasszikus élperkoláció $p_c = 1/N$ kritikus küszöb értéke), ami teljesen konzisztens azzal a megállapítással, hogy az élek átlagos száma kétszer annyi az irányított Erdős–Rényi-gráf esetében az irányítatlan esethez képest azonos élbekötési valószínűség mellett.

Az irányítatlan esethez hasonlóan az átalakulás rendparaméterének a (2.20)–(2.21) egyenletekben definiált Φ vagy Ψ mennyiségeket választhatjuk. A 2.6a-b ábrákon a numerikus szimulációkból kapott Φ illetve Ψ értékeket mutatjuk p/p_c^{ir} függvényében $k=4$ esetén, $N=50$ és $N=1600$ között változó rendszerméret mellett. Hasonlóan az irányítatlan esethez, Φ egy lépcsőfüggvényhez tart növekvő rendszerméret mellett, míg Ψ ehelyett egy, a kritikus p_c^{ir} felett nulla és egy között folytonosan növekvő függvényhez konvergál. A 2.6c ábrán a Ψ rendparaméterhez tartozó χ szuszceptibilitást tüntettük fel, melyet az irányítatlan esethez hasonlóan a (2.23) egyenlet definiál. Látható, hogy növekvő rendszerméret esetén χ egy egyre élesedő maximumot mutat, mely ezzel párhuzamosan egyre közelebb kerül a $p/p_c^{\text{ir}} = 1$ ponthoz. A χ maximumhelyét kezelhetjük úgy, mint a kritikus élbekötési valószínűségre vonatkozó numerikus eredményt. A 2.6d ábra szerint a maximumhelyből kapott p_c^{num} numerikus eredmény, illetve a (2.28) egyenletben levezetett elméleti p_c^{ir} növekvő rendszerméret esetén egymáshoz tart, hiszen $p_c^{\text{num}}/p_c^{\text{ir}} \simeq 1 + cN^{-\kappa}$, ahol c egy konstans és $\kappa \simeq 1/2$.

2. FEJEZET: k -klikkperkoláció véletlen gráfokban

2.6. ábra. Az irányított k -klikkperkoláció numerikus vizsgálata Erdős–Rényi-gráfban. A görbék minden esetben több szimulációra kapott átlagértékeket mutatnak, a szimulációk száma a rendszermérettől függően 4 és 100 között mozgott. a) A Φ rendparaméter (a legnagyobb k -klikkperkolációs klaszter csúcsainak száma osztva N -nel), a $p/p_c(k)^{\text{ir}}$ függvényében, ahol $p_c(k)^{\text{ir}}$ -t a (2.28) összefüggés alapján számítottuk ki. b) A Ψ rendparaméter (a legnagyobb k -klikkperkolációs klaszter k -klikkjeinek száma osztva az összes k -klikkek számával), a $p/p_c(k)^{\text{ir}}$ függvényében. c) A (2.23) egyenletben definiált szuszceptibilitás a $p/p_c(k)^{\text{ir}}$ függvényében. Ennek maximumhelye tekinthető a kritikus élbekötési valószínűség numerikus eredményének. d) A szuszceptibilitás maximumhelyéből kapott numerikus p_c^{num} és a (2.28) egyenletben megadott elméleti $p_c(k)^{\text{ir}}$ hányadosa az inverzrendszerméret, $1/N$ függvényében. Látható, hogy nagy rendszerméret esetén ez a hányados 1-hez tart. (Az ábra forrása a [T6] publikáció).

3. fejezet

Hálózati csoportkeresés k -klikkperkolációval

Amint azt a 2. fejezetben említettük, a k -klikkperkolációs klaszterek feltárása egyben egy természetes csoportkeresési módszert is nyújt hálózatokban. Ennek a megközelítésnek a különlegességét az adja, hogy a feltárt csoportok egymással át is fedhetnek. A hálózati csoportok (más néven modulok, csoportosulások, klaszterek) sűrű, erősen összekötött részgráfoknak felelnek meg egy hálózatban [187, 190, 68, 111, 81, 147, 182, T5], mint például egy család vagy baráti kör az emberi kapcsolatok hálózatában [187, 217, T7], vagy azonos témakörhöz tartozó weblapok sűrűn egymásra linkelt csoportja a Világhálón [T6]. Hasonló egységek előfordulnak fehérje-kölcsönhatási hálózatokban (pl. fehérjefunkciós csoportok) [175, 195], pénzügyi illetve gazdasági hálózatokban (pl. ipari szektorok) [157, 96], valamint játékelméleti modellekben is [200, 209, 199]. A csoportok feltérképezése nagyban segítheti az említett hálózatok szerkezetének, működésének alaposabb megismerését, mélyebb megértését. A fejezet bevezetéseként röviden áttekintjük a hálózati csoportkeresés legfontosabb korábbi eredményeit.

A hálózati csoportok gyakorlati fontossága miatt rendkívül sokféle csoportkereső módszer született az utóbbi egy évtizedben (egy részletes, átfogó leírását a területnek S. Fortunato adta meg a [72] publikációban), és a csoportoknak a mai napig nincs egy általánosan elfogadott, egyértelmű matematikai definíciója. A témakörnek messzire nyúló gyökerei vannak a gráfelméletben, hiszen például a klasszikus minimális elvágó csúcshalmaz vagy elvágó élhalmaz megtalálásának problémája is tekinthető egyfajta csoportkeresésnek. Ennél a problémánál a cél azon minimális számú csúcs, illetve él megtalálása, melyek elvételével a gráf szétesik két egyforma méretű izolált komponensre. Talán a legismertebb gyakorlati módszerek erre a Kernigham–Lin-algoritmus [109] illetve a spektrális biszekció [17], melyeket például rekurzívan alkalmazva egyre kisebb és kisebb részekre lehet vágni egy adott hálózatot, minden egyes vágásnál törekedvén a lehető legkevesebb él eltávolítására. Ezen algoritmusokat, melyek az 1970-es illetve 80-as években jelentek meg, olyan gyakorlati problémák is motiválták, mint például az, hogy miként lehet egy sok összetevőből álló nyomtatott áramkört szétosztani több lapra úgy, hogy a lapok között a lehető legkevesebb vezeték húzódjon.

3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval

A hálózati csoportkeresésre tekinthetünk úgy is mint az általános adatklaszterezés problémakör egyik speciális esetére. Ebben a kontextusban természetesen felmerülnek olyan széles körben alkalmazott módszerek is, mint például a hierarchikus klaszterezés [94] vagy a k -középklaszterezés [126]. E módszerek könnyűszerrel adaptálhatóak a hálózati csoportkeresés problémájára, amennyiben definiálunk egy hasonlóság- vagy távolságmértéket a csúcsok között. Egy egyszerű megoldás erre például a szomszédsági mátrix¹ két megfelelő sorának koszinusz hasonlóságát vagy korrelációs együtthatóját használni. Azonban a tapasztalatok szerint egy ilyen fajta megközelítés általában gyenge minőségű hálózati csoportokat ad, például a hierarchikus klaszterezés esetén gyakran előfordul, hogy két szomszédos csoport „belső magja” hamarabb összeolvad egymással, minthogy az egyes magokhoz hozzákapcsolódnának az adott csoportok periférikus tagjai.

Az első komoly áttörés a hálózati csoportkeresés területén Michelle Girvan és Mark E. J. Newman nevéhez kötődik [81]. Az általuk bevezetett algoritmus egyik fő motivációja pont a fent említett, túl korai csoportmag összeolvadás kiküszöbölése volt. A megközelítés lényege, hogy a csoportokat az élek sorozatos eltávolításával állítjuk elő a hálózathoz, ahol az élek sorrendjét a köztesség² határozza meg. Az egyszerűség kedvéért tegyük fel, hogy a vizsgált hálózat egyetlen összefüggő komponensből áll. Az említett, köztességen alapuló éleltávolítási folyamat ezt előbb-utóbb két izolált részgráfra bontja, ilyenkor az eredményül kapott két komponensen külön-külön rekurzívan folytatjuk az élek eltávolítását, míg a folyamat végén minden csúcs izolálttá nem válik. A hierarchikus klaszterezéshez hasonlóan így itt is előáll egy csoportdendrogram (melynek gyökere az egész hálózat, levelei pedig az egyes csúcsok), azonban jelen esetben ez „felülről lefelé” történik. Az egyetlen nyitott kérdés ezek után az, hogy a kapott dendrogram melyik szintje felel meg az optimális csoportfelosztásnak?

Girvan és Newman egy nagyon frappáns válasszal álltak elő a modularitás mennyiségének bevezetésével [153]. A modularitás egy adott csoportfelbontás minőségét hivatott kvantitatíven megadni, pusztán a hálózati topológia (és természetesen a csúcsok csoportbesorolása) alapján. Lényege, hogy a csoportokon belüli élsűrűséget egy olyan referenciamodellnél kapható csoporton belüli élsűrűséggel hasonlítjuk össze, melynél az éleket véletlenszerűen kötögetjük át a hálózatban a csúcsok fokszámának megtartásával. A modularitás képletét a

$$Q = \sum_J Q_J = \sum_J \frac{l_J}{M} - \left[\sum_{i \in J} \frac{d_i}{2M} \right]^2 \quad (3.1)$$

alakban lehet megadni, ahol J a csoportokon fut végig, Q_J a J csoport modularitása, l_J a csoport belső (csoporttagok közt húzódó) éleinek száma, d_i az i csoporttag fokszáma, M pedig a hálózatban található élek teljes száma. A Q mennyiséget (némi más formában, de ekvivalens módon) Girvan és Newman elsőként arra javasolták, hogy a köztességen alapuló éleltávolítás során előálló dendrogramban megkeressük a legjobb csoportfelbontást adó szintet, mely esetén Q maximális [153]. Néhány valós rendszert reprezentáló hálózaton,

¹Egy hálózat vagy gráf $\mathbf{A}$ szomszédsági mátrixa egy négyzetes mátrix, melyben minden sor és oszlop egy-egy csúcsnak felel meg. A mátrixelemek a szomszédsági viszonyokat tükrözik, azaz $A_{ij} = 1$, ha az i csúcsból mutat egy él a j csúcsba, és $A_{ij} = 0$ egyébként.

²Egy él köztessége a rajta áthaladó legrövidebb utak számával arányos.

valamint számítógéppel generált véletlen gráfokon végzett tesztek alapján így lényegesen jobb csoportfelbontást kapunk, mint a hierarchikus klaszterezés révén. (A „jobb csoportfelbontás” alatt itt azt kell érteni, hogy a számítógépes hálózatok esetén a generálásnál „beplántált” csoportokkal sokkal jobb egyezést mutattak a eredmények, valamint egyes valószínű hálózatok esetén szintén ismert volt a „helyes” csoportfelbontás, melyekkel megint csak jobb egyezést mutattak a Girvan–Newman-algoritmus eredményei).

A modularitás használata gyorsan nagyon nagy népszerűsége tett szert, és a mai napig a legelfogadottabb csoportminőség mérésére szolgáló mennyiségnek számít. Ennek kapcsán természetesen adódik az az ötlet, hogy lehetne csoportkeresést folytatni a modularitás direkt maximalizálásával is az éleltávolításon alapuló Girvan–Newman-algoritmus helyett. A modularitás egzakt maximumának megkeresése egy NP-teljes probléma (ennek bizonyítása Ulrik Brandes és munkatársai nevéhez fűződik [36]), ezért a gyakorlatban csak heurisztikus módszerek jöhetnek szóba, melyek célja egy viszonylag jó „lokális” maximum megtalálása lehetőleg minél gyorsabban. Erre számos alternatív megközelítést dolgoztak ki az évek során, itt csak a legismertebb módszereket soroljuk fel.

Az első modularitás maximalizáló algoritmus szintén Mark E. J. Newman nevéhez fűződik [148], mely egy egyszerű „mohó” eljárás, ami viszonylag gyors, viszont meglehetősen pontatlan, az általa talált lokális modularitásmaximumoknál többnyire lényegesen jobbak adódnak például szimulált hőkezelés útján, melyet először Roger Guimerá és Luis A. Nunes Amaral használtak csoportkeresésre metabolikus hálózatokban [91]. További, mindenképp említésre méltó modularitásmaximalizáló módszerek még a Jordi Duch és Alex Arenas által bevezetett extrém optimalizáció [58], a rendkívül gyors (és ezáltal nagyon nagy hálózatok elemzését lehetővé tevő) LOUVAIN-algoritmus³ [27], valamint a spektrális optimalizáció [150, 149, 212, 179, 197], ahol többnyire egy ún. modularitásmátrixra kell rekurzívan alkalmazni a spektrális biszekció módszerét.

Összességében elmondhatjuk, hogy a modularitásmaximalizálás manapság az egyik legnépszerűbb csoportkeresési megközelítés a hálózatkutatásban. Azonban fontos leszögezni azt is, hogy ennek is megvannak a maga korlátai. Az egyik jelentős hátránya a modularitásmaximalizáló módszereknek az ún. felbontási határ problémájában rejlik, melyre Santo Fortunato és Marc Barthélemy hívták fel a figyelmet [73]. Az effektus lényege, hogy egy összefüggő hálózatban $\mathcal{O}(\sqrt{M})$ -nél kevesebb élből álló csoportokat nem tudunk feltárni (ahol M a teljes hálózat élszáma). A másik nagy hátránya a modularitás maximalizálásán alapuló módszereknek, hogy csak izolált csoportokat adnak és nem engedik meg a csoportok közti átfedéseket. Amint látni fogjuk, az általunk javasolt k -klikkperkolációs csoportok egy természetes megoldást jelentenek ezekre a problémákra.

A modularitásmaximalizálás fent említett hátulütői miatt számos alternatív csoportkeresési módszer került bevezetésre, melyek egyáltalán nem, vagy csak részben függenek a modularitástól. Ezek közül kiemelkednek az INFOMAP [185] és az OSLOM [118] algoritmusok: az előbbi egy információelméleti megközelítésen alapuló módszer, mely képes

³Ennél a módszernél az egyes csoportok modularitását lokálisan próbáljuk növelni a környező csúcsok véletlen sorrendben történő hozzáadásával vagy tagok elvételével, majd minden csoportot egy-egy csúccsal helyettesítünk, és az így kapott (sokkal kisebb méretű) hálózatban rekurzívan folytatjuk a modularitás lokális maximalizálását.

3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval

akár hierarchikusan egymásba ágyazott csoportokat is feltárni, míg az utóbbi egy lokális csoportfitness-maximalizáláson alapuló módszer. Egy további érdekes irányvonalat képviselnek a sztochasztikus blokkmodelleken alapuló módszerek [50, 161, 164, 165, 166, 4], melyeknél egy „belül” sűrű, és egymáshoz csak gyengébben kapcsolódó blokkokból álló gráfmodell paramétereit illesztjük a vizsgált hálózathoz, és a legjobb illesztés alapján következtetünk a hálózat csoportszerkezetére.

Ezekén felül fontos megemlíteni az ún. dinamikus módszereket is, melyeknél q lehetséges állapottal rendelkező spineket rendelünk hozzá a csúcsokhoz, és a Potts-modelhez [221] hasonló dinamikát tételezve fel a szomszédok között a spinek rendeződése alapján határoljuk be a csoportokat. Az első, ebbe az irányba mutató módszer Marcelo Blatt és munkatársai nevéhez fűződik [26], mely egy általános adatklaszterező eljárásnak felel meg. A kifejezetten a hálózati csoportkeresés problémájára fókuszáló dinamikus módszert Jörg Reichardt és Stefan Bornholdt vezették be [177], a következő Hamilton-függvény segítségével:

$$\mathcal{H} = -\mathcal{J} \sum_{i,j} A_{ij} \delta(\sigma_i, \sigma_j) + \gamma \sum_{s=1}^q \frac{n_s(n_s - 1)}{2}, \quad (3.2)$$

ahol A_{ij} a hálózat szomszédsági mátrixa, σ_i az i -edik csúcson lévő spin állapota, δ a Kronecker-delta, n_s az s -edik állapotban lévő spinek száma, $\mathcal{J}$ a ferromágneses csatolás erőssége a spinek között, γ pedig egy további paraméter. A megközelítés lényege, hogy a csoportokat a rendszer alapállapotában az azonos irányba mutató spinek által kijelölt csúcsoknak feleltetjük meg. A 3.2 egyenletben megadott Hamilton-függvény első tagjának szempontjából az a kedvező, ha minél több spin mutat ugyanabba az irányba, míg a második tag ezzel ellentétesen a spinek egyenletes elosztását preferálja a lehetséges állapotok között. Az, hogy az alapállapotban mekkora (és hány darab) csoportot találunk, a hálózati topológia mellett a $\mathcal{J}/\gamma$ aránytól függ. Természetesen idővel több eltérő Hamilton-függvény használata is felmerült [103, 193], és egy későbbi cikkben Reichardt és Bornholdt azt is megmutatták, hogy megfelelő Hamilton-függvény választásával a spinrendszer alapállapotának megkeresése ekvivalensé tehető a modularitás maximalizálásával [178].

A továbbiakban az általunk bevezetett, k -klikkperkoláción alapuló csoportkeresési módszer részleteit és legfontosabb eredményeit mutatjuk be. Ennek a megközelítésnek a legfontosabb újítása az volt, hogy olyan hálózatokban is lehetővé tette a csoportok feltárását, melyeknél nagyon jelentősek a csoportátfedések. Nem sokkal korábban már megjelent az irodalomban a Girvan–Newman-algoritmus egy olyan módosítása, mely esetenként megengedi, hogy egy csúcs egyszerre több csoportnak is tagja legyen [220]. Azonban ez a módszer csak akkor működik megbízhatóan, ha az ilyen csúcsok ritkák, és a csoportok döntő többsége izolált. Ezzel szemben a k -klikkperkoláció olyankor is jó eredményeket ad, ha a csoportok túlnyomó része átfed más csoportokkal, és a csúcsok jelentős része szerepel csoportátfedésben. Ennek egyik hozadéka, hogy megnyílik az út olyan, korábban nem tanulmányozott statisztikák vizsgálata előtt, melyek nem értelmese (vagy triviálisak) akkor, ha csak elvétve fordulnak elő csoportátfedések. (Ezen statisztikák viselkedését valós rendszereket reprezentáló hálózatok esetén a 3.3.2. alfejezetben mutatjuk be).

A k -klikkperkolációs módszer hamar nagy népszerűsége tett szert, és a mai napig az egyik meghatározó csoportkeresési módszernek számít olyan esetekben, ahol átfedő csoport-

tokra szeretnénk bontani a vizsgált hálózatot. Természetesen idővel több további alternatív módszer is megjelent az irodalomban, melyek szintén az átfedő csoportkeresés problémájára keresték a megoldást. Az egyik kézenfekvő ötlet ezzel kapcsolatban a modularitás általánosítása átfedő csoportokra, amire több eltérő javaslat is született [156, 189, 46, 119]. Mindenesetre az így bevezetett mennyiségek használhatók átfedő csoportok keresésére a modularitásmaximalizálás megközelítésében. Egy másik lehetőséget kínál a Y. Y. Ahn és munkatársai által javasolt élklasszterezésen alapuló algoritmus [3], ahol a hálózat éleit soroljuk be diszjunkt csoportokba, megengedvén, hogy a csúcsok viszont egyszerre több csoportban is részt vegyenek. Egy érdekes további alternatívát nyújt a Csermely Péter vezette csoport által kifejlesztett ModuLand módszer család, melyeknél az átfedő csoportok bizonyos centralitásfüggvények⁴ lokális maximumhelyei körül alakulnak ki [114, 202]. Ezek mellett fontos megemlíteni, hogy az INFOMAP módszernek is született átfedéseket megengedő verziója [65], az OSLOM módszer is képes a csoportátfedések kezelésére [118], valamint a Nepusz Tamás és munkatársai által javasolt fuzzy-klaszterező eljárásnál is megengedett, hogy egy csúcs egyszerre több csoportban vegyen részt [142, 141]. Az átfedő csoportok feltárását megcélzó módszerekről egy áttekintő jellegű publikációt Jierui Xie és munkatársai jelentettek meg [223].

A fejezet a disszertáció szerzőjének első szerzőségével megjelent *Nature* és *New Journal of Physics* cikkeken alapszik [T5, T6], illetve az anyag jelentős része szerepel a k -klikkperkolációval kapcsolatos eredményeket összegyűjtő könyvfejezetben is [S2]. Amint már a 2. fejezetben említettük, a k -klikkperkoláció alapötlete a [T5] publikáció szerzői (Palla Gergely, Derényi Imre, Farkas Illés és Vicsek Tamás) közt zajló diszkussziók eredménye, melyeken a hálózati csoportkereséssel kapcsolatos alkalmazások kezdetektől fogva a legfontosabb szempontként szerepeltek. A később bevezetésre kerülő, irányított k -klikkperkolációs csoportok megalapozásában a fent említett kutatók mellett már Pollner Péter is aktívan részt vett. A több szálon folyó kutatásokat Vicsek Tamás irányította és koordinálta, magát a csoportkereső (k -klikkperkolációs klaszterkinyerő) algoritmust irányítatlan esetre a disszertáció szerzője és Derényi Imre közösen dolgozták ki. Az irányított k -klikkperkolációra vonatkozó csoportkereső algoritmust a disszertáció szerzője adta meg. Az említett két módszer implementálása, a vizsgált hálózatokon való lefuttatása, az eredmények kiértékelése és a különböző csoportstatisztikák elkészítése szintén a disszertáció szerzőjének feladata volt. A vizsgálatokat megelőzően a hálózatos adatbázisokat Farkas Illés készítette elő.

3.1. A k -klikkperkoláció mint átfedő csoportkeresési módszer

A fejezet bevezetőjében említettük, hogy a hálózati csoportokat általában sűrű, erősen összekötött részgráfoknak szokás megfeleltetni. Ezek megkereséséhez a maximális klikkek jó kiindulási pontnak tűnnek, hiszen lokálisan annál sűrűbb már nem lehet egy hálózat, mint egy klikk belsejében. Természetesen ha csak egy bizonyos méret feletti maximális

⁴A centralitási jellemzők azt hivatottak mérni, hogy egy adott csúcs mennyire tölt be központi szerepet a hálózatban. Az egyik legegyszerűbb centralitás maga a fokszám, de ilyen például a korábban említett köztesség is.

3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval

klikkekre szorítjuk le a csoportok fogalmát, az egy túl erős megkötésen alapuló definíció lenne a legtöbb valódi rendszert reprezentáló hálózat esetén.

Egy rugalmas és megengedőbb csoportdefiníciót kínálnak a k -klikkperkolációs klaszterek (egy adott k érték mellett), melyeknél már nem kell mindenkinek mindenkivel közvetlen összeköttetésben lenni egy csoporton belül, viszont továbbra is „garantált” sűrűségű részegységekből épülnek fel a csoportok. Ennek a csoportdefiníciónak a legnagyobb előnye az, hogy megengedi a csoportok közötti átfedést, azaz egy csúcs egyszerre akár több csoportnak is tagja lehet. Ez az aspektus különösen fontos az emberi kapcsolatok hálózatának elemzésekor, valamint biológiai hálózatok esetén, hiszen általában egy adott személy esetén a család, baráti kör, munkahelyi közösség mind külön csoportot képeznek, és hasonlóan számos olyan fehérje ismert, mely több funkciós csoportban is szerepet játszik. Egy további előny az, hogy a csoportdefiníció lokális⁵, emiatt például nem lépnek fel felbontási határ problémák. Végezetül amennyiben irányított hálózatot kívánunk elemezni, áttérhetünk az irányított k -klikkperkolációs klaszterekre, melyek átfedő, irányított csoportokat eredményeznek.

3.1.1. A csoportkeresési algoritmus

A k -klikkkeresés egy gráfban polinomiális komplexitású probléma, ezért egy algoritmus, mely megkeresi a vizsgált hálózat összes k -klikkjét, majd azokból felépíti a k -klikkperkolációs klasztereket szintén garantáltan polinomiális komplexitású. Azonban ha a hálózat tartalmaz nagy méretű maximális klikkeket, akkor azokban exponenciálisan sok kis méretű k -klikk található, ezért a gyakorlatban az összes k -klikk egyenkénti feltárása, eltárolása sokszor nem kivitelezhető reális időkorlát- és tárolókapacitás mellett. Szerencsére erre nincs is feltétlen szükség, hiszen amennyiben ismerjük a maximális klikkeket, azok alapján a k -klikkperkolációs klaszterek egyszerűen felépíthetők.

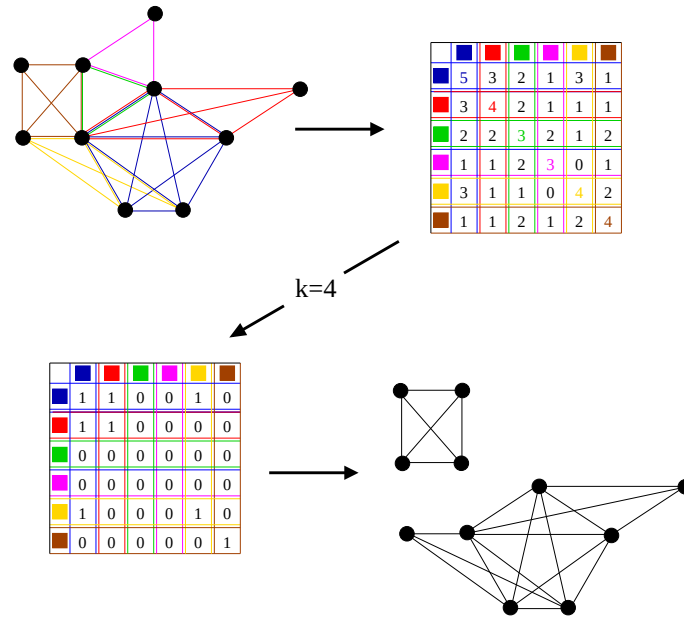
Ennek révén az általunk a [T5] publikáció kiegészítő anyagában közölt csoportkeresési algoritmus a maximális klikkeken alapszik. (Az algoritmust Derényi Imre és a disszertáció szerzője közösen dolgozták ki, a módszer implementálását teljes egészében a disszertáció szerzője végezte el). A maximális klikkek megkeresése egy közismerten NP-teljes probléma, ennél fogva ez a megközelítés egy exponenciális komplexitású algoritmust eredményez, ami elsőre nem tűnik előnyösnek gyakorlati szempontból. Szerencsére a heurisztikus maximálisklikk-kereső eljárások a legtöbb valós hálózaton meglehetősen gyorsak, ezért ez ritkán okoz tényleges problémát az alkalmazások terén.

A k -klikkperkolációs csoportok feltárásának első lépése tehát az adott hálózat összes maximális klikkjének megkeresése. Az általunk erre használt módszer lényegében a Bron–Kerbosch-algoritmus [38] egy saját implementációjú verziója, mely nemcsak a maximális klikkeket, hanem azok egymással való átfedéseit is eltárolja. A csoportokat ezen átfedé-

⁵Egy lokális csoportdefiníció pontosan előírja, hogy egy adott részgráf milyen feltételek teljesülése mellett tekinthető csoportnak. Ezzel szemben a globális módszerek általában egy „globális” paramétert próbálnak optimalizálni, és amit az eljárás eredményeként kapunk, azok a csoportok (ilyen pl. a modularitásmaximalizálás). Ennek hátránya, hogy a hálózat kis lokális perturbációja (mint pl. egy él kivágása), akár a zavar környezetétől nagyon távol is megváltoztathatja a csoportokat (ami egy lokális módszer esetén viszont nem fordulhat elő).

3.1. A k -klikkperkoláció mint átfedő csoportkeresési módszer

sek révén lehet felépíteni a következő módon: Adott k esetén minden legalább k méretű maximális klikk része egy (és csakis egy), k -klikkperkolációs klaszternek. Amennyiben két maximális klikk legalább $k - 1$ csúcsban átfed egymással, úgy egyazon k -klikkperkolációs klaszterhez tartoznak, hiszen az átfedő részen át tudunk „gördülni” az egyik maximális klikkből a másikba a k -klikkszomszédtság definíciója alapján. Ennek alapján a csoportokat megkaphatjuk úgy, hogy definiálunk egy hálózatot a maximális klikkek között, melyben akkor köt össze két maximális klikket, ha átfedésük legalább $k - 1$ csúcsot tartalmaz. A k -klikkperkolációs klaszterek ezen hálózat összefüggő komponenseinek felelnek meg.

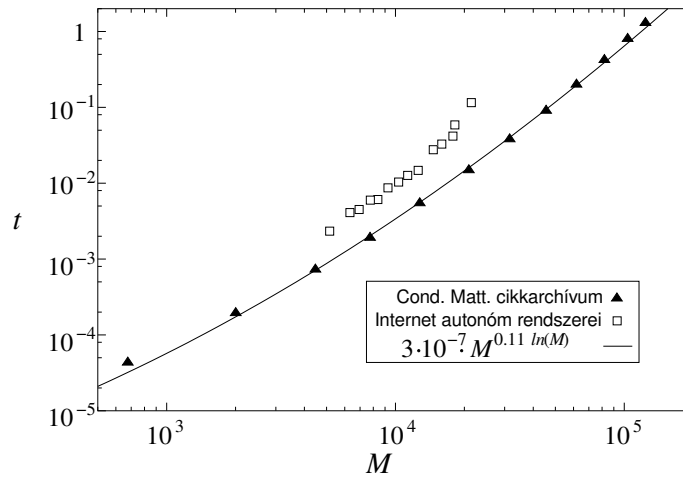


3.1. ábra. A k -klikkperkolációs csoportkeresési algoritmus szemléltetése. Az ábra bal felső részén látható, 10 csúcsból álló hálózat csoportjainak feltárását mutatjuk be $k = 4$ esetén. A hálózat összesen 6 maximális klikket tartalmaz, melyeket a különböző színek jelölnek, a maximális klikkek átfedési mátrixa az ábra jobb felső részén látható. Ha töröljük belőle a $k = 4$ -nél kisebb diagonális, valamint a $k - 1 = 3$ -nál kisebb nemdiagonális elemeket, megkapjuk a $k = 4$ -hez tartozó csoportokat megadó maximálisklikk-gráf szomszédsgái mátrixát, ahogy azt az ábra bal alsó részén láthatjuk. Ennek összefüggő komponensei adják a keresett csoportokat, ami alapján két (egymással átfedő) csoportot kapunk, amint azt az ábra jobb alsó része mutatja. (Az ábra forrása a [T5] publikáció kiegészítő része.)

A fenti eljárást szemlélteti a 3.1. ábra, melyen egy kis méretű hálózatban tárjuk fel a csoportokat $k = 4$ esetén. A hálózatból először kinyerjük a maximális klikkek átfedési mátrixát, melyben a diagonális elemek az adott maximális klikk méretét adják meg, a nemdiagonális elemek pedig a maximális klikkek átfedésének méretét mutatják. (Ennélfogva a mátrix szimmetrikus). Ebből egy adott k -hoz tartozó csoportokat úgy lehet megkapni, hogy a k -nál kisebb diagonális elemeket és a $k - 1$ -nél kisebb nemdiagonális elemeket töröljük, majd a megmaradó elemeket 1-gyel helyettesítjük, és előáll azon maximálisklikk-gráf

3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval

szomszédsági mátrixa, melynek összefüggő komponensei kiadják a keresett csoportokat.



3.2. ábra. A k -klikkperkolációs csoportkeresési algoritmus futási ideje. A futási időt órában mértük és a vizsgált hálózat élszámának függvényében ábrázoltuk, az eredmények egy 2005-ben egy átlagosnak számító PC-n születtek. A fehér négyzetek az Internet autonóm rendszerei közötti hálózaton kapott futtatási időket jelölik, míg a sötét háromszögek a Cornell Egyetem könyvtárának internetes Condensed Matter cikkarchívum társzerzőségi hálózatára vonatkoznak. Ez utóbbi rendszer esetén az eredményekre egy $t = aM^{b \ln(M)}$ alakú függvényt illesztettünk (folytonos vonal), az illesztési paraméterekre $a = 3 \cdot 10^{-7}$ és $b = 0.11$ adódtak. (Az ábra forrása a [T5] publikáció kiegészítő része).

A 3.2. ábrán láthatjuk a csoportkeresési algoritmus futási idejét két valós hálózatra abban az esetben, mikor a hálózatban található élek száma fokozatosan növekedett. Mind a Cornell Egyetem könyvtárának internetes Condensed Matter cikkarchívumán [213] alapuló társszerzőségi hálózat esetén, mind az Internet autonóm rendszereit leíró hálózat⁶ esetén a növekvő élszám az adott rendszer időbeli fejlődésének következménye, azaz a különböző adatpontok az adott rendszer által különböző időpontokban mutatott hálózatokon kapott eredményeknek felelnek meg. A 3.2. ábrán a $k = 6$ -os k -klikkméret esetén tapasztalt futási időket mutatja órában mérve. Látható, hogy a növekvő élszám növekvő futási időt eredményez és az adatokra viszonylag jól illeszkedik a $t = aM^{b \ln(M)}$ függvény, ahol t a futási idő, M az élek száma, az a és b illesztési paraméterek. Megjegyezzük, hogy a teljes társszerzőségi hálózatra (kb. 127 000 él) minden lehetséges k érték figyelembevételével kevesebb, mint 2 óra alatt lefutott az algoritmus egy 2005-ben átlagosnak számító PC-n, illetve születtek eredmények k -klikkperkolációs csoportokról többmilliós méretű hálózatokban is, amint ezt a 4. fejezetben bemutatjuk.

Érdemes megemlíteni, hogy a kutatócsoport által közölt publikációk nyomán Jussi M. Kumpula és munkatársai egy alternatív, a k -klikkek direkt feltárásán alapuló eljárást javasoltak a k -klikkperkolációs csoportok megkeresésére [116]. Ennek a módszernek előnye,

⁶Az Internet autonóm rendszereit leíró időfejlődő hálózat adatait a <http://www.cosin.org/extra/data/internet/nlanr.html> oldalról töltöttük le.

hogy könnyűvé teszi a csoportok felépítéséhez használható k -klikkek szűrését valamilyen kritérium (pl. súly) szerint. A hátránya viszont az (amint azt az alfejezet elején is említettük), hogy a nagy maximális klikkek jelenléte a gyakorlati kivitelezést nagyon erőforrás-igényessé teheti.

3.1.2. Az irányított csoportkeresési algoritmus

Az irányított k -klikkperkolációs klaszterek feltárása hasonló elvek szerint zajlik, mint az előző alfejezetben vázolt irányítatlan csoportok megkeresése. (Az algoritmust a disszertáció szerzője dolgozta ki és a [T6] publikáció függelékében került közlésre.) Elsőként bevezetjük az irányított maximális klikk fogalmát, mely egy olyan irányított k -klikknek felel meg, mely nem részgráfja egy még nagyobb irányított k -klikknek. Az irányítatlan esettel teljesen analóg módon egy irányított k -klikkperkolációs klaszter egy rögzített k esetén előáll úgy, mint maximális irányított klikkek uniója, melyek legalább k méretűek, és el lehet jutni köztük bármelyikből bármelyikbe legalább $k - 1$ méretű átfedések sorozatán keresztül.

A maximális irányított k -klikkek feltárását a következő iteráció segítségével lehet elvégezni:

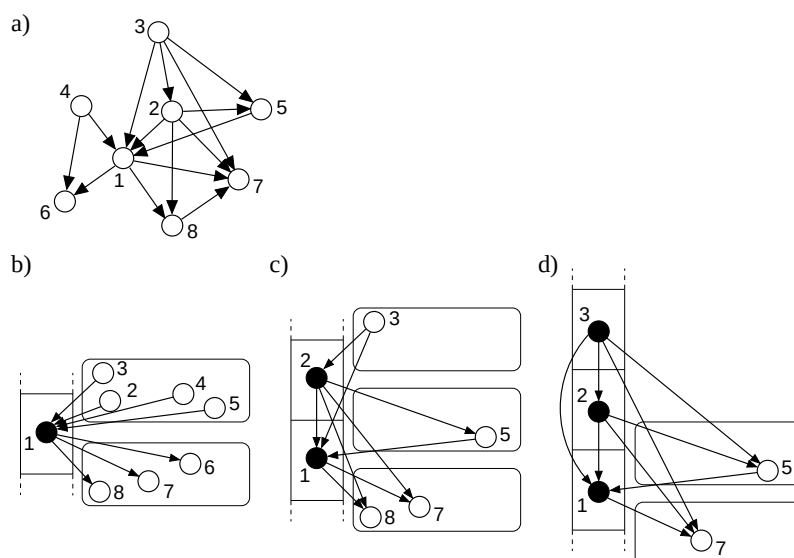
1. Megkeressük egy adott csúcs összes maximális irányított klikkjét.
2. Az adott csúcsot eltávolítjuk a hálózathoz.

(Természetesen az iteráció addig fut, amíg az összes csúcs el nem fogy.) Egy adott i csúcs maximális irányított klikkjeinek feltárását egy visszalépéses algoritmus segítségével lehet elvégezni, melyet a 3.3. ábrán szemléltetünk. Először létrehozunk egy rendezett tárolót, mely a maximális irányított klikk csúcsait fogja tartalmazni a megfelelő sorrendben, a kiinduló állapotban csak az i csúcsot tartalmazza. Emellett további tárolókra is szükség van, melyekből lépésenként emelünk be egy-egy csúcsot az irányított klikkbe. Ezen tárolók esetén nyilvántartjuk, hogy a kialakulóban lévő klikk csúcsaihoz képest hol helyezkednek el az élek iránya által megszabott sorrendben. A kezdeti állapotban két ilyen tárolónk van: az egyik az i elődjeit (a bejövő élek kezdőpontjait), a másik az i utódjait (a kimenő élek végpontjait) tartalmazza, ahogy azt a 3.3b ábrán láthatjuk. (Amennyiben egy csúcs az i -nek egyszerre elődje és utódja, úgy mind a két tárolóban szerepel). Az elődöket tartalmazó tároló az i előtt (vagy felett) van, hiszen ha innen emelünk be egy csúcsot az irányított klikkbe, akkor az az i elé kerül a sorrendben, míg az utódokat tartalmazó tároló az i mögött (vagy alatt) van, mert ha egy innét származó csúccsal bővítjük az irányított klikket, akkor az az i mögé (vagy alá) kerül.

A rekurzív iterációk során minden lépésnél egy csúcsot kiválasztunk valamelyik tárolóból, majd a tárolóban elfoglalt helyének sorszáma alapján beemeljük az irányított klikk csúcsait gyűjtő tároló megfelelő helyére. Ezután a potenciális klikktagokat tartalmazó tárolókat megszüntjük, annak érdekében hogy az irányított klikk ezen új tagjára is teljesüljön a következő két feltétel:

- az adott csúcs felett található tárolókban található csúcsok mind az adott csúcs elődjei,

3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval



3.3. ábra. Az irányított maximális klikk keresési algoritmus szemléltetése. a) Az 1-es csúcs lokális környezete egy elképzelt irányított hálózatban. b) Kiinduláskor az 1-es bekerül egy rendezett tárolóba, melyben később az irányított klikk csúcsait fogjuk kigyűjteni. További két tárolót is inicializálunk, egyet 1-es elődjeinek (a bejövő élek kezdőpontjainak), mely az 1-es fölé (elé) kerül, és egyet az 1-es utódjainak (a kimenő élek végpontjainak), mely az 1-es alá (mögé) kerül. c) A 2-es csúcsot beemeljük az irányított klikkbe, ahol az 1-es fölé (elé) kerül. A 4-es és 6-os csúcsok nem szomszédosak a 2-essel, ezért ezeket eltávolítjuk a tárolókból. Ezen felül egy új tárolót kell létrehoznunk az 5-ös számára, hiszen 2-esnek utódja, az 1-esnek viszont elődje. d) A 3-as csúcsot is behelyezzük az irányított klikkbe, ahol a 2-es fölé (elé) kerül, ezáltal az 5. tárolója meg is szűnik. Az alsó tárolóból a 8-as csúcsot törölni kell, mert nem szomszédja 3-asnak. (Az ábra eredeti verziója a [T6] publikáció függelékében jelent meg, itt apróbb módosításokkal mutatjuk be.)

- az adott csúcs alatt található tárolókban található csúcsok mind az adott csúcs utódjai.

Ha szükséges, új tárolókat is létrehozunk, például a 3.3c ábrán látható esetben a 2-es csúcs választásával az 5-ös csúcsot (mely a 2-esnek utódja, az 1-esnek elődje) egy új tárolóba kell helyezni, mely a 2-es és az 1-es közé kerül. Ennek révén bármely csúcsot is választjuk majd a következő iterációban a tárolókból, mindig teljesülni fog, hogy az adott csúcs pozíciója az irányított klikken belül megegyezik a tároló pozíciójával, melyből kiválasztottuk.

A fenti eljárást addig folytatjuk amíg az összes tároló ki nem ürül, és így megkapjuk az i csúcs egy maximális irányított klikkjét. Ezután következik a visszatérés az algoritmus egy olyan pontjához, ahol több potenciális új klikktag közül is választhattunk és az összes lehetőséget még nem próbáltuk ki. A lehetséges elágazásokat sorban végigkövetve egymás után megkapjuk i összes maximális irányított klikkjét.

Amennyiben az összes csúcs összes maximális irányított klikkjét sikerült feltárni, az

irányított k -klikkperkolációs csoportok megadása teljesen hasonló módon zajlik, mint az irányítatlan esetben. Itt is a maximális irányított klikkek átfedési mátrixából indulunk ki, melyből adott k választása esetén a $k - 1$ -nél kisebb átfedéseket és a k -nál kisebb maximális klikkeket töröljük. A visszamaradó mátrixban a nullától különböző nemdiagonális elemeket 1-el helyettesítjük, majd egy hagyományos klaszterező eljárással megkeressük a maximális irányított klikkek összefüggő komponenseit, melyek már a keresett irányított k -klikkperkolációs klasztereknek felelnek meg.

3.2. Lokális csoportszerkezet

Mivel a k -klikkperkolációs klasztereket eredetileg súlyozatlan hálózaton definiáltuk a 2.1. alfejezetben, első ránézésre úgy tűnhet, hogy az imént vázolt csoportkeresési módszer csak súlyozatlan hálózatokra alkalmazható. Ellenben minden súlyozott hálózat könnyűszerrel átalakítható súlyozatlanra, egyszerűen az élsúlyok elhagyásával. Sőt, általános gyakorlat, hogy egy valós hálózat vizsgálatánál először egy adott w^* súlyküszöbnél gyengébb éleket kiszűrjük és eltávolítjuk a rendszerből, majd a kapott szűrt hálózatot már súlyozatlanként kezeljük. Egy ilyen jellegű szűrés mindenféleképp indokolt ha meglehetősen zajos az adatforrás (mint pl. biológiai rendszerek esetén), és a gyenge kapcsolatok jó eséllyel csak mérési hibák miatt jelennek meg.

A fentiek alapján két paramétert kell megválasztanunk, ha egy súlyozott hálózatban szeretnénk csoportokat keresni a k -klikkperkoláció segítségével: egyfelől a k -klikkméretet, másrészt az élek szűréséhez használt w^* súlyküszöböt. A következőekben körüljárjuk, hogy miként érdemes ezeket a paramétereket beállítani ahhoz, hogy a feltárt csoportszerkezet a számunkra leghatékonyabb legyen. A részletek ismertetése előtt megemlítjük, hogy az idők során a k -klikkperkolációs módszernek született egy olyan verziója is, melynél az egyes élek helyett a k -klikkeket szűrjük meg az élsúlyok alapján [S3]. Ennél a megközelítésnél azon k -klikkeket vesszük figyelembe a csoportok felépítésénél, melyekben az élek mértani közepe elér egy általunk megszabott küszöböt. Mivel az erre vonatkozó eredmények első sorban Farkas Illés és Ábel Dániel munkáját dicsérik, itt nem kerülnek bemutatásra, és a továbbiakban olyan esetekre koncentrálunk, ahol először megszűrjük az éleket egy w^* súlyküszöb alapján, majd a kapott hálózatra alkalmazzuk vagy az eredeti, vagy az irányított k -klikkperkolációs módszert.

A w^* és k paraméterek változtatása távolról hasonlít egy mikroszkóp felbontásának beállításához. Ha viszonylag magas súlyküszöböt használunk, csak a legerősebb, és ezáltal valószínűleg legfontosabb, „legrobustusabb” élek maradnak meg a hálózatban, ezért a feltárt csoportok kisebbek lesznek ugyanolyan k méret mellett egy alacsonyabb súlyküszöbhez viszonyítva. Hasonló módon, ha növeljük a k -klikkméretet rögzített súlyküszöb mellett, egy egyre erősebb feltételt szabunk meg a csoportok számára, hiszen az egyre nagyobb k -klikkek egyre ritkábban fordulnak elő egy hálózatban. Ennélfogva a viszonylag nagy k -klikkméret megkövetelése is azzal jár, hogy a talált csoportok lényegesen kisebbek lesznek egy alacsony k értékhez viszonyítva (ugyanolyan súlyküszöb mellett). Összességében azt mondhatjuk, hogy a magas súlyküszöb és k -klikkméret kicsi, viszont nagyon erős csoportokat eredményez, melyeket súlyos élek hálózhatnak be nagyon sűrű módon.

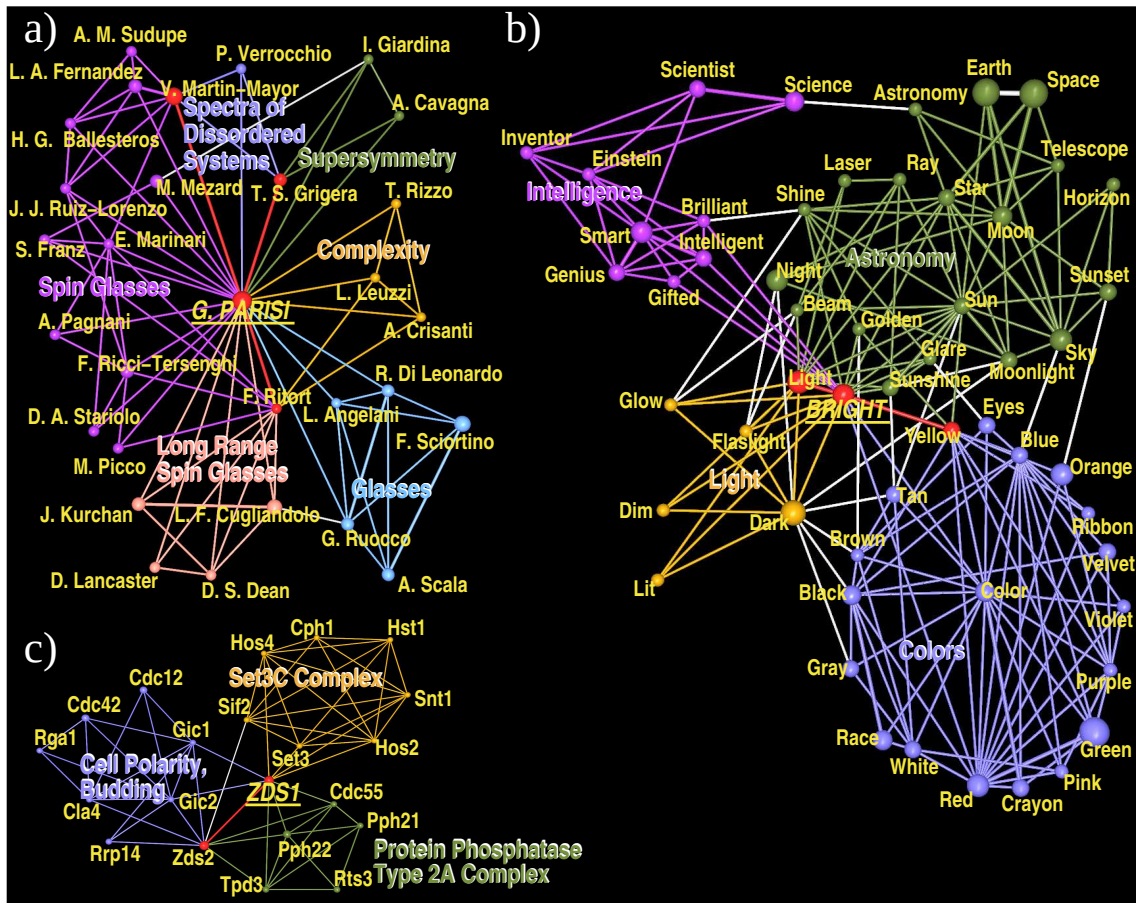
3. FEJEZET: *Hálózati csoportkeresés k -klikkperkolációval*

Ezzel szemben ha elkezdjük csökkenteni a súlyküszöböt, a hálózatban egyre több és több élt fogunk meghagyni, így egy rögzített k paraméter mellett várhatóan egyre nagyobb csoportokat fogunk találni (hiszen a gyengébb élekből álló részek sorban „hozzáépülnek” az erős csoportmaghoz). Hasonló effektussal jár a k -klikkméret csökkentése is, hiszen ezáltal gyengítjük a csoportok alap-építőelemeivel szemben támasztott kritériumot, melynek révén egyre több és több k -klikk kapcsolódik be a csoportokba. Alacsony súlyküszöb és k -klikkméret mellett gyakran előfordul, hogy a legnagyobb csoport mérete már az egész hálózat méretével összevethető, azaz egy perkoláló klaszternek felel meg, ami a rendszer jelentős részére kiterjed.

Amennyiben egy adott csúcs közvetlen környezetében található csoportokra vagyunk kíváncsiak, úgy érdemes több w^* és k paraméter beállítást is kipróbálni, hogy a kapott eredmények közül ki lehessen választani a legérdekesebb csoportokat. Erre mutat példákat a 3.4. ábra, mely a [T5] publikációban leírt, társszerzőségi, szóasszociációs és fehérje-kölcsönhatási hálózatokon végzet vizsgálataink során kapott csoportok közül ábrázol néhányat. A 3.4a ábrán Giorgio Parisi csoportjait láthatjuk, melyeket a Cornell Egyetem könyvtárának internetes Condensed Matter cikkarchívuma [213] alapján felépített (30 739 csúcsból álló) társszerzőségi hálózatban találtunk a k -klikkperkoláció segítségével. (Az adatokat Farkas Illés gyűjtötte és dolgozta fel; a csoportkeresést, csoportelemzést és az ábrakészítést a disszertáció szerzője végezte el). A hálózat élsúlyait a következő módon definiáltuk: minden n szerzős cikk $1/(n-1)$ -el növeli a szerzői közti kapcsolatok erősségét. E mögött az a megfontolás áll, hogy míg például egy kétszerzős cikk esetén egy szoros együttműködés tételezhető fel a szerzőpáros között, addig egy több tucat vagy akár több száz szerzővel rendelkező publikációnál a páronkénti kapcsolatok erőssége valószínűleg jóval kisebb.

A 3.4a ábrán Giorgio Parisi csoportjait $k = 4$ és $w^* = 0.75$ paraméter értékek mellett mutatjuk be. A csoportokat a különböző színek jelölik, a csoportok közti átfedések pirosak. A csúcsok mérete („térfogata”) azon csoportok számával arányos, melyeknek az adott csúcs tagja. (Természetesen az ábrán Giorgio Parisit leszámítva a többi csúcs egyéb csoportjai nem látszanak). Az élekekhez tartozó publikációk címe és kivonata alapján sikerült megállapítani, hogy a csoportok jól megfeleltethetők egy-egy kutatási témának Giorgio Parisi széles érdeklődési köréből, melyeken többnyire más és más társszerzőkkel dolgozott együtt.

A 3.4b ábrán a South Florida Free Association Norms List [140] alapján kapott szóasszociációs hálózatban talált csoportok közül a „BRIGHT” szó csoportjainak részgráfját tüntettük fel. A nyers adatok ennél a rendszernél egyetemistákkal kitöltetett űrlapokon alapultak, ahol a lapon található hívószavak mellé oda kellett írniuk az első szót, ami az adott hívószóról eszükbe jutott. A kapott válaszok mint új hívószavak szerepeltek a későbbi űrlapokon, így az adatbázisban szereplő szavak száma az adatgyűjtésben résztvevő egyetemisták számával fokozatosan emelkedett. Az eredmény egy 10 617 szóból álló súlyozott, irányított hálózat lett, melyben az élek iránya a hívószóból mutat a kapott válasz felé, míg az élsúly az adott szóasszociáció gyakoriságára utal. Mi az egyszerűség kedvéért itt figyelmen kívül hagytuk az élek irányítottságát és összeadtuk a két élsúlyt egy adott szópáros esetén, ha a kapcsolat mindkét irányban megvolt. A 3.4b ábránál egy $w^* = 0.025$ -ös súlyküszöböt és $k = 4$ -es k -klikkméretet használtunk. (Az adatokat Farkas Illés gyűjtötte



3.4. ábra. Átfedő csoportok egy adott csúcs környezetében. A különböző csoportokat különböző színekkel jelöltük, az átfedések pirossal vannak kiemelve, a csúcsok térfogata az adott csúcs csoportjainak számával arányos. Mindhárom esetben $k=4$ -es k -klikkméretet használtunk. a) Giorgio Parisi csoportjai a Cornell Egyetem könyvtárának internetes Condensed Matter cikkarchívuma alapján készült társszerzéségi hálózatban ($w^*=0.75$ súlyküszöbnél). A csoportok különböző témaköröknek, kutatási területeknek felelnek meg. b) A „BRIGHT” szó csoportjai a South Florida Free Association Norms List alapján készült szóasszociációs hálózatban ($w^*=0.025$ súlyküszöb mellett). A csoportok a szó különböző jelentéseivel kapcsolatos egyéb szavakból állnak össze. c) A ZDS1 fehérje csoportjai az élesztő (*S. cerevisiae*) fehérje-kölcsönhatási hálózatában a Database of Interacting Proteins adatai alapján. A különböző csoportok itt különböző fehérjefunkcióknak felelnek meg. (Az ábra forrása a [T5] publikáció).

össze, a csoportkeresést és csoportelemzést a disszertáció szerzője végezte el). A csoportok látható módon a „BRIGHT” szó négyféle jelentése körül szerveződnek, például a lila színű csoportban intelligenciával kapcsolatos szavakat találunk, a zöld a csillagászathoz kötődik, stb.

3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval

Végül a 3.4c ábrán a ZDS1 fehérje csoportjait mutatja az élesztő (*S. cerevisiae*) fehérje-kölcsönhatási hálózatában, mely a Database of Interacting Proteins [222] adatain alapult. Ez egy irányítatlan és súlyozatlan hálózat, melyben két csúcs közti kapcsolat azt jelzi, hogy az adott fehérjepárra vonatkozóan olyan konkrét kísérleti vagy egyéb viszonylag erős közvetett eredmények ismertek, melyek arra utalnak, hogy képesek egymással kölcsönhatni (pl. összekapcsolódni). A másik két hálózathoz hasonlóan itt is $k = 4$ esetén mutatjuk a csoportokat, melyek a ZDS1 fehérje különféle funkcióinak felelnek meg: Vizsgálataink szerint a talált csoportok jelentős részénél található egy fehérjefunkció, mellyel az adott csoporton belül a fehérjék túlnyomó része rendelkezik, és ezáltal az adott funkció jól jellemzi a csoportban résztvevő fehérjék biológiai szerepét. (A csoportkeresést a disszertáció szerzője végezte el, míg az adatgyűjtés, a csoportok elemzése és a fehérjefunkciók beazonosítása Farkas Illés érdeme). A csoportokhoz a legjellemzőbb funkció hozzárendelése a Gene Ontology Term Finder [35] segítségével történt. Ez a program az adott funkcióhoz egy szignifikancia szintet is képes rendelni, mely megmutatja, hogy mennyi a valószínűsége, hogy ha véletlenszerűen választanánk ki a csoport méretével megegyező számú fehérjét az adatbázisból, akkor azok között pont annyi rendelkezne a kérdéses funkcióval, mint ahány fehérje esetén a feltárt csoportban tapasztaltuk.

A 3.4b ábrán bemutatott szóasszociációs hálózatot később újra megvizsgáltuk az irányított k -klikkperkoláció segítségével is, hiszen ez a hálózat eredendően irányított. Az élek irányítottsága révén az eredményül kapott csoportokban különbséget tudunk tenni a csoporthoz főleg kimenő élekkel kapcsolódó csoporttagok és a csoporthoz főleg bejövő éllel kötődő tagok között. Az első fajta csoporttag egy fajta „forrást” képez a csoportban, míg a másik inkább a „nyelő” szerepét tölti be. Természetesen a csoporttagok többsége általában valahol a két véglet között helyezkedik el.

Mindenesetre ezen tulajdonság mérésére bevezethetjük az i csúcs J csoportra vonatkoztatott relatív bejövő fokszámát és a relatív ki-fokszámát [T6],

$$\hat{d}_{i,be}^J = \frac{d_{i,be}^J}{d_{i,be}^J + d_{i,ki}^J}, \quad (3.3)$$

$$\hat{d}_{i,ki}^J = \frac{d_{i,ki}^J}{d_{i,be}^J + d_{i,ki}^J}, \quad (3.4)$$

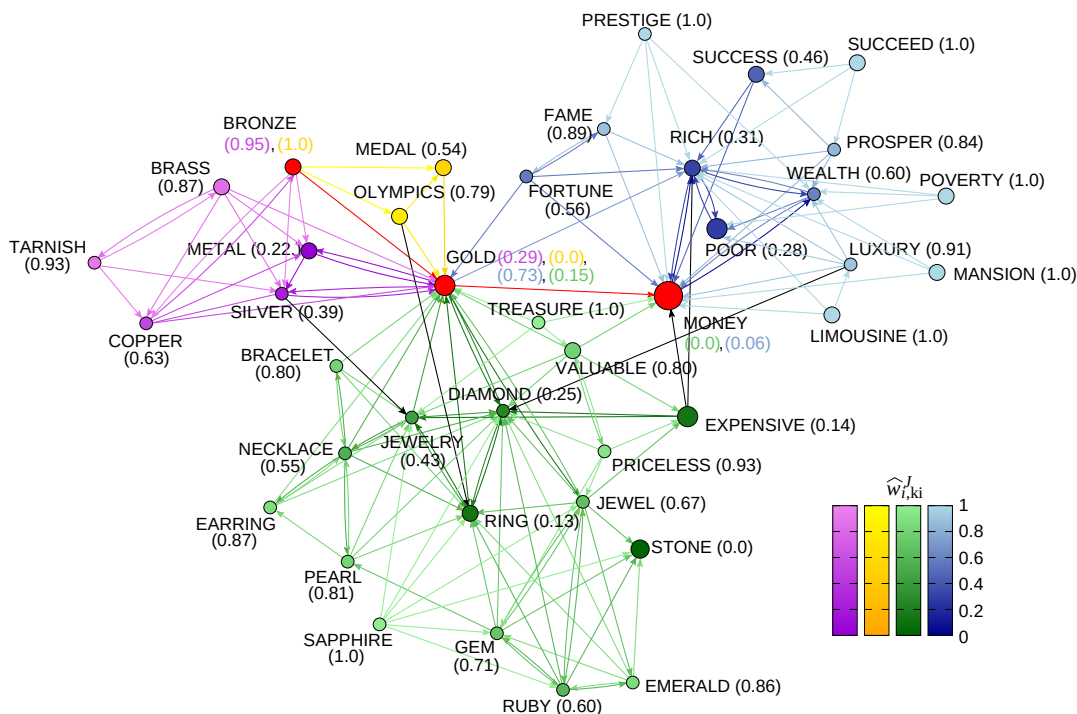
ahol $d_{i,be}^J$ és $d_{i,ki}^J$ az i csúcs J csoportban található elődjeinek illetve utódjainak számát adják meg. Természetesen $\hat{d}_{i,be}^J$ és $\hat{d}_{i,ki}^J$ mindig 0 és 1 közé esnek, valamint összegük 1-et ad. Súlyozott hálózatok esetén a (3.3-3.4) természetes módon általánosíthatók a

$$\hat{w}_{i,be}^J = \frac{w_{i,be}^J}{w_{i,be}^J + w_{i,ki}^J}, \quad (3.5)$$

$$\hat{w}_{i,ki}^J = \frac{w_{i,ki}^J}{w_{i,be}^J + w_{i,ki}^J}, \quad (3.6)$$

alakban, ahol $w_{i,be}^J$ és $w_{i,ki}^J$ az i csúcs J csoportban található elődjeihez illetve utódjaihoz

kapcsolódó éleinek összsúlyát adják meg. (A (3.3–3.6) egyenletekben definiált mutatók bevezetését a disszertáció szerzője javasolta).



3.5. ábra. Átfedő irányított csoportok a „GOLD” szó környezetében a szóasszociációs hálózatban. Ezúttal figyelembe vettük az élek irányítását is a South Florida Free Association Norms List alapján készült szóasszociációs hálózatban. Az irányított k -klikkperkoláció segítségével talált csoportokat láthatjuk $w^* = 0.023$ -as súlyküszöb és $k = 4$ -es irányított k -klikkméret esetén. A csúcsok mérete a csoportjaik számával arányos. A csúcsok mellett feltüntettük a hozzájuk tartozó $\hat{w}_{i,ki}^J = w_{i,ki}^J / (w_{i,be}^J + w_{i,ki}^J)$ értéket is, ami a csúcs színárnyalatában is megjelenik. A magas $\hat{w}_{i,ki}^J$ értékkel rendelkező csúcsok (pl. „SAPPHIRE”) általában különleges, ritkán használt szavak, míg az alacsony kifelé mutató súlyal rendelkező csúcsok (pl. „MONEY”) többnyire általános, gyakran használt szavak. (Az ábra forrása a [T6] publikáció.)

A 3.5. ábrán a „GOLD” szó közvetlen környezetében mutatjuk be az irányított és átfedő csoportokat $w^* = 0.023$ -as súlyküszöb és $k = 4$ -es irányított k -klikkméret választás mellett. A sárga csoport az olimpiai érmekkel kapcsolatos, a lila fémeket tartalmaz, a zöld ékszerekhez kötődik, míg a kék a sikerrel és gazdagsággal kapcsolatos szavakból áll. A csúcsok mellett feltüntettük az adott csoportra vonatkozó $\hat{w}_{ki}^J$ értéket is, amit vizuálisan a csúcs árnyalata is megjelenít. A képen látszik, hogy a speciális, ritkán használt szavak (pl. „SAPPHIRE”) többnyire nagy relatív kimenő súllyal rendelkeznek, hiszen rájuk ritkábban asszociálunk. Ezzel szemben az általános, gyakran használt szavak (pl. „MONEY”) alacsony relatív kimenő súlyt kapnak, mert rájuk gyakran asszociálunk és ezért az élek inkább feléjük mutatnak,

mint hogy tőlük induljanak.

Azonban érdekes megfigyelni azt, hogy például a „GOLD” szónak 0 a relatív kimenő súlya a sárga csoportban (azaz 1 a relatív bejövő súlya), tehát ott „nyelő”, míg például a kék színű csoportban már egy viszonylag magas, $\hat{w}_{ki}^J = 0.73$ -as kimenő súllyal rendelkezik. Ez azt mutatja, hogy egy fontos híd szerepet tölt be a két csoport között, mintegy átvezetve az asszociációk fonalát az olimpiai érmekről a siker, és tágabb értelemben a gazdagság felé. Ez az egyszerű példa rávilágít arra, hogy az imént bevezetett csoportra vonatkoztatott relatív kimenő illetve bejövő súly (vagy foksám) hasznos eszköz lehet arra, hogy megtaláljuk a csoportok közötti információáramlásért felelős csúcsokat egy általános irányított hálózat esetén.

További példákat a lokális szerkezet feltárására a C.1. függelékben adunk meg. Ezeknél egyrészt a szóasszociációs hálózatban vizsgáljuk meg további, több jelentéssel bíró szavak közvetlen környezetét, másrészt egy informatikához kapcsolódó hálózat lokális csoportszerkezetét mutatjuk be.

3.3. Teljes csoportszerkezet

3.3.1. A paraméterek optimális beállítása

A 3.4-3.5 ábrákon bemutatott csoportok jól illusztrálják, hogy a k -klikkperkolációs módszerrel értelmes csoportok tárhatók fel a legkülönbébb hálózatokban. Azonban még tisztáznunk kell azt a kérdést, hogy ha például nem egy adott csúcs közvetlen környezetében vagyunk kíváncsiak a csoportokra, hanem a teljes hálózat csoportszerkezetét fel szeretnénk tárni, akkor miként érdemes a paramétereket beállítani. Ehhez azt érdemes végiggondolni, hogy mi történik a csoportszerkezettel nagy vonalakban akkor, ha egy rögzített k paraméter mellett kiindulunk egy minden élsúlynál magasabb w^* súlyküszöböl, majd w^* értékét fokozatosan nullára csökkentjük. Kezdetben (az élek hiánya miatt) egyáltalán nincsenek csoportok, majd a legerősebb élek megjelenésével először kicsi, többnyire izolált csoportok keletkeznek. A folyamat során aztán a figyelembe vett élek számának növekedésével a csoportok mérete is nő, és idővel csoport-összeolvadások is előfordulhatnak. Amennyiben k nem túl magas, a végállapotban könnyen előfordulhat, hogy kialakul egy perkoláló óriás csoport, mely a folyamat során a korábbi csoportok nagy részét sorra magába olvasztotta.

A kérdés az, hogy egy ilyen folyamatnál mi jelöli ki az optimális súlyküszöböt általános esetben? Természetesen a cél az, hogy a csoportfelbontás lehetőleg minél gazdagabb struktúrát mutasson, hiszen ha a csoportokat használjuk később valamely vizsgálat bemenő adataként, akkor várhatóan így nyerhető ki belőlük a legtöbb információ. Ennélfogva a túl magas w^* , melynél még csak elvétve találunk csoportokat biztosan nem optimális. Ha viszont olyan alacsonyra állítjuk a súlyküszöböt, hogy megjelenik a perkoláló k -klikk-klaszter, akkor a kapott csoportfelbontás szintén nem hordoz túl sok információt. Ezek alapján az optimális w^* választás adott k paraméter érték mellett a k -klikkperkoláció kritikus élsúlyküszöbénél egy picit magasabb érték [T5]: így egyfelől biztosítjuk a lehetőleg minél gazdagabb csoportszerkezet létrejöttét, másfelől meggátoljuk, hogy megjelenjen egy perkoláló k -klikk-klaszter (egy óriás csoport), mely bekebelezná a többi kisebb csoport

jelentős részét. (Elsőként Vicsek Tamás javasolta ezt a módszert az optimális paraméterértékek kiválasztására.)

Előfordulhatnak olyan rendszerek, ahol egy adott (általában nem túl nagy) k esetén el lehet ugyan érni a k -klikkperkoláció kritikus pontját felülről egy alkalmasan magas w^* választásával, ez viszont az élek jelentős hányadának kidobásával jár. Ezért célszerű az élsúlykülöbségről áttérni a figyelembe vett élek hányadára: Az éleket először súly szerint sorba rendezzük, majd a k -klikkperkolációs átalakuláshoz a leggyengébb élektől kezdve sorban vesszük ki az éleket a hálózathoz, közben nyilvántartva a bennmaradó élek r hányadát. Természetesen így is pontosan annyi élet fogunk elhanyagolni a kritikus pontnál, mint az élsúlykülöbsős módszernél, viszont ha azt tapasztaljuk, hogy kritikus pontban mért r^* hányada a figyelembe vett éleknek túl alacsony, akkor érdemes egy magasabb k értékre áttérni és annál a k értéknél újra megkeresni a kritikus pontot. Az általunk használt kritérium a még elfogadható r^* élhányadra az volt, hogy vegyük figyelembe legalább az élek felét, azaz r^* legyen legalább $r^* = 0.5$.

A fenti kritériumok alapján például a 3.4a ábrán bemutatott társszerzői hálózat esetén a teljes csoportszerkezet feltárásának szempontjából optimális paraméter beállítás a $k=6$ és $r^*=0.93$ értékeknek felel meg. Hasonló módon, a 3.4b ábrán szereplő társszerzői hálózathoz a $k=4$ és $r^*=0.67$ értékek adódtak optimálisnak, míg a 3.4c ábrán szereplő súlyozatlan fehérje-kölcsönhatási hálózathoz a $k=4$ -es k -klikkméret-választás a legkedvezőbb.

3.3.2. Csoportstatisztikák

Amint korábban már jeleztük, a k -klikkperkolációs megközelítés legfontosabb újítása a korábbi módszerekhez képest az volt, hogy nagy skálán tette lehetővé a csoportszerkezet feltárását olyan esetekben is, amikor a csoportok közti átfedések nagyon gyakoriak, többszörös átfedések is előfordulnak, és az átfedő régiók kiterjedtebbek is lehetnek (azaz nem pusztán egy-egy csúcsban fed át két vagy akár több csoport). Ennek révén lehetőség adódott olyan, korábban nem vizsgált statisztikák vizsgálatára, melyeknek nincs értelme vagy triviális eredményt adnak teljesen izolált, vagy csak ritkán és kevésbé átfedő csoportok esetén.

Az egyik ilyen statisztika annak eloszlása, hogy egy adott csoport hány másik csoporttal van átfedésben. Ez egyben annak a hálózathoz a fokszámoszlása, melynek csúcsai az eredeti hálózat csoportjainak felelnek meg, és két csúcs akkor szomszédos, ha az adott csoportok átfednek egymással. Természetesen a csoportok közt úgy is lehet hálózatot definiálni, hogy elegendőnek tekintjük két csoport összekötöttségéhez azt, ha van legalább egy él az eredeti hálózatban az egyik csoportból a másikba. Az általunk használt, átfedéseken alapuló definíciója a csoportok közti hálózathoz egy sokkal erősebb követelményen alapszik. Ennek előnye, hogy valószínűsíthetően tényleg csak azok a csoportok kapcsolódnak össze a csoport-hálózatban, melyek között lényegi, erős kapcsolat áll fent a közös tagok révén. A csoport-hálózat fokszámoszlásával kapcsolatos legérdekesebb kérdés az, hogy vajon ez is skálafüggetlen-e, mint a csúcsok fokszámoszlása?

Két további, az átfedéseken alapuló statisztika, melyet érdekes megvizsgálni az átfedések méreteeloszlása, valamint az egyes csúcsok csoportszámának eloszlása, ahol a csoportszám megadja, hogy az adott csúcs hány csoportnak tagja párhuzamosan. (A felsorolt

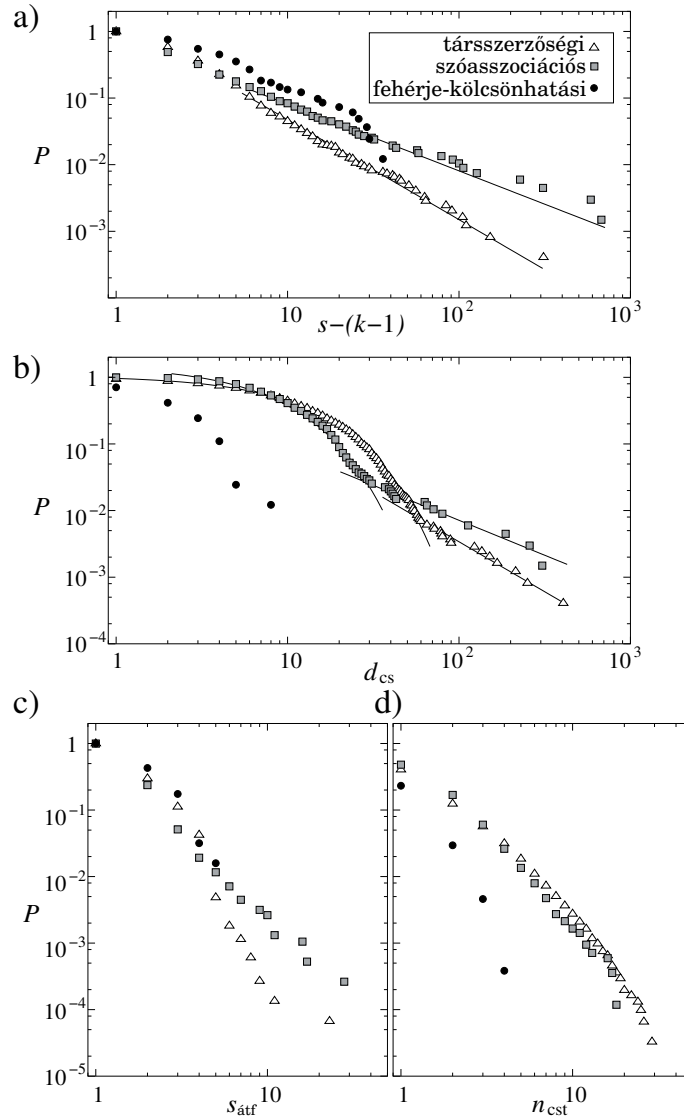
3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval

három csoportátfedésen alapuló statisztika bevezetését a disszertáció szerzője javasolta.) A továbbiakban megvizsgáljuk, hogy milyen eredményeket mutatnak ezek a statisztikák a 3.4. ábrán már bemutatott társszerzőségi, szóasszociációs és fehérje-kölcsönhatási hálózatok esetén. Az elemzésbe bevonjuk a csoportok méreteloszlását is, mivel ez egy olyan alapvető csoportstatisztika melyet széles körben szokás vizsgálni még a nemátfedő csoportkereső módszerek esetén is.

A három hálózat csoportszerkezetére vonatkozó statisztikákat a 3.6. ábrán láthatjuk. Mind a négy vizsgált mennyiség esetén a kiegészítő eloszlásfüggvényt ábrázoltuk. (A statisztikákat és magát az ábrát annak idején disszertáció szerzője készítette el a [T5] publikációhoz). A 3.6a ábrán a csoportok méreteloszlása látható. Mivel a minimális csoportméret k , az eloszlásokat az $s - (k - 1)$ függvényében ábrázoltuk, ahol s a csoportméret. (Emiatt a minimális méretű csoportok a vízszintes tengelyen az 1-es értékhez kerülnek mindhárom hálózat esetén). Az ábra tanúsága szerint a csoportok méreteloszlása hatványszerűen cseng le. A társszerzőségi hálózat (fehér háromszögek) és a szóasszociációs hálózat (szürke négyzetek) esetén egy -1 -es illetve egy -1.6 -os kitevővel rendelkező hatványfüggvény illesztést láthatunk folytonos vonallal feltüntetve az adatpontok mellett. Nemátfedő csoportok esetén már korábban is ismert volt, hogy a csoportméret-eloszlás lehet hatványszerű [170, 148]. Átfedő csoportokról viszont a [T5] publikációban jelent meg először ilyen eredmény. Természetesen ez szorosan összefügg azzal a feltétellel, mely szerint az optimális paraméterbeállítás a k -klikkperkoláció kritikus pontjánál van: A kritikus ponttól távol a perkolációs klaszterek (csoportok) méreteloszlása az óriás klasztert nem számítva gyorsan levág. Azonban a kritikus pontnál, ahol épp azon a határon van a legnagyobb klaszter, hogy „kiváljon” a többi klaszter eloszlásából és óriásivá növekedjen, a klaszterméret- (csoportméret-) eloszlás szükségszerűen egy lassan lecsengő, „vastag farkú” alakot kell felvegyen.

A 3.6b ábrán a csoportok fokszámoszlását láthatjuk (ahol egy adott csoport d_{cs} fokszáma a vele átfedő csoportok számát adja meg). A társszerzőségi és a szóasszociációs hálózatok esetén az eloszlás nemtriviális alakot vesz fel, mely két részből tevődik össze: egy bizonyos rendszerfüggő értékig exponenciálisan cseng le, $P(d_{cs}) \propto \exp(-d_{cs}/d_0)$, majd ezt követi egy hatványszerűen lecsengő rész, ahol $P(d_{cs}) \propto (d_{cs})^{-\gamma}$. Ez utóbbi, a nagy d_{cs} tartományra érvényes skálafüggetlen viselkedés szemléletes magyarázatot nyer, ha feltesszük, hogy egy csoporton belül minden csúcshoz egymástól függetlenül nagyjából egyforma esélye van arra, hogy egy másik csoportnak is tagja legyen és ezáltal egy csoportátfedés jöjjön létre. Ennek révén a nagyméretű csoportoknál a csoport foksámát jól becsülhetjük a csoport méretének, s -nek és egy δ arányossági tényezőnek a szorzatával, $d_{cs} \simeq s\delta$.

A 3.6b ábra szerint ez valóban teljesül, ugyanis a társszerzőségi és a szóasszociációs hálózatok esetén a csoportok fokszámoszlására a nagy foksámok tartományában ugyanolyan kitevővel rendelkező hatványfüggvényt lehet illeszteni, mint a csoportméret-eloszlásra. (Az adatpontok melletti egyenes vonalak ezeket az illesztett hatványfüggvényeket mutatják). Egy kicsit más a helyzet a kisméretű csoportok esetén: a minimális, k méretű csoportokból, illetve az ennél csak néhány csúccsal nagyobb csoportokból van a legtöbb a rendszerben, ezért a csoportok fokszámoszlásában megjelenik egy karakterisztikus $d_0 \simeq k\delta$ skála, és az eloszlásban a $k\delta$ érték körüli fluktuációk hatását látjuk megjelenni egy exponenciális lecsengésként. (Az adatpontok melletti görbült vonalak egy exponenciálisan lecsengő függvény



3.6. ábra. A k -klikkperkolációs csoportok statisztikái a 3.4. ábrán bemutatott hálózatok esetén. A kiegészítő eloszlásfüggvényeket a $k = 6$ és $r^* = 0.93$ értékek mellett láthatjuk a társszerzői hálózathoz (fehér háromszögek), $k = 4$ és $r^* = 0.67$ esetén a szóasszociációs hálózathoz (szürke négyzetek), és $k = 4$ -nél a fehérje-kölcsönhatási hálózathoz (fekete körök). a) A csoportméret-eloszlás. b) A csoportok foksámeloszlása. c) A csoportátfedések méreteloszlása. d) Az egyes csúcsok csoportszámának eloszlása. (Az ábra forrása a [T5] publikáció.)

illesztését mutatják).

A 3.6c ábrán a csoportátfedések méreteloszlását láthatjuk, ahol $s_{\text{átf}}$ az adott csoportok közös csúcsainak számát adja meg. Bár egy adott ponton két csoport maximum egy

3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval

$(k-2)$ -klikkben fedhet át, elvileg semmi sem tiltja, hogy csoportok több helyen is átfedjenek egymással, és ezáltal az összesített átfedésük nagy értéket vegyen fel. Az ábrán az eloszlások messze nem olyan szélesek, mint a csoportméret vagy a csoportfokszám esetében, de nem jelenik meg egy karakterisztikus skálája a csoportátfedéseknek, az adatpontok a dupla logaritmikus ábrázoláson durván egyeneseket követnek. Végül megvizsgálhatjuk azt is, hogy az egyes csúcsok egyszerre hány csoportban vesznek részt, ennek eloszlását a 3.6d ábrán mutatjuk be (itt n_{cst} egy adott csúcs esetén azt adja meg, hogy összesen hány csoportban számít tagnak). Ezek az adatpontok is nagyjából egy-egy egyenest követnek dupla logaritmikus ábrázolásnál a társszerzőségi és a szóasszociációs hálózat esetén. Ezért itt sem beszélhetünk egy tipikus skáláról, ami megadja, hogy egy csúcs nagyjából hány csoporthoz fog tartozni. A fehérje-kölcsönhatási hálózat ez esetben némileg kilóg a sorból, hiszen esetében a legtöbb csoportban felbukkanó csúcs is csak 4 csoportban vesz részt. Ez összecseng azzal, hogy már a csoportok fokszámeloszlása is jelentősen keskenyebb volt a másik két rendszerhez viszonyítva.

Az átfedéseken alapuló csoporthálózatokról további áttekintő jellegű statisztikákat gyűjtöttünk össze a 3.1 táblázatban. Ezek közül kiemelnénk a viszonylag magas klaszterezettségi együtttható értékeket, melyek azt mutatják, hogy ha egy csoport átfed két másik csoporttal, akkor jó eséllyel az a másik két csoport is átfed egymással. Érdekes továbbá a táblázat utolsó oszlopa is, mely rávilágít arra, hogy a vizsgált rendszerekben nagyon fontos a csoportok közti átfedések szerepe, hiszen a csoporttagok jelentős hányada (a társszerzőségi és szóasszociációs hálózatnál több, mint a fele) vesz részt legalább egy másik csoporttal való átfedésben.

Alap hálózat	N_{cs}	$\langle d_{\text{cs}} \rangle$	$\langle K_{\text{cs}} \rangle$	$\langle o \rangle$
társszerzőségi	2450	12.10	0.44	58%
szóasszociációs	670	11.33	0.56	72%
fehérje-kölcsönhatási	82	1.54	0.17	26%

3.1. táblázat. A csoporthálózatok alapvető statisztikái. A csoportok számát N_{cs} adja meg, $\langle d_{\text{cs}} \rangle$ az átlagos csoportfokszámnak felel meg, $\langle K_{\text{cs}} \rangle$ a csoporthálózat átlagos klaszterezettségi együttthatója. Végül $\langle o \rangle$ az utolsó oszlopban azt adja meg, hogy átlagosan egy csoport mekkora hányada vesz részt valamely másik csoporttal való átfedésben. (A táblázat forrása a [T5] publikáció.)

A csoportstatisztikákkal kapcsolatos további eredményeink megtalálhatók a C.2. függelékben. Itt többek között megvizsgáljuk, hogy miként változnak meg a statisztikák akkor, ha a vizsgált hálózat éleit véletlenszerűen átkötögetjük a csúcsok fokszámának megtartásával. Ezen felül két további hálózatra terjesztjük ki az elemzéseket, melyek közül az egyik egy informatikához kapcsolódó hálózat, melyben a csúcsok egy program változóinak felelnek meg, a másik pedig egy magyar nyelvű szinonima hálózat.

3.3.3. A csoportok hálózata

A csoportok közti hálózatra tekinthetünk úgy, mintha egy kis felbontású, durva kép tárulna elénk a hálózatról, melyben a rendszert egy magasabb szerveződési szinten vizsgálhatjuk. Ennek előnye, hogy sokkal nagyobb részét tudjuk áttekinteni a hálózatnak, mint amikor minden egyes csúcsot külön ábrázolunk. Ezt illusztrálja a 3.7. ábra, ahol az élesztőre (*S. cerevisiae*) vonatkozó, a Database of Interacting Proteins [222] adatbázisa alapján felépített fehérje-kölcsönhatási hálózatot ábrázoltuk ilyen módon. A hálózat relatíve kis mérete megengedi, hogy a teljes csoport-hálózat elférjen egy ábrán. A csoportok egy részének belső szerkezete teljes részletességgel is látható az ábra alsó felében. Amint a 3.4c ábránál már említettük, ennél a rendszernél a talált k -klikkcsoportok többnyire fehérje-komplexeknek, vagy fehérjefunkciós csoportoknak felelnek meg. Az adott csoportokhoz társítható fehérje-komplexek vagy fehérjefunkciók beazonosítása a GO-TermFinder programcsomaggal [35], illetve a Saccharomyces Genome Database [47] online eszközeivel történt.

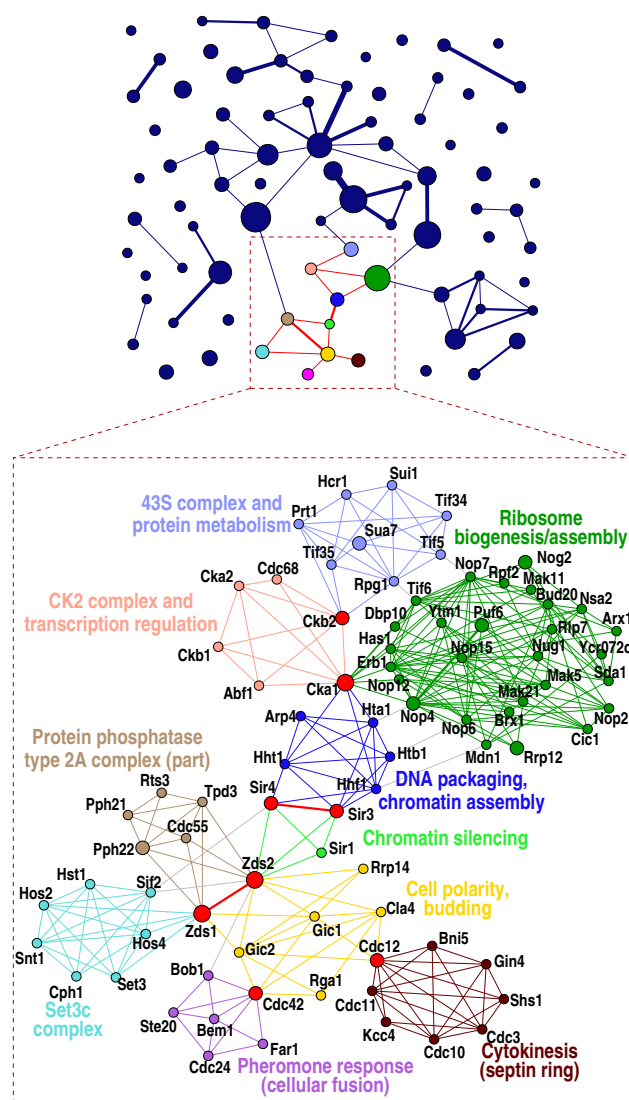
A csoportstatisztikák vizsgálatánál bemutattuk, hogy a csoport-hálózat fokszámoszlása hatványfüggvényszerűen esik nagy csoportfokszámok esetén, hasonló módon az alaphálózat skálafüggetlen viselkedéséhez. Azonban az alacsonyabb csoportfokszámok tartományában a csoportfokszám-eloszlás egy exponenciális eloszlásba megy át. Ez a viselkedés konzisztens a többszintű komplex rendszerekről alkotott általános képpel, mely szerint egy-két alap szerveződési szabály a rendszer minden szintjén azonos módon fellelhető, megengedve emellett néhány egyéb, szintenként specifikus tulajdonság kialakulását.

Ahogy a Bevezetésben már említettük, a skálafüggetlen fokszámoszlás kialakulásának egyik egyszerű és plauzibilis magyarázata a csúcsok szintjén a preferenciális kapcsolódási szabály, mely szerint a rendszer növekedése során az új csúcsok becsatlakozásakor a régi csúcsok a már meglévő foksámukkal arányos valószínűséggel kötődnek az új csúcshoz [14, 7, 130]. A preferenciális kapcsolódási szabály jelenlétét több olyan valódi rendszert reprezentáló hálózaton is ellenőrizték, melyek időfejlődéséről rendelkezésre állt elegendő adat [15, 104, 143]. Egy érdekes kérdés ehhez kapcsolódóan, hogy vajon a csoportok szintjén is megjelenik-e a preferenciális csatolódás, azaz vajon a csoport-hálózat fejlődése is hasonló elvek alapján zajlik, mint azt a csúcsok szintjén láttuk? Ezt a problémát vizsgáltuk az [S4] publikációban a 3.4. és 3.6. ábrákon bemutatott társszerzőségi hálózatban. Az eredmények szerint (melyek elsősorban Pollner Péter munkájának köszönhetők), a csoport-hálózat kialakulását a nagy csoportfokszámok tartományán olyan folyamatok vezérlik, melyek nagyon hasonlítanak a csúcsok hálózatának növekedésénél feltárt jelenségekhez.

A fejezet zárásaként megemlítjük, hogy a k -klikkperkoláción alapuló csoportkeresési módszer az idők során jelentős népszerűsége tett szert, például a [T5] publikáció 2016 márciusáig több mint 2000 független hivatkozást kapott. Ez többek között annak is köszönhető, hogy kutatási célra szabadon elérhető az interneten a CFinder programcsomag⁷, mely nemcsak megkeresi egy adott hálózat k -klikkperkolációs klasztereit, hanem képes az eredmények vizualizálására is. (A CFinder-hez a csoportkereső programkódot a disszertáció szerzője írta, a megjelenítésért felelős részt Adamcsek Balázs és Ábel Dániel). A C.3. függelékben néhány érdekes példát mutatunk be olyan kutatási eredményekből, melyeknél

⁷<http://www.cfinder.org>

3. FEJEZET: Hálózati csoportkeresés k -klikkperkolációval



3.7. ábra. A k -klikkperkolációs csoportok hálózata az élesztő (*S. cerevisiae*) fehérje-kölcsönhatási hálózatában. Az adatok a Database of Interacting Proteins adatbázisából származtak, a csoportokat a $k = 4$ -es k -klikkméret mellett tártuk fel. Az csúcsok mérete a csoportmérettel arányos. Az ábra alsó részében kinagyítottunk néhány csoportot, melyeknek így a belső szerkezete is láthatóvá válik. (A csoportokat eltérő színekkel, az átfedéseket pirossal jelöltük). Itt az egyes csúcsok mérete az adott csúcs csoportjainak számával arányos. (Az ábra forrása a [T5] publikáció.)

a k -klikkperkolációs módszert alkalmazták különféle rendszerekben a molekuláris biológiai hálózatoktól kezdve szociológiai hálózatokon keresztül a pénzügyi hálózatokig.

4. fejezet

Időfejlődő csoportdinamika

Az 1. fejezetben a topologikus fázisátalakulások tárgyalásánál már említésre került, hogy a valódi hálózatok időben változóak. Ez különösen igaz az emberi kapcsolatok hálózatára, hisz élete során mindenki előfordul, hogy új ismeretségeket köt, illetve, hogy valamely régebbi ismerőseivel idővel megszűnik a kapcsolata. E folyamatoknak köszönhetően az emberi kapcsolatok hálózata egy dinamikus, folyamatosan átstrukturálódó rendszer, melyben új csúcsok és élek jelennek meg, és ezzel párhuzamosan már létező csúcsok és élek tűnhetnek el. Ebben a fejezetben két, viszonylag nagy mérettel rendelkező emberi kapcsolathálózatban vizsgáljuk a hálózati topológia időfejlődésével együtt járó változásokat a k -klikkperkolációs csoportokban.

Ez a kutatás szorosan kapcsolódik a szociológia azon új területéhez, melynél az emberi kapcsolatok hálózatát olyan adatbázisok révén vizsgálják, melyeket eredetileg nem kifejezetten szociológiai kutatások céljából hoztak létre. A szociológiában hagyományosan erős gyökerei vannak a hálózatos megközelítésnek, hiszen már az 1930-as évektől kezdve tanulmányozták emberek kisméretű kapcsolathálóit, és alapvető felismeréseket tettek [219, 85, 214]. (Itt fontos megemlíteni a világhírű Mérei Ferenc nevét [70].) Ezek a vizsgálatok kérdőíves felméréseken alapultak, melynek nagy előnye, hogy a kapcsolatokról nagyon részletes információkat lehet kapni: honnan van az ismeretség, milyen erős a kapcsolat, mennyire kölcsönös, érzelmileg hogyan viszonyulnak egymáshoz a vizsgált személyek, stb. Ezzel szemben viszont nagy hátrány, hogy az ilyen módon tanulmányozható minta mérete erősen korlátozott, továbbá a válaszokból a szubjektivitást nem lehet teljesen kiszűrni.

Az ezredforduló környékén nagy változások álltak be az emberi kapcsolatokat leíró hálózatok elemzésében, amikor is elkezdték a különféle nagyméretű, automatizáltan gyűjtött adatbázisok használatát a vizsgált hálózatok rekonstruálására [158, 159, 216]. Jó példa erre különféle tudományos társszerzőségi hálózatok vizsgálata [88, 87, 144, 15], e-mailhálózatok vizsgálata [60, 61, 56], vagy mobiltelefon-hívási hálózatok vizsgálata [5, 158, 159]. Természetesen ezek az adatrendszerek csak nagyon korlátozott információt szolgáltatnak a kapcsolatok természetéről egy kérdőíves felméréshez képest, de egyrészt több nagyságrenddel nagyobb hálózatok elemzését teszik lehetővé, másrészt esetenként lehetőség nyílik a kapcsolatok erősségének objektívebb mérésére (pl. két személy között a telefonhívások számának segítségével).

4. FEJEZET: *Időfejlődő csoportdinamika*

Az egyik első, a szociológia szempontjából is fontos új eredmény, amit ezek a nagy skálájú vizsgálatok hoztak, a Granovetter-hipotézis igazolása volt egy több, mint 4 millió felhasználót tartalmazó mobiltelefon-hívási hálózatban [158, 159]. Mark Granovetter, amerikai szociológus még az 1970-es években fogalmazta meg a következő, híres hipotézist: Két ember kapcsolatának erőssége az egymásnak szentelt idő, anyagi ráfordítás, érzelmi intenzitás, bizalom és kölcsönös segítség/szívéesség kombinációja, és ez a kapcsolaterősség monoton növekedő függvénye a két ember relatív átfedésének (mely a közös ismerősök száma osztva a két személy összes ismerőseinek számával). A fent említett, Kertész János társszerzőségével megjelentetett publikációkban egyrészt a telefonhívások számát, másrészt a hívások összesített hosszát használták élsúlyként két személy között. Az eredmények egy, az élek 95%-áig határozottan emelkedő relatív átfedési görbét mutattak az élsúly függvényében, azaz a hipotézis beigazolódott [158, 159]. Ezen felül a hálózat perkolációs tulajdonságainak vizsgálata erősen alátámasztotta Granovetter egy másik híres hipotézisét, az ún. gyengeél-hipotézist. Ennek lényege, hogy az emberikapcsolat-hálóban az erős és gyenge élek jól elkülönülő szerepet játszanak, mely szerint az erős élek leginkább csoportok, közösségek belsejében találhatók, és ezeket hivatottak egyben tartani, szemben a gyenge élekkel, melyek első sorban a különböző csoportok között teremtenek kapcsolatot, és az egész hálózat összetartásáért felelősek.

A fejezetben elemzésre kerülő két hálózat egyike megegyezik a fent említett [158, 159] publikációkban vizsgált mobiltelefon-hívási hálózattal. A vizsgálatokba bevontunk ezen felül a Cornell Egyetem könyvtárának internetes cikkgyűjteményén alapuló társszerzőségi hálózatot is, melyet a 3. fejezetben már bemutatunk. A kutatás a két hálózatban feltárható k -klikkperkolációs csoportok időfejlődésére fókuszált. Pusztán az emberi kapcsolatokat leíró hálózatok időfejlődését (csoportelemzés nélkül) már több adatbázis alapján is vizsgálták korábban [123, 15, 99, 59, 211, 224, 154], és néhány tanulmány külön felhívta a figyelmet arra, hogy a csoportok időfejlődésének tanulmányozása nagyban segítheti a társadalom önszervező folyamatainak mélyebb megértését [89, 101, 90, 122, 113]. Ennek fényében az itt bemutatásra kerülő eredmények fontosságát az adja, hogy az irodalomban ezek nyújtották az első példát időfejlődő csoportok szisztematikus előállítására és statisztikus elemzésére nagyméretű emberikapcsolat-hálózatokban.

A fejezet alapjául a disszertáció szerzőjének első szerzőségével közölt [T7] publikáció szolgált, mely a *Nature*-ben jelent meg. Az eredményekből később született még két cikk [S5, S6], és szerves részét képezik egy könyvfejezetnek is [S7]. Magyar nyelven a *Természet Világában* jelent meg egy ezen kutatásokkal foglalkozó népszerűsítő cikk [S8], valamint a *Fizikai Szemlében* is szerepelt egy közlemény [S9], mely az említett mobilhívási hálózat esetén együtt mutatta be a [158, 159] cikkek valamint a [T7] publikáció legfontosabb eredményeit.

A mobiltelefon-hívási hálózat nyers adataihoz Barabási Albert-Lászlónak köszönhetően férünk hozzá; ezek alapfeldolgozását Szabó Gábor végezte el. A társszerzőségi hálózat esetén a 3. fejezetben bemutatott vizsgálatoknál használt adatokra támaszkodtunk, melyeket Farkas Illés gyűjtött össze. A kiinduló adatbázisokból az időfüggő élsúlyok és az időfejlődő hálózatok előállításának módszerét a disszertáció szerzője dolgozta ki. Ennek segítségével időlépésként feltárta a tanulmányozott hálózatok csoportszerkezetét, majd az így kinyert, pillanatfelvétellelhez hasonló statikus csoportokat összefűzte dinamikus, idő-

ben változó csoportokká. Ezt követte az időfejlődő csoportok statisztikai elemzése, mely szintén a disszertáció szerzőjének feladata volt. Ennek során bevezette a csoportok időbeli változékonyságára jellemző stacionaritást, valamint egy érdekes összefüggést tárt fel a stacionaritás, a csoportméret és a csoport várható élettartama közt. A csoportmegszűnés jóslására vonatkozó vizsgálatok ötletét Barabási Albert-László vetette fel, az erre vonatkozó elemzéseket megint csak a disszertáció szerzője végezte el. A kutatásokat Vicsek Tamás irányította és koordinálta.

4.1. A hálózatok és a csoportok előállítása

4.1.1. Súlyozott, időfejlődő hálózatok

Amint azt a 3.2. alfejezetben már említettük, az általunk vizsgált társszerzőségi hálózat esetén a kiindulási adatbázist a Cornell Egyetem könyvtárának internetes cikkgyűjteményében havi bontásban megjelenő, szilárdtestfizikához sorolt cikkek képezték (angol néven ez a *cond-mat arXiv*) [213]. Ez összesen több, mint 30 ezer szerző által közölt publikációk adatait tartalmazta, melyek 142 hónapnyi időtartamot töltöttek ki. A mobiltelefon-hívási hálózat esetén egy több, mint 4 millió ügyféllel rendelkező szolgáltató által lebonyolított hívások számát, idejét és költségét összegezték kéthetes periódusokban, összesen 52 héten át [158, 159]. (Adatvédelmi okok miatt természetesen a felhasználókat anonimizálák az eljárás során). Annak érdekében, hogy a nememberi felhasználókat, call-centereket, stb. kiszűrjük, csak azokat a hívásokat vettük figyelembe a feldolgozás során, melyeknél a két fél között mindkét irányban volt hívás az adatbázisban.

Egy közösen publikált cikk, illetve egy telefonhívás azt mutatják, hogy az érintett személyek között valamilyen kapcsolat alakult ki, és hálózatos szempontból az ilyen kapcsolatokat időfüggő élekkel reprezentálhatjuk. Mindkét rendszer esetén feltettük, hogy az emberek közti interakciók, ismeretségek már az említett, adatbázisban rögzített események bekövetkezése előtt elkezdődtek, és hasonló módon, az események bekövetkezése után is egy ideig még fennmaradtak. (Pl. egy cikk megírása általában intenzív együttműködést és kommunikációt kíván a cikk szerzői között, ami sokszor a cikk befejezése után is megmarad egy darabig).

Természetesen ugyanazon személyek egymás után többször is publikálhatnak együtt, vagy hívhatják egymást telefonon, és joggal tehetjük fel azt, hogy minél gyakrabban fordulnak elő ezek az események, annál erősebb az érintett személyek közti kapcsolat. Emellett azt is érdemes figyelembe venni, hogy az egyes publikációkhoz és telefonhívásokhoz is lehet súlyt rendelni, hiszen egy sokszerzős cikk esetén feltehetjük, hogy páronként kevesebb együttműködés volt a szerzők között, mint egy néhány szerzős cikknél. Erre egy természetes megoldás a 3.2. alfejezetben bevezetett súlyozás, melynél egy n szerzős cikk egy $1/(n-1)$ -es súlynak felel meg minden szerzőpár esetén a szerzőlistában. Ezzel párhuzamosan a telefonhívások esetén az egyes hívások költségét kezeltük úgy, mint az adott i esemény w_i súlyát.

A következő fontos kérdés az, hogy miként vegyük figyelembe a rendszer időfejlődését az egyes kapcsolatok szintjén? Egy egyszerű ötletnek tűnik bevezetni egy rögzített szélességű

4. FEJEZET: *Időfejlődő csoportdinamika*

időablakot, és egy adott időpillanatban kizárni azon eseményeket az élsúly kiszámításából, melyek az időpillanatra illesztett időablakon kívül esnek. Ezzel az a probléma, hogy így az események fontosságának (súlyának) nincs befolyása arra, hogy időben milyen hosszan tart az esemény „hatása”, ami szembe megy az emberi intuícióval. Például egy kétszerzős cikk megjelentetése után valószínűsíthetően jóval hosszabb ideig tartják egymással a kapcsolatot a szerzők, mint egy 100 szerzős cikk esetén. Ennél fogva időablakok alkalmazása helyett olyan élsúlyokat vezettünk be, melyek az eseményektől időben távolodva exponenciálisan lecsengenek, és a lecsengés exponense függ az adott esemény súlyától. Ezek alapján az élsúlyt a következő módon definiáltuk egy t időpontban valamely a és b csúcsok között:

$$w_{a,b}(t) = \sum_i w_i \exp\left(-\lambda \frac{|t - t_i|}{w_i}\right), \quad (4.1)$$

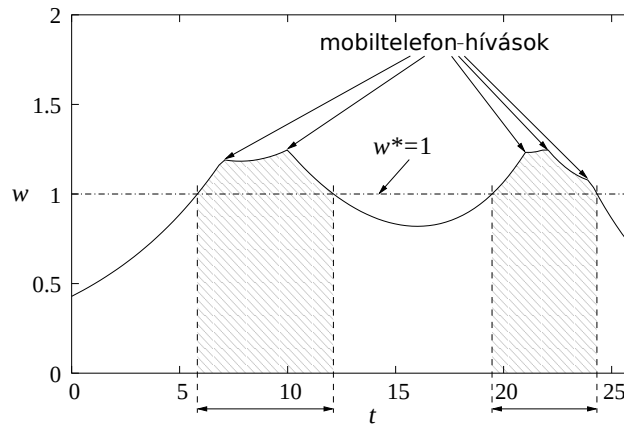
ahol az összegzés azon eseményeken (cikkeken illetve telefonhívásokon) fut végig, melyekben a és b érintettek, t_i jelöli az i esemény bekövetkezésének időpontját, w_i pedig az esemény súlyát. A (4.1) egyenletben bevezetésre került λ paraméter segítségével lehetőség nyílik az élsúlylecsengést lassítani vagy gyorsítani az egész hálózatban. Erre olyankor lehet szükség, ha például átskálazzuk az eseménysúlyokat (pl. a telefonhívások költségét más pénznemben kezdjük mérni) és a túl nagy (vagy túl kicsi) w_i értékek túl gyors (vagy túl lassú) lecsengést okoznak általánosan. Továbbá az is látható a (4.1) kifejezésben, hogy a w_i eseménysúlyok nem csak az exponensben szerepelnek, hanem emellett szorzófaktorként is megjelennek. Erre azért van szükség, hogy a $t = t_i$ speciális esetben is legyen különbség a fontos (nagy súlyú) eseményekből származó és a kis súlyú eseményekből származó élek között.

A fenti megközelítésben a hálózat időfejlődése az élsúlyok folytonos változásában manifesztálódik. Azonban egy súlyküszöb alkalmazásával, törölve azokat az éleket, melyek gyengébbek, mint egy rögzített w^* érték, egy valóban átstrukturálódó hálózat tárul elénk, melyben az élek eltűnnek, ha már egy ideje nem történt publikációs vagy telefonhívási esemény az adott csúcsok között, de ugyanakkor később újra megjelenhetnek, ha ismét van közös publikáció vagy telefonhívás a vonatkozó személyek részvételével. Ezt demonstrálja a 4.1. ábra, melyen a mobiltelefon-hívási hálózat egy véletlenszerűen választott élének időfejlődését láthatjuk. (A (4.1) egyenleten alapuló módszert a súlyozott hálózat időfejlődésének figyelembe vételére a disszertáció szerzője dolgozta ki).

4.1.2. A statikus csoportok

Miután előállítottuk az időfejlődő hálózatokat a 4.1.1. alfejezetben vázolt módon, minden időlépésnél megkerestük bennük a k -klikkperkolációs csoportokat. (Bár a (4.1) egyenletben definiált, időben folytonos élsúlyok megengedik, hogy akár két időlépés között is megvizsgáljuk a hálózatot, egy természetes választás volt az, hogy az adatbázisokban rögzített események által megadott időlépéseknél elemezzük a két rendszert). A csoportkeresésnél a k -klikkek mérete $k = 3$, az élsúlyküszöb $w^* = 0.1$ voltak a társszerzőségi hálózat esetén, illetve $k = 4$ valamint $w^* = 1.0$ a mobiltelefon-hívási hálózat esetén. (A paramétereket a 3.3. alfejezetben bemutatott módon állítottuk be).

A 4.2. ábrán a lokális csoportszerkezetet mutatjuk be egy adott időlépésnél, egy-egy véletlenszerűen választott csúcs néhány lépéses környezetében a két hálózat esetén. Az



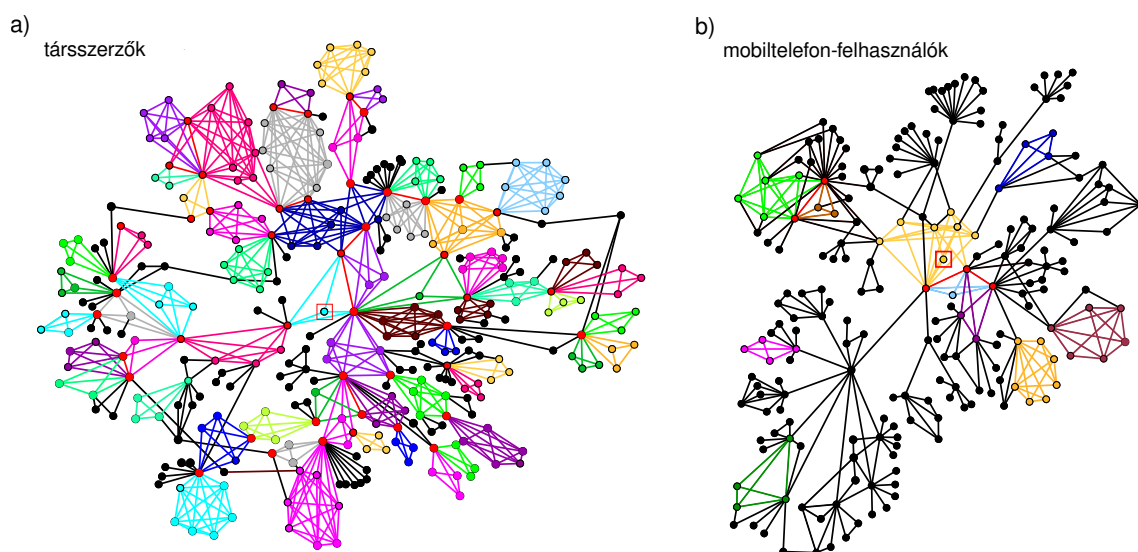
4.1. ábra. Az élsúly időfejlődése két véletlenszerűen választott felhasználó között a mobiltelefon-hívási hálózatban. Ha bevezetünk egy $w^* = 1$ nagyságú súlyküszöböt, ez az él csak a satírozott időszakokban lesz jelen. (Az ábra forrása a [T7] publikáció kiegészítő része).

eltérő csoportokat különböző színek jelölik, ahol a piros szín a csoportok közti átfedéseknek van fenntartva. A feketével jelölt csúcsok és élek nem kerültek bele egyik csoportba sem. Az ábrán is mutatott, adott időlépésnél kinyert csoportokra tekinthetünk úgy, mint a csoportszerkezet álló pillanatfelvételére. Később a 4.1.4. alfejezetben mutatjuk be, hogy miként lehet ezeket a statikus állóképeket egymás után fűzni ahhoz, hogy előálljanak az időfejlődő csoportok.

A 4.2. ábrán az látható, hogy a két hálózat elég eltérő lokális szerkezetet mutat. A társ-szerzőségi hálózatra tekinthetünk úgy is, mint egy páros gráf¹ vetületére. Emiatt érthető, hogy ez egy lokálisan elég sűrű hálózat, melyben gyakran fordulnak elő kisebb-nagyobb klikkek (hiszen minden cikk az összes szerzőjét összeköti egymással). A kapott csoportok lefedik a hálózat jelentős részét és nagyon gyakran fednek át egymással. Ezzel szemben a mobiltelefon-hívási hálózat egy jelentősen ritkább képet mutat, melyben a csoportok jobban elkülönülnek egymástól, ritkábbak az átfedések, és jóval több a csoportokon kívüli csúcs, illetve a csoportok között húzódó él.

Ezen eltérések kialakulásában vélhetően szerepe van a kapcsolatok eltérő természetének a két hálózat esetén. A telefonhívások pillanatszerű eseményeknek tűnnek, ha mondjuk egy tudományos publikáció megírásának folyamatához hasonlítjuk őket. Noha ezeket a pillanatszerű, gyors eseményeket kéthetes periódusokra aggregálták, még így is valószínűsíthető, hogy időben szorosabban követik az emberi kapcsolatok változásait, mint a társ-szerzőségi hálózat, hiszen egy cikk megírása általában több hónapos, vagy akár egy-két éves együttműködést is megkívánhat. Ezen felül, a társ-szerzőségi hálózatnál a cikkek túlnyomó részénél nem igazi páros interakciókat látunk, hanem nagyobb csoportok közös munkájának

¹Ez a páros gráf a szerzők és a cikkek között feszül ki, ahol az élek csak eltérő „típusú” csúcsok között húzódhatnak. Ezt úgy vetíthetjük rá egy már pusztán szerzőkből álló gráfra, hogy minden olyan szerző párost összekötünk akinek van közös szomszédja a cikkek között a páros gráfban.



4.2. ábra. a) A lokális csoportszerkezet egy véletlenszerűen választott csúcs szűk környezetében a társzerzőségi hálózat esetén. A kiindulásul választott csúcsot egy piros keretbe foglaltuk. A különböző csoportokat az eltérő színek jelölik, a csoportok közti átfedéseket pirossal emeltük ki. b) Ugyanaz az ábra a mobiltelefon-hívási hálózat esetén. (Az ábra forrása a [T7] publikáció).

eredményét. Ezzel szemben a telefonhívási hálózatnál az élek ilyen szempontból függetlenek egymástól, és minden esetben az adott két személy között lezajlott páros interakciók eredményének tekinthetők.

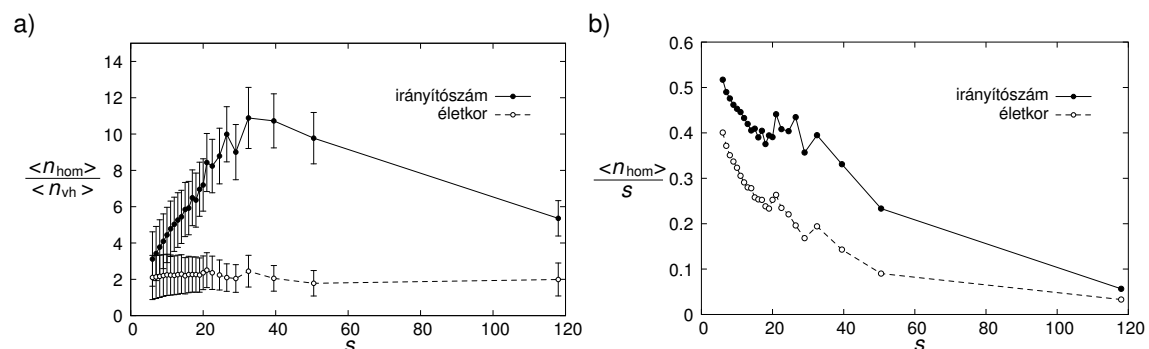
4.1.3. A csoportok ellenőrzése

Mielőtt nekilátnánk összefűzni az időlépésenként kinyert statikus csoportokat időfejlődő csoportokká, célszerű megvizsgálni, hogy vajon a statikus csoportok mutatnak-e olyan jeleket, melyek alátámasztják azt a feltételezésünket, hogy ezek jelentős részben ténylegesen létező emberi közösségek lenyomatai. Az egyik egyszerű vizsgálat, amit elvégeztünk, az a csoportokon belüli élek átlagos súlyának összehasonlítása volt a csoportok között húzódó élek átlagos súlyával. A két mennyiség arányára 2.9 adódott a társzerzőségi hálózat esetén, és 5.9 a mobiltelefon hívási hálózat esetén. Ez azt jelenti, hogy a csoportokon belüli együttműködés illetve kommunikáció lényegesen intenzívebb a csoportokon kívüli kapcsolatokhoz viszonyítva. (Ez az eredmény szintén szépen egybevághat a Granovetter-féle gyengeél-hipotézissel).

A mobiltelefon-hívási hálózat esetén néhány kiegészítő információ mint például az irányítószám és az életkor is rendelkezésre állt a (más szempontból teljesen anonim) felhasználókról. E két adattípus segítségével megvizsgálhatjuk, hogy mennyire homogének az általunk feltárt csoportok. Ennek érdekében minden csoportnál megkerestük a tagok legnagyobb olyan részalmazát, melyen belül az irányítószámok megegyeznek. Összehasonlítás-

ként megcsináltuk ugyanezt a csoportokkal megegyező méretű, véletlenszerűen összeválogatott halmazokra is. A 4.3a ábra a csoportok legnagyobb, irányítószám alapján homogénnek tekinthető részhalmazának $\langle n_{\text{hom}} \rangle$ átlagos méretét mutatja az s csoportméret függvényében, a véletlenszerű halmazokra kapott $\langle n_{\text{vh}} \rangle$ átlagértékkel leosztva. A k -klikkperkolációs csoportok esetén kapott jelentősen nagyobb értékek azt mutatják, hogy ezek a csoportok többnyire olyan személyekből tevődnek össze, akik viszonylag közel laknak egymáshoz. A 4.3a ábrán mutatott görbének maximuma van nagyjából az $s \simeq 35$ -ös csoportméret-nél, ami alapján elsőre úgy tűnhet, mintha az ilyen méretű csoportok lennének földrajzi értelemben a leghomogénebbek.

Azonban ennél összetettebb a helyzet, ahogy azt a 4.3b ábra mutatja, melyen az $\langle n_{\text{hom}} \rangle / s$ arányt ábrázoltuk a csoportméret függvényében. Ennek alapján a kisebb csoportok a leghomogénebbek, bár továbbra is észlelhető egy kisebb csúcs az $s \simeq 30-35$ tartományban. A 4.3. ábrán az irányítószám mellett az életkorra kapott hasonló eredményeket is feltüntettük, ahol $\langle n_{\text{hom}} \rangle$ a legnagyobb, egy három éves időablakon belül azonos életkorú felhasználókból álló részhalmaz átlagos méretének felel meg adott méretű csoportokon belül, míg $\langle n_{\text{vh}} \rangle$ a csoportokkal megegyező méretű véletlenszerű halmazokban kapott átlagérték. Az, hogy $\langle n_{\text{hom}} \rangle / \langle n_{\text{vh}} \rangle > 1$ azt mutatja, hogy a csoportok nagyobb eséllyel tartalmaznak azonos generációhoz tartozó személyeket, mint a véletlenszerű halmazok. A 4.3b ábrán az is látszik, hogy életkor alapján is a kisebb csoportok mutatnak nagyobb homogenitást a nagyobb csoportokhoz képest.

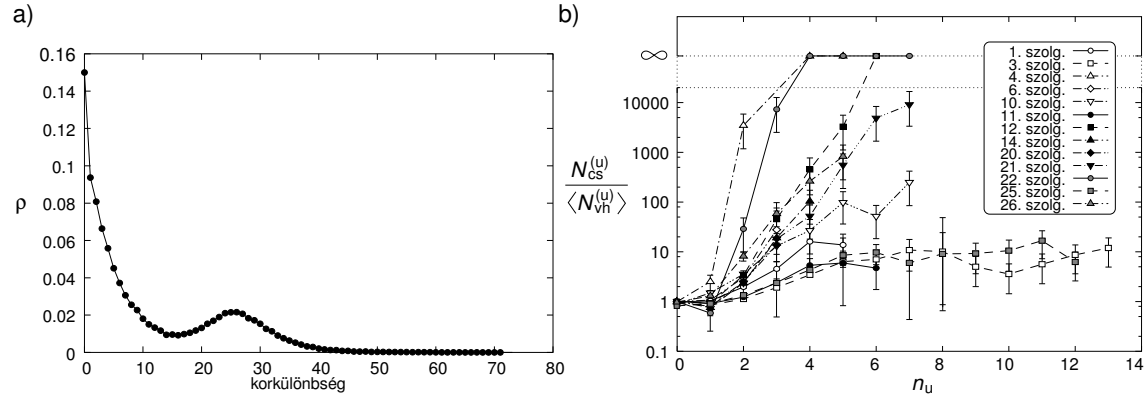


4.3. ábra. a) A teli szimbólumok a csoportok legnagyobb, azonos irányítószámmal rendelkező részhalmazának $\langle n_{\text{hom}} \rangle$ átlagos méretét mutatják az s csoportméret függvényében, ahol a telefonhálózat csoportjaira kapott eredményt leosztottuk az ugyanolyan méretű, véletlenszerűen választott felhasználókból álló halmazokra kapott $\langle n_{\text{vh}} \rangle$ értékkel. Hasonló módon, az üres szimbólumok a csoportok legnagyobb, egy három éves időablakon belül azonos életkorral rendelkező részhalmazának átlagos méretét mutatják a véletlenszerűen összerakott, ugyanolyan méretű halmazokra kapott eredménnyel leosztva, szintén s függvényében. Az alsó hibahatár $\langle n_{\text{hom}} \rangle / (\langle n_{\text{vh}} \rangle + \sigma_{\text{vh}})$, a felső pedig $\langle n_{\text{hom}} \rangle / (\langle n_{\text{vh}} \rangle - \sigma_{\text{vh}})$, ahol σ_{vh} a véletlenszerűen összerakott halmazokra kapott szórás. b) Az $\langle n_{\text{real}} \rangle / s$ az s csoportméret függvényében az irányítószám (fekete szimbólumok) és az életkor (fehér szimbólumok) esetén. (Az ábra forrása a [T7] publikáció).

Ezen felül a 4.3. ábrából az is leszűrhető, hogy a csoportok nagyobb homogenitást

4. FEJEZET: Időfejlődő csoportdinamika

mutatnak az irányítószám szempontjából az életkorhoz hasonlítva. Egy kézenfekvő magyarázat erre az, hogy a szülők és gyerekeik között általában a gyerekek felnőtté válása után is megmarad az erős kapcsolat, melynek révén egyes csoportok könnyen tartalmazhatnak eltérő generációkhoz tartozó felhasználókat. Ezt alátámasztja a 4.4a ábra, melyen a csoporttagok életkorkülönbségének eloszlását mutatjuk be. Az empirikus sűrűségfüggvény maximuma nullánál van, viszont megjelenik egy jól látható kisebb maximum is 25 évnél, ami nagyjából megfelel egy tipikus szülő-gyerek életkorkülönbségnek.



4.4. ábra. a) A csoporttagok életkorebeli különbségének empirikus sűrűségfüggvénye, ρ , a mobiltelefon hívási hálózatban. A két legtöbbször előforduló érték a 0 és a 25. Ez azt jelenti, hogy ha veszünk két felhasználót egyazon csoportból, akkor a két legvalószínűbb eset az, hogy azonos az életkoruk, vagy egy generációnyi eltérés van köztük. b) Azon csoportok $N_{cs}^{(u)}(n_u)$ száma, melyben az adott szolgáltatást igénybe vevő tagok száma n_u , leosztva azon véletlenszerűen összerakott halmazok átlagos $\langle N_{vh}^{(u)}(n_u) \rangle$ számával, melyekben szintén n_u felhasználó vette igénybe a szolgáltatást. A véletlenszerű halmazok méreteloszlása megegyezett a csoportok méreteloszlásával. (Az ábra forrása a [T7] publikáció kiegészítő része).

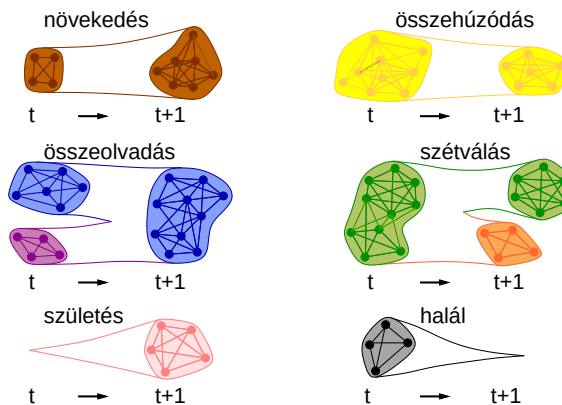
Az irányítószámon és az életkoron felül rendelkezésre álltak bizonyos adatok a felhasználók által igénybe vett különféle szolgáltatásokról is. A kiindulási adatbázisban például szerepelt, hogy a 34-féle lehetséges szolgáltatást hányszor vette igénybe egy-egy felhasználó az adott kéthetes időszakokban. Azonban fontos leszögezni azt, hogy a felhasználók döntő többsége egyáltalán nem vett igénybe egy ilyen extra szolgáltatást sem, azaz egy véletlenszerűen választott felhasználó esetén bármely szolgáltatás használatának esélye nagyon kicsi volt. Emiatt az irányítószám és az életkor esetén alkalmazott módszer helyett itt minden szolgáltatás esetén azon csoportok $N_{cs}^{(u)}(n_u)$ számát vizsgáltuk, melyek n_u olyan taggal rendelkeztek, akik legalább egyszer használták az adott szolgáltatást. Összehasonlításként minden csoporthoz legyártottunk egy megegyező méretű véletlen halmazt is, majd meghatároztuk azon halmazok $N_{vh}^{(u)}(n_u)$ számát, melyek szintén n_u olyan taggal rendelkeztek, akik használták az adott szolgáltatást. A véletlenszerű mintát összesen 10.000-szer gyártottuk le és a kapott értékeket átlagoltuk.

A 4.4b ábrán 13 különböző szolgáltatás esetén ábrázoltuk az $N_{cs}^{(u)}(n_u)$ és az $\langle N_{vh}^{(u)}(n_u) \rangle$ arányát az n_u függvényében. Látható, hogy az általunk megtalált csoportok esetén $N_{cs}^{(u)}(n_u)$ jelentősen nagyobb lehet, mint a véletlenszerűen összeválogatott halmazokra vonatkozó $\langle N_{vh}^{(u)}(n_u) \rangle$. Sőt, bizonyos esetekben $N_{cs}^{(u)}(n_u) / \langle N_{vh}^{(u)}(n_u) \rangle$ végtelenné is válik, ami annak tudható be, hogy egyetlen egy véletlen halmaz sem tartalmazott olyan magas n_u számú, az adott szolgáltatást igénybe vevő felhasználót, mint a valódi k -klikkperkolációs csoportok. Ezek az eredmények azt jelzik, hogy noha ritkák azok a felhasználók, akik extra szolgáltatást vesznek igénybe, jellemző rájuk, hogy olyan más felhasználókkal vannak egy csoportban, akik szintén használják az adott szolgáltatást.

Összességében elmondhatjuk, hogy a felhasználókról rendelkezésre álló egyéb adatok vizsgálata megerősítette azt a feltételezést, mely szerint a feltárt csoportok valódi emberi közösségek lenyomatainak tekinthetők. Az általunk megadott k -klikkperkolációs csoportok többnyire egymáshoz közel lakó, hasonló életkorú személyekből állnak, illetve esetenként plauzibilis lehet szülő-gyerek kapcsolatot feltételezni a csoporttagok között az életkor különbsége alapján. Ezen felül azon meglehetősen ritka felhasználók, akik extra szolgáltatásokat vesznek igénybe, jelentősen gyakrabban fordulnak elő egyszerre nagyobb számban egyszerre egy-egy csoporton belül, mint ahogy azt egy véletlenszerű leosztás alapján várnánk. (Az alfejezetben bemutatott vizsgálatokat a disszertáció szerzője végezte el.)

4.1.4. Időfejlődő csoportok

Ahhoz, hogy a 4.1.2. alfejezetben kinyert, ez egymást követő időlépésekhez tartozó csoportokat össze tudjuk fűzni, ki kell dolgozni egy szisztematikus eljárást, mely adott „pillanat-felvétel” esetén képes megkeresni egy csoport előképét az előző időlépéshez tartozó statikus csoportok között. A 4.5. ábrán felsoroljuk azon elemi eseményeket, melyek a csoportfejlődés



4.5. ábra. Alapvető események egy csoport élete folyamán. Ha új tagok jelennek meg, a csoport növekszik, míg ha régebbi tagok távoznak, a csoport mérete csökken. Különböző csoportok összeolvadhatnak egymással, illetve egy nagyobb csoport széteshet kisebb csoportokra. Ezen felül teljesen új csoportok jelenhetnek meg, és régebbiek szűnhetnek meg. (Az ábra forrása a [T7] publikáció).

4. FEJEZET: *Időfejlődő csoportdinamika*

során bekövetkezhetnek. Ha egy csoport új tagokat toboroz magának, a mérete növekedni fog, míg ha régi tagok hagyják el, a mérete csökken. Két (vagy akár több) csoport egymással összeolvadhat, és hasonló módon előfordulhat, hogy egy nagyobb csoport több kisebb csoportra esik szét. Végül teljesen új csoportok is formálódhatnak „a semmiből”, és már létező csoportok szűnhetnek meg teljesen.

Egy hosszabb életű csoport esetén könnyen előfordulhat, hogy idővel a tagok teljesen lecserélődnek. Erre a hétköznapi életből jó példát nyújthat egy nagy történelmi múlttal rendelkező sportklub szurkolóinak „kemény magja”: Száz évvel ezelőtt is létezett egy nagyon lelkes csoport, mely rendszeresen látogatta a mérkőzéseket, és melyen belül az emberek jól ismerték egymást, és most is létezik ilyen, viszont ennek tagjai között hiába keresnénk a száz évvel ezelőtti szurkolókat. Ennek ellenére az eltelt évek során a csoport végig létezett, csak az összetétele változott folyamatosan az idővel.

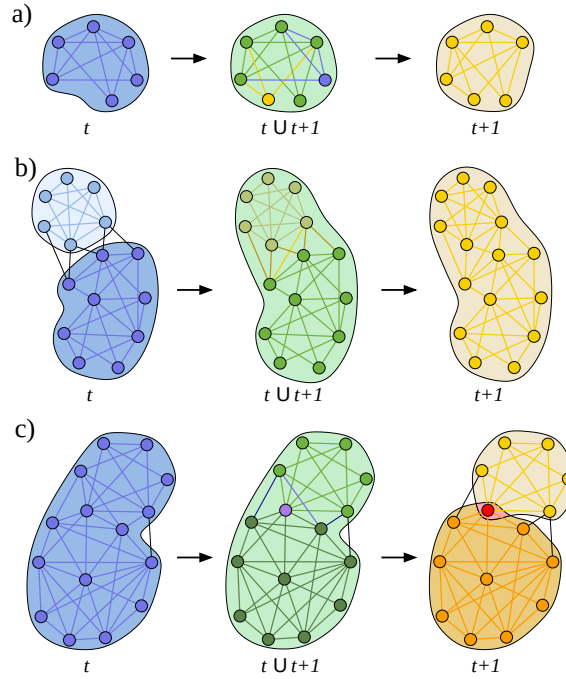
Visszatérve az alfejezet központi problémájához vizsgáljuk meg, hogy miként lehetne megtalálni egy adott statikus csoport előképét a megelőző időlépésnél feltárt csoportok között. Mivel a tanulmányozott hálózatok és a bennük lelhető csoportok száma igen nagy, ez a feladat nemtriviális és igen számításigényes lehet. (A problémát tovább nehezíti, hogy a csoportok át is fedhetnek egymással). Talán a legtermészetesebb alapötlet az, hogy két szomszédos időlépésnél a csoportokat a relatív átfedés alapján feleltetjük meg egymásnak. Valamely A és B csoportok relatív átfedését az

$$R(A, B) \equiv \frac{|A \cap B|}{|A \cup B|} \quad (4.2)$$

alakban definiálhatjuk, ahol $|A \cap B|$ az A és B közös tagjainak száma, míg $|A \cup B|$ az A és B uniójában található csúcsok száma. Azonban a csoportok közti átfedés alááshatja ezt az egyszerű csoportmegfeleltetési módszert: Amennyiben egy kezdetben kisméretű A csoport mérete nagyot nő t és $t+1$ között, és $t+1$ -ben átfed egy kisebb $B = B_t = B_{t+1}$ csoporttal, mely nem változott az időlépés alatt, úgy előfordulhat, hogy a (4.2) kifejezéssel megadott relatív átfedés nagyobb lesz A_{t+1} és B_t között, mint A_{t+1} és A_t között.

Ezt a problémát a 4.6. ábrán vázolt módon küszöböltük ki. A megközelítés lényege, hogy két szomszédos t és $t+1$ időlépés esetén a vonatkozó két hálózat uniójában is elvégezzük a csoportkeresést. (Ezt a lehetőséget Derényi Imre vetette fel először). A k -klikkperkolációs csoportoknak megvan az a jó tulajdonsága, hogy ha egy hálózatba új éleket húzunk be úgy, hogy ezzel párhuzamosan már létező éleket vagy csúcsokat nem távolítunk el, akkor az eredetileg már meglévő csoportok növekedhetnek, összeolvadhatnak, vagy változatlanok maradhatnak, de nem zsugorodhatnak, nem eshetnek szét kisebb csoportokra, és nem szűnhetnek meg. Ebből az következik, hogy akár a t , akár a $t+1$ időlépésnél választunk egy csoportot, azt pontosan egy csoport fogja magában foglalni a két hálózat uniójából kinyert csoportok közül.

Jelöljük $\mathbf{A}$ -val a t -nél kapott csoportok halmazát, illetve $\mathbf{B}$ -vel a $t+1$ -nél kapott csoportok halmazát, valamint $\mathbf{V}$ -vel a két időlépéshez tartozó hálózatok uniójából kapott csoportok halmazát. A fentiek alapján bármely $A_i \in \mathbf{A}$ vagy $B_j \in \mathbf{B}$ esetén pontosan egy olyan $V_k \in \mathbf{V}$ létezik, mely lefedi az adott csoportot. Az $\mathbf{A}$ és $\mathbf{B}$ elemeit úgy feleltethetjük meg egymásnak, hogy először minden $V_k \in \mathbf{V}$ esetén megkeressük azon $A_i^k \in \mathbf{A}$ és $B_j^k \in \mathbf{B}$



4.6. ábra. Néhány egyszerű csoportfejlődési „forgatókönyv” $k = 4$ -nél. A t időlépésnél a csoportok kék színezést kaptak, az ezt követő $t + 1$ időlépésnél pedig sárgát. A két időlépéshez tartozó hálózatok uniójában is elvégezhetjük a csoportkeresést, ennek eredményét mutatják a zöldre színezet csoportok a két időlépés között. Az a) panel esetén a csoport egyszerűen csak „propagál”, a b) panelnél két csoport összeolvad, míg a c) panel esetén egy kis csoport kiválik egy nagyobb kezdeti csoportból. (Az ábra forrása a [T7] publikáció kiegészítő része).

csoportok teljes listáját, melyeket V_k lefed (azaz $A_i^k \subseteq V_k$ és $B_j^k \subseteq V_k$). (Természetesen bármelyik lista lehet teljesen üres is). Ezek után kiszámítjuk a relatív átfedést minden lehetséges (A_i^k, B_j^k) pár között,

$$R_{ij}^k = \frac{|A_i^k \cap B_j^k|}{|A_i^k \cup B_j^k|}, \quad (4.3)$$

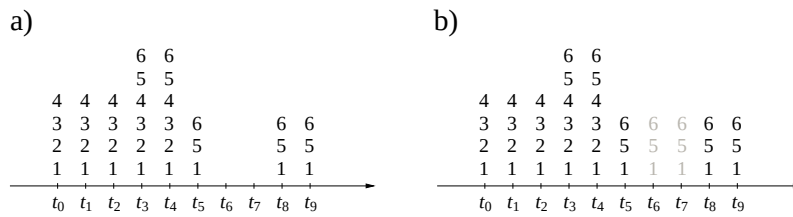
és a csoportpárokat a relatív átfedések sorrendjében feleltetjük meg egymásnak.

A 4.6. ábra egy egyszerű illusztrációt nyújt a fenti eljáráshoz. A 4.6a esetén mind az A_i^k és B_j^k összesen egy-egy csoportból áll, ezért ezeket egyből meg is lehet feleltetni egymásnak (akár a relatív átfedés kiszámítása nélkül is). Egy sokkal összetettebb a szituáció a 4.6b esetben, ahol A_i^k már két csoportból áll, melyek közül a kisebb A_1^k csoport 6, a nagyobb A_2^k csoport pedig 9 csúcsot tartalmaz. A B_j^k továbbra is csak egy csoportból áll (jelöljük ezt B_1^k -val), melynek összesen 15 tagja van. A relatív átfedésekre $R_{1,1}^k = 2/5$, illetve $R_{2,1}^k = 3/5$ adódik. Mivel $R_{2,1}^k > R_{1,1}^k$, a B_1^k előképe az A_2^k csoport lesz, ami által az A_1^k csoport megszűnik önálló csoportként tovább élni, és beleolvad a másik csoportba. Ezzel ellentétes

4. FEJEZET: Időfejlődő csoportdinamika

folyamat figyelhető meg a 4.6c ábrán. Itt A_i^k pusztán egy csoportból áll (legyen ez A_1^k), melynek 15 tagja van. Viszont $t+1$ -nél B_j^k most két csoportból áll, melyek közül B_1^k mérete 6 csúcs, míg B_2^k mérete 10 csúcs. A relatív átfedések ebben az esetben $R_{1,1}^k = 2/5$, illetve $R_{1,2}^k = 2/3$, emiatt az A_1^k csoportot B_2^k -vel fűzzük össze, melynél fogva B_1^k -t úgy kezeljük, mint egy új csoportot, mely A_1^k -ről vált le. Általánosan, ha V_k több csoportot fed le $\mathbf{A}$ -ból, mint ahányat $\mathbf{B}$ -ből, akkor azon A_i^k csoportok melyekhez nem tudunk párt biztosítani a B_j^k csoportok közül, megszűnnek létezni a t időlépés után. Hasonló módon, ha V_k több csoportot fed le $\mathbf{B}$ -ből, mint $\mathbf{A}$ -ból, akkor azon B_j^k csoportokat, melyekhez nem találunk előképet az A_i^k csoportok közül, úgy kell kezelnünk, mint egy $t+1$ -ben létrejött új csoportot. (A csoportok nyomon követésének itt bemutatott módját Derényi Imre ötlete nyomán a disszertáció szerzője dolgozta ki és alkalmazta a vizsgált hálózatokon.)

A tanulmányozott két hálózat esetében többször előfordult, hogy miután egy csoport megszűnt, néhány időlépés elteltével újra felbukkant. Valószínűnek tűnik, hogy ilyenkor igaziból nem esik szét, majd formálódik újra az adott csoport, pusztán egy hosszabb szünet következik be a csoporthoz kapcsolódó publikációs tevékenységben, illetve mobiltelefon hívásokban. Ennélfogva minden olyan esetben, amikor egy újonnan megszülető csoport teljes mértékben lefedett egy korábban már megszűnt csoportot, feltettük, hogy a csoport igaziból létezett egész végig, csak éppen az adatrögzítés sajátosságai miatt erről nem volt direkt információnk. Ilyen esetekben az újonnan létrejövő csoportot úgy kezeltük, mint a korábban megszűnt csoport folytatását, kitöltve a korábban észlelt megszűnés és az újraformálódás közti rést, ahogy azt a 4.7. ábra illusztrálja.



4.7. ábra. a) Egy csoport megszűnik a t_5 időlépés után, majd újjászületve ismét megjelenik a t_8 időlépésnél. b) Az ilyen eseteket úgy kezeljük, mintha a csoport a köztes t_6 és t_7 időlépéseknél is jelen lett volna ugyanazon tagokkal, mint t_5 -nél. (Az ábra forrása a [T7] publikáció kiegészítő része).

4.2. A csoport-időfejlődés statisztikus tulajdonságai

4.2.1. Alapvető statisztikák

A csoportszerkezethez tartozó egyik legegyszerűbb statisztika a csoportok összesített lefedése, ami megadja, hogy a csúcsok mekkora hányada került besorolásra legalább egy csoportba. A társszerzőségi hálózatnál ez valamivel 59% felett volt, ami egy elfogadható aránynak számít a k -klikkperkolációs módszer esetén. Ezzel szemben jelentősen alacsonyabb értéket kaptunk a mobiltelefon-hívási hálózatnál. Ennek az alapvető oka a rendkívül

nagy rendszerméret volt: Annak érdekében, hogy az időfejlődő csoportok előállítását kezelhető idő alatt lefusszon, az időlépésenként feltárt statikus csoportok számát csökkentenünk kellett. Emiatt magasabb k és w^* értékeket állítottunk be ($k = 4$, illetve $w^* = 1.0$), és ezen felül kénytelenek voltunk bevezetni egy olyan korlátozást, mely szerint csak azon csoportokat tartottuk meg, melyek legalább $s = 6$ felhasználóból álltak. E paraméterek mellett a csúcsok mindössze 11%-a került be a csoportokba. Mindazonáltal ez még mindig több, mint 400 000 felhasználót jelent átlagosan, ami nagyon jó eséllyel egy reprezentatív mintát szolgáltat az egész rendszerről. Ha levisszük a k -klikkméretet $k = 3$ -ra, a csoportokba sorolható csúcsok hányada felmegy 43%-ra.

Megjegyezzük, hogy mindkét hálózat esetén nagyon sok olyan csúcs van, ami ugyan nem tagja egy k -klikknek sem, viszont az összes élével egyazon k -klikkperkolációs csoporthoz kapcsolódik, ami által egyértelműen hozzásorolhatjuk az adott csoporthoz, ha tovább szeretnénk növelni a csoportok által lefedett csúcsok arányát. Ezen eljárást végrehajtva a csoportokba sorolt csúcsok aránya 72%-ra szökkent fel a társszerzőségi hálózat esetén, illetve szintén 72%-ra a $k = 3$ -as valamint 61%-ra a $k = 4$ -es paraméterű csoportok esetén a mobiltelefon-hívási hálózatonál.

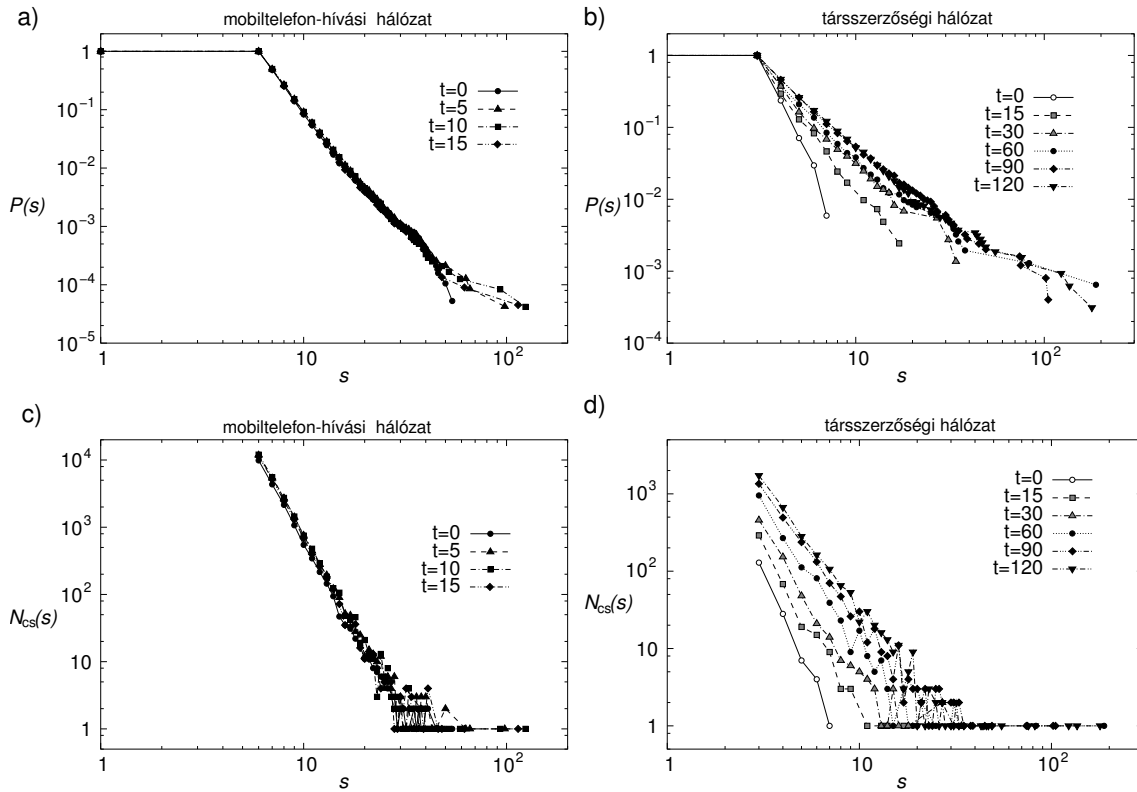
A csoportokhoz köthető másik alapvető statisztika a csoportok méreteloszlása. A 4.8a ábrán a statikus csoportokhoz tartozó méreteloszlást mutatjuk különböző időlépéseknél a mobiltelefon-hívási hálózatban. Látható, hogy az eloszlás egy nagy exponensú hatványfüggvényhez hasonlít. Hasonló módon, a 4.8b ábrán a társszerzőségi hálózat csoportméreteloszlását mutatjuk különböző időlépéseknél. Mivel ennél a rendszerenél a $t=0$ időlépés megegyezik az egész rendszer létrejöttének időpontjával, az egész hálózat meglehetősen kisméretű az első néhány időlépésnél. Ennek megfelelően a kezdeti időlépéseknél a csoportméreteloszlás is gyorsan lecseng, nincsenek igazán nagyméretű csoportok. Azonban az idő előrehaladtával egyre szélesebbé válik az eloszlás és beáll egy hatványszerűen lecsengő alak. A 4.8c-d ábrákon az adott csoportmérethez tartozó csoportok $N_{cs}(s)$ számát ábrázoltuk az s csoportméret függvényében a két rendszerben, különböző időlépéseknél. A mobiltelefon-hívási hálózat esetén $N_{cs}(s)$ az idővel nem igazán változik (4.8c ábra). Ezzel szemben a társszerzőségi hálózatonál a rendszer növekedése maga után vonja, hogy a feltárt csoportok száma is növekszik, ami jól látható a 4.8d ábrán.

Az időfejlődő csoportok esetén elsőként két mennyiséget vizsgáltunk, melyek a legalapvetőbb jellemzését adják egy csoportnak: a csoportok s méretét és τ korát, mely a csoport születése óta eltelt időlépések számával egyezik meg. A 4.9a ábra tanúsága alapján ez a két mennyiség erősen összefügg egymással, mely szerint az idősebb csoportok átlagosan nagyobb méretűek, mint a fiatalabbak. Ez teljesen természetes, hiszen egy csoport általában kis induló mérettel jön létre, és eltarthat egy darabig, amíg sikerül annyi tagot toboroznia, hogy már nagyméretű csoportnak számítsa.

Egy további természetes ötlet a csoportok időfüggő autokorrelációját tanulmányozni, hiszen ez megmutatja, hogy mennyit változott a csoport összetétele az eltelt idő alatt. Egy időfejlődő $A(t)$ csoport esetén az autokorrelációs függvényt a következő módon definiálhatjuk:

$$C_A(t) \equiv \frac{|A(t_0) \cap A(t_0 + t)|}{|A(t_0) \cup A(t_0 + t)|}, \quad (4.4)$$

4. FEJEZET: Időfejlődő csoportdinamika

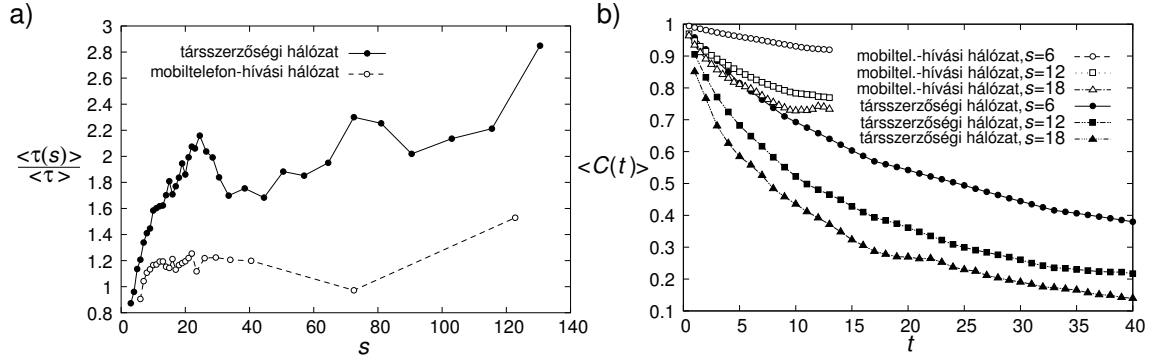


4.8. ábra. a) A csoportméret (kumulatív) eloszlásfüggvénye a statikus csoportok esetén, különböző időlépéseknél a mobiltelefon-hívási hálózatonál. b) Ugyanaz, mint az a) ábra társaszerzőségi hálózat esetén. c) Adott s csoportmérettel rendelkező csoportok $N_{cs}(s)$ száma különböző időlépéseknél a mobiltelefon-hívási hálózatban. d) Ugyanaz, mint a c) ábra a társaszerzőségi hálózat esetén. (Az ábra forrása a [T7] publikáció kiegészítő része).

ahol $|A(t_0) \cap A(t_0 + t)|$ megadja $A(t_0)$ és $A(t_0 + t)$ közös tagjainak számát, míg a nevezőben szereplő $|A(t_0) \cup A(t_0 + t)|$ egyszerűen az $A(t_0)$ és $A(t_0 + t)$ uniójában található csúcsok száma. A 4.9b ábrán az átlagos autokorrelációs függvény lecsengését mutatjuk be a két hálózatban, több csoportméret esetén. Az ábrából leolvasható, hogy mindkét hálózatban a nagyobb csoportok autokorrelációs függvénye gyorsabban cseng le, ami azt jelenti, hogy a nagyobb csoportok többnyire gyorsabban változtatják az összetételüket, mint a kisméretű csoportok.

4.2.2. Stacionaritás és élettartam

A 4.2.1. alfejezet eredményei alapján egy eltérés figyelhető meg a nagy- és kisméretű csoportok változékonyságában. Annak érdekében, hogy matematikai eszközökkel is jellemezzük egy csoport átlagos változékonyságát, bevezetjük a stacionaritás mennyiségét, mely a



4.9. ábra. a) Egy adott s csoportmérethez tartozó csoportok átlagos $\langle \tau(s) \rangle$ életkora az összes csoport átlagos életkorával, $\langle \tau \rangle$ -al leosztva a csoportméret, s függvényében. b) Különböző méretű csoportok átlagos autokorrelációs függvénye, $\langle C(t) \rangle$ az idő függvényében, ahol egy hónapot választottunk időegységnek. (Az ábra forrása a [T7] publikáció).

szomszédos időlépések közti autokorreláció átlaga a csoport élete során:

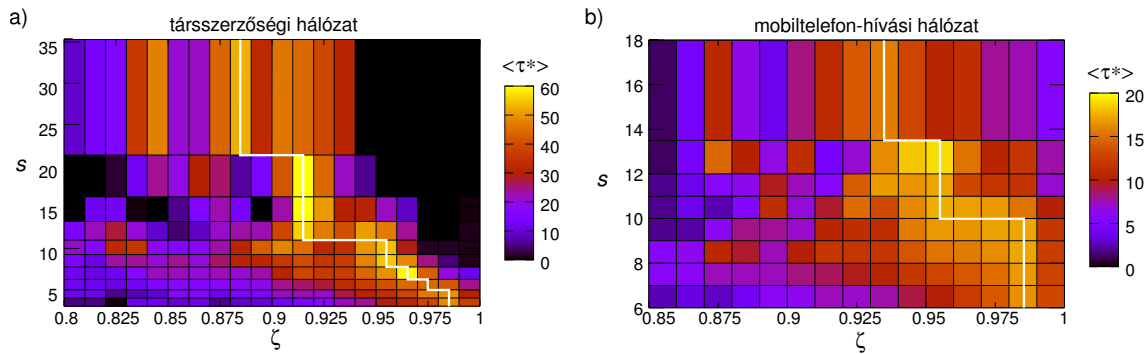
$$\zeta \equiv \frac{\sum_{t=t_0}^{t_{\max}-1} C(t, t+1)}{t_{\max} - t_0 - 1}, \quad (4.5)$$

ahol t_0 a csoport létrejöttének időpontja, $t_{\max}$ pedig az utolsó időlépés, melyben a csoport még létezik. Az elnevezés oka abban rejlik, hogy minél nagyobb a ζ értéke, annál kevesebbet változik a csoport átlagosan, hiszen $1 - \zeta$ az egy időlépés alatt megváltozó csoporttagok átlagos számának felel meg.

Az általunk tanulmányozott két hálózatban egy nagyon érdekes összefüggés figyelhető meg a csoportok élettartama, τ^* (a csoport létrejötte és megszűnése között eltelt időlépések száma), valamint az imént bevezetett stacionaritás és a csoportméret között. A csoport élettartamára tekinthetünk úgy is, mint egy „fitnesz” értékre: A fittebb, életrevalóbb csoportok tovább maradnak fent, míg a kevésbé fitt csoportok hamar szétesnek, vagy beolvadnak egy nagyobb csoportba. A 4.10. ábrán a $\langle \tau^* \rangle$ átlagos élettartamot a stacionaritás, ζ , és a csoportméret, s függvényében ábrázoltuk a tanulmányozott két hálózat esetén. Mindkét rendszerben a kisméretű csoportoknál az élettartam maximuma a nagy stacionaritás értékeknél található. Ez azt jelenti, hogy egy kisméretű csoport számára az az optimális, ha az összetétele csak alig változik az időben. Ezzel szemben a $\langle \tau^* \rangle$ maximuma a nagy csoportok esetén egy jelentősen alacsonyabb stacionaritás érték felé tolódik el, mely szerint a nagy csoportok számára viszont inkább az tűnik a túlélés zálogának, ha folyamatosan változtatják a tagösszetételüket. (Például a kis csoportoknál optimális magas ζ értékkel rendelkező nagyméretű csoportok átlagos élettartama nagyon alacsony).

A nagy- és kisméretű csoportok ezen eltérő viselkedését a 4.11. ábrán illusztráljuk, ahol négy csoport életét követhetjük végig a társszerződési hálózatból. A 4.11a ábrán mutatott kisméretű csoport csak alig változik, és ennek megfelelően viszonylag hosszú az élettartama. A mellékelt néhány pillanatfelvétel alapján a csoport központi „magját” három csúc alkotja, melyek végig kitartanak a csoport mellett, és ő hozzájuk csapódik egy időre egy

4. FEJEZET: Időfejlődő csoportdinamika

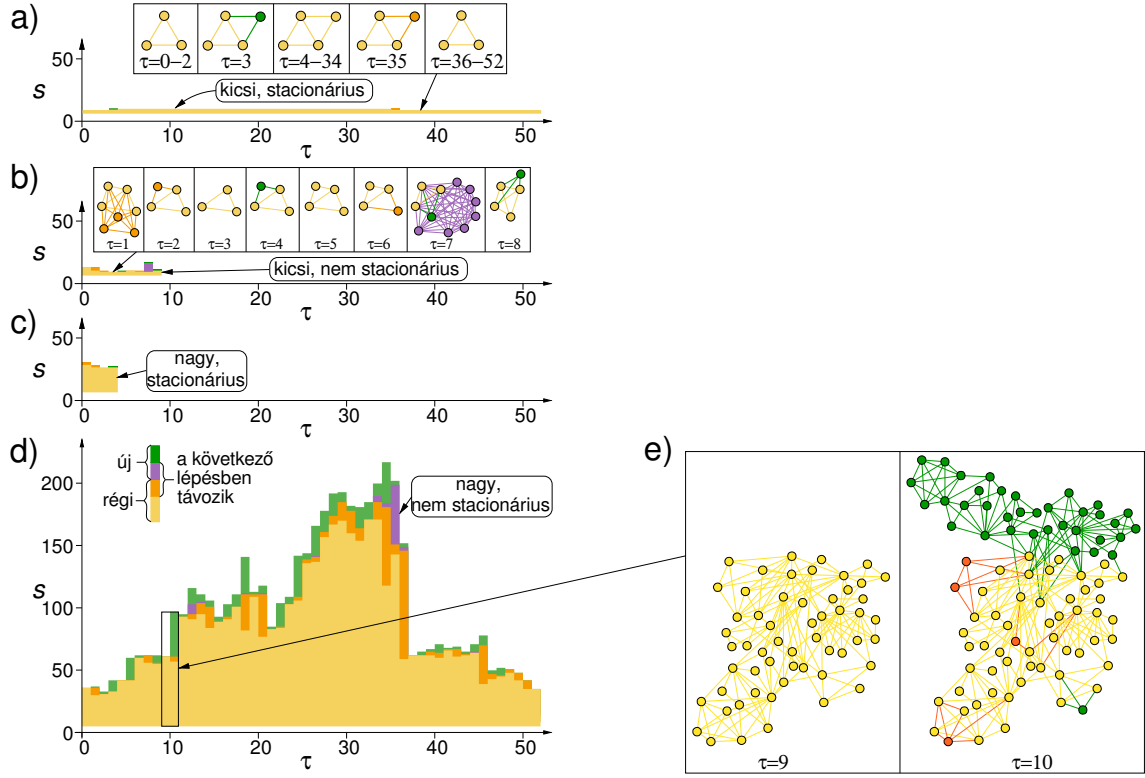


4.10. ábra. a) A csoportok átlagos élettartama, $\langle \tau^* \rangle$ a stacionaritás, ζ és a csoportméret, s függvényében a társszerzőségi hálózat esetén. A színárnyalatokkal jelzett görbe maximumhelyét egy fehér vonallal jelöltük. Látható, hogy a maximum nagy ζ értékeknél van, ha kicsi a csoportméret, míg az alacsonyabb stacionaritás felé tolódik el a nagyobb csoportok esetén. b) Ugyanezek az eredmények a mobiltelefon hívási hálózat esetén. (Az ábra forrása a [T7] publikáció).

negyedik csoporttag, de ezen kívül más változás nem következik be. Ezzel szemben lényegesen több változás figyelhető meg a 4.11b ábrán mutatott, szintén kisméretű csoport esetén, ami pusztán 9 időlépésen keresztül tud fennmaradni. Ezekkel ellentétes viselkedést láthatunk a nagyméretű csoportok esetén, hiszen például a 4.11c ábrán feltüntetett, nagyméretű és stacionárius csoport nagyon hamar megszűnik, miközben a 4.11d ábra nagyméretű és nagyon dinamikusan változó csoportja hosszú ideig létezik. Látható, hogy annak ellenére, hogy néha igen drámai változások következnek be a csoportösszetételében, a csoport „átvészeli” ezeket és nem szűnik meg egyhamar.

4.2.3. A csoportmegszűnés jóslása

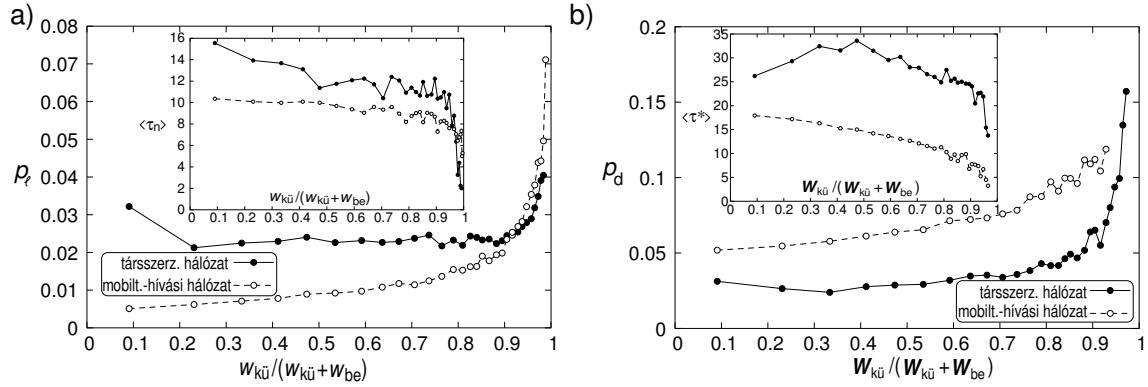
Az előző alfejezetben bemutatott eltérések a nagy- és kisméretű csoportok viselkedésében felvetik azt a kérdést, hogy vajon meg lehet-e jósolni egy csoport megszűnését bizonyos statisztikák alapján? Ennek vizsgálatát Barabási Albert-László javasolta először, a bemutatásra kerülő elemzéseket a disszertáció szerzője végezte el. A felvetett probléma kapcsán minden csoporttag esetén meghatároztuk a saját csoportjának többi tagjához kapcsolódó „belső” éleinek összesített súlyát (jelölje ezt w_{be}), és hasonló módon a csoporton kívül eső, egyéb csúcsokhoz kapcsolódó „külső” élek összesített súlyát (jelölje ezt $w_{kü}$). A 4.12a ábrán azt a p_ℓ valószínűséget ábrázoltuk $w_{kü}/(w_{be} + w_{kü})$ függvényében, mely megadja, hogy milyen eséllyel hagyja el csoportját az adott csoporttag a következő lépésben. Látható, hogy minél nagyobb a külső kapcsolatok relatív súlya, annál valószínűbbé válik, hogy az adott csúcs el fogja hagyni a csoportot. Ezzel párhuzamosan a 4.12a betétábrája alapján az is látszik, hogy a külső kapcsolatok relatív súlyának csökkenő függvénye a csúcs által eltöltött átlagos idő a csoporton belül. Ez egyben azt is jelenti, hogy azon csoporttagok maradnak a leghosszabb ideig hűségesek a csoporthoz, melyek relatíve sok időt és energiát fektetnek a többi taggal való együttműködésbe, illetve kommunikálásba.



4.11. ábra. Négy csoport időfejlődése a társaszerzőségi hálózatban. Az oszlopok magassága a méretnek felel meg, melyen belül a sárga szín a „rég” tagok arányát mutatja (ezek a csúcsok már az előző időlépésben is tagjai voltak a csoportnak), míg az újonnan jött tagok arányát a zöld szín jelöli. A narancs és lila színek azon tagok arányát jelölik, kik a következő lépésben el fogják hagyni a csoportot, ezen belül a narancs szín a régi, a lila az új tagoknak felel meg (ezek a csúcsok pusztán egy időlépésre csatlakoznak a csoporthoz). Felülről lefelé haladva bemutatunk egy kicsi és stacionárius csoportot (a), egy kicsi és nem stacionárius csoportot (b), egy nagy és stacionárius csoportot (c), végül egy nagy és nem stacionárius csoportot (d). Ez utóbbi csoport életéből két szomszédos időlépést külön is megjelenítettünk az (e) panelben. (Az ábra forrása a [T7] publikáció)

A fenti vizsgálatot könnyen kiterjeszthetjük a teljes csoportokra is. A csoportokon belüli élek összesített súlya (melyet W_{be} -vel jelölünk), megmutatja, hogy a csoport mint közösség mennyi időt, energiát szán „saját magára”, míg ezzel párhuzamosan a csoportot a hálózat többi részével összekötő külső élek összesített súlya (melyet $W_{kü}$ -vel jelölünk) a csoport által a csoporton kívüli csúcsokra fordított erőforrásokkal kapcsolatos. A 4.12b ábra szerint annak a p_d valószínűsége, hogy a csoport a következő lépésben megszűnik, egy emelkedő függvénye a $W_{kü}/(W_{be} + W_{kü})$ aránynak. Ezzel párhuzamosan a 4.12b betétábrája tanúsága szerint a csoport átlagos élettartama csökkenő tendenciát mutat $W_{kü}/(W_{be} + W_{kü})$ függvényében. Ezek alapján azoknál a csoportoknál számíthatunk hosszabb élettartamra, melyek elsősorban magukra fókuszálnak. Ezzel párhuzamosan azonban érdemes felfigyelni a 4.12b

4. FEJEZET: Időfejlődő csoportdinamika



4.12. ábra. a) A csoport elhagyásának p_ℓ valószínűsége a $w_{k\ddot{u}}/(w_{be} + w_{k\ddot{u}})$ arány függvényében, ahol a számláló a csoport tag külső kapcsolatainak erősségét adja meg, a nevező pedig a teljes összesített kapcsolat erősségének felel meg (külső plusz belső). A betétábra a csoportban eltöltött $\langle \tau_n \rangle$ átlagos időt mutatja szintén $w_{k\ddot{u}}/(w_{be} + w_{k\ddot{u}})$ függvényében. b) A csoport megszűnésének p_d valószínűsége a $W_{k\ddot{u}}/(W_{be} + W_{k\ddot{u}})$ arány függvényében, ahol a számláló a csoportból a hálózat egyéb részeihez kapcsolódó élek összesített súlyát adja meg, míg a nevező a csoport összes élének összesített súlyával egyenlő (külső plusz belső). A betétábra a $\langle \tau^* \rangle$ átlagos csoport élettartamot mutatja szintén $W_{k\ddot{u}}/(W_{be} + W_{k\ddot{u}})$ függvényében. (Az ábra forrása a [T7] publikáció kiegészítő része.)

betétábráján a társzerzőségi hálózatra vonatkozó görbe maximumhelyére, mely nem a minimális $W_{k\ddot{u}}/(W_{be} + W_{k\ddot{u}})$ értéknél van, mint a mobiltelefonhívási hálózat esetén, hanem egy közepes értéknél, nagyjából a megfigyelt teljes $W_{k\ddot{u}}/(W_{be} + W_{k\ddot{u}})$ intervallum felénél. Ez azt jelenti, hogy a társzerzőségi hálózathoz azok lesznek a leghosszabb ideig fennmaradó csoportok, melyeknél nagyjából egyensúlyban van a csoporton belüli együttműködések és a csoporton kívüli kutatókkal folytatott együttműködések összesített súlya. Összességében azt mondhatjuk, hogy mind az egyéni csoporttagok, mind a teljes csoport szintjén az adott csoport felé illetve a kifelé irányuló ráfordítások nyomon követése jelentősen segítheti a csoport jövőjének megjölését.

5. fejezet

Multifraktál alapú véletlengráf-modell

A természetben nagyon gyakran fordul elő fraktálgeometria, melynek legjellemzőbb tulajdonsága az ön hasonlóság és a törtértékű dimenzió [128]. Egy fraktálgeometria által meghatározott fizikai problémánál a skalár mennyiségek (mint pl. az elektromos térerősség, koncentráció, stb.) eloszlása többnyire egy multifraktált követ [208]: végtelen sok szinguláris pontot találunk, melyek közelében az eloszlás hatványszerűen divergál, viszont a hatványkitevő mindig a lokális környezettől függ, így pontonként más és más értéket vesz fel. Ilyenkor az adott hatványkitevővel jellemezhető szinguláris pontok szintén egy-egy fraktálon helyezkednek el, melyek jellemzői (pl. a fraktáldimenzió) más és más a különböző hatványkitevők esetén.

Ebben a fejezetben a statisztikus fizikában jól ismert multifraktáloknak egy olyan hálózati elméleti alkalmazását mutatjuk be, melynek segítségével egy nagyon általános véletlengráf-generáló algoritmust lehet megalkotni. A Bevezetésben már említettük, hogy a különböző véletlengráf-modellek mint például az Erdős–Rényi-modell, a Watts–Strogatz-modell vagy a Barabási–Albert-modell szerves részét képezik a komplex hálózatok területének. Egy-egy ilyen modell ideális eszközt nyújt arra, hogy egy végletekig leegyszerűsített képen tanulmányozzuk a komplex hálózatokat és ez által megérthessük egy kiszemelt mennyiség (pl. klaszterezettség, fokszámoszlás, stb.) viselkedését. A modellek segítségével különféle hipotézisek is tesztelhetők olyan mérettartományokban, melyeket a valós adatok oldaláról nem érhetünk el. A fent említett három kiemelt fontosságú modell mellett számos további véletlengráf-modell került bevezetésre az utóbbi mintegy 15 évben. Ezek azonban többnyire vagy egy kiszemelt tulajdonságra koncentráltak (mint például skálafüggetlenség), vagy egy viszonylag jól meghatározott hálózattípust vizsgáltak (mint például az emberek közti kapcsolathálókat). Emiatt az ezen modellek keretein belül előállítható hálózatok tulajdonságai csak viszonylag szűk tartományon változtathatók, a modellezni kívánt rendszertől gyökeresen eltérő tulajdonságokkal rendelkező hálózatot nem lehet velük generálni.

A fenti okok miatt 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)

5. FEJEZET: Multifraktál alapú véletlengráf-modell

szempontjából. A legegyszerűbb modell, melynek alapgondolata ebbe az irányba mutat az ún. konfigurációs modell, mely tetszőleges fokszámoszlású véletlen hálózat generálását teszi lehetővé [135, 136]. Amennyiben a fokszámoszlást egy valós rendszert reprezentáló hálózatról mintázzuk, a véletlen gráfot a gyakorlatban az élek véletlenszerű átkötögetésével állítjuk elő az eredeti hálózatból, mely folyamat során természetesen minden csúcsnál megőrizzük a fokszámot. Ezt az eljárást nem nehéz módosítani úgy, hogy a véletlen átkötögetések során egy általunk választott tulajdonság (pl. klaszterezettség) egy előre meghatározott módon változzon [188].

Bár egy ilyen megközelítés könnyen célra vezethet, nem igazán segít feltárni az adott tulajdonságokkal rendelkező hálózatok esetleges rejtett összefüggéseit. Ennek révén van létjogosultsága a következőekben röviden vázolt generatív módszereknek, ahol élkötögetés helyett előzményektől mentesen, az alapoktól indulva generálunk le egy véletlen hálózatot valamilyen szabályok alapján. E módszerek közös tulajdonsága, hogy ahelyett, hogy egy adott tulajdonságra (pl. skálafüggetlen fokszámoszlás) vagy egy adott rendszertípusra koncentrálnának (pl. szociológiai hálózatok), általános eljárást adnak meg véletlen gráfok előállítására.

Ezen a vonalon egy figyelemre méltó megközelítést nyújt a rejtett változókon alapuló modellek családja [42, 28]. Az alapötlet itt az, hogy minden csúcsához társítunk egy z rejtett paramétert, melyet egy előre definiált $\rho(z)$ eloszlásból húzunk. Meg kell még adni ezen felül egy szimmetrikus $p(z, z')$ függvényt is, melynek segítségével egy adott i, j csúcspár összekötésének valószínűsége a $p(z_i, z_j)$ alakban áll elő. Ennek révén egy előre megadott fokszámoszlással és foksám-korrelációkkal rendelkező, véletlen hálózat generálható megfelelő $\rho(z)$ és $p(z, z')$ választása esetén. Ezen felül bizonyos nemegyensúlyi, növekvő hálózatok is megfeleltethetők a vázolt, rejtett paraméteren alapuló modellnek, amennyiben z a csúcs korával van összefüggésben. Továbbá G. Bianconi rávilágított arra is, hogy rögzített fokszámoszlású, rögzített foksám-korrelációjú, vagy rögzített csoportszerkezetű véletlengráf-sokaságok entrópiájának tanulmányozása is összekapcsolható a rejtettparaméter-moddellel [22].

Egy alapvetően eltérő alternatívát kínál a P. Mahavedan és munkatársai által javasolt, dK -sorozatokon alapuló módszer, melynek kiinduló pontja a hálózati topológiák szisztematikus vizsgálata [127]. A dK -sorozatok olyan eloszlások sorozatainak felelnek meg, melyek meghatározzák egy d méretű részgráfon belül tapasztalható foksám-korrelációkat: a OK visszaadja az átlagos fokszámot, az $1K$ reprodukálja a fokszámoszlást, a $2K$ a szomszédos csúcspárok együttes fokszámoszlását, stb. A [127] publikáció több módszert is felsorol egy előre definiált dK -sorozattal rendelkező véletlen gráf legyártására (tipikusan $d \leq 3$ esetére).

A dK -sorozatokhoz hasonló módon az exponenciális véletlengráf-modell is azon a gondolon alapszik, hogy egy hálózatot jellemezhetünk megadott topológiájú részgráfok előfordulási gyakoriságaival (mely részgráfok egy növekedő sorozatot alkotnak a bennfoglalt csúcsok száma szerint) [74, 215, 183]. Ebben a megközelítésben minden figyelembe vett $\mathcal{G}$ részgráf (pl. két szomszédos csúcs, egy közös csúcson elhelyezkedő élpár, melyet „két csillagnak” szokás hívni, egy háromszög, stb.) kap egy $\vartheta_{\mathcal{G}}$ paramétert, mely az adott részgráf gyakoriságával kapcsolatos. Ezek alapján a teljes hálózat egy adott konfigurációjának valószínűsége az $\exp(\sum \vartheta_{\mathcal{G}} n_{\mathcal{G}})$ mennyiséggel arányos, ahol $n_{\mathcal{G}}$ az adott részgráf előfordulá-

sainak számával egyenlő. A ϑ paramétereket egy adott hálózat esetén általában maximum likelihood módszerrel szokták becsülni. A „két csillagokra” koncentráló speciális esetben J. Park és M. E. J. Newman egy nemperturbatív analitikus megoldással álltak elő, mely érdekes fázisátalakulást mutatott az alacsony és a magas élsűrűségű hálózatok között [160].

A dK -sorozatok és az exponenciális véletlengráf-modell alulról építkező megközelítések: Az első lépésben az előforduló lehető legegyszerűbb struktúrára koncentrálunk, mely nem más, mint egy csúcspár között húzódó él. Ha az élek számát sikerült reprodukálni, akkor továbblépünk, és egy fokkal bonyolultabb struktúra előfordulási gyakoriságát is megpróbáljuk visszaadni az adott modell keretein belül, stb. Az egymás után figyelembe vett, fokozatosan növekvő és egyre bonyolultabb szerkezetű részgráfok egyfajta hierarchiába rendeződnek, hiszen az összetettebb struktúrák gyakoriságának reprodukálása nem ronthatja el a korábban már helyesen visszaadott egyszerűbb részgráfok gyakoriságait. Azonban egy valós hálózat elemzése során tipikusan csak az első néhány tagot vesszük figyelembe a fenti sorozatokból, hiszen egyfelől a leglényegesebb tulajdonságokat már ezek is reprodukálják, másfelől az egyre magasabb rendű tagok figyelembe vétele egy bizonyos ponton túl nagyon számításigényessé válik.

Természetesen olyan általános véletlengráf-modellek is születtek az irodalomban, melyeknél a strukturális hierarchia nem alulról felfelé, hanem felülről lefelé építkezik, azaz az első lépésben egy nagy méretskálán adjuk meg a hálózat szerkezetét, majd ezt követően a közepes, végül a kis léptékű struktúrákat definiáljuk. Ezeknél a módszereknél a különböző méretskálákon ugyanolyan, vagy nagyon hasonló szabályok határozzák meg a szerkezetet, amit például a természetben gyakran előforduló önhasonló struktúrák, fraktálok is motiválnak. A hierarchikus illetve önhasonló szerkezet és a fraktálszerű tulajdonságok vizsgálata a hálózatkutatásban a generatív modelleken túl is egy fontos területet képez [175, 16, 194, 169, 49, 207, 205, 137, 52, 51, S10, S11, S12].

Az egyik figyelemre méltó, inherensen hierarchikus általános véletlengráf-generáló algoritmust V. A. Avetisov és munkatársai publikálták [10]. A megközelítésük lényege, hogy a gráf szomszédsági mátrixát egy p -edik randomizált, lokálisan konstans Parisi-mátrixnak feleltetjük meg, mely a spinüvegelmélet egyik központi objektuma [133]. Ez egy hierarchikus belső szerkezettel rendelkező szimmetrikus mátrix, melynek elemei Bernoulli-eloszlású valószínűségi változók, melyek értéke g_γ valószínűséggel 1, és $1 - g_\gamma$ valószínűséggel 0, ahol γ adja meg, hogy az adott él a hierarchia melyik szintjéhez tartozik. E konstrukció egyik következménye, hogy egy adott γ -hoz tartozó hierarchiaszinten lévő élek által kifestett részgráf ekvivalens egy Erdős–Rényi-gráffal. Ennek ellenére a teljes hálózat fokszámeloszlása lehet skálafüggetlen.

Szintén hierarchikus szerkezetű szomszédsági mátrixok szolgálgják a J. Leskovec és társai által bevezetett Kroneckergráf-módszer alapját is [120]. Ebben az esetben egy kisméretű szomszédsági mátrixból indulunk ki, melynek elemeit minden iterációnál helyettesítjük az adott elemnek és magának a kezdeti mátrixnak a szorzatával (ezt hívják Kronecker-szorzásnak). Ennek révén a szomszédsági mátrix mérete az l -edik iteráció után a kezdeti mátrix méretének l -edik hatványával egyenlő. Az algoritmus egy sztochasztikus verziójában a kezdeti mátrix eredetileg 0 vagy 1 értéket felvevő elemeit a $[0,1]$ intervallumból származó valós számokra cserélve a végeredményül kapott mátrix elemeit értelmezhetjük

5. FEJEZET: *Multifraktál alapú véletlengráf-modell*

úgy is, mint csúcspáronként definiált élbekötési valószínűségeket. Az ez alapján legyártott véletlen gráfok (megfelelő kezdeti mátrix választása esetén) képesek a valós rendszereket leíró hálózatok több tulajdonságát is reprodukálni (pl. lassan lecsengő fokszámeloszlás, kisvilág-tulajdonság, stb.). Ezen felül a [121] publikációban J. Leskovec és C. Faloutsos megadtak egy gyakorlati módszert a modell valós rendszereket reprezentáló hálózatokhoz való illesztésére is.

Az előző két megközelítésnél alkalmazott élvalószínűségi mátrixok a hálózati csoportkeresés szempontjából is fontos objektumok, hiszen nagyon hasonló mátrixok előfordulnak például a 3. fejezetben már említett sztochasztikus blokk-modellek esetén [50, 161, 164, 165, 166, 4], illetve a Nepusz Tamás és munkatársai által javasolt módszer esetén is [142, 141] (melyet a Szemerédi-féle regularitási lemma inspirált [203]). Ebben a kontextusban az élvalószínűségi mátrix diagonális blokkjai felelnek meg a csoportokon belüli éleknek, a nem diagonális blokkok pedig a csoportok közti éleknek.

A fejezet további részében az általunk javasolt, multifraktálokra alapuló általános véletlengráf-modell részleteit ismertetjük. A bemutatásra kerülő eredmények a *PNAS* folyóiratban jelentek meg a disszertáció szerzőjének első szerzőségével [T8]. Az általunk bevezetett modell egyfelől hasonlít a Parisi-mátrixra illetve a Krocnecker-szorzáson alapuló megközelítésekre abban, hogy a hálózat struktúráját kódoló objektum hierarchikus, mely egy rekurzív eljárás során keletkezik, és ez által önhasznós a szerkezete. Ugyanakkor a módszerünk nagyon erősen támaszkodik Lovász László és munkatársainak gráfsorozatok konvergenciájával kapcsolatos korábbi eredményeire [125, 34, 124], illetve közeli rokonságot mutat a Bollobás Béla és munkatársai által bevezetett véletlengráf-generáló eljárással is [31, 33]. Ennek megfelelően az 5.1. alfejezetben röviden vázoljuk a konvergens gráfsorozatok és a $[0,1]$ intervallumon értelmezett szimmetrikus kétdimenziós függvények közti kapcsolatot, mielőtt rátérnénk a módszerünk részletes ismertetésére az 5.2.-5.3.4. alfejezetekben. A multifraktálok használatát véletlen gráf generálásra elsőként Vicsek Tamás javasolta. A modell részleteinek kidolgozását, a módszer implementálását, valamint a különféle hálózatjellemzők analitikus és numerikus meghatározását a disszertáció szerzője végezte el. A modell termodinamikai határesetével kapcsolatos analitikus eredmények Lovász Lászlónak köszönhetők.

5.1. Gráfsorozatok termodinamikai határesete

Az általunk javasolt általános véletlengráf-modell alapötletét Lovász László és Szegedy Balázs gráfsorozatok konvergenciájával kapcsolatos, a [125] publikációban közölt vizsgálataik adták, ezért ebben az alfejezetben röviden vázoljuk a gráfsorozatok termodinamikai határesetével kapcsolatos legfontosabb eredményeket. A termodinamikai határeset tanulmányozása nagyon fontos például statisztikus fizikai rendszerek esetén, ugyanis számos jelenség, tulajdonság ebben a határesetben jelentkezik a legletisztultabb formájában. Hasonló módon, a gráfsorozatok végtelen méretű határesete egy fajta „plátói” gráfnak felel meg a hálózatelméletben, mely a gráfsorozat tagjainak közös tulajdonságait a lehető legtisztább módon mutatja meg. De vajon milyen feltételek mellett mondhatjuk azt egy gráfsorozatról, hogy egy értelmes (és nemtriviális) objektumhoz konvergál? Egy egyszerű és

intuitív feltétel az, hogy a hálózatok jellemzésére használt statisztikáknak mint például a fokszámoszlás, klaszterezettség, stb. konvergálniuk kell ahhoz, hogy a sorozat konvergens lehessen.

Mint látni fogjuk, a konvergencia feltétele ennél jóval erősebb, ugyanis azt gráf-homeomorfizmusokon keresztül adták meg. (A gráf-homeomorfizmus egy olyan leképezés, mely megtartja a szomszédsági viszonyokat). Jelöljük az $\mathcal{F}$ és $\mathcal{G}$ egyszerű gráfok esetén $\text{hom}(\mathcal{F}, \mathcal{G})$ -vel az $\mathcal{F}$ csúsaiból a $\mathcal{G}$ csúcsaira képező homeomorfizmusok számát. Ennek alapján a $\varrho(\mathcal{F}, \mathcal{G})$ homeomorfizmus-sűrűség definíciója a létező homeomorfizmusok és az $\mathcal{F}$ csúcshalmazából a $\mathcal{G}$ csúcshalmazába képező leképezések lehetséges számának hányadosa,

$$\varrho(\mathcal{F}, \mathcal{G}) = \frac{\text{hom}(\mathcal{F}, \mathcal{G})}{|V(\mathcal{G})|^{|V(\mathcal{F})|}}, \quad (5.1)$$

ahol $V(\mathcal{F})$ és $V(\mathcal{G})$ az $\mathcal{F}$ illetve a $\mathcal{G}$ csúcsainak halmazát jelölik (és ennek alapján pl. $|V(\mathcal{F})|$ az $\mathcal{F}$ csúcsainak számával egyenlő). Egy $\mathcal{G}_n$ gráfsorozat konvergens, ha a $\varrho(\mathcal{F}, \mathcal{G}_n)$ sorozat minden véges egyszerű $\mathcal{F}$ gráfra konvergens. Egy leegyszerűsítő megfogalmazásban ez azt jelenti, hogy bármely véges részgráf előfordulási valószínűsége konvergens a sorozatban.

A konvergens gráfsorozatok és az egységnégyzeten értelmezett szimmetrikus függvények nagyon érdekes kapcsolatban állnak egymással [125, 34, 124]. Egyfelől egy szimmetrikus és mérhető $0 \leq L(x, y) \leq 1$ függvény segítségével könnyedén generálhatunk egy konvergens gráfsorozatot a következő módon. Egy adott N gráfméret esetén véletlenszerűen és egyenletesen elosztunk N pontot a $[0, 1]$ intervallumon, majd a pontokat képzeletben felmásoljuk az egységnégyzet vízszintes és függőleges oldalára is. Minden pont egy csúcsnak felel meg a legyártandó gráfban, és egy adott pontpár közti él valószínűsége az $L(x, y)$ -nal egyenlő a pontok által meghatározott x, y koordinátáknál. Az $N \rightarrow \infty$ határesetben az ilyen módon előállított gráfsorozat konvergens lesz. Ennél még meglepőbb, hogy a dolog fordítva is igaz: minden konvergens gráfsorozathoz tartozik egy, az egységnégyzeten értelmezett mérhető, szimmetrikus $L(x, y)$ függvény, mely pontosan az adott sorozathoz tartozó részgráfgyakoriságokat generálja a fenti módon.

Egy adott $L(x, y)$ függvény alapján a fenti módon generált véletlen hálózat átlagos foksámát a

$$\langle d \rangle = N \int \int L(x, y) dx dy \quad (5.2)$$

összefüggéssel lehet kiszámolni, mely azt mutatja, hogy az $N \rightarrow \infty$ határesetben kapott gráf sűrű. Ezzel szemben a valós rendszereket leíró hálózatok általában ritkák. Bollobás Béla és munkatársai emiatt úgy módosították az egységnégyzeten értelmezett szimmetrikus mérhető függvény alapján történő véletlengráf-generálás eljárását, hogy az élbekötési valószínűség $L(x, y)/N$ -el legyen egyenlő [31, 33]. Ennek révén a generált gráfok átlagos foksáma méretfüggetlenné válik, és megfelelő $L(x, y)$ választásával a ritka gráfok egy széles spektruma legyártható.

5.2. Véletlengráf-generálás multifraktállal

A multifraktál alapú véletlengráf-generálásnak három fő fázisa van: Elsőként definiálunk egy ún. generáló mértéket az egységnégyzeten, majd ezt egy rekurzív eljárás keretében néhány iteráción keresztül önhasonló módon átranszformáljuk a konkrét élbekötési valószínűségeket megadó mértékké. Végül legyártjuk a véletlen hálózatot az adott élvalószínűségek mellett, hasonló módon, mint ahogy azt az 5.1. alfejezetben vázoltuk az $L(x, y)$ függvény esetén. A generáló mérték megadásánál az egységnégyzet x és y tengelyeit meg egyező módon felosztjuk m darab, nem feltétlen egyforma méretű intervallumra. Ezáltal az egységnégyzetet m^2 darab téglalapra bontottuk, melyek a főátlóra szimmetrikusan helyezkednek el. Minden egyes téglalaphoz hozzárendelünk egy p_{ij} valószínűséget, ahol az $i, j \in [1, m]$ indexek a téglalapok sor- és oszlopindexeinek felelnek meg. E p_{ij} valószínűségekre kikötjük, hogy normáltak és szimmetrikusak legyenek, azaz a $\sum p_{ij} = 1$ és $p_{ij} = p_{ji}$ azonosságoknak teljesülnie kell.

Az élvalószínűségi mérték előállításánál q darab iteráción keresztül minden egyes négyzetet rekurzívan megszorozunk magával a generáló mértékkel, ahogy azt az 5.1. ábra szemlélteti. Ennek révén a kapott élvalószínűségi mérték ekvivalens a generáló mérték q -adik tenzoriális hatványával. Az eredményül kapott, m^{2q} téglalaphból álló élvalószínűségi mértéknél minden esetben összesen q tényező szorzata adja meg az adott mezőhöz társított valószínűséget a

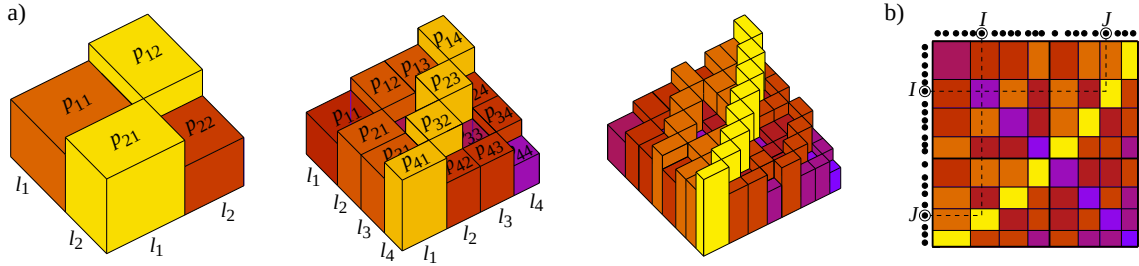
$$p_{ij}(q) = \prod_{h=1}^q p_{i_h j_h} \quad (5.3)$$

formában, ahol természetesen a szorzótényezők mind a generáló mértéknél megadott p_{ij} valószínűségek valamelyikének felelnek meg. Azt, hogy pontosan melyiknek, az i_h és j_h indexek határozzák meg az

$$i_h = \left\lfloor \frac{(i-1) \prod_{r=1}^{h-1} m^{q-r}}{m^{q-h}} \right\rfloor + 1 \quad (5.4)$$

alakban, ahol $\lfloor a/b \rfloor$ az a/b egész részét jelöli, valamint $\prod_{r=1}^{h-1} m^{q-r}$ az m^{q-r} -nel való osztás maradékának kiszámítását jelenti rekurzívan ismételve. (Pl. $h = 1$ esetén az (5.4) kifejezés az $i_h = \lfloor (i-1)/m^{q-1} \rfloor + 1$ összefüggéssé egyszerűsödik.) Látható, hogy a $q = 1$ eset magának a generáló mértéknek felel meg, és természetesen egy, az (5.4) kifejezéssel teljesen analóg formula írható fel j_h -ra is.

Ez az eljárás teljesen megegyezik egy multifraktál előállításával, melynek során megálunk a q -adik iterációnál. Az 5.1a ábrán látható, egyre durvuló felület esetén minden mező a generáló mértékben megadott mintázatot követve bomlik kisebb téglalapokra a következő iteráció során. Ennek megfelelően ezen kisebb téglalapokhoz társított valószínűségek megegyeznek a kiinduló mezőhöz tartozó valószínűség és a generáló mérték megfelelő elemének szorzatával. (Ezzel párhuzamosan az új kisebb téglalapok és a kiinduló mező területének aránya szintén a generáló mérték megfelelő mezőjének területével azonos). Ez a konstrukció egyfelől biztosítja, hogy a kialakuló struktúra önhasonló legyen, másfelől azt eredményezi,



5.1. ábra. A multifraktál alapú véletlengráf-generáló eljárás szemléltetése. a) Egy, az egységnégyzeten definiált, $m \times m$ téglalaphból álló generáló mértékből indulunk ki, mely az ábra bal oldalán látható. A bemutatott példában $m = 2$, a téglalapok oldalhosszait megadó intervallumok l_1 és l_2 , a téglalapokhoz társított p_{ij} valószínűségeket a téglalapok fölé rajzolt hasábok magassága és a színe kódolja. A generáló mértékkel rekurzívan megszorozunk minden téglalapot q iteráción keresztül, így áll elő az $m^q \times m^q$ téglalaphból álló élvalószínűségi mérték. A hasábok magasságának szórása nő az iterációk számával, ezért az élvalószínűségi mérték egy egyre durvuló felülethez hasonlít, mely ugyanakkor végig szimmetrikus és önhasonló marad. b) Minden csúcs kap egy véletlenszerűen választott koordinátát a $[0,1]$ intervallumon, és egy adott I, J csúcspár összekötésének valószínűsége az élvalószínűségi mértéknek a két koordináta által meghatározott helyen felvett értékével egyenlő. (Az ábra forrása a [T8] publikáció.)

hogy az egyes mezőkhöz társított élbekötési valószínűségek szórása az iterációk számával egyre nő, azaz a kapott „felület” egyre durvább.

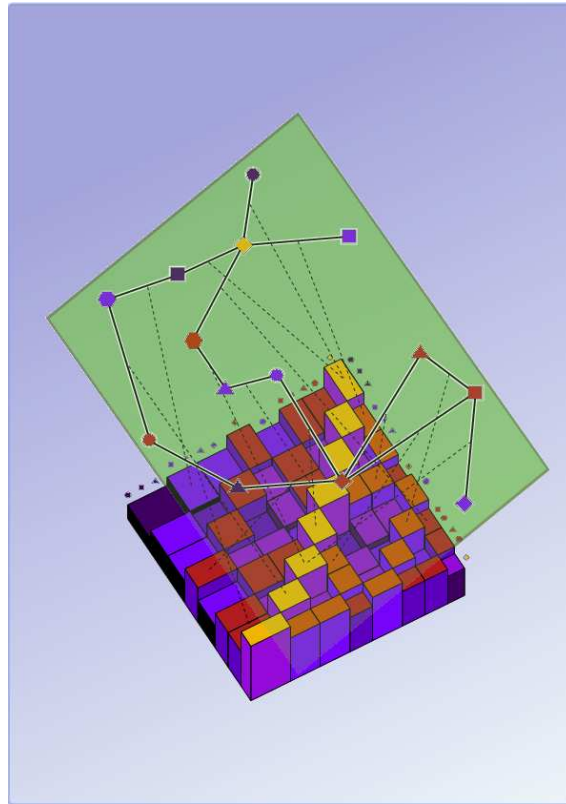
A véletlen gráf generálásának utolsó lépése megegyezik az 5.1. alfejezetben említett, kétváltozós szimmetrikus $L(x, y)$ függvényen alapuló hálózatgenerálással: minden egyes csúcs kap egy, a $[0,1]$ intervallumon véletlenszerűen választott koordinátát, majd a csúcspárokat a megfelelő koordinátáknál található $p_{ij}(q)$ valószínűséggel kötjük össze. Ezt szemlélteti az 5.1b és az 5.2. ábra. Továbbá az 5.3. ábrán bemutatunk egy 500 csúcsból álló véletlen gráfot, melyet ezzel a módszerrel generáltunk, a hozzá tartozó multifraktállal együtt.

Módszerünk még általánosabbá tehető, ha a „standard”, mezőnként konstans multifraktált lecseréljük egy szimmetrikus, az egységnégyzeten értelmezett $0 \leq L(x, y) \leq 1$ függvényre, és ennek képezzük a k -szoros tenzoriális szorzatát. (Erre a lehetőségre Lovász László hívta fel a figyelmet). Bár az eredményül kapott $L(x_1, \dots, x_k, y_1, \dots, y_k)$ függvény értelmezési tartománya $[0,1]^{2k}$ a szokásos $[0,1]^2$ helyett, ez a probléma könnyen áthidalható egy mértékmegőrző bijekcióval a $[0,1]$ és a $[0,1]^q$ között. Ennek révén lényegében ugyanúgy lehetne véletlen gráfokat generálni a tenzorszorzat eredményeként kapott függvénnyel, mint a fent bemutatott multifraktállal.

Fontos megjegyezni, hogy az általunk javasolt konstrukcióban a generált hálózat átlagos fokszáma a

$$\langle d \rangle = N \sum_{i=1}^{m^q} \sum_{j=1}^{m^q} p_{ij}(q) a_{ij}(q) \quad (5.5)$$

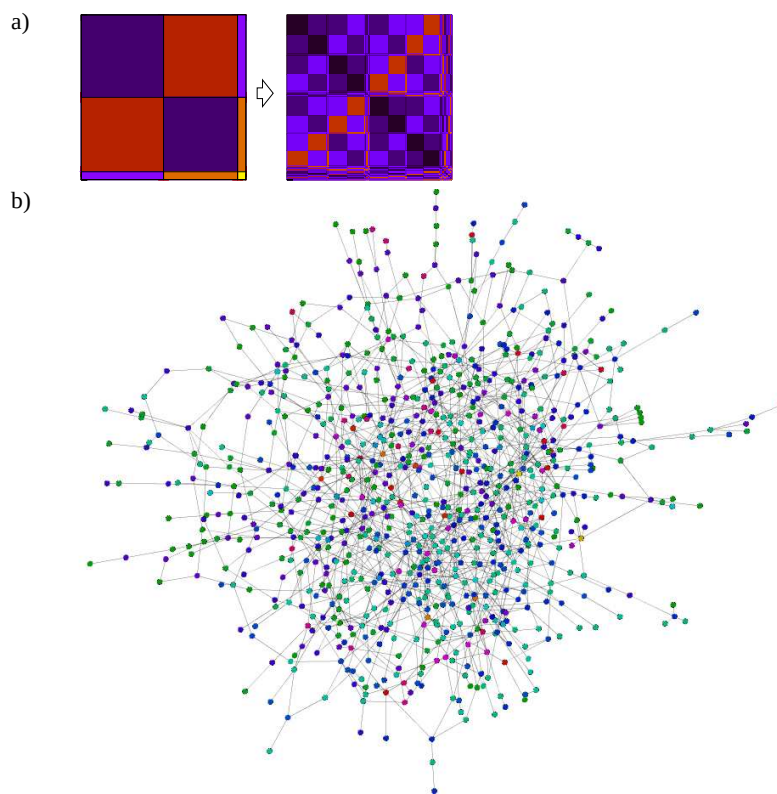
formulával adható meg, ahol $a_{ij}(q)$ az i, j indexekkel rendelkező mező területe a q -adik ite-



5.2. ábra. A véletlen hálózat előállítása a multifraktál alapján. A színes hasábok az egység-négyzeten kialakuló élvalószínűségi mértéket jelenítik meg, mely megfelel egy multifraktálnak. A multifraktál felett lévő zöld síkban mutatjuk a kapott hálózatot, melynek éleit a multifraktál által meghatározott valószínűségekkel húzzuk be az adott csúcspárok közé.

rációnál. Amennyiben a generáló mérték csupa egyforma alapterületű mezővel rendelkezik, úgy az élvalószínűségi mérték is egyforma területű mezőkből fog állni, melyekre az $a_{ij}(q) = m^{-2q}$ összefüggés lesz érvényes. Ilyenkor a $p_{ij}(q)$ normáltságából fakadóan az 5.5 egyenlet a $\langle d \rangle = Nm^{-2q}$ összefüggésre egyszerűsödik. Ebből azt szűrhetjük le, hogy amennyiben ritka hálózatot szeretnénk generálni (ami a valós rendszerek modellezésénél alapkövetelmény), illetve ha még pontosabban a hálózat átlagos fokszámát konstansnak szeretnénk tartani az iterációk során, úgy a generált hálózat méretét exponenciálisan növelni kell az iterációk számával.

Összefoglalásul elmondhatjuk, hogy a módszerünk nagyon szorosan kapcsolódik az egység-négyzeten értelmezett, szimmetrikus $0 \leq L(x, y) \leq 1$ függvényeken alapuló véletlengráf-generáló módszerekhez [125, 34, 31, 33]. Egy fontos újdonság azonban az itt bevezetett modellnél, hogy az iterációk számának növelésével egyre többféle $p_{ij}(q)$ valószínűség szerepel az élvalószínűségi mértékben, melynek révén a generált hálózat is egyre struktúráltabbá válik az iterációk során. Ebben a tekintetben tehát módszerünk a Kronecker-szorzáson [120, 121] illetve a Parisi-mátrixokon alapuló megközelítésekre hasonlít [10], hi-



5.3. ábra. Egy multifraktál alapján generált véletlen hálózat. a) A generáló mérték és a belőle származtatott élvalószínűségi mérték $q = 3$ iteráció esetén. b) Egy 500 csúcsból álló hálózat, melyet az a) ábrán látható élvalószínűségi mérték alapján állítottunk elő. A csúcsok színezése a hozzájuk tartozó véletlen pozíciójú pont koordinátájára utal: az élbehúzási mérték minden sorához (és oszlopához) egy külön színt rendeltünk (mely véletlenszerű, és nem egyezik meg az a) ábra színekódolásával), és ezt a színt „örökölték” a csúcsok is. Ennek révén az azonos sorba (oszlopba) eső csúcsok azonos színnel rendelkeznek. (Az ábra forrása a [T8] publikáció.)

szen a multifraktál előállításánál használt rekurzió lényegében megfeleltethető a Kronecker-szorzásnak. Természetesen egy rögzített élvalószínűségi mérték esetén módszerünk tekinthető a rejtettparaméter-modellek egy speciális esetének is, ahol a csúcsokhoz társított koordináták felelnek meg a rejtett paraméternek és a $p_{ij}(q)$ élvalószínűségi mérték az $p(z, z')$ függvénynek. Azonban a q növelésével a mi modellünkben ez a $p(z, z')$ függvény egyre komplexebbé válik.

5.3. Statisztikák meghatározása analitikus úton

A multifraktál alapú véletlengráf-modell egyik előnye, hogy a legenerált hálózatok legfontosabb statisztikáit (mint pl. a fokszámeloszlás) viszonylag egyszerű analitikusan meghatároz-

5. FEJEZET: Multifraktál alapú véletlengráf-modell

ni. Az első fontos megállapítás ezzel kapcsolatban az, hogy az élvalószínűségi mérték azonos mezőjébe eső csúcsok statisztikailag ekvivalensek egymással, azaz például a foksám vagy a klaszterezettségi együttható várható értéke rájuk vonatkozóan ugyan az. Ennek megfelelően a teljes hálózatra vonatkozó eloszlások is az egyes sorokhoz (oszlopokhoz) tartozó eloszlások összegeként állnak elő.

5.3.1. Fokszámeloszlás

A fentiek alapján a fokszámeloszlás a

$$\rho^{(q)}(d) = \sum_{i=1}^{m^q} \rho_i^{(q)}(d) l_i(q), \quad (5.6)$$

alakban adható meg, ahol $\rho_i^{(q)}(d)$ felel meg az i -edik sor csúcsaihoz tartozó eloszlásnak és $l_i(q)$ adja meg az adott sor szélességét (és ezzel együtt az adott sorba eső csúcsok arányát az összes csúcshoz viszonyítva). A $\rho_i^{(q)}(d)$ eloszlásokat a következő módon lehet meghatározni. A véletlen gráf legyártása során az i -edik sorból a j -edik sor felé összesen $n_j(q)$ -szor próbálunk éleket bekötni $p_{ij}(q)$ valószínűséggel, ahol $n_j(q)$ a j -edik sorba eső csúcsok száma, $n_j(q) = N l_j(q)$. Ha n_j elég nagy, az i -edik sorból a j -edik sorba menő élek eloszlását egy Poisson-eloszlással lehet közelíteni,

$$\rho_{ij}^{(q)}(d) = \frac{\langle d_{ij}(q) \rangle^d}{d!} e^{-\langle d_{ij}(q) \rangle}, \quad (5.7)$$

ahol $\langle d_{ij} \rangle$ az i -edik sorból a j -edik sorba menő élek átlagos száma, $\langle d_{ij} \rangle = n_j(q) p_{ij}(q)$. Egy i -edik sorban lévő csúcs foksáma a $d_i(q) = \sum_j d_{ij}(q)$ alakban írható fel, ahol $d_{ij}(q)$ természetesen a j -edik sor felé menő élek számát jelenti. Ennek alapján a $\rho_i^{(q)}(d)$ -hoz tartozó generátorfüggvény az egyes $\rho_{ij}^{(q)}(d)$ -hez tartozó generátorfüggvények szorzata:

$$G_i^{(q)}(x) = \prod_j G_{ij}^{(q)}(x), \quad (5.8)$$

ahol

$$G_{ij}^{(q)}(x) = \sum_d \rho_{ij}^{(q)}(d) x^d = \sum_{d=0}^{\infty} \frac{\langle d_{ij}(q) \rangle^d}{d!} e^{-\langle d_{ij}(q) \rangle} x^d = e^{\langle d_{ij}(q) \rangle (x-1)}. \quad (5.9)$$

Ha behelyettesítjük az (5.9) kifejezést az (5.8) egyenletbe akkor a

$$G_i^{(q)}(x) = \prod_j e^{\langle d_{ij}(q) \rangle (x-1)} = e^{(x-1) \sum_j \langle d_{ij}(q) \rangle} = e^{(x-1) \langle d_i(q) \rangle} \quad (5.10)$$

eredményre jutunk, ahol kihasználtuk, hogy az élek függetlensége miatt az i -edik sor csúcsainak várható foksáma $\langle d_i \rangle = \sum_j \langle d_{ij} \rangle$. Az (5.10) kifejezést visszatranszformálva megkapjuk

az i -edik sor csúcsainak fokszámoszlását a

$$\rho_i^{(q)}(d) = \frac{\langle d_i(q) \rangle^d}{d!} e^{-\langle d_i(q) \rangle} \quad (5.11)$$

alakban. Mivel $\langle d_{ij} \rangle = n_j(q)p_{ij}(q)$, az i -edik sor csúcsainak várható fokszáma a fenti kifejezésben a

$$\langle d_i(q) \rangle = N \sum_j p_{ij}(q) l_j(q) \quad (5.12)$$

formában adható meg.

Az (5.6-5.11) egyenletek megfeleltethetők a rejtettparaméter-modellek esetén a [28] publikációban levezetett eredményeknek. Bár az élvalószínűségi mérték egyes soraira vonatkozó fokszámoszlások egyenként Poisson-eloszlást követnek, a teljes hálózat fokszámoszlása könnyűszerrel vehet fel nemtriviális alakot, amint azt később konkrét példákön keresztül is demonstráljuk.

5.3.2. Klaszterezettség együttható

A klaszterezettség együttható esetén is abból indulunk ki, hogy a mennyiség várható értéke konstans az élvalószínűségi mérték egyazon sorába eső csúcsokra. Az i -edik sor csúcsainak klaszterezettségét megkaphatjuk úgy, hogy a csúcsokat tartalmazó háromszögek számát elosztjuk a csúcsokból induló összes élpár számával. Az említett háromszögek szintén az adott csúcsokból induló élpároknak felelnek meg azzal az extra megkötéssel, hogy az élpár másik végeit egy harmadik élnek kell összekötnie. Ez alapján a q -adik iterációnál az i -edik sor csúcsainak várható klaszterezettség együtthatója a

$$\langle K_i(q) \rangle = \frac{\frac{1}{2} \sum_{j=1}^{m^q} [l_j(q)]^2 [p_{ij}(q)]^2 p_{jj}(q) + \sum_{j=1}^{m^q} \sum_{h=j+1}^{m^q} l_j(q) l_h(q) p_{ij}(q) p_{ih}(q) p_{jh}(q)}{\frac{1}{2} \sum_{j=1}^{m^q} [l_j(q)]^2 [p_{ij}(q)]^2 + \sum_{j=1}^{m^q} \sum_{p=j+1}^{m^q} l_j(q) l_p(q) p_{ij}(q) p_{ip}(q)} \quad (5.13)$$

alakban írható fel. A számláló és nevező első tagja azon i -beli csúcsból induló élpároknak felel meg, melyek másik vége ugyanazon j sorban van, míg a második tagok azon élpárokat adják, melynél az élpár másik két vége különböző sorokba esik.

5.3.3. Szomszédok átlagos fokszáma

A foksám-korrelációk jellemzésének egyik alapvető eszköze a $d_{\text{ksz}}(d)$ függvény, mely megadja, hogy mi lesz egy d foksámú csúcs közvetlen szomszédainak átlagos fokszáma. Természetesen ez a mennyiség is analitikusan számítható az általunk javasolt modellben. A k -adik iterációnál az élvalószínűségi mérték i -edik sorába eső csúcsok szomszédainak átlagos foksámát a

$$d_{\text{ksz},i}^{(q)} = \frac{\sum_{j=1}^{m^q} p_{ij}(q) l_j(q) \langle d_j(q) \rangle}{\sum_{j=1}^{m^q} p_{ij}(q) l_j(q)} \quad (5.14)$$

5. FEJEZET: Multifraktál alapú véletlengráf-modell

alakban lehet felírni. A d fokszámú csúcsok szomszédainak átlagos fokszámát az élvalószínűségi mérték soraira való összegzéssel lehet megadni, ahol a fenti $d_{\text{ksz},i}^{(q)}$ értékek azon $p^{(q)}(i|d)$ feltételes valószínűséggel vannak megszorozva, mely megadja, hogy az adott csúcs az i -edik sorba esik, feltéve, hogy a fokszáma d :

$$d_{\text{ksz}}^{(q)}(d) = \sum_{i=1}^{m^q} p^{(q)}(i|d) d_{\text{ksz},i}^{(q)}. \quad (5.15)$$

E feltételes valószínűség a következő módon határozható meg. Az i -edik sorba eső, d fokszámú csúcsok száma $n_i(q)\rho_i^{(q)}(d)$, míg az összes d fokszámú csúcsok száma $n\rho^{(q)}(d)$. Annak valószínűsége, hogy egy csúcs az i -edik sorba esik, feltéve, hogy d a fokszáma ez alapján

$$p^{(q)}(i|d) = \frac{n_i(q)\rho_i^{(q)}(d)}{n\rho^{(q)}(d)} = \frac{l_i(q)\rho_i^{(q)}(d)}{\rho^{(q)}(d)}. \quad (5.16)$$

Az (5.16) és az (5.14) kifejezéseket az (5.15) egyenletbe helyettesítve a

$$d_{\text{ksz}}^{(q)}(d) = \frac{1}{\rho^{(q)}(d)} \sum_{i=1}^{m^q} l_i(q)\rho_i^{(q)}(d) \frac{\sum_{j=1}^{m^q} p_{ij}(q)l_j(q)\langle d_j(q) \rangle}{\sum_{j=1}^{m^q} p_{ij}(q)l_j(q)}. \quad (5.17)$$

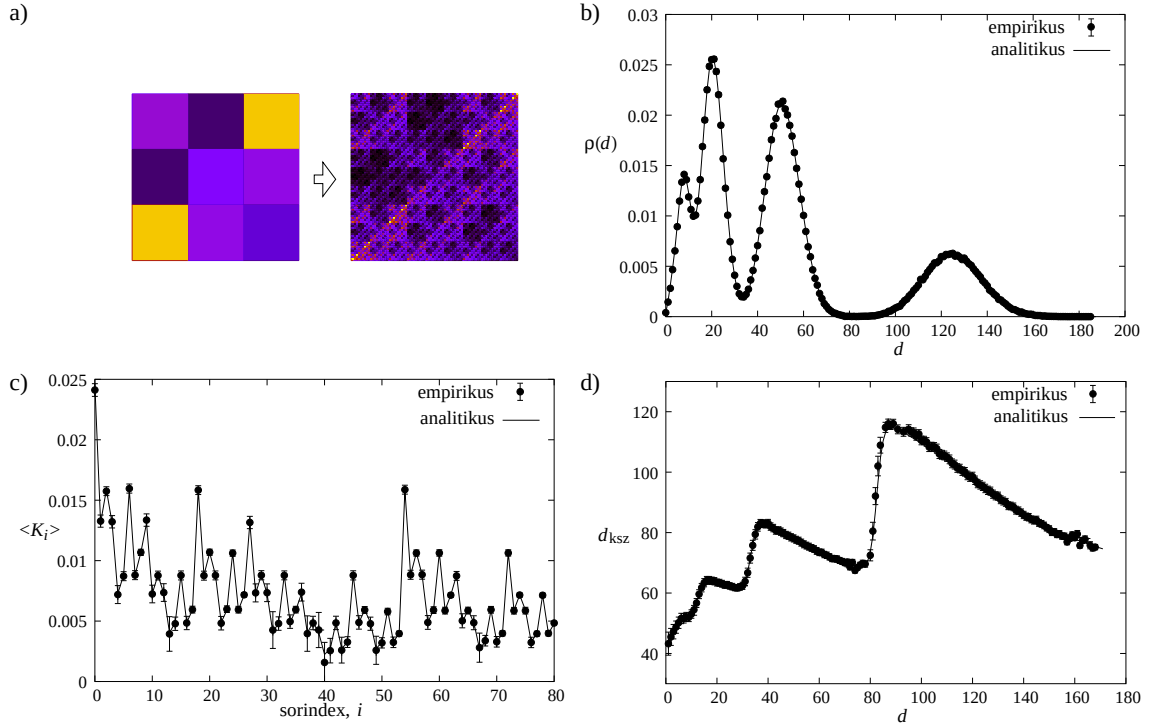
eredményre jutunk.

Az 5.4. ábrán egy véletlenszerű generáló mérték alapján legyártott hálózat fokszám-eloszlásán, klaszterezettségi együtthatóján és fokszám-korrelációin keresztül ellenőrizzük a felsorolt statisztikákra vonatkozó analitikus formulák helyességét. Az egyszerűség kedvéért a generáló mértéket 9 darab egyforma alapterületű mezőre osztottuk, a hozzájuk társított p_{ij} valószínűségeket véletlenszerűen választottuk. Ezek után a $q = 4$ iteráció után kapott élvalószínűségi mérték alapján generált, $N = 5000$ csúcsból álló hálózatok esetén hasonlítottuk össze az empirikusan mért fokszám-eloszlást, klaszterezettségi együtthatót és fokszám-korrelációt az analitikus formulákkal. Az ábra alapján az analitikus eredmények nagyon pontosan egyeznek az empirikus adatokkal.

A fentiek alapján a legfontosabb hálózati statisztikák egyszerűen és pontosan számíthatók pusztán az élvalószínűségi mérték és a hálózat mérete alapján. A modell ezen tulajdonsága nagyon előnyös akkor, amikor például adott fokszám-eloszlású hálózatra szeretnénk optimalizálni a generáló mértéket, hiszen az eloszlások fent részletezett kiszámítása sokkal gyorsabb, mint kellő számú minta legenerálása, majd az eloszlások empirikus meghatározása. Ilyen jellegű eljárásra konkrét példát az 5.4. alfejezetben mutatunk be.

5.3.4. Az izolált csúcsok problémája

Érdekes külön figyelmet fordítani a multifraktál alapú hálózatgeneráló termodinamikai háttérre, melynél $N \rightarrow \infty$. Amint azt az 5.2. alfejezetben említettük, ha növekvő iteráció szám mellett azonos sűrűségű hálózatokat szeretnénk generálni, a csúcsok számát exponenciálisan kell növelni a q függvényében. Azonban látni fogjuk, hogy ilyen körülmények között $N \rightarrow \infty$ esetén a csúcsok döntő többsége izolálttá válik, míg a hálózat „érdekes”



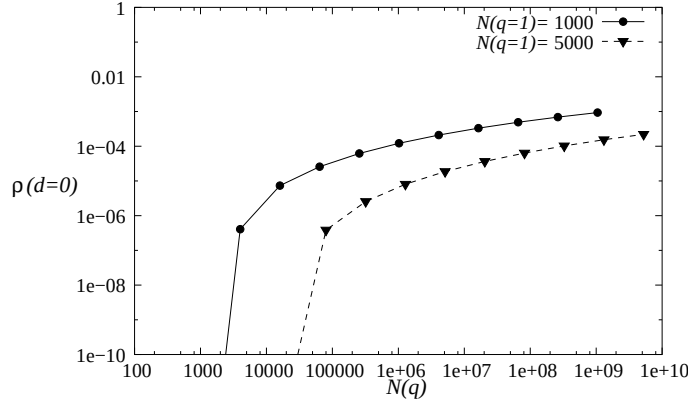
5.4. ábra. Az analitikus eredmények ellenőrzése egy véletlenszerű generáló mérték esetén. a) A generáló mérték és a belőle származó élvalószínűségi mérték $q=4$ iteráció esetén. b) Az élvalószínűségi mérték alapján legyártott, $N = 5000$ csúcsból álló hálózat $\rho(d)$ foksámeloszlása. A folytonos vonal az (5.6–5.11) egyenleteken alapuló analitikus eredményt mutatja, a szimbólumok 100 darab legenerált minta empirikus átlagának felelnek meg. Az empirikus szórásnak megfelelő hibahatárok kisebbek a szimbólumméretnél, ezért nem látszanak az ábrán. c) Az élvalószínűségi mérték egy adott sorába eső csúcsok átlagos klaszterezettségi együttthatója a sor indexének függvényében. A folytonos vonal az (5.13) egyenlet eredményét mutatja, a szimbólumok az empirikus eredményeknek felelnek meg. d) A közvetlen szomszédok átlagos fokszáma, d_{ksz} analitikus számítások alapján (folytonos vonallal), és a numerikus eredmények szerint (szimbólumokkal) a foksám függvényében. (Az ábra forrása a [T8] publikáció.)

része, melyben az élek találhatók, az egész rendszer egyre kisebb hányadát tölti ki. Szerencsére ez az effektus csak extrém nagy rendszerméreteknél jelentős, a valós rendszereket leíró hálózatok mérettartományában lényegében elhanyagolható, ezért a módszer esetleges gyakorlati alkalmazását nem befolyásolja.

A modellünk termodinamikai határesetére vonatkozó analitikus számításokat Lovász László végezte el, az erre vonatkozó eredményeket a D függelék vázolja röviden. Itt az izolált csúcsok problémájához kapcsolódó numerikus eredményeket ismertetjük, melyek a valós rendszereket leíró hálózatok mérettartományára vonatkoznak. Véletlenszerű generáló mértékekből indultunk ki, melyek összesen 4×4 egyforma méretű mezőből álltak, az ite-

5. FEJEZET: Multifraktál alapú véletlengráf-modell

rációk számát $q = 1$ -től $k = 11$ ig növeltük. A csúcsok kezdő száma $q = 1$ -nél 1000, illetve 5000 volt, a további iterációk során pedig N -et az (5.5) egyenlet alapján úgy állítottuk be, hogy az átlagos fokszám konstans maradjon. Az 5.5. ábrán mutatjuk be a futtatások



5.5. ábra. Az izolált csúcsok hányada a generált hálózatokban véletlenszerű generáló mértékek esetén $q = 1$ -től $q = 11$ -ig, 1000 mintára átlagolva. A körökkel jelölt mintáknál a csúcsok száma $q = 1$ -nél $N(q=1) = 1000$ volt, míg a háromszögekkel jelölt esetekben $N(q=1) = 5000$.

eredményeit. Látható, hogy annak ellenére, hogy az utolsó iterációknál a hálózatok mérete gyakran meghaladta 10^9 csúcsot, az izolált csúcsok átlagos hányada nagyon alacsony maradt. Ennek révén az izolált csúcsok térhódítása nem veszélyezteti komolyan a multifraktál alapú véletlengráf-generáló olyan lehetséges alkalmazásait, melyek a valós rendszereket leíró hálózatok mérettartományára koncentrálnak.

Végül megjegyezzük, hogy egy későbbi publikációban (melyet a disszertáció szerzője Pollner Péterrel és Vicsek Tamással közösen közölt) sikerült egy olyan módosítást kidolgozni a multifraktál alapú véletlengráf-modellnek, melynél az izolált csúcsok aránya konvergált az iterációk számának (vagy ekvivalens módon a rendszerméret) függvényében [S13]. A kapcsolódó vizsgálatok során világossá vált, hogy az izolált csúcsok problémája szoros összefüggésben van azzal, hogy a fokszámeloszlást a multifraktál egy kitüntetett irányba eső projekciója határozza meg. A multifraktál elforgatásával ez a projekció már nem lesz kitüntetett, és a szimulációk alapján ez orvosolja a problémát.

5.4. A gráfgeneráló optimalizálása

A generáló mértéktől, az iterációk és a csúcsok számától függően széles skálán mozoghatnak a legyártott hálózatok statisztikus tulajdonságai. Azonban ha egy előre meghatározott tulajdonságokkal rendelkező véletlen gráfot szeretnénk a multifraktál alapú hálózatgenerálóval legyártani, valahogyan optimalizálni kell az említett paramétereket. Tegyük fel, hogy a hálózat csúcsainak száma rögzített. Ebben az esetben a következő paraméterek merülnek fel: a generáló mértékben található különböző mezők száma, m^2 , valamint az iterációk száma, q . Ha ezeket rögzítettük, a generáló mértékhez tartozó p_{ij} valószínűségek,

illetve a generáló mérték különböző mezőit elválasztó határvonalak pozíciói automatikusan beállíthatók a következő módon.

Jelölje $\mathcal{F}$ azt a mennyiséget, melyikre optimalizálni kívánunk (pl. $\mathcal{F}$ lehet a fokszámeloszlás vagy a klaszterezettség együttható, illetve bármely egyéb hálózatjellemző). Általános esetben $\mathcal{F}$ függ a p_{ij} , l_i , q és N mennyiségek mindegyikétől (és természetesen implicit módon m -től is, a generáló mértéket megadó mezők pontos elhelyezkedésén keresztül). Mivel feltételezésünk szerint az N , m és q értéke rögzített, ezek külön jelölését elhagyjuk, és ennek révén $\mathcal{F}$ egy adott p_{ij} és l_i beállításához tartozó „értékét” $\mathcal{F}(p_{ij}, l_i)$ -vel jelöljük. (Az esetek többségében $\mathcal{F}(p_{ij}, l_i)$ egy valós szám helyett egy magasabb dimenziós objektumot jelöl, pl. a fokszámeloszlást).

A célunk, hogy a legyártott véletlen hálózat esetén az $\mathcal{F}$ lehetőleg minél közelebb essen egy $\mathcal{F}^*$ „értékhez”. Ehhez először be kell vezetnünk egy távolság- vagy hasonlóságmértéket, mely megadja, hogy az aktuális $\mathcal{F}(p_{ij}, l_j)$ milyen messze van a céltól (vagy mennyire hasonlít rá). A p_{ij} és l_i értékeket egy szimulált hőkezelésen alapuló eljárással lehet optimalizálni, melynél az említett távolságfüggvény (vagy hasonlóságfüggvény) játssza az energia szerepét, melyet általánosan $E[\mathcal{F}(p_{ij}, l_i), \mathcal{F}^*]$ -gal jelölhetünk. Természetesen ezen energiafüggvény konkrét alakja nagyban függ attól, hogy éppen milyen mennyiségre szeretnénk optimalizálni a multifraktál alapú hálózatgenerálót. Amennyiben például a fokszámeloszlásra szeretnénk optimalizálni, egy kézenfekvő választás a sűrűségfüggvények összegzett relatív különbsége,

$$E[\mathcal{F}(p_{ij}, l_i), \mathcal{F}^*] = E[\rho^{(k)}(d), \rho^*(d)] = - \sum_d \frac{|\rho^{(k)}(d) - \rho^*(d)|}{\max(\rho^{(k)}(d), \rho^*(d))}, \quad (5.18)$$

ahol d a lehetséges fokszámon fut végig, $\rho^{(k)}(d)$ a ténylegesen kapott sűrűségfüggvény értéke d -nél, és $\rho^*(d)$ a célul kitűzött sűrűségfüggvény értéke szintén d esetén.

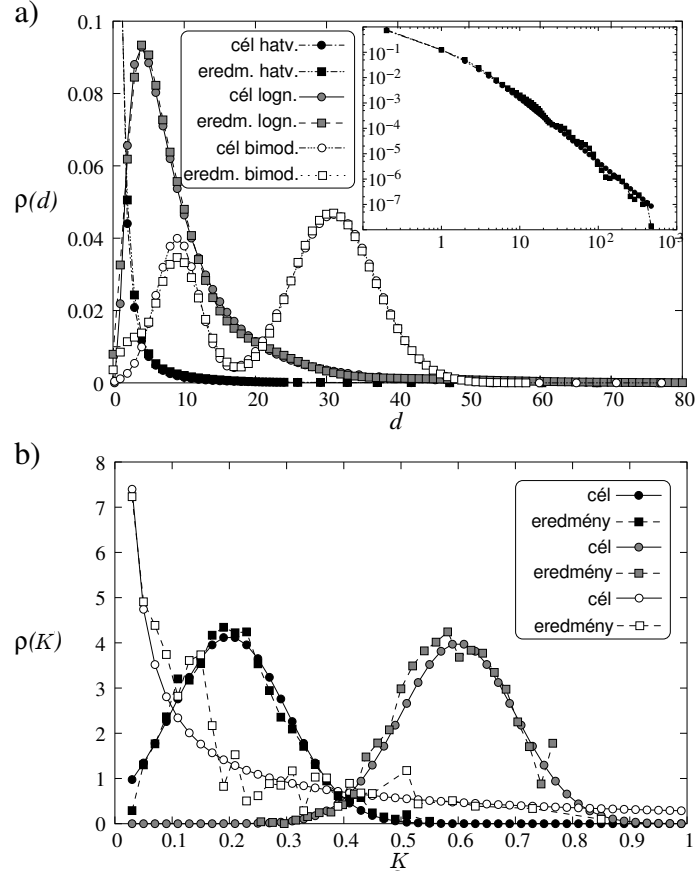
A szimulált hőkezelés során emellett definiálunk egy T effektív hőmérsékletet is, melynek értékét a folyamat során lassan csökkentjük. A hőkezelés maga sok elemi Monte-Carlo-lépésből áll, melyeknél vagy az egyik p_{ij} valószínűséget, vagy az egyik mezőhatárt próbáljuk kis mértékben megváltoztatni a generáló mértékben. A lépések elfogadása vagy elvetése a Metropolis–Hastings-algoritmus szerint történik [132, 95]: Ha az új generáló mérték alapján kapható E_2 energia kisebb, mint a lépés előtt számított E_1 energia, akkor a lépést mindenképp elfogadjuk, az ellenkező esetben viszont csak $P = \exp[-(E_2 - E_1)/T]$ valószínűséggel.

Ezt az eljárást lehet általánosítani több mennyiség párhuzamos optimalizálására is. Egy alternatív lehetőség emellett a maximum likelihood technikák alkalmazása (mint pl. a KronFit algoritmus a [121] publikációban), melyeknél magára a hálózatszerkezetre lehet optimalizálni a származtatott mennyiségek helyett. A mi esetünkben azonban egy ilyen módszer formalizálása és implementálása lényegesen bonyolultabb lenne, mint a vázolt, szimulált hőkezelésen alapuló megközelítés. Az egyszerűség kedvéért ezért itt a különféle hálózatjellemzőkre külön optimalizálunk, két példán keresztül mutatva meg a módszer működését.

Az 5.6. ábrán a fokszámeloszlásra és a klaszterezettség együttható eloszlására történő optimalizálás során kapott eredményeket mutatjuk be. (Ahogy azt már kifejtettük, az

5. FEJEZET: *Multifraktál alapú véletlengráf-modell*

eljárás során a csúcsok száma, N , a generáló mérték sorainak és oszlopainak száma, m , illetve az iterációk száma, q , konstans volt, és csak a p_{ij} valószínűségek és a generáló mérték sorainak és oszlopainak szélességét megadó l_i értékek változtak.) Az állandó értéken tartott paraméterek tipikus értéke a szimulációk során $N \in [10000, 20000]$, $m \in [3, 4]$ és $q \in [3, 5]$ voltak. Az 5.6a ábra esetén annak ellenére, hogy meglehetősen különböző fokszámeloszlásokat tettünk meg célfüggvénynek (egy skálafüggetlen, egy lognormális és egy bimodális fokszámeloszlást), a módszer sikeresen kalibrálta a p_{ij} és l_i értékeket úgy, hogy a multifraktál alapú hálózatgenerálóval kapott hálózatok fokszámeloszlása nagyon pontosan lekövesse a célfüggvényeket. Hasonlóan, az 5.6b ábra szerint a klaszterezettségi együttható eloszlására történő optimalizálás is elfogadható eredményekkel zárult, hiszen az eredményül kapott eloszlásokkal sikerült megközelíteni mindhárom, egymástól meglehetősen különböző célfüggvényt.



5.6. ábra. A generáló mérték optimalizálása különféle célfüggvények esetén. a) Különböző fokszámeloszlásokra történő optimalizálás. A célfüggvényeket kör alakú szimbólumok mutatják, az eredményül kapott sűrűségfüggvényeket a négyzetek jelenítik meg. A fekete színnel jelölt esetről a célfüggvény egy skálafüggetlen fokszámeloszlás volt (a betétábra ezt mutatja logaritmikus skálán), a szürke színnel jelölt esetről egy lognormális eloszlás, míg a fehér színnek megfelelő esetről egy bimodális eloszlás. b) A klaszterezettségi együttható különböző eloszlásaira történő optimalizálás. Az előző panelhez hasonlóan a kör alakú szimbólumok jelölik a célfüggvényeket, az eredményül kapott eloszlások négyzetekkel lettek ábrázolva, és a különböző színek különböző alakú célfüggvényekkel végrehajtott kísérleteknek felelnek meg. (Az ábra forrása a [T8] publikáció.)

Összefoglalás

A dolgozat a szerző hálózatkutatással kapcsolatos legfontosabb eredményeit foglalta össze, melyek közös vonása, hogy nagyon erősen kötődnek a statisztikus fizikához. A komplex hálózatok területén a statisztikus fizikai megközelítés már a kezdetektől fogva nagyon erős befolyást szerzett, és még manapság is számtalan fontos, hálózatokkal kapcsolatos kutatás zajlik világszerte statisztikus fizikai háttérrel rendelkező kutatók vezetésével. Ezek alapján a disszertáció szerzőjének eredményei szervesen illeszkednek a komplex hálózatok területéhez, hozzájárulván annak az utóbbi 10-15 évben tapasztalható, meglehetősen intenzív fejlődéséhez.

Az 1. fejezet a topológiai fázisátalakulásokkal foglalkozott. Ezek kapcsán a disszertáció szerzője az egyrészcskés energiafüggvényekkel leírható rendszereket tanulmányozta a kanonikus sokaságban egyrészt analitikus úton, másrészt az egzakt leszámlálás numerikus technikájának segítségével. Az eredmények megmutatták, hogy az ilyen rendszerekben az energiafüggvény-választástól függően elsőrendű és folytonos fázisátalakulások is előfordulnak. Egy külön érdekesség, hogy az $E = - \sum d_i \ln d_i$ energia esetén a folytonos átalakulás kritikus pontjánál a hálózat skálafüggetlenné válik (azaz a fokszámeloszlás hatványszerűen cseng le a nagy foksámok tartományán).

A 2. fejezet a k -klikkperkolációval kapcsolatos elméleti eredményekre koncentrált. A disszertáció szerzője megadott egy generátorfüggvényeken alapuló levezetést a k -klikkperkoláció kritikus élbekötési valószínűségére az Erdős–Rényi-gráfban. Emellett meghatározta az irányított k -klikkperkoláció kritikus pontját is egy heurisztikus megközelítés segítségével. Az irányítatlan Erdős–Rényi-gráf esetén a disszertáció szerzője által végzett numerikus szimulációk maximálisan alátámasztották a k -klikkperkolációra vonatkozó analitikus eredményeket¹.

A k -klikkperkoláción alapuló, átfedéseket megengedő hálózati csoportkereső módszer a 3. fejezetben került bemutatásra. A csoportkereső algoritmust (mely feltérképezi egy hálózat k -klikkperkolációs klasztereit), a disszertáció szerzője és Derényi Imre közösen dolgozták ki. A módszert a disszertáció szerzője implementálta, és segítségével feltárta több nagyméretű, valós rendszert reprezentáló hálózat csoportszerkezetét. Egy-egy kiválasztott csúcs lokális környezetében a csoportok részletes elemzése megmutatta, hogy ez a megközelítés értelmes, az emberi intuíciónak megfelelő csoportokat tárt fel a vizsgált hálózatokban (pl. a társszerzőségi hálózat csoportjaiban egy kiválasztott kutató a környezetében található

¹Az irányított esetben is hasonlóan jó egyezés volt tapasztalható, azonban ott Farkas Illés végezte el a szimulációkat.

többi kutatóval a megfelelő kutatási területek szerint lett besorolva, a fehérje-kölcsönhatási hálózatban kiválasztott fehérje csoportjai különféle fehérjefunkcióknak feleltek meg, stb.)

A csoportátfedések statisztikai jellemzésére a disszertáció szerzője három, az irodalomban korábban nem vizsgált eloszlást vezetett be. Ezek közül a legérdekesebb egy véletlenszerűen választott csoporttal átfedő többi csoport számának eloszlása volt. Ez megegyezik azon hálózat fokszámoszlásával, melyben a csúcsok az eredeti hálózat csoportjainak felelnek meg, és két ilyen csúcs akkor van összekötve, ha az adott csoportpár átfed egymással. Az így kapott fokszámoszlás (a csúcsok közti „alap” hálózat fokszámoszlásához hasonlóan) hatványszerűen csengett le a nagy csoportfokszámok tartományán több vizsgált rendszerben is. Érdekes módon a kisebb csoportfokszámok tartományán viszont a hatványfüggvény helyett inkább az exponenciális eloszláshoz hasonló alakot vett fel a csoport-hálózat fokszámoszlása. Ennek alapján elmondhatjuk, hogy a tanulmányozott rendszerekben a csoportszintű hálózatstruktúra egyes vonásai hasonlítottak az alaphálózat szerkezetéhez, viszont emellett fontos eltéréseket is tapasztaltunk. 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.

A 4. fejezet a hálózati csoportok időfejlődésével foglalkozott két nagyméretű, emberi kapcsolatokat ábrázoló hálózatban. 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. A disszertáció szerzője kidolgozott egy módszert, melynek segítségével az időlépésenként rögzített eseményekből kiindulva olyan folytonosan változó élsúlyok állíthatók elő, melyek az egyes események (cikkek vagy aggregált telefonhívások) fontossága mellett figyelembe veszik azt is, hogy egy emberi kapcsolat erőssége időben általában lecseng, ha megszűnnek a két fél közötti interakciók. Ezen időfejlődő élsúlyok által definiált két hálózat esetén a disszertáció szerzője először az egyes időlépéseknek megfelelő állapotokban tárta fel a k -klikkperkolációs csoportokat. A mobiltelefon-hívási hálózat esetén a felhasználókról rendelkezésre álló egyéb adatok csoportszintű elemzése azt mutatta, hogy a kapott 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.

Ahhoz, hogy csoportdinamikát lehessen tanulmányozni, az egyes időlépésekhez tartozó statikus csoportokat még össze kellett fűzni időben változó, dinamikus csoportokká. Az ezen műveletet lehetővé tévő módszert Derényi Imre ötlete alapján a disszertáció szerzője dolgozta ki és programozta be, majd az ezáltal kinyert időfejlődő csoportokat több szempontból is elemezte. A csoportok időbeli változékonyságára bevezetett egy új mennyiséget, a stacionaritást, mely megegyezik a szomszédos időlépésekben tapasztalt tagösszetételek relatív átfedésének átlagával a csoport élete során. A disszertáció szerzője egy érdekes összefüggést tárt fel a stacionaritás, a csoportok mérete és várható élettartama között: Az eredmények szerint az átlagos élettartam maximuma alacsonyabb stacionaritás értékeknél volt a nagyméretű csoportok esetén, mint a kisebb csoportoknál. 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.

Végül az 5. fejezetben egy multifraktál alapú véletlengráf-modell került bemutatásra. Az alapötlet, mely szerint egy véletlen gráf előállítása során az élek valószínűségét egy multifraktál segítségével állítjuk elő, Vicsek Tamástól származik, és az alkalmazott megközelítés erősen támaszkodik Lovász László és munkatársainak gráfsorozatok konvergenciájával kapcsolatos eredményeire. A modell részleteit a disszertáció szerzője dolgozta ki és programozta be. Emellett analitikus úton meghatározta a legfontosabb hálózatjellemzők mint például a foksámeloszlás és a klaszterezettségi együttható alakját a paraméterek függvényében. A multifraktál alapú véletlengráf-modell legfontosabb előnye, hogy az általa generált hálózatok tulajdonságai nagyon széles skálán változhatnak, a foksámeloszlás lehet például skálafüggetlen, vagy Poisson-eloszlású. A disszertáció szerzője kidolgozott egy szimulált hőkezelésen alapuló módszert, mely lehetővé teszi, hogy egy előre meghatározott tulajdonságokkal rendelkező véletlen gráf generálásához illesszük a modell paramétereit. 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.

Köszönetnyilvánítás

Szeretném köszönetemet kifejezni Vicsek Tamásnak, aki kutatócsoport-vezetőként lehetőséget adott arra, hogy a disszertációban bemutatott témákon dolgozhassak, és aki ezekben a munkákban tevékenyen részt véve végig irányította és koordinálta a kutatásokat. Az általa hozott rendkívül inspiráló ötletek és az intuitív látásmód, a barátságos légkör és a rendkívül jó munkakörülmények egy egyedülálló kutatóhelyet nyújtottak. Köszönöm Pollner Péternek, hogy amellet, hogy bizonyos munkákban társszerzőként is közreműködött, biztosította a kutatócsoport informatikai eszköztárának zavartalan működését, folyamatos fejlesztését és karbantartását. Szeretném megköszönni Derényi Imrének és Farkas Illésnek a topológiai fázisátalakulások és a k -klikkperkoláció vizsgálata során kialakult nagyon konstruktív és motiváló együttműködést, ami meghatározó pontjává vált kutatói pályafutásomnak. Továbbá szeretnék köszönetet mondani többi társszerzőmnek is, Barabási Albert-Lászlónak és Lovász Lászlónak, akik közreműködése nélkül nem jöhettek volna létre a disszertáció anyagának alapjául szolgáló publikációk.

Köszönettel tartozom Vattay Gábornak és Cserti Józsefnek, akik PhD- és diplomamavezetőként segítettek előre korai pályafutásomat és bevezettek a kutatómunka alapjaiba. Szeretném megköszönni egyetemi oktatóimnak, közép- és általános iskolai tanárainak az áldozatos munkájukat, közülük kiemelten hálás vagyok Édes Péternek, Varga Lászlónénak és Honyek Gyulának, akik megszeretették velem a fizikát és a fizikusi gondolkodásmódot. Nagyon köszönöm szüleimnek, hogy stabil háttérrel és jó példákkal elindítottak azon az életpályán ami a kutatói hivatáshoz vezetett. Köszönöm húgomnak és családjának a támogatást. Feleségemnek és két kislányomnak nagyon hálás vagyok az utóbbi időszakban a türelmükért, biztatásukért és a nyugodt, vidám családi háttérért.

A függelék

Komponensszintű Hamilton-függvények

Az 1. fejezetben tárgyalt egyrészcskés Hamilton-függvények mellett szintén természetes ötlet megvizsgálni azt, hogy milyen fázisátalakulások figyelhetők meg egy gráfsokaságban akkor, ha az energiafüggvény csak az összefüggő komponensek méretétől függ, a csúcsok egyenkénti fokszámától nem. Az itt bemutatásra kerülő analitikus számításokat Derényi Imre végezte el, a vonatkozó Monte-Carlo-szimulációkat pedig Farkas Illés futtatta le és értékelte ki.

A.1. Perkolációs átalakulás

Amint korábban említettük, az Erdős-Rényi-modell egyik érdekes tulajdonsága, hogy amennyiben az élbekötési valószínűség eléri a $p = p_c = 1/N$ kritikus értéket, a hálózatban megjelenik egy óriás komponens, melynek relatív mérete még $N \rightarrow \infty$ esetén sem tart nullához [64, 30]. Az átlagos fokszám a kritikus pontban az Erdős-Rényi-gráf méretétől függetlenül $\langle k \rangle = Np = 1$, és általánosan, ha $\langle k \rangle < 1$ egy Erdős-Rényi-gráfban, akkor az a „rendezetlen” fázisban van, azaz nem tartalmaz óriás komponenst. Minthogy az általunk vázolt élettrendezőzési dinamikánál a $T \rightarrow \infty$ esetben a hálózat egy Erdős-Rényi-gráffal ekvivalens, ha a sokaságunkban az átlagos fokszám $\langle k \rangle = 2M/N < 1$, akkor a fentiek alapján az szintén nem tartalmazhat óriás komponenst. Egy természetes kérdés, ami felmerül ezzel kapcsolatban, hogy lehet-e az Erdős-Rényi-modell perkolációs átalakulásához hasonló fázisátalakulást indukálni a rendszer hűtésével, ha megfelelően választjuk az energiafüggvényt?

Egy ilyen átalakulás nyomon követésére a $\varphi_s = M_1/M$ rendparaméter a legalkalmasabb, mely a legnagyobb összefüggő komponensben található élek és a teljes élszám hányadosa. Természetesen egy olyan energiafüggvényt kell választanunk, mely preferálja a nagy összefüggő komponensek kialakulását. Az egyszerűség kedvéért tegyük fel, hogy az energia az $E = f(M_1)$ alakban áll elő, ahol f egy monoton csökkenő függvény. Ilyen esetekben az energia csak a legnagyobb komponens méretétől függ, a kisebb komponensek méreteloszlása, struktúrája, stb. semmilyen befolyással nincs rá.

A FÜGGELÉK: *Komponensszintű Hamilton-függvények*

Ahhoz, hogy a szabadenergia entrópiával kapcsolatos tagját is fel tudjuk írni, vezető rendben meg kell határozzuk azon gráfok számát, melyek legnagyobb összefüggő komponense M_1 méretű. Ehhez először meg kell számoljuk, hogy általánosságban hányféle konfigurációja lehet egy m méretű összefüggő komponensnek. Az egyszerűség kedvéért felteesszük, hogy a komponens faszerű. Mivel $\langle k \rangle < 1$ esetén a hurkok valószínűsége $1/N$ szerint csökken a rendszermérettel [155], ez egy jogos közelítés a termodinamikai határesetben. A lehetséges m méretű irányítatlan fák számát a következő módon becsülhetjük. Kiindulunk egy véletlenszerű realizációból, mely természetesen m darab élből és $m+1$ csúcsból áll. Az egyik csúcsot kinevezzük a fa gyökerének. Ezek után képzeletben irányítást vezetünk be a fán a gyökértől kifelé indulva, egészen a gyökértől legtávolabb lévő levelekig. Ez által a gyökeret leszámítva a fában minden csúcsnak pontosan egy darab bejövő éle lesz. Ezután az m méretű fa egy másik realizációját a következő módon tudjuk előállítani: választunk véletlenszerűen egy csúcsot, majd a bejövő élének másik végét átkötjük a fának egy másik csúcsára, mint ahonnan eredetileg indult. Csak arra kell vigyázni, hogy ez az új kiindulópont ne legyen a választott csúcs leszármazottja, mert olyan esetben egy hurkot hoznánk létre és a fa szétesne két izolált részgráfra. Ha m elég nagy, akkor ez a megkötés csak a csúcsok elenyésző részét érinti súlyosan. Ennél fogva vezető rendben a fa közel m csúcsának közel m -féle új közvetlen felmenőt választhatunk, melynek révén becsülünk az m méretű, faszerű összefüggő komponens összes lehetséges konfigurációjára m^m .

Visszatérve az M_1 méretű legnagyobb összefüggő komponens tartalmazó állapotok entrópiájára, az ide tartozó gráfok $\mathcal{N}_{\text{komp}}$ számában az $M_1^{M_1}$ faktorhoz társul még egy $\binom{N}{M_1}$ szorzó, hiszen ennyiféle módon lehet a komponens csúcsait kiválasztani N -ből. A legnagyobb komponensből kimaradó $M - M_1$ élt a szintén kimaradó $N - M_1$ csúcs közé bárhogyan be lehet kötni, feltéve, hogy nem hoznak létre egy M_1 -nél nagyobb komponenset. Mivel $M_1 = \varphi_s M$ egy extenzív mennyiség, és a kimaradó élekből létrejövő komponensek mérete tipikusan lassabban nő, mint M , ez a megszorítás nem igazán befolyásolja az eredményeket. Ennek révén a kimaradó csúcsok és élek járulékát szimplán egy $\binom{(N-M_1)^2/2}{M-M_1}$ faktorialis figyelembe. Ezek alapján a termodinamikai határesetben (amikor $N, M \rightarrow \infty$, $\langle k \rangle = \text{konst.}$), az $M_1 = \varphi_s M$ azonosságot használva az

$$\mathcal{N}_{\text{komp}} \approx (\varphi_s M)^{\varphi_s M} \binom{N}{\varphi_s M} \binom{(N - \varphi_s M)^2/2}{M - \varphi_s M} \quad (\text{A.1})$$

eredményre jutunk.

Mivel azt tettük fel, hogy az energia lényegében a rendparaméter függvénye, az $F(\varphi_s, T)$ feltételes szabadenergiát az

$$e^{-F(\varphi_s, T)/T} = \mathcal{N}_{\text{komp}}^{-f(\varphi_s)/T} = e^{-[f(\varphi_s) - T \ln \mathcal{N}_{\text{komp}}]/T} \quad (\text{A.2})$$

összefüggés alapján az

$$F(\varphi_s, T) = f(\varphi_s) - T \ln \mathcal{N}_{\text{komp}} \quad (\text{A.3})$$

alakban fejezhetjük ki. Az (A.1) egyenlet alapján, a Stirling-formula segítségével (mely szerint $l! \approx (l/e)^l \sqrt{2\pi l}$), az $\mathcal{O}(\ln N)$ nagyságrendű tagok elhanyagolásával a fenti kifejezésben

az $\ln \mathcal{N}_{\text{komp}}$ tagot az

$$\ln \mathcal{N}_{\text{komp}} \approx \text{konst.} + \left[1 + \ln \frac{2M}{N} - \frac{2M}{N} \right] \varphi_s M + \left[\frac{3M}{2N} - \frac{1}{2} - \frac{M^2}{N^2} \right] \varphi_s^2 M \quad (\text{A.4})$$

alakban közelíthetjük. Kihasználva, hogy $2M/N = \langle k \rangle$, a fenti közelítés alapján az (A.3) egyenletet az

$$F(\varphi_s, T) \approx f(\varphi_s M) + MT \left\{ [\langle k \rangle - 1 - \ln \langle k \rangle] \varphi_s + [\langle k \rangle^2 - 3 \langle k \rangle + 2] \frac{\varphi_s^2}{4} \right\} \quad (\text{A.5})$$

formába írhatjuk át.

A legegyszerűbb energiafüggvény-választás az

$$f(M_1) = -M_1 = -\varphi_s M, \quad (\text{A.6})$$

ahol az energia szimplán a legnagyobb összefüggő komponens mérete, negatív előjellel. Ha ezt behelyettesítjük az (A.5) kifejezésbe, akkor abból kiderül, hogy amennyiben

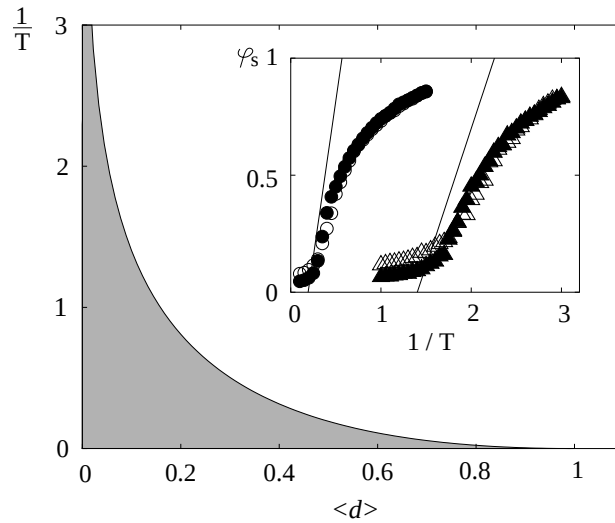
$$T > T_c(\langle k \rangle) = \frac{1}{\langle k \rangle - 1 - \ln \langle k \rangle}, \quad (\text{A.7})$$

a szabadenergia minimuma $\varphi_s = \varphi_s^*(T) = 0$ -nál van, azaz a hálózatban nincs óriás komponens. Ezzel szemben ha a hőmérséklet $T_c(\langle k \rangle)$ alá csökken, a minimum elmozdul $\varphi_s = 0$ -tól és megjelenik egy óriási összefüggő komponens. A rendparaméter értékét a kritikus hőmérséklet közelében az (A.7) egyenlet alapján becsülhetjük meg, melyből a

$$\varphi_s^*(T) = 2 \frac{T^{-1} - T_c^{-1}(\langle k \rangle)}{\langle k \rangle^2 - 3 \langle k \rangle + 2}, \quad (\text{A.8})$$

kifejezés adódik. Ez alapján a tapasztalt átalakulás egy folytonos fázisátalakulásnak felel meg, analóg módon az Erdős–Rényi-moddal perkolációs átalakulásával.

A kritikus hőmérséklet $\langle k \rangle$ függését az (A.7) egyenlet adja meg. Látható, hogy ha $\langle k \rangle \rightarrow 1$, akkor $T_c \rightarrow \infty$, ami teljes összhangban van azzal, hogy az Erdős–Rényi-gráf esetén a perkoláció kritikus pontja $\langle k \rangle = 1$ -nél van. Ennek alapján ha $\langle k \rangle > 1$, akkor a hálózatot hűteni sem kell ahhoz, hogy megjelenjen benne egy óriás komponens, hiszen ahogy azt az alfejezet elején említettük, ilyenkor már $T \rightarrow \infty$ esetén is tartalmaznia kell egy óriás komponensnek. A kritikus hőmérséklet inverzének értékét az átlagos foksám függvényében az A.1. ábrán tüntettük fel. A görbe alatti szürke terület felel meg a „rendezetlen” fázisnak, melynél a hálózat csupa kicsi, N -hez képest elhanyagolható méretű komponensből áll, míg a görbe felett megjelenik a rendszerben egy óriás komponens. A betétábra a Monte–Carlo-szimulációk eredményét mutatja. Látható, hogy a rendparaméter szakadás nélkül, egy véges meredekséggel kezd el nőni a kritikus pontban, ami szintén alátámasztja azt, hogy egy folytonos fázisátalakulással van dolgunk.



A.1. ábra. Az $E = -M_1$ energiához tartozó fázisdiagram. A fázishatárt az (A.7) egyenletben megadott kritikus hőmérséklet szabja meg, itt most $1/T$ került ábrázolásra a $\langle k \rangle$ átlagos fokszám függvényében. A szürke színnel jelölt terület felel meg a „rendezetlen” fázisnak, melyben nincs óriás komponense a hálózatnak, a fehér pedig a „rendezett” fázis, ahol megjelenik egy óriás komponens. A betétábra Monte–Carlo-szimulációk eredményét mutatja a $\varphi_s = M_1/M$ rendparaméterre vonatkozóan az $1/T$ függvényében. A kör alakú adatpontok esetén a hálózat átlagos fokszáma $\langle k \rangle = 0,5$ volt, míg a háromszögek esetén $\langle k \rangle = 0,1$, az üres szimbólumok $N = 500$ -as, a teli szimbólumok $N = 1000$ -es rendszerméretnek felelnek meg. Az adatok 10 futásra lettek átlagolva, futásonként a rendparaméter időátlaga a $t = 100N$ és $t = 500N$ közti Monte–Carlo-lépésekre lett kiszámítva. (Az ábra forrása a [T1] publikáció).

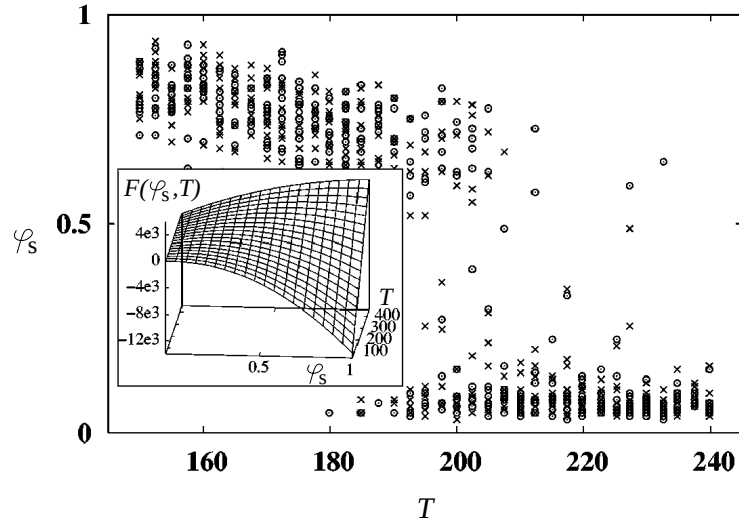
A.2. Elsőrendű fázisátalakulások

Természetesen elsőrendű fázisátalakulásokat is meg lehet figyelni egy klasztermérettől függő energiafüggvénnyel bíró gráfsokaságban, amennyiben megfelelő módon választjuk meg az $E = f(M_1)$ konkrét alakját. Egy egyszerű példa az, amikor $f(M_1)$ nulla vagy pozitív meredekséggel indul, viszont nagyobb M_1 értékekre továbbra is csökkenő tendenciát mutat. Természetesen a $T \rightarrow \infty$ határesetben ilyenkor is az entrópiával kapcsolatos tagok dominálják a feltételes szabadenergiát és az (A.5) minimuma $\varphi_s = 0$ -nál van, ami az óriás komponens nélküli, „rendezetlen” állapot. Hasonló módon a korábbiakhoz, $T \rightarrow 0$ esetén viszont az energia fog dominálni, ami miatt $F(\varphi_s, T)$ minimuma a $\varphi_s = 1$ -nél lesz, ami egy minden élt magába foglaló óriás komponenssel rendelkező állapotnak felel meg.

Azonban a korábbiakkal ellentétben találunk egy olyan hőmérséklet-tartományt, ahol az (A.5) által megadott $F(\varphi_s, T)$ a $\varphi_s = 0$ -nál emelkedve indul (hiszen az entrópiával kapcsolatos tagok közül a φ_s -ben lineáris tag dominál ebben a tartományban), majd egy adott φ_s -nél eléri maximumát és onnantól viszont csökkenő tendenciát mutat. Ennek révén a feltételes szabadenergiának két versengő minimumhelye van φ_s szerint, az egyik metastabil, a másik pedig globálisan stabil. Egy globálisan stabil és egy metastabil minimumhely

párhuzamos létezése az elsőrendű fázisátalakulások jellemzője.

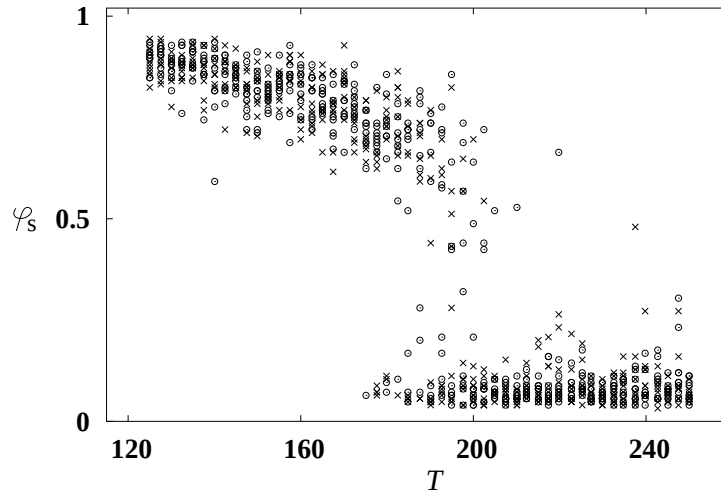
Egy egyszerű energiafüggvény, ami megfelel a fenti feltételnek az $E = -M_1^2$. Az ezen energiafüggvényre vonatkozó numerikus szimulációk eredményeit mutatja be az A.2. ábra egy $N = 500$ csúcsból és $M = 125$ élből álló hálózat esetén. Látható, hogy egy hiszterézis



A.2. ábra. Szimulációs eredmények a $\varphi_s = M_1/M$ rendparaméter a hőmérséklet függésére (ahol M_1 a legnagyobb komponens éleinek számát jelöli), az $E = -M_1^2$ energiafüggvény esetén egy $N = 500$ csúcsból és $M = 125$ élből álló hálózat esetén. A kör alakú adatpontok esetén a szimulációk kiinduló állapota egy Erdős–Rényi-gráf volt, míg az x-szel jelölt adatpontok esetén egy csillagszerű topológia, ahol minden él egy azonos csúcshoz kapcsolódik. A φ_s rendparaméter értéke a $t = 490N$ és $t = 500N$ Monte–Carlo-lépések között lett kiátlagolva. A betétábra a vizsgált sokaság $F(\varphi_s, T)$ feltételes szabadenergiáját mutatja az (A.5) kifejezés alapján. Az a hőmérséklet tartomány, melyben a feltételes szabadenergiának két minimuma van, teljesen konzisztens a numerikus szimulációk eredményében látható hiszterézissel. (Az ábra forrása a [T1] publikáció.)

alakul ki a φ_s rendparaméter hőmérséklet függésében: Ha a szimulációkat alacsony hőmérsékleten egy óriás komponenssel rendelkező állapotból indítjuk (mely az egyszerűség kedvéért egy csillagszerű topológiának felel meg), majd emeljük a hőmérsékletet, akkor az óriás komponens akár $T = 230$ -ig is megmaradhat a rendszerben. Ezzel szemben ha magas hőmérsékleten egy Erdős–Rényi-gráffal ekvivalens állapotból indítjuk a szimulációkat (melyben a $\langle k \rangle = 0.5$ révén nincs óriás komponens), akkor a hűtés során előfordulhat, hogy az óriás komponens csak $T = 170$ környékén jelenik meg. Ez a jelenség teljesen konzisztens azzal, hogy az $f(M_1) = -M_1^2$ kifejezést az (A.5) egyenletbe helyettesítve azt kapjuk, hogy a $80 < T < 430$ hőmérséklet-tartományban a feltételes szabadenergiának két minimuma is van a $[\varphi_s = 0, \varphi_s = 1]$ intervallumban. Természetesen a numerikus szimulációk csak egy szűkebb tartományon jelzik a hiszterézis jelenlétét, hiszen ha nagyon erős az eltérés a metastabil és a stabil minimum mélysége között, akkor a metastabil állapot megjelenésének

a szimulációkban nagyon csekély az esélye.

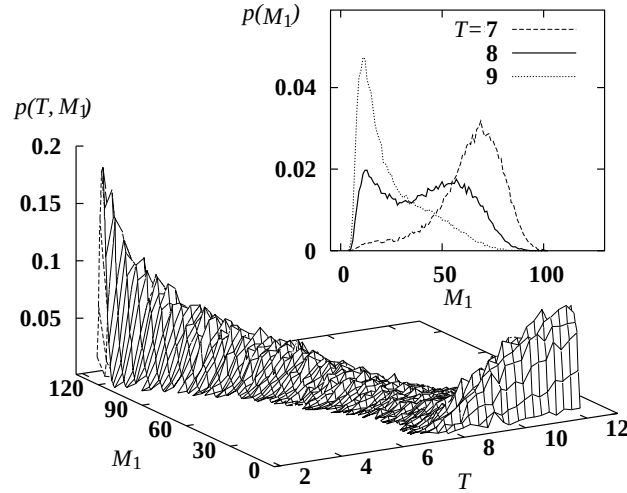


A.3. ábra. A Monte–Carlo-szimulációk eredményei a $\varphi_s = M_1/M$ rendparaméter a hőmérséklet függésére, az $E = -\sum_i M_i^2$ energiafüggvény esetén egy $N=500$ csúcsból és $M=125$ élből álló hálózat esetén. A kör alakú adatpontok esetén a szimulációk kiinduló állapota egy Erdős–Rényi-gráf volt, míg az x-el jelölt adatpontok esetén egy csillagszerű topológia, ahol minden él egy azonos csúcshoz kapcsolódik. A φ_s rendparaméter értéke a $t=200N$ Monte–Carlo-lépésnél került kiolvasásra. Látható, hogy a köztes hőmérséklet tartományban az adatpontok két viszonylag stabil állapot körül csoportosulnak, nagyjából $\varphi_s \approx 0.6 - 0.9$ illetve $\varphi_s \approx 0 - 0.15$ környékén, míg ezen értékek közé eső φ_s rendparaméter érték csak ritkán fordul elő. (Az ábra forrása a [T1] publikáció.)

Az (A.5) egyenleten alapuló analitikus számítások és numerikus szimulációk alapján hasonlóan viselkedik az $E = -M_1 \ln M_1$ energiafüggvény által definiált gráfsokaság is. Hasonlóan az előző esethez, ennél a sokaságnál is létezik egy hőmérséklet tartomány, ahol a feltételes szabadenergiának két minimuma is van φ_s szerint a $[0,1]$ tartományban. Ennek révén egy hiszterézis alakul ki a $\varphi_s(T)$ -ben hűtés illetve melegítés során, ami alapján kijelenthetjük, hogy az átalakulás elsőrendű.

A komponensektől függő energiafüggvényekkel kapcsolatos vizsgálatokat kiterjeszthetjük olyan esetekre is, ahol az energia már nem csak a legnagyobb komponens méretétől függ, hanem figyelembe veszi az összes többi komponens méreteit is. Az energiafüggvényt ilyenkor az $E = \sum_j f(M_j)$ alakban írhatjuk fel, ahol az összegzés az összefüggő komponensekre történik és M_j a j -edik komponens éleinek számát jelöli. Mivel az élek összesített száma időben állandó, ahhoz, hogy az energiafüggvény preferálja az élek „kondenzálódását”, $f(M_j)$ -nek a lineárisnál gyorsabban kell csökkennie. Ha megjelenik a rendszerben egy óriás komponens, mely magába foglalja az élek döntő többségét, akkor abban a tartományban az energia jól közelíthető úgy, mint $E = \sum_j f(M_j) \approx f(M_1)$, azaz a korábban részletezett számítások megismételhetők az adott $f(M_j)$ függvény behelyettesítésével.

Az $E = -\sum_i M_i^2$ és az $E = -\sum_i M_i \ln M_i$ energiafüggvényekhez tartozó sokaságok vizsgálata megmutatta, hogy ezekben az esetekben például egy elsőrendű fázisátalakulás tapasztalható a magas hőmérsékletekhez tartozó, óriás komponens nélküli „rendezetlen” állapot és az alacsony hőmérsékleten megjelenő, óriás komponenssel rendelkező „rendezett” állapot között. A numerikus szimulációk eredményeit az A.3. ábra mutatja be az $E = -\sum_i M_i^2$ esetén, illetve az A.4. ábra az $E = -\sum_i M_i \ln M_i$ esetén. Ez utóbbi ábrán



A.4. ábra. A numerikus szimulációk eredménye egy $N=500N$ csúcsból és $M=125$ élből álló, $E = -\sum_i M_i \ln M_i$ energiafüggvénnyel rendelkező sokaság esetén a legnagyobb komponens méretének eloszlására. Adott T esetén $p(T, M_1)$ az M_1 méretű legnagyobb komponens megfigyelésének valószínűségét adja. A komponensméretek a $t=200N$ és $t=400N$ Monte-Carlo-lépések között lettek összesítve összesen 10 futásra. A betétábra a $T=7$, $T=8$ és $T=9$ esetén mért legnagyobb komponens méret eloszlást mutatja 200 futásra átlagolva. (Az ábra forrása a [T1] publikáció.)

a legnagyobb komponens méretének eloszlását láthatjuk M_1 és a hőmérséklet, T függvényében. Alacsony hőmérsékleten az eloszlás maximuma nagy M_1 értékeknél van, hiszen a hálózat rendelkezik egy óriás komponenssel. A T növelésével ez a maximum a kisebb értékek felé tolódik el, és az eloszlás ki is szélesedik. Magas hőmérsékleten az eloszlás maximuma már alacsony értékeknél van, hiszen ilyenkor már nem találunk a hálózatban óriás komponenset. A betétábra nagyobb részletességgel mutatja a legnagyobb komponens méretének eloszlását $T=7$, $T=8$ és $T=9$ esetén. A $T=7$ -hez tartozó eloszlás maximuma még viszonylag magas M_1 értéknél van, ezzel szemben $T=9$ -nél a maximum hely már közel csúszik az origóhoz. A legérdekesebb azonban a $T=8$ -hoz tartozó eloszlás, mely bimodális alakot mutat. Ez azt jelzi, hogy ennél a hőmérsékletnél a feltételes szabadenergiának két minimuma van, melyek nagyjából egyforma mélységűek, emiatt két, a legnagyobb komponens méretében jelentősen eltérő állapot fordul elő viszonylag nagy valószínűséggel.

A FÜGGELÉK: *Komponensszintű Hamilton-függvények*

B függelék

További eredmények a k -klikkperkolációval kapcsolatban

B.1. A kritikus pont levezetése heurisztikus módon

A kritikus élbekötési valószínűségnek egy egyszerű heurisztikus módon történő levezetését Derényi Imre adta meg a 2.1. alfejezetben bemutatott k -klikksablon-görgetési kép segítségével. Az alapötlet a következő: a kritikus p_c esetén azon szomszédos k -klikkek számának várható értéke, ahová tovább gördíthetjük a sablont az aktuális pozíciójából eggyel kell egyenlő legyen. Ha ugyanis ez a várható érték egynél kisebb, akkor az azt jelenti, hogy csak nagyon kicsi k -klikkperkolációs klasztereket tudunk felfedezni a sablon görgetésével, hiszen ahogy egyre tovább próbáljuk gördíteni a sablont, általában nagyon gyors megállásra kényszerülünk. Ezzel szemben ha ez a várható érték egynél nagyobb, akkor a sablon görgetésénél nem csak, hogy mindig tovább tudunk menni, hanem még elágazásokra is van lehetőség, ami egy óriás k -klikkperkolációs klasztert vetít elénk.

Itt természetesen feltettük, hogy a k -klikkszomszédosági gráf a kritikus pontban faszerű. Ez egy plauzibilis feltételezés, hiszen ha visszanyúlunk a hagyományos élperkolációhoz, ott közismert, hogy az Erdős–Rényi-gráf esetén a kritikus pontban az óriás komponens faszerű. Ennek analógiájára joggal tehetjük fel, hogy amikor a p élbekötési valószínűséggel alulról közelítjük meg a k -klikkperkoláció kritikus pontját, a kis k -klikkperkolációs klaszterek úgy állnak össze egy óriás klaszterre p_c -nél, hogy csak elenyésző számban formálnak hurkokat a k -klikkszomszédosági gráfban.

A szomszédos k -klikkekre vonatkozó fenti várható értéket a következő módon becsülhetjük meg: A sablon k csúcsa közül $k-1$ -et választhatunk ki arra, hogy tovább gördüljünk az adott csúcs (egy szomszédos k -klikkre történő) áthelyezésével. A gráf N csúcsa közül durván $N-(k-1)$ potenciális „célpont” van ahová gördülhet a sablon áthelyezendő csúcsa, viszont ahhoz, hogy tényleg oda tudjunk gördülni, az adott csúcsból $k-1$ él kell induljon a sablon jelenlegi helyzetének megfelelő k -klikkhez. Ezeket a faktorokat figyelembe véve a továbbgördítés célpontjaként felmerülő k -klikkszomszédok számának várható értékét a

$$(k-1)[N-(k-1)]p^{k-1} \simeq (k-1)Np^{k-1} \quad (\text{B.1})$$

B FÜGGELÉK: További eredmények a k -klikkperkolációval kapcsolatban

kifejezéssel becsülhetjük, ahol az élek függetlensége miatt az egyes élek létezésének p valószínűsége összeszorozódik, valamint a nagy N esetekre koncentrálva elhanyagoltuk N mellett a $k-1$ -et. A kritikus élbekötési valószínűség esetén a fenti kifejezés eggyel egyenlő, amiből meghatározhatjuk $p_c(k)$ -t [T4]:

$$(k-1)Np_c(k)^{k-1} = 1, \quad p_c(k) = \frac{1}{[(k-1)N]^{\frac{1}{k-1}}}. \quad (\text{B.2})$$

(Ez az eredmény megegyezik a 2. fejezetben megadott (2.19) kifejezéssel).

B.2. Parciális differenciálegyenlet megközelítés

Ráth Balázs és Tóth Bálint egy általánosabb modell keretein belül tanulmányozták a k -klikkperkoláció kritikus pontját [174]. Ebben a megközelítésben az Erdős–Rényi-gráf a következő sztochasztikus folyamat révén áll elő: Egy kiinduláskor teljesen üres gráfhoz az idő múlásával $1/\sqrt{N}$ rátával adunk éleket véletlenszerűen. Így a t időpontban az összes lehetséges élhelynek $1 - e^{-t/\sqrt{N}}$ hányada lesz betöltve, és $N \rightarrow \infty$ esetén bármely t -re a kapott gráf ekvivalens egy Erdős–Rényi-gráffal, melyben az élbekötési valószínűség $p = t/\sqrt{N}$.

A fenti modellt természetes módon lehet kiterjeszteni olyan esetekre is, ahol a kiinduló állapot nem egy üres gráf. A Ráth Balázs és Tóth Bálint által bemutatott levezetése a k -klikkperkoláció kritikus élbekötési valószínűségének parciális differenciálegyenleteken alapszik, és alkalmazható tetszőleges kezdeti állapotú gráfokra is, a következő megközelítésekkel:

- Ha véletlenszerűen kiválasztjuk az élek egy kis halmazát, akkor az ezekhez tartozó perkolációs klaszterek méreteloszlása $N \rightarrow \infty$ esetén aszimptotikusan független a többi perkolációs klaszter méreteloszlásától.
- A kezdeti állapotban a k -klikkperkolációs klaszterek faszerűek a k -klikkszomszédssági gráfban.

A kritikus pont levezetéséhez az m élből álló k -klikkperkolációs klaszterek $\mathcal{C}_m$ számának változását kell megbecsülni t és $t+dt$ között. Bár ez a megközelítés tetszőleges k -ra alkalmazható, a [174] publikáció szerzői az egyszerűség kedvéért pusztán a $k=3$ esetre koncentráltak. Először bevezettek még két további mennyiséget,

$$c_m(t) \equiv \lim_{N \rightarrow \infty} \frac{\mathcal{C}_m(N, t)}{\frac{1}{2}N^{\frac{3}{2}}}, \quad (\text{B.3})$$

$$v_m(t) \equiv m \cdot c_m(t), \quad (\text{B.4})$$

(ahol $\mathcal{C}_m(N, t)$ természetesen az m élből álló k -klikk perkolációs klaszterek számát jelöli a t

időpontban). A $c_m(t)$ és $v_m(t)$ Laplace-transzformáltja a következő alakban adható meg:

$$V(t, x) \equiv \sum_{m \in \mathbb{N}} v_m(t) \cdot e^{-m \cdot x}, \quad (\text{B.5})$$

$$C(t, x) \equiv \sum_{m \in \mathbb{N}} c_m(t) \cdot e^{-m \cdot x}. \quad (\text{B.6})$$

Bevezetve továbbá az

$$E(t) \equiv \lim_{N \rightarrow \infty} \frac{|E(N, t)|}{\frac{1}{2}N^{\frac{3}{2}}}, \quad (\text{B.7})$$

mennyiséget megmutatható, hogy az átlagtér-közelítésben $N \rightarrow \infty$, $dt \rightarrow 0$ mellett $C(t, x)$ a következő egyenletnek tesz eleget:

$$\frac{\partial}{\partial t} C(t, x) = e^{V(t, x)^2 - E(t)^2 - x} - 2V(t, x) \cdot (t)^2. \quad (\text{B.8})$$

A (B.8) egyenlet karakterisztikák segítségével oldható meg, mely a következő eredményre vezet [174]:

$$t_c^2 \cdot (2ab + 1) + t_c \cdot (b + a \cdot (2ab + 1)) = \frac{1}{2}, \quad (\text{B.9})$$

ahol $a = V(0, 0) = \sum_m v_m(0)$, $b = -\frac{\partial}{\partial x} V(0, 0)$ és t_c az óriás k -klikkperkolációs klaszter megjelenésének időpontját jelöli. Az üres kiindulási gráf speciális esetében $a = b = 0$, amiből $t_c = 1/\sqrt{2}$ adódik, valamint

$$p_c(k=3) = \frac{t_c}{\sqrt{N}} = \frac{1}{\sqrt{2N}}, \quad (\text{B.10})$$

ami teljes mértékben megegyezik az általunk levezetett (2.19) eredménnyel. A (B.8) összefüggés alapján Ráth Balázs és Tóth Bálint levezettek egy további egyenletet az óriás k -klikkperkolációs klaszter $v_\infty(t)$ relatív méretére vonatkozóan, ahol a méretet a tartalmazott élek számában adták meg. Bevezetvén a

$$W(t, x) \equiv e^{V(t, x)^2 - E(t)^2 - x}, \quad (\text{B.11})$$

$$\widehat{V}(t, w) \equiv V(t, \widehat{X}(t, w)), \quad (\text{B.12})$$

jelölést, ahol a $\widehat{X}(t, w)$ függvény a $W(t, x)$ inverzének felel meg az x változó szerint, a következő összefüggésre jutottak [174]:

$$v_\infty(t) = \widehat{V}(0, 1) + t - \widehat{V}(0, W(t, 0)) - tW(t, 0). \quad (\text{B.13})$$

B.3. Egzakt eredmények

A k -klikkperkolációval kapcsolatban egzakt matematikai eredmények is születtek, melyek Bollobás Béla és Oliver Riordan munkásságának köszönhetők [32]. A k -klikkszomszédságot általánosabban definiálták, mely szerint két k -klikk szomszédosnak tekinthető, ha legalább l csúcsuk közös, azaz átfednek egy l -klikkben. (Az általunk javasolt eredeti definíció megfelel az $l = k - 1$ speciális esetnek).

B FÜGGELÉK: További eredmények a k -klikkperkolációval kapcsolatban

Az Erdős–Rényi-gráfban egy adott l -klikket tartalmazó k -klikkek várható számát a

$$\binom{N}{k-l} p^{\binom{k}{2} - \binom{l}{2}} \quad (\text{B.14})$$

formában adhatjuk meg, hiszen az l -klikken kívül egymástól függetlenül $\binom{k}{2} - \binom{l}{2}$ további él létezése szükséges ahhoz, hogy egy k -klikkben benne legyen, és az összes N csúcsból $\binom{N}{k-l}$ -féle módon választhatjuk ki, hogy melyek legyenek a tartalmazó k -klikk további csúcsai. Képzeljük el, hogy adott k és l esetén feltárunk egy k -klikkperkolációs klasztert a 2.1. alfejezetekben bemutatott sablongördítős módszerrel. Ez lényegében egy elágazási folyamatnak felel meg, ahol egy adott lépésnél feltárt k -klikkekből a következő lépésben tovább lépünk az összes szomszédos, még feltáratlan k -klikkbe. A B.14 formula alapján egy adott k -klikkből várhatóan

$$v = v(N, p) = \left(\binom{k}{l} - 1 \right) \binom{N}{k-l} p^{\binom{k}{2} - \binom{l}{2}} \quad (\text{B.15})$$

szomszédos k -klikkbe tudunk tovább menni, hiszen $\binom{k}{l}$ darab l -klikk van egy k -klikkben, amik közül egyet nem használhatunk tovább gördülésre, hiszen „azon keresztül” érkeztünk meg a k -klikkbe az előző lépésben. A

$$\binom{k}{2} - \binom{l}{2} = \frac{k(k-1) - l(l-1)}{2} = \frac{k^2 - k - l^2 + l}{2} = \frac{(k-l)(k+l-1)}{2} \quad (\text{B.16})$$

azonosságot felhasználva (B.15) a

$$v(N, p) = \left(\binom{k}{l} - 1 \right) \binom{N}{k-l} p^{\frac{1}{2}(k-l)(k+l-1)} \quad (\text{B.17})$$

alakra hozható.

A k -klikkperkoláció kritikus pontja intuitíven a $v(N, p) = 1$ egyenletből meghatározott élbekötési valószínűségnek felel meg, hiszen ilyenkor képzelhető el, hogy akár végtelen sok lépésen keresztül tudunk újabb és újabb k -klikkeket feltárni, míg $v(N, p) < 1$ esetén az elágazási folyamat szükségszerűen hamar véget ér. Amennyiben a (B.17) jobb oldalán a második binomiális tagot vezető rendben $\binom{N}{k-l} \simeq N^{k-l}$ -el közelítjük, a következő heurisztikus egyenletet kapjuk a (most már l -től is függő) $p_c(k, l)$ kritikus élbekötési valószínűsége:

$$\left(\binom{k}{l} - 1 \right) N^{k-l} p_c(k, l)^{\frac{1}{2}(k-l)(k+l-1)} = 1, \quad (\text{B.18})$$

ami alapján

$$p_c(k, l) = \left[\binom{k}{l} - 1 \right]^{-\frac{2}{(k-l)(k+l-1)}} N^{-\frac{2}{k+l-1}}. \quad (\text{B.19})$$

Ez a formula az $l = k - 1$ esetben visszaadja a 2. alfejezetben a disszertáció szerzője által levezetett (2.19) kifejezést.

A következőekben rátérünk arra, hogy a B.17 egyenletben megadott $v(N, p)$ segítségével mi mondható ki egzaktul a k -klikkperkolációról. Ehhez még először be kell vezetnünk a

$$\sigma(v) = 1 - (1 - v(N, p))^{\binom{k}{l}} \quad (\text{B.20})$$

mennyiséget, melynek szemléletes jelentése a következő: Tegyük fel, hogy egy adott k -klikkből indítjuk el a fent említett, elágazási folyamathoz hasonló feltárását egy k -klikkperkolációs klaszternek. Mivel a kiinduló k -klikk összesen $\binom{k}{l}$ darab l -klikket tartalmaz, igazából $\binom{k}{l}$ független elágazási folyamatunk zajlik párhuzamosan, melyekre mind vonatkoznak a (B.15–B.17) egyenleteknél leírtak. Amennyiben $v(N, p) < 1$ (azaz a kritikus pont alatt vagyunk), egy adott elágazási folyamatnál annak esélye, hogy a következő lépésben leáll $1 - v(N, p)$, így a (B.20) egyenletben definiált $\sigma(v)$ megfelel az összesített feltárási folyamat túlélési valószínűségének. A Bollobás Béla és Oliver Riordan által közölt [32] publikáció fő eredménye (informális megfogalmazásban) az, hogy amennyiben $v(N, p)$ aszimptotikusan egy, úgy a véletlen gráf legnagyobb k -klikkperkolációs klasztere nagyon nagy valószínűséggel tartalmazza az összes k -klikk $\sigma(v) + o(1)$ hányadát.

Továbbá az Erdős–Rényi-gráf összes k -klikkjeinek száma, $\mathcal{N}(k)$, esetén a várható értéket könnyűszerrel felírhatjuk az

$$\langle \mathcal{N}(k) \rangle = \binom{N}{k} p^{k(k-1)/2} \quad (\text{B.21})$$

formában, hiszen k különböző csúcs összesen $\binom{N}{k}$ módon választható, és ezek akkor formálnak egy k -klikket ha a köztük lévő $k(k-1)/2$ él mind létezik (egymástól függetlenül p valószínűséggel). Ennek segítségével a következő állítás mondható ki:

B.1. Tétel (Bollobás Béla és Oliver Riordan). *Egy rögzített $1 \leq l < k$ mellett tegyük fel, hogy egy Erdős–Rényi-gráfban a p élbekötési valószínűséget N függvényében úgy állítjuk be, hogy a (B.17) egyenletben megadott $v(N, p)$ -re $v = \Theta(1)$ teljesüljön¹. Ilyenkor minden $\epsilon > 0$ mellett $N \rightarrow \infty$ esetén annak valószínűsége egyhez tart, hogy a legnagyobb k -klikkperkolációs klaszter k -klikkjeinek $\mathcal{N}_1(k)$ száma*

$$(\sigma(v) - \epsilon) \langle \mathcal{N}(k) \rangle \leq \mathcal{N}_1(k) \leq (\sigma(v) + \epsilon) \langle \mathcal{N}(k) \rangle,$$

illetve a második legnagyobb k -klikkperkolációs klaszter k -klikkjeinek száma $\mathcal{N}_2(k) \leq \epsilon \langle \mathcal{N}(k) \rangle$.

Bollobás Béla és Oliver Riordan ezt a tételt a [32] publikációban mondták ki és bizonyították be (a bizonyítás részleteinek ismertetésétől itt eltekintünk). Ugyanitt közöltek egzakt eredményeket az irányított k -klikkperkolációról is, a továbbiakban ezeket vázoljuk.

Hasonlóan az irányítatlan esethez, az irányított k -klikkperkoláció esetén is az általánosabb k -klikkszomszédság definíciót használták, ahol két irányított k -klikk akkor tekinthető szomszédosnak, ha l csúcsuk közös. Kiinduló pontként vizsgáljuk meg, hogy miként lehet egy irányított k -klikkperkolációs klaszter elágazási folyamat szerű feltárása során megadni

¹Általánosan $f(n) = \Theta(g(n))$ azt jelenti, hogy egyrészt $\exists c > 0, q \geq 0 : \forall n > q : f(n) \geq cg(n)$, másrészt $\exists z \in \mathbb{R}, \ell \geq 0 : \forall n > \ell : f(n) \leq zg(n)$

B FÜGGELÉK: *További eredmények a k -klikkperkolációval kapcsolatban*

azt, hogy egy adott lépésnél várhatóan hány új irányított k -klikkbe lehet tovább lépni. Teljesen hasonlóan a (B.15) egyenlethez, ezt a

$$v^{\text{ir}} = v^{\text{ir}}(N, p) = \frac{k!}{l!} \left(\binom{k}{l} - 1 \right) \binom{N}{k-l} p^{\binom{k}{2} - \binom{l}{2}} \quad (\text{B.22})$$

formában lehet megadni, hiszen a (B.15) kifejezéshez képest még biztosítani kell azt is, hogy az új k -klikkel közös részgráfban megfelelőek legyenek az élek irányultságai. Maga az állítás is nagyon hasonlít a B.1. tételhez, ugyanis lényegében azt mondja ki, hogy ha v^{ir} aszimptotikusan egy, akkor a legnagyobb irányított k -klikkperkolációs klaszter tartalmazza az összes irányított k -klikk $\sigma(v^{\text{ir}}) + o(1)$ részét, ahol σ -t a (B.20) egyenlet definiálja. Az irányított k -klikkek várható számát a gráfban a

$$\langle \mathcal{N}^{\text{ir}}(k) \rangle = \binom{N}{k} p^{k(k-1)/2} k! \quad (\text{B.23})$$

formában lehet megadni, hiszen a (B.21) egyenletnél tárgyalt irányítatlan esethez képest itt még a k -klikkeken belül a csúcsok lehetséges $k!$ -féle sorrendjét is figyelembe kell venni.

A fentiek alapján a következő állítás mondható ki:

B.2. Tétel (Bollobás Béla és Oliver Riordan). *Egy rögzített $1 \leq l < k$ mellett tegyük fel, hogy egy irányított Erdős–Rényi-gráfban a p élbekötési valószínűséget N függvényében úgy állítjuk be, hogy a (B.22) egyenletben megadott $v^{\text{ir}}(N, p)$ -re $v^{\text{ir}} = \Theta(1)$ teljesüljön. Ilyenkor minden $\epsilon > 0$ mellett $N \rightarrow \infty$ esetén annak valószínűsége egyhez tart, hogy a legnagyobb irányított k -klikkperkolációs klaszter k -klikkjeinek $\mathcal{N}_1^{\text{ir}}(k)$ száma*

$$(\sigma(v^{\text{ir}}) - \epsilon) \langle \mathcal{N}^{\text{ir}}(k) \rangle \leq \mathcal{N}_1^{\text{ir}}(k) \leq (\sigma(v^{\text{ir}}) + \epsilon) \langle \mathcal{N}^{\text{ir}}(k) \rangle.$$

Bollobás Béla és Oliver Riordan ezt a tételt a [32] publikációban mondták ki és bizonyították be (a bizonyítás részleteinek ismertetésétől megint csak eltekintünk).

C függelék

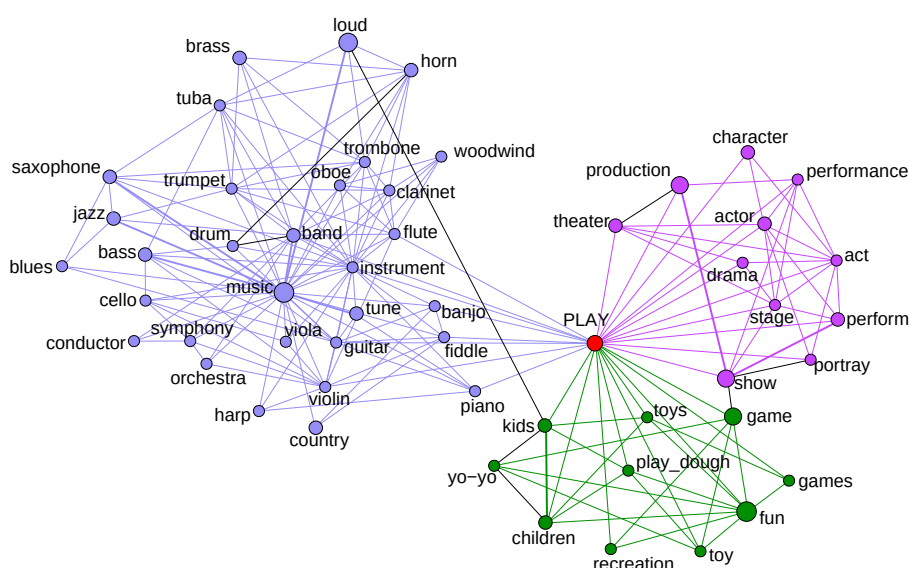
A csoportkereséssel kapcsolatos további eredmények

C.1. További példák lokális csoportszerkezetre

A 3.2. alfejezetben a k -klikkperkolációs csoportok lokális szerkezetét vizsgáltuk. Itt további két példát mutatunk arra, hogy miként néz ki egy-egy kiragadott csúcs közvetlen környezetében található átfedő csoportok szerkezete a South Florida Free Association Norms List [140] alapján felépülő szóasszociációs hálózatban. Az egyszerűség kedvéért a 3.4b ábrához hasonlóan itt sem vettük figyelembe az élek irányultságát, és mindkét ábra esetén $w^* = 0.025$ -ös súlyküszöböt és $k = 4$ -es k -klikkméretet használtunk a csoportok feltárásánál. A C.1. ábrán a „PLAY” csoportjai láthatók, melyek a zene, a színház, illetve a gyerekjátékok köré szerveződnek. Ezzel párhuzamosan a C.2. ábrán a „DAY” csoportjait tüntettük fel. A zöld csoport itt a munkanapokhoz kapcsolódik, bár hiányzik a „THURSDAY”, melynek csak két szomszédja van a hálózatban, így nem lehet része egyetlen csoportnak sem $k = 4$ -nél. (Megemlítjük, hogy a „SATURDAY” és „SUNDAY” egy hétvégéhez köthető csoportban van, a „FRIDAY”-el, a „NIGHT”-al, a „WEEK”-el és magával a „WEEKEND”-el együtt). A lila csoport a napszakoknak felel meg, a sárga időjárással kapcsolatos szavakból áll, melyek gyakran fordulnak elő olyan jelzős szerkezetekben, mint például az esős nap, míg a kék csoport néhány naptárhoz kötődő szóból áll.

A C.3 ábrán egy informatikához kapcsolódó hálózat lokális csoportszerkezetéből mutatunk egy kisebb részletet, melynél a csúcsok a wu-ftp program változóinak felelnek meg¹. Ennél a hálózatnál két változó közt akkor van él, ha együtt szerepelnek valahol egy függvényhívásban vagy utasításban. A változók neve két részből tevődik össze: az első specifikusan magára a változóra vonatkozik, míg a második, „@” szimbólummal elválasztott rész a változó érvényességi körét adja meg, ami tipikusan egy függvény. Mivel egy változó csak az érvényességi körén belül hozzáférhető, az ebben a hálózatban feltárt csoportok túlnyomórészt függvényeknek (szubrutinoknak) felelnek meg. A „@glb” végződésű változók globálisak, vagyis minden függvény számára hozzáférhetőek, és ezáltal több függvényhí-

¹A hálózatot a <http://www.wu-ftp.org> oldalról töltöttük le.

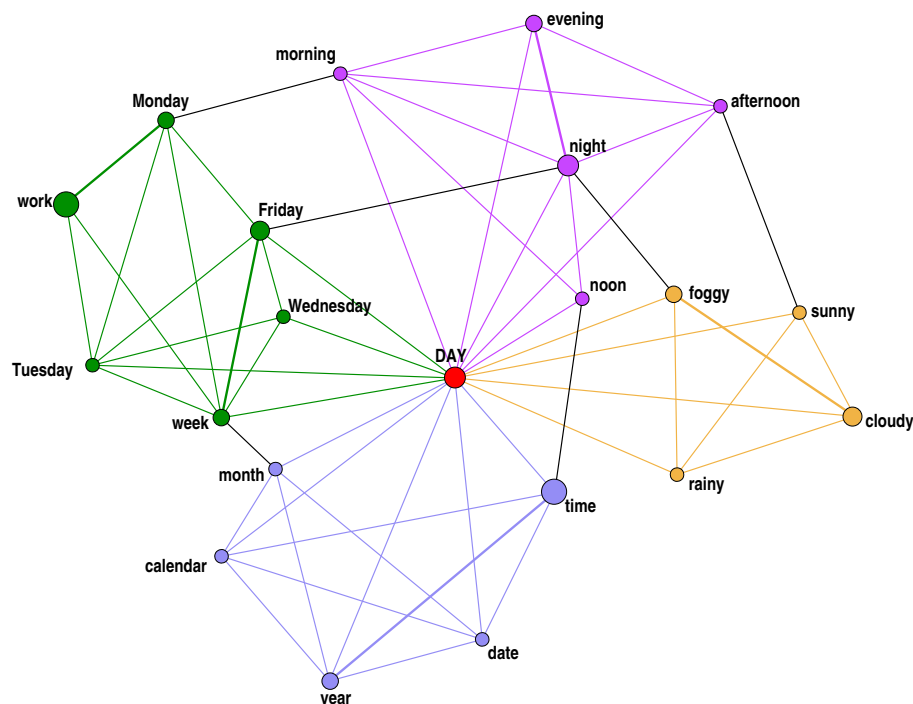


C.1. ábra. Átfedő csoportok a „PLAY” szó környezetében a szóasszociációs hálózatban. Hasonlóan a 3.4b ábrához, a hálózat a South Florida Free Association Norms list alapján épült fel és a csoportokat egy $w^* = 0.025$ -ös súlyküszöb és $k = 4$ -es k -klikkméret alkalmazásával tártuk fel. A kék színű csoport a zenéhez kötődik, a lila a színházhoz kapcsolódik, míg a zöld a gyerekekkel, gyerek játékokkal kapcsolatos. (Az ábra forrása a [T5] publikáció kiegészítő része.)

vásban, illetve eltérő függvényeken belüli utasításokban is szerepelhetnek. Ennek révén jó esélyük van arra, hogy a csoportok közti átfedésekbe kerüljenek. Ezt igazolja a C.3 ábra, ahol néhány, a hálózatban feltárt csoport látható $k = 5$ -ös k -klikkméret választása esetén.

C.2. További csoportstatisztikák

A 3.6. ábrán, illetve a 3.1. táblázaton alapuló eredmények ellenőrzéseként elvégeztük ugyanazokat az elemzéseket véletlenszerűen áthuzalozott hálózatokra is. Az élek átkötögetése során ügyeltünk arra, hogy a csúcsok fokszáma ne változzon meg, csak a szomszédsági viszonyok legyenek véletlenszerűen keverve. Ezáltal lényegében azt vizsgáltuk meg, hogy milyen csoportszerkezetet adnak a tanulmányozott rendszerek a (5. fejezet bevezetőjében már említett) konfigurációs modellben. A három hálózat közül a szóasszociációs hálózatra kapott eredményeket láthatjuk az eredeti hálózaton kapott eredményekkel összehasonlítva a C.4. ábrán. A véletlenszerű hálózatban nagyon gyér csoportszerkezet lelhető csak fel, kisméretű és kevés, egymással maximum egy csúcsban átfedő csoportokkal. A másik két rendszer esetén is hasonlóan drasztikus eltéréseket tapasztaltunk az eredeti hálózatban és a konfigurációs modellben kapható csoportszerkezetek között. Ezek alapján kijelenthetjük, hogy a 3.6. ábrán, illetve a 3.1. táblázaton alapuló eredmények nem a k -klikkperkolációs csoportkeresési módszer melléktermékei, hanem a vizsgált hálózatok tényleges csoportszer-

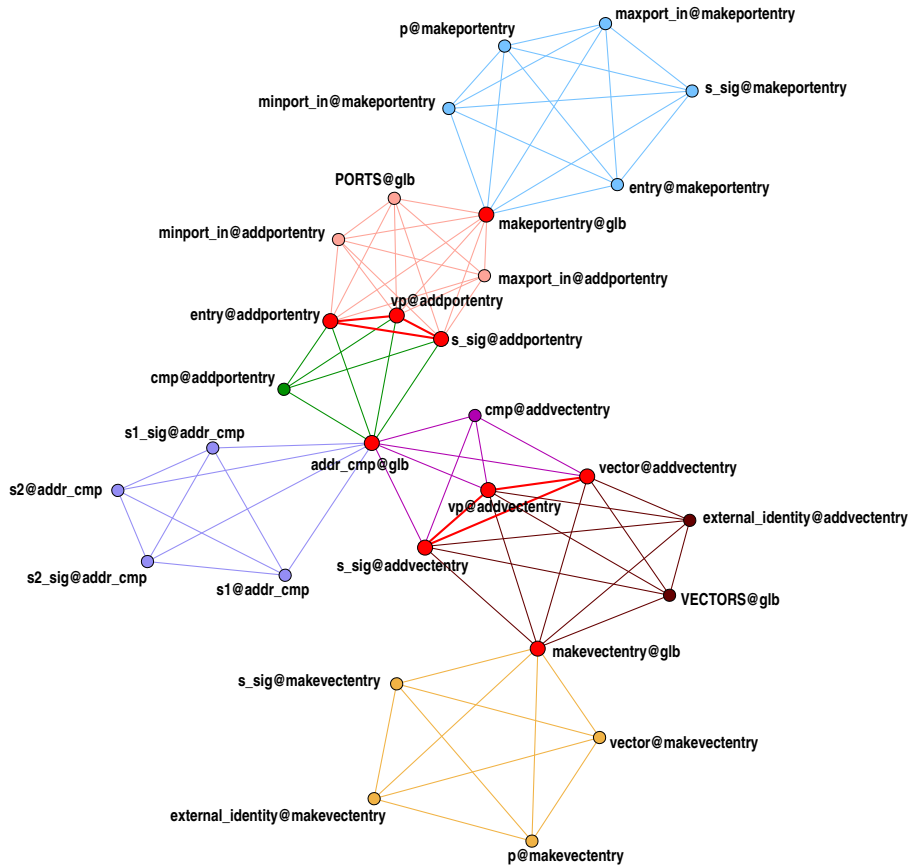


C.2. ábra. Átfedő csoportok a „DAY” szó környezetében a szóasszociációs hálózatban. A 3.4b és a C.1 ábrákkal azonos módon a csoportokat egy $w^* = 0.025$ -ös súlyküszöb és $k = 4$ -es k -klikkméret alkalmazásával kaptuk meg a South Florida Free Association Norms list alapján készített hálózatban. A zöld színű csoport a munkanapoknak felel meg, a lila a napszakokat sorolja fel, a sárga az időjáráshoz kapcsolódik (pl. esős nap), míg a kék a naptárhoz kötődik. (Az ábra forrása a [T5] publikáció kiegészítő része.)

kezetének tulajdonságait tükrözik.

Megemlítjük, hogy a társszerzőségi hálózat esetén a k -klikkperkoláció kritikus pontján alapuló optimális paraméterbeállítás $k = 5$ esetén is megvalósítható, természetesen egy alacsonyabb, $r^* = 0.75$ -ös élhányad mellett. Az ehhez a paraméterválasztáshoz tartozó eredmények a C.5 ábrán láthatók, ahol összehasonlításként a korábbi, $k = 6$ és $r^* = 0.93$ paramétereknél kapott eredményeket is feltüntettük. Látható, hogy a $k = 5$ -höz és a $k = 6$ -hoz tartozó eloszlások nagyon közel vannak egymáshoz, lényeges különbség nem tapasztalható köztük. Egy kis eltérés látható a csoportok foksámeloszlásának exponenciális részénél, az eloszlás ezen részéhez tartozó karakterisztikus csoportfokszám egy kicsit magasabb a $k = 6$ -os esetben. Egy minimális eltérés az átfedés méreteloszlásában is látható, viszont például a csúcsok csoportszámának eloszlása szinte teljesen fedi egymást a két esetben, illetve a csoportméret-eloszlások is nagyon szorosan követik egymást. Vizsgálataink szerint nem csak a csoportstatisztikák hasonlóak, hanem maguk a csoportok is: a 6-klikkcsoportok 44%-a megjelenik 5-klikkcsoportként is, illetve a 6-klikkcsoportok 70%-ához találunk olyan 5-klikkcsoportot, mely a csoporttagok kevesebb, mint 10%-ában tér el. Ezek alapján le-
szögezhetjük, hogy a vizsgált csoportstatisztikák által feltárt tulajdonságok a rendszerre

C FÜGGELÉK: A csoportkereséssel kapcsolatos további eredmények

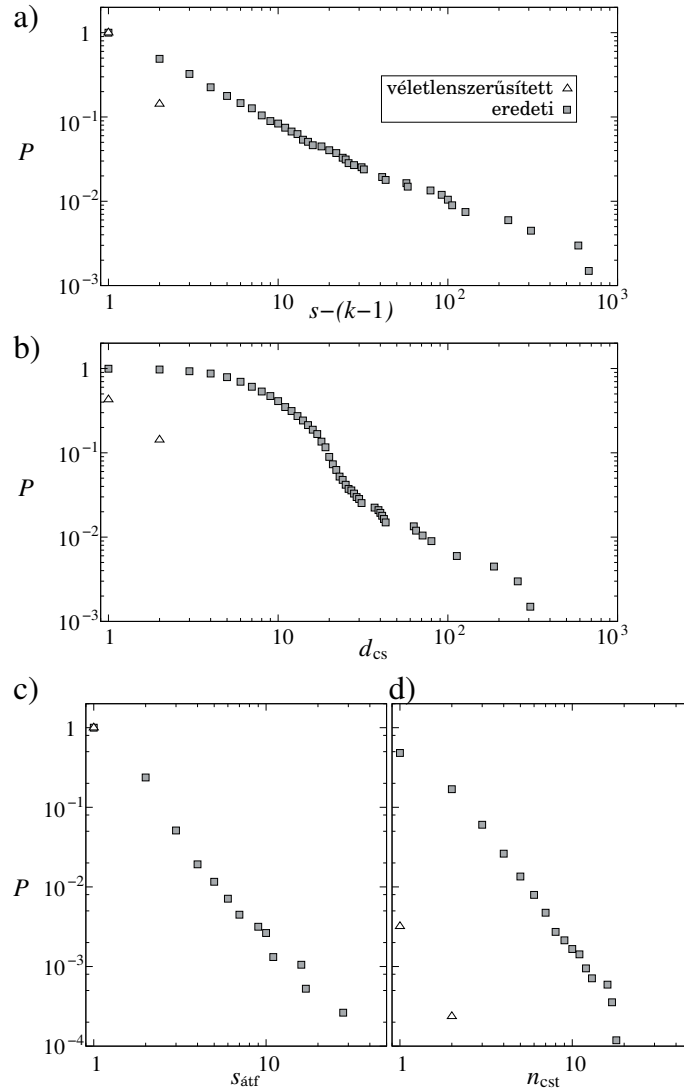


C.3. ábra. Átfedő csoportok a wu-ftp program változóinak hálózatában $k = 5$ -ös k -klikkméret esetén. A csoportok itt általában programfüggvényeknek (szubrutinoknak) felelnek meg. A „@glb” végződésű változók globálisak, azaz minden függvény számára láthatók, és több függvényhívásban is szerepelhetnek. Éppen ezért nagyobb eséllyel szerepelnek a csoportok közti átfedésekben, mint például ezen az ábrán a „makevectentry@glb”, az „addr_cmp@glb”, vagy a „makeportentry@glb” változók. (Az ábra forrása a [T5] publikáció kiegészítő része.)

jellemzők és csak kevésbé függenek a választott k paraméter értékétől.

Az alfejezet zárásaként az eddig tanulmányozott csoportstatisztikákat megvizsgáljuk két informatikához kapcsolódó hálózatban is. Az egyik példa a C.1. alfejezetben bemutatott wu-ftp program változói közt kapott hálózat, míg a másik az OpenOffice szövegszerkesztő program magyar nyelvű szinonimahálózata². Ez utóbbi hálózatban a csúcsok szavaknak felelnek meg, melyek akkor szomszédosak, ha a program szinonimaként ismeri fel őket. A hálózatok mérete $N = 1,886$ és $N = 20,139$ csúcs, illetve $M = 6,001$ és $M = 100,427$ él volt a wu-ftp illetve a szinonimahálózat esetén. A C.6. ábrán a hálózatok csoportstatisztikáit láthatjuk $k = 5$ esetén. Az eredmények meglehetősen hasonlítanak a 3.6. ábrán látottakhoz,

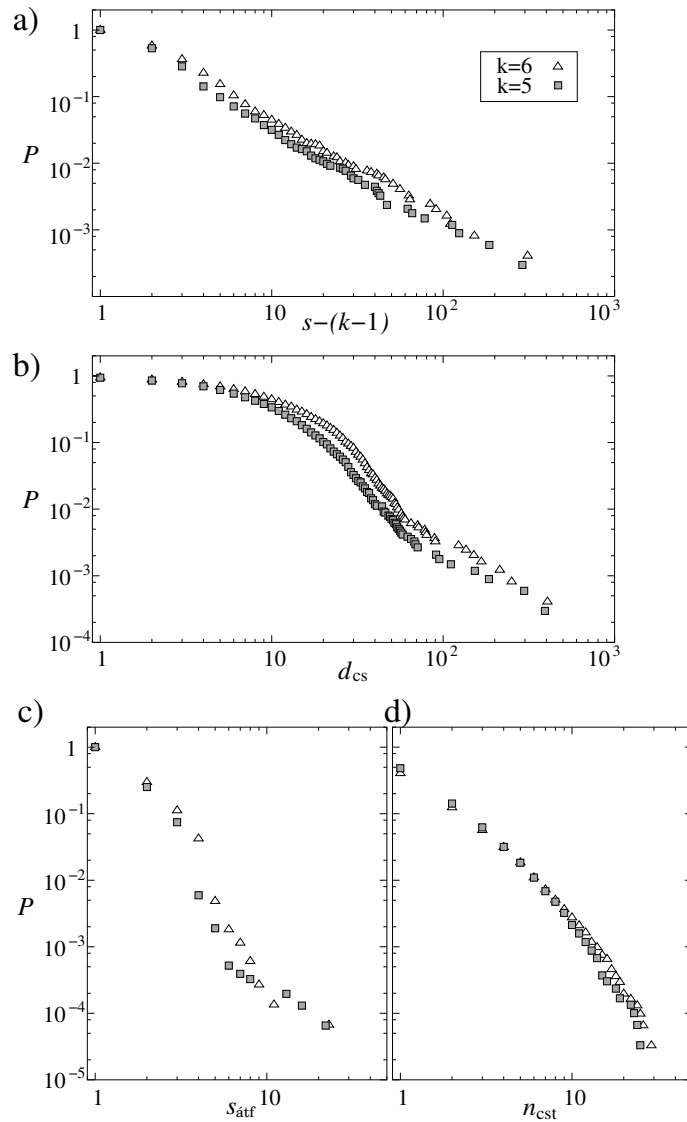
²<http://www.openoffice.org/>



C.4. ábra. A k -klikkperkolációs csoportok statisztikái a véletlenszerűen áthuzalozott szóasszociációs hálózatban. Az elemzéseket a 3.6. ábrához hasonlóan $k = 4$ és $r^* = 0.67$ paraméterbeállítás mellett végeztük el. Az eredeti hálózat statisztikáit a szürke négyzetek jelölik, a véletlenszerűen átkötögetett hálózatét a fehér háromszögek. a) A csoportméret-eloszlás. b) A csoportok foksámeloszlása. c) A csoportátfedések méreteloszlása. d) Az egyes csúcsok csoportszámának eloszlása. (Az ábra forrása a [T5] publikáció kiegészítő része.)

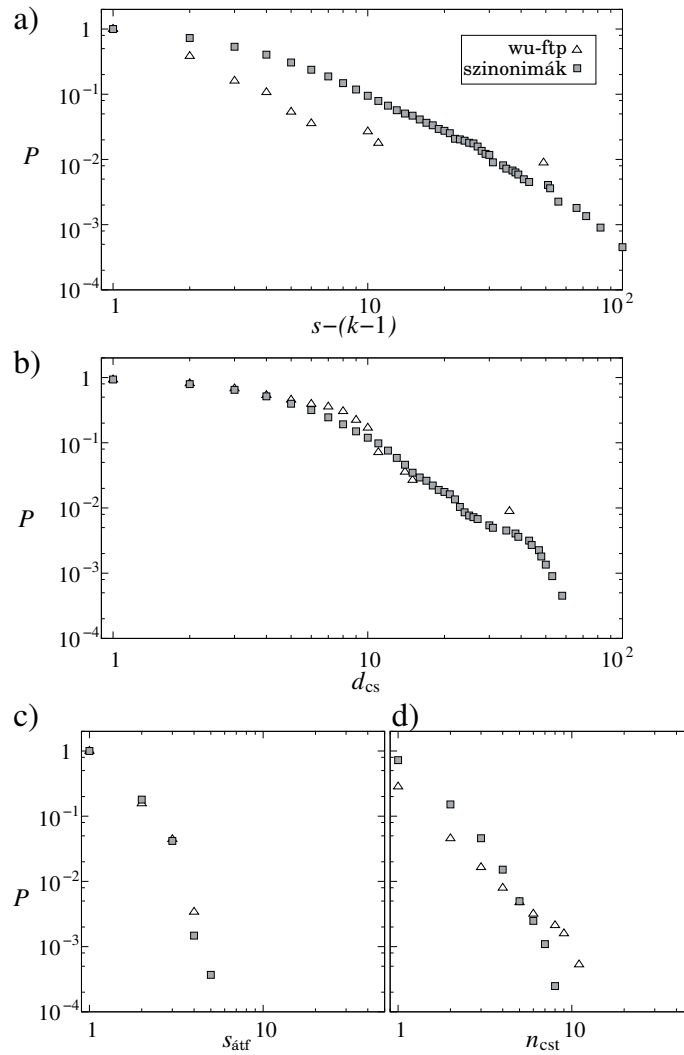
némi nem túl jelentős eltéréssel. A csoportméret-eloszlások (C.6a ábra) itt is hatványszerűen csengenek le, viszont nem olyan széles tartományban, mint például a társszerzőségi vagy a szóasszociációs hálózat esetén. A csoportok foksámeloszlása exponenciális lecsengést mutat a kis csoportfokszámok tartományában ugyanúgy, mint a 3.6b ábra esetén. Viszont a nagyobb d_{cso} tartományra jellemző hatványszerű lecsengés ezeknél a hálózatok-

C FÜGGELÉK: A csoportkereséssel kapcsolatos további eredmények



C.5. ábra. A k -klikkperkolációs csoportok statisztikái a társszerzőségi hálózatban $k=5$ -nél. A $k=5$ és $r^*=0.75$ paraméterbeállításához tartozó adatokat a szürke négyzetek mutatják. Összehasonlításként feltüntettük a 3.6. ábrán szereplő $k=6$ és $r^*=0.93$ paramétereknél kapott eredményeket is (fehér háromszögek). a) A csoportméret-eloszlás. b) A csoportok fokszámeloszlása. c) A csoportátfedések méreteloszlása. d) Az egyes csúcsok csoportszámának eloszlása. (Az ábra forrása a [T5] publikáció kiegészítő része.)

nál nem annyira hangsúlyos, mint a társszerzőségi vagy a szóasszociációs hálózat esetén. Például a wu-ftp hálózatban csak egy csoportnak van kiugróan magas fokszáma, amit magyarázhat az, hogy az egész hálózat mérete itt lényegesen kisebb, mint a társszerzőségi vagy a szóasszociációs hálózat esetén. Ennél fogva a csoportok száma is kevesebb, és így



C.6. ábra. A k -klikkperkolációs csoportok statisztikái a wu-ftp és a szinonimahálózatokban. Mindkét esetben $k = 5$ -ös k -klikkméretnél tártuk fel a csoportokat, a fehér háromszögek a wu-ftp hálózatra kapott eredményeket mutatják, míg a szürke négyzetek az OpenOffice program magyar nyelvű szinonimahálózatának felelnek meg. a) A csoportméret-eloszlás. b) A csoportok fokszámeloszlása. c) A csoportátfedések méreteloszlása. d) Az egyes csúcsok csoportszámának eloszlása. (Az ábra forrása a [T5] publikáció kiegészítő része.)

a maximálisan lehetséges csoportfokszám sem lehet túl nagy. A csoportátfedések méreteloszlása is gyorsabban cseng le, mint a 3.6c ábrán, a legnagyobb átfedés pusztán 5 csúcsot tartalmaz a szinonimák esetén és 4-et a wu-ftp hálózathál. Hasonló viselkedést mutat a csúcsok csoportszámának viselkedése is (C.6d ábra), mely szintén hamarabb lecseng, mint a társszerzőségi illetve a szóasszociációs hálózat esetén.

C.3. A k -klikkperkolációs módszer alkalmazásai

Ebben az alfejezetben áttekintjük a k -klikkperkolációs módszer fontosabb alkalmazásait valódi rendszereket reprezentáló hálózatok elemzésében. A módszert 2005-ös megjelenése óta hálózatok széles skáláján alkalmazták csoportkeresésre, a rákmetasztázishoz kötődő fehérjehálózatoktól szociológiai hálózatokon keresztül weblapok hálózatáig. A jelen esetben igyekszünk csak a lényegesebb, illetve érdekesebb példákra szorítkozni.

Molekuláris biológiai hálózatok

A biológiában, azon belül a bioinformatikában és a rendszerbiológiában a hálózatos megközelítés nagyon nagy népszerűsége tett szert az utóbbi időben [93, 12, 6, 21, 129, 227]. A k -klikkperkolációs megközelítés, különösen a CFinder³ kutatási célból ingyenesen letölthető szoftveres implementációnak [S14] köszönhetően egy rugalmas, könnyen kezelhető megoldást kínál az ezen a területen felmerülő hálózatokban a csoportkeresésre. Mint azt korábban említettük, az egyik nagyon fontos hálózat típus, melyben a k -klikkperkolációs módszerrel kapott eredmények a további elemzések szempontjából érdekes csoportokat adnak, a fehérje-kölcsönhatási hálózatok. Ezeknél a rendszereknél mind a saját [T5, S14], mind egy tőlünk független kutatócsoportok eredményei [226] azt mutatták, hogy a talált csoportok fehérjekomplexeknek, illetve fehérjefunkciós csoportoknak felelnek meg. Ennek alapján a módszer használható fehérjefunkció jóslására is, hiszen ha egy ismeretlen funkciójú fehérje csoportjában a többi fehérjének van egy jól meghatározott közös funkciója, akkor ez alapján erősen felmerül a gyanú, hogy az ismeretlen funkciójú fehérje is rendelkezhet ezzel a funkcióval.

Ezen a vonalon elindulva P. F. Jonsson és munkatársai az általuk vizsgált fehérje-kölcsönhatási hálózat éleinek megbízhatóságát ellenőrizték a k -klikkperkolációs csoportok alapján [107]. A különböző fajok esetén elég széles skálán mozog az adott fajra vonatkozó fehérje-kölcsönhatási hálózatot direktben vizsgáló kísérletek száma, emiatt néhány faj esetén a hálózat felépítéséhez egyéb módszereket is felhasználnak. Ilyenek például a fehérje-homológián (DNS hasonlóságon) alapuló módszerek, melyeknél egy adott fehérjepár esetén jó eséllyel tételezhetünk fel kapcsolatot, ha egy másik fajnál a megfelelő homológ párok esetén kísérleti bizonyítékunk is van a kapcsolat jelenlétére. Ennek révén egy kapcsolaterősség egy fehérje-kölcsönhatási hálózatban több faktorból tevődhet össze: kísérleti eredmények, homológ párok és egyéb források járulhatnak hozzá. A kutatók általában igyekeznek minél több forrást fölhasználni egy adott fehérjehálózat felépítésénél, melynek során általában egy súlyfüggvény kerül bevezetésre, mely megmondja, hogy a különböző források jóslatát egy adott kapcsolat erősségére milyen súllyal kell figyelembe venni. Ezen súlyfüggvény kalibrálása, illetve igazolása egy fontos probléma, amire egy lehetséges megoldást kínálnak a csoportok: Ha jól állítjuk be a súlyfüggvényt, akkor feltehetően magasabb élsúlyokat fogunk kapni a csoportokon belül található élekre, mint az eltérő csoportokat összekötő élekre. Ezt a feltételezést igazolta P. F. Jonsson-nak és munkatársainak tanulmánya, mely a patkány fehérje-kölcsönhatási hálózatát vizsgálta, és kimutatta, hogy a csoportokon belüli élsúlyok

³<http://www.CFinder.org>

szignifikánsan magasabbak, mint a csoportok közti élek súlya [107].

Az idézett publikáció egyik fontos célja volt a rákmetasztázissal kapcsolatos fehérjecsoportok automatikus lokalizációja. A metasztázis során a rákos sejtek le tudnak szakadni az elsődleges tumorról, majd áttéteket hoznak létre más szervekben, ezáltal nagyban nehezítve a rák kezelését. A metasztázis molekuláris biológiai háttere még nem teljesen tisztázott, az viszont általánosan elfogadott, hogy ezeknek a sejteknek megnövekedett motilitásuk és invazív képességeik vannak. Ezek, a sejt korábbi állapotához képest új viselkedési formák kapcsolatban kell legyenek a fehérje-kölcsönhatási hálózat megváltozásával, melynek feltérképezése nagyban segítheti egy hatásos kezelési módszer kifejlesztését. A [107] publikáció egyik fő eredménye, hogy a k -klikkperkolációs módszerrel feltárt csoportok segíthetnek a metasztázis szempontjából kulcsfontosságú fehérjék megtalálásában.

Egy következő, a fentihez szorosan kapcsolódó publikációban P. F. Jonsson és P. A. Bates a metasztázishoz kapcsolódó fehérjék topologikus tulajdonságait vizsgálta az emberi fehérje-kölcsönhatási hálózatban [106]. A hálózat csoportszerkezetét ismét a k -klikkperkolációs módszerrel tárták fel. Az elemzések során kiderült, hogy a metasztázishoz köthető fehérjék gyakrabban szerepelnek a csoportok közti átfedésekben, mint amit az átlagos gyakoriságuk alapján várnánk. Mivel az eltérő csoportok általában különböző sejtbeli folyamatokhoz köthetők, az átfedésekben szereplő fehérjéknek ez által több folyamatra is hatásuk lehet és egyfajta kapocs és szabályozó szerepet is betölthetnek eltérő biokémiai folyamatok között⁴. Például a [106] publikációban bemutatott eredmények között szerepelt egy négy csoportból álló formáció, amit néhány, a rák kialakulásában fontos szerepet játszó fehérje tartott össze, melyek funkciói a kémiai jelátviteltől a reguláción keresztül a sejtnövekedésig és sejthalálig nagyon széles skálán mozogtak. Ezen felül a nagyobb k -klikkméretválasztás mellett feltárt csoportokban magasabb volt a rákhoz köthető fehérjék aránya, és általában, a rákhoz kötődő fehérjék többnyire a nagyobb méretű csoportokban vettek részt. Ennek lehetséges magyarázata az, hogy a rák kialakulásáért és terjedéséért felelős fehérjék feltehetően összetettebb, bonyolultabb folyamatokban vesznek részt, és ezeknek nagyobb csoportok felelnek meg a fehérje-kölcsönhatási hálózatban.

A rákkutatáshoz kapcsolódnak G. Finocchiaro és munkatársainak eredményei is, melyeknél szintén felhasználták a k -klikkperkolációs módszert csoportkeresésre [71]. Ezúttal a hálózat emberi génexpressziós-adatok alapján épült fel: egy bizonyos szignifikancia szint felett az együtt expresszáldó gének csoportjain belül (maximum 10 tagig) minden gént minden másikkal összekötöttek. Az így kapott hálózat csoportszerkezetét több módszerrel is vizsgálták, köztük a k -klikkperkolációs módszerrel is. Az eredményül kapott csoportok sok sejtciklust, apoptózist, foszforizációs kaszkádokat, extracelluláris mátrixot, immun- és interferonválaszt szabályozó fehérjét tartalmaztak. A csoportok többségéről bebizonyosodott, hogy biológiailag releváns, koregulált géneket tartalmaznak, melyek expressziója megváltozik rákos daganat esetén. Ennek révén a talált csoportok új marker géneket szolgáltathatnak, melyekről további vizsgálatok alapján könnyen kiderülhet, hogy esetleg egy

⁴Távolról kapcsolódik ezekhez az eredményekhez a kutatócsoportunk egy olyan publikációja, melyben irányított k -klikkperkolációs csoporttagok centralitását vizsgáltuk valódi rendszereket leíró hálózatokban [S15]. Az elemzések azt mutatták, hogy az átfedésekben szereplő csúcsok általában nagyobb centralitással rendelkeznek a többi csúcshoz viszonyítva.

eddig ismeretlen jelátviteli útvonalat szabályoznak, és ez által új lehetőségeket nyitnak a gyógyszeres kezelések számára.

Szociológiai hálózatok

A k -klikkperkolációs módszert sikeresen alkalmazták szociológiai hálózatokban is, melyekben a csúcsok egyes személyeknek, az élek pedig valamilyen társas kapcsolatnak, ismeretségnek felelnek meg. Ahogy azt a 4. fejezet bevezetőjében is említjük, az ilyen hálózatok tanulmányozásának régi hagyománya van a szociológiában, ahol a kezdetek-kezdetén kézzel kitöltött kérdőlapok és személyes beszélgetések alapján térképezték fel a vizsgált (általában kis méretű) hálózatot. Manapság az informatika, információtechnológia rendkívül gyors fejlődésének köszönhetően szinte automatikusan gyűlnek rendkívül nagy adatbázisok az emberek különféle kapcsolatairól. Egy fontos, úttörő példáját adják ennek az új típusú szociológiai hálózatelemzésnek a [158, 159] publikációkban Kertész János társszerzőségével közölt eredmények, melyekről bővebben szintén a 4. fejezet bevezetőjében teszünk említést.

A szociológiai hálózatok csoportjai általában családoknak, baráti köröknek, munkahelyi csoportoknak, vagy esetleg egy nagyobb társadalmi egységnek felelnek meg. Az előbb felsorolt egységek fontos tulajdonsága, hogy egymással nagyon is átfedhetnek, emiatt célszerű olyan csoportkereső módszert alkalmazni, mint a k -klikkperkolációs módszer, mely megengedi a csoportok közti átfedéseket. A M. C. González és munkatársai által közölt [83] publikációban például amerikai középiskolák tanulói között fellelhető átfedő csoportokat vizsgáltak ezzel a módszerrel. Az elemzések 84 iskola tanulói között (saját bevallás alapján) fellelhető baráti kapcsolatokra terjedtek ki, az alapadatok az Add Health adatbázisából⁵ származtak. A talált csoportok $k = 3$ -as k -klikkméretnél lefedték majdnem az összes tanulót, és a csoportok hálózata egy komplex, viszonylag nagy élsűrűségű szerkezetet mutatott. Ezzel szemben $k = 4$ -nél már csak a diákok kevesebb, mint 20%-a vett részt legalább egy csoportban, és a csoport-hálózat is nagyon ritkává vált. Ezzel párhuzamosan viszont $k = 4$ -nél a különböző etnikai csoportokhoz köthető csoportok száma kiegyensúlyozottá vált még akkor is, ha a diákok között jelentős eltérés volt a különböző etnikumok arányai között. (Ez az effektus $k = 3$ -nál még nem látszott, ott általában a többségben lévő etnikai csoportnak több, és összességében nagyobb csoportjai voltak). Egy lehetséges magyarázat erre az, hogy a kisebbségben lévő diákok erősebb, sűrűbben behálózott csoportokat alkotnak, melyek ezáltal relatíve nagyobb számban maradnak láthatók a magasabb k -klikkméreteknél.

Egy másik érdekes eredménye a [83] publikációnak az volt, hogy míg a diákok közti, barátságokon alapuló hálózat asszortatív (azaz a sok baráttal rendelkező egyének egymás társaságát is kedvelik), addig a csoportok hálózata már disszortatív. Ez megint csak egy arra utaló jel, hogy egy több szerveződési szinttel rendelkező komplex rendszer esetén az egyes szintek nem mindenben hasonlítanak egymásra, hanem ehelyett bizonyos, szintekre specifikusan jellemző tulajdonságokat találhatunk.

A szociológiai hálózatokkal kapcsolatos új eredmények rövidesen új hálózati modellek bevezetését is inspirálták [204, 117, 84]. R. Toivonen és munkatársai [204], valamint J. M.

⁵<http://www.cpc.unc.edu/addhealth>

Kumpula és munkatársai [117] két eltérő kapcsolódási mechanizmus közti egyensúlyra vezetik vissza a szociológiai hálózatok sajátos szerkezetének kialakulását. Az egyik mechanizmus (angolul *cyclic closure*) szerint az új kapcsolatok a már amúgy is sűrűn kapcsolt szomszédságok környezetében jönnek létre, ahogy két ember, akiknek nagyon sok közös ismerőse van, az várhatóan előbb utóbb egymással is ismeretséget fog kötni. A másik mechanizmus (angolul *focal closure*) viszont független a csúcsok környezetétől vagy a hálózatban mérhető távolságától, ehelyett valamilyen közös tevékenység, közös aktivitás (pl. közös hobbi) révén jön létre. A két eltérő mechanizmus által létrehozott élek relatív gyakoriságát változtatva más és más lesz a hálózat szerkezete: Ha az előbbi mechanizmus dominál, akkor lokálisan sűrű lesz a hálózat, jól definiált erős csoportokkal, míg ha a második mechanizmusa a döntő szerep, akkor egy homogén hálózatot kapunk, nem túl jelentős csoportszerkezettel. A két határeset közötti átmenetnél a csoportszerkezetet a k -klikkperkolációs módszerrel tárták fel [204, 117]. Az eredmények szerint a két mechanizmus megfelelő arányú keverésével a korábban tanulmányozott mobiltelefon-hívási hálózat lényeges tulajdonságai jól reprodukálhatók.

Ettől eltérően M. C. González és munkatársai egy olyan hálózatmodellel álltak elő, mely véges méretű, periodikus határfeltételeknek kitett, egymással időnként ütköző részecskék kölcsönhatásainak eredményeként áll elő [84]. Minden ütközés egy új élt köt be az ütközésben résztvevő részecskék közé, és a sebességek frissítésének szabályai az ütközés után függenek a részecskéhez kapcsolódó élek számától. Az ütközési szabályok megfelelő beállításával, valamint a részecskék számára véges élettartam bevezetésével (és természetesen az „elbomlott” részecskék pótlásával), olyan kvázistacionárius állapotba áll be a hálózatnak, mely viszonylag pontosan reprodukálja a korábban vizsgált középiskolás diákok közti hálózatok statisztikus tulajdonságait. Ezen vizsgált tulajdonságok egyike például a k -klikkperkolációs módszerrel kapható csoportok méreteloszlása.

Végül érdekességként megemlíjtjük még R. D. Smith munkáját is [192], melyben rapperek közti együttműködési hálózatokat vizsgált. A hálózat élei itt közös albumoknak, közös számoknak, vagy közös fellépéseknek feleltek meg. R. D. Smith a hálózat csoportszerkezetét több módszerrel tanulmányozta, többek között a k -klikkperkolációs módszerrel is.

Egyéb rendszerek

Gazdasági hálózatok esetén is használható csoportkeresésre a k -klikkperkolációs módszer, amint azt T. Heimo és munkatársai tették a [96] publikációban. Vizsgálatuk tárgya a new yorki tőzsdéhez kapcsolható, részvények közötti hálózat volt. Ebben a hálózatban a részvények árfolyamai közötti korreláció alapján húzzuk be az éleket (azaz ha egy bizonyos értéket meghalad a korreláció, akkor van kapcsolat, ha alatta marad, akkor nincs). A használt korrelációs mátrixot spektrális módszerekkel, a kapott hálózatot a k -klikkperkolációs módszerrel elemezték. Az eredmények szerint a korrelációs mátrix első néhány sajátvektora ugyan a csoportokon lokalizálódik, viszont nem adnak olyan világos klaszterhatárokat, mint a csoportkeresés. Ezután ahogy haladunk végig a sajátvektorokon a sajátértékek csökkenő sorrendjében, a sajátvektorok egyre kevésbé lesznek lokalizáltak a korrelációs értéken alapuló hálózaton. Ennek alapján az erősen összekapcsolt, egymáshoz szorosan kötődő részvénycsoportokat érdemesebb hálózati csoportkereséssel lokalizálni a tisztán spektrális

megközelítés mellett.

A k -klikkperkolációs módszer dokumentumklaszterezésre is használható W. Gao és K. F. Wong munkája alapján [78]. A dokumentumokat itt egy hasonlósági hálózat kapcsolta össze (az egymásra jobban hasonlító dokumentumok között erősebb élek voltak). Az eredmények szerint a k -klikkperkolációs módszerrel kapott csoportok bizonyos esetekben jobb eredményeket adnak néhány sztenderd dokumentumklaszterező algoritmusnál és közelebb állnak a természetes dokumentumklaszterezéshez.

Egy további érdekes alkalmazását láthatjuk a k -klikkperkolációs módszernek a [45] publikációban, melyben X. Castelló és munkatársai egymással versengő vélemények dinamikáját tanulmányozták. A híres voter modellben az ágensek állapota (véleménye) vagy A vagy B [98], míg az ún. AB modellben egy harmadik, átmeneti AB állapot is megengedett [44]. Az ágensek hálózatát a korábban említett, új keletű szociológiai hálózatmodellek [204, 117] segítségével hozták létre. A szimulációk során időlépésenként egy véletlenszerűen választott ágens állapotát (véleményét) a közvetlen szomszédainak állapotától függő valószínűséggel változtatták meg. A vélemények véletlenszerű kiinduló állapotából indulva mind a voter modellben, mind az AB modellben a rendszer konszenzushoz konvergált, melynél már minden csúcs azonos véleményen volt. Ellenben az AB modell esetén a konszenzus eléréséhez szükséges iterációk átlagos száma nagy mértékben függött az ágensek hálózatának topológiájától. Az eredmények szerint amennyiben a hálózatban a k -klikkperkolációs módszerrel feltárt csoportok egy kiterjedt, gazdag csoportszerkezetet mutatnak, a konszenzustól eltérő metastabil állapotok élettartam eloszlása hatványfüggvényyszerű alakot vesz fel. Ennek révén általában akadnak a rendszerben kiemelkedően magas élettartamú metastabil állapotok, melyekben a véleménydinamika hosszú időre „csapdába eshet”. Ezek a skálázó metastabil állapotok egyértelműen a mezoszkopikus hálózatszerkezettel, azaz a csoportokkal vannak szoros kapcsolatban.

D függelék

A multifraktál alapú gráfgeneráló termodinamikai határeset

Az itt bemutatásra kerülő, a multifraktál alapú hálózatgenerálás $N \rightarrow \infty$ határesetével kapcsolatos analitikus számításokat Lovász László végezte el [T8]. Tekintsük általánosan az egység négyzeten definiált, szimmetrikus, mérhető függvények $L_1(x, y), L_2(x, y), \dots, L_q(x, y)$ sorozatát (ahol $0 \leq L_k(x, y) \leq 1$ tetszőleges q -ra), és definiáljuk az x pozíciójú csúcs $p_q(x)$ átlagos élbekötési valószínűségét úgy, mint

$$p_q(x) = \int_0^1 L_q(x, y) dy. \quad (\text{D.1})$$

Hasonló módon, jelölje ω_q az egész hálózat átlagos élbekötési valószínűségét, melyet az

$$\omega_q = \int_0^1 p_q(x) dx \quad (\text{D.2})$$

alakban lehet megadni. A csúcsok N_q számát válasszuk meg úgy, hogy a generált hálózat átlagos fokszáma egy nullánál nagyobb konstans $\langle d \rangle$ értékhez konvergáljon $q \rightarrow \infty$ esetén, azaz

$$N_q \omega_q \rightarrow \langle d \rangle. \quad (\text{D.3})$$

(Ez azt jelenti, hogy az élek száma valahol $\langle d \rangle / 2$ körül található.) Az x pozíciójú csúcs fokszámeloszlása binomiális,

$$\rho(d, x) = \binom{N_q}{d} p_q(x)^d [1 - p_q(x)]^{N_q - 1 - d}. \quad (\text{D.4})$$

A termodinamikai határesetben ezt egy Poisson-eloszlással lehet közelíteni a

$$\rho(d, x) \simeq \frac{[N_q p_q(x)]^d}{d!} e^{-N_q p_q(x)} \quad (\text{D.5})$$

alakban. A teljes hálózat fokszámeloszlását a $\rho(d, x)$ integrálásával kaphatjuk meg, mely a

$$\rho(d) = \frac{1}{d!} \int_0^1 [N_q p_q(x)]^d e^{-N_q p_q(x)} dx \quad (\text{D.6})$$

D FÜGGELÉK: *A multifraktál alapú gráfgeneráló termodinamikai határesetek*

eredményre vezet. Ennek alapján egy izolált csúcs valószínűsége (mely a nulla fokszámnak felel meg),

$$\rho(d=0) = \int_0^1 e^{-N_q p_q(x)} dx. \quad (D.7)$$

A (D.2) és (D.3) alapján a $p_q(x)$ átlagértéke $\langle d \rangle / N_q$ körül van. Amennyiben $p_q(x)$ független x -től, úgy $p_q(x) = \langle d \rangle / N_q$ és

$$\rho(d=0) \simeq e^{-\langle d \rangle}. \quad (D.8)$$

Azonban ha $p_q(x)$ olyan alakot vesz fel, mely esetén tipikus értéke sokkal kisebb az átlagánál, akkor többnyire $e^{-N_q p_q(x)} \simeq 1$, melynek hozományaként

$$\rho(d=0) \simeq 1, \quad (D.9)$$

ami egyben azt is jelenti, hogy a csúcsok többsége izolálttá válik. Ezen degeneráció akkor kerülhető el, ha

$$\int_0^1 e^{-N_q p_q(x)} dx < c, \quad (D.10)$$

ahol $c < 1$ egy konstans.

A multifraktál alapú hálózatgenerálásnál, a fenti feltétel csak akkor teljesül, ha $p_q(x)$ független x -től. Amint azt az 5.2. alfejezetben említettük, a modellünk általánosítható egy, a $[0,1]$ és $[0,1]^q$ között definiált mérték megőrző bijekció segítségével, melynek révén áttérhetünk a $L_q \equiv L^{\otimes q} = L \otimes \cdots \otimes L$ tenzorokra, ahol

$$L(x_1, \dots, x_q, y_1, \dots, y_q) = L(x_1, y_1) \cdots L(x_q, y_q). \quad (D.11)$$

A (D.1) kifejezéssel analóg peremeloszlások ebben a reprezentációban a

$$\begin{aligned} p_k(x_1, \dots, x_q) &= \int_{[0,1]^q} L(x_1, y_1) \cdots L(x_q, y_q) dy_1 \dots dy_q \\ &= p(x_1) \cdots p(x_q), \end{aligned} \quad (D.12)$$

alakban írhatók fel, ahol $p(x) = \int_0^1 L(x, y) dy$. Hasonlóan, a (D.2) összefüggés az

$$\omega_q = \int_{[0,1]^q} p_q(x_1, \dots, x_q) dx_1 \dots dx_q = \omega^q, \quad (D.13)$$

egyenletbe megy át, ahol $\omega = \int_0^1 p(x) dx$. A (D.3) alapján így a rendszerméretet az $N_q \simeq \langle d \rangle / \omega^q$ módon kell változtatnunk.

Sajnos ezek a függvények csak akkor elégítik ki a (D.10) feltételt, ha $p(x)$ konstans. Valóban, ha $(x_1, \dots, x_k, y_1, \dots, y_q)$ egy véletlenszerű pont a $[0,1]^q$ -ban, akkor a nagy számok törvénye alapján annak valószínűsége, hogy

$$\ln p_k(x_1, \dots, x_q) = \ln p(x_1) + \cdots + \ln p(x_q) \sim k \int_0^1 \ln p(x) dx \quad (D.14)$$

egyhez tart. Legyen $\alpha = \exp(\int \ln p(x) dx)$. Ekkor $\alpha < \omega$ a Jensen-egyenlőtlenség alapján (kivéve ha p konstans), és a p_q értéke majdnem mindig közel esik α^q -hoz, míg az átlaga ω^q -val egyenlő. Mivel $(\alpha/\omega)^q \rightarrow 0$ ha $q \rightarrow \infty$, a (D.10) egyenlőség csak akkor állhat fent, ha $p(x)$ konstans.

Másfelől, ha $p(x) = \omega$ tetszőleges x -re, akkor a csúcsok várható fokszáma függetlenné válik a pozíciójuktól, és a fokszámeloszlás egy Poisson-eloszláshoz konvergál hasonló módon, mint egy Erdős–Rényi-gráfban. Ebben az esetben könnyű meghatározni az L_q alapján generált bármely $\mathcal{G}_q$ hálózatban egy tetszőleges, l csúcsból álló, összefüggő $\mathcal{F}$ gráf másolatainak számát. Összesen $N_k(N_q-1) \cdots (N_q-l+1) \sim N_q^l$ -féle módon lehet $V(\mathcal{F})$ -et injektíven $[n_q]$ -ba képezni, és minden leképezés esetén annak valószínűsége, hogy egy homomorfizmust kaptunk $\varrho(\mathcal{F}, L_q) = \varrho(\mathcal{F}, L)^q$. Ennek alapján

$$E(\text{inj}(\mathcal{F}, \mathcal{G}_q)) \sim N_q^l \varrho(\mathcal{F}, L)^q = \langle d \rangle^l \left(\frac{\varrho(\mathcal{F}, L)}{\omega^l} \right)^q. \quad (\text{D.15})$$

Mivel egy ritka gráffal állunk szemben, ezt N_q -val normáljuk, így $\mathcal{F}$ másolatainak normált száma

$$E(s(\mathcal{F}, \mathcal{G}_q)) = \frac{\text{hom}(\mathcal{F}, \mathcal{G}_q)}{N_q} \sim N_q^{l-1} \varrho(\mathcal{F}, L)^q = \langle d \rangle^{l-1} \left(\frac{\varrho(\mathcal{F}, L)}{\omega^{l-1}} \right)^q. \quad (\text{D.16})$$

Ennek alapján például a háromszögek normált száma

$$\frac{1}{6} s(\mathcal{K}_3, \mathcal{G}_q) \sim \langle d \rangle^2 \left(\frac{\varrho(\mathcal{K}_3, L)}{\omega^2} \right)^q. \quad (\text{D.17})$$

Könnyű belátni, hogy ha $p(x) = \omega$ konstans, akkor

$$\varrho(\mathcal{F}, L) \leq \omega^{l-1}, \quad (\text{D.18})$$

ahol egyenlőség akkor és csak akkor áll fent ha $\mathcal{F}$ fa-szerű vagy L egy olyan ekvivalencia reláció, hogy létezik egy $S_1 \cup \dots \cup S_m = [0, 1]$ partíció és

$$L(x, y) = \begin{cases} 1, & \text{ha } x, y \in S_i \text{ valamely } i\text{-re,} \\ 0, & \text{egyébként.} \end{cases} \quad (\text{D.19})$$

A (D.16) és a (D.18) alapján a következőket írhatjuk fel:

$$E(s(\mathcal{F}, \mathcal{G}_q)) \rightarrow \begin{cases} \langle d \rangle^{l-1}, & \text{ha } \mathcal{F} \text{ fa-szerű vagy } L \text{ egy ekvivalencia reláció,} \\ 0, & \text{egyébként.} \end{cases} \quad (\text{D.20})$$

Megmutatható, hogy ezen konvergencia majdnem mindig bekövetkezik. A (D.20) alapján az is látszik, hogy a $\mathcal{G}_q$ sorozat 1 valószínűséggel konvergens a Benjamini–Schramm-értelmeben.

A vázolt probléma kiküszöbölésére több irányban is el lehet indulni, azonban ezek tárgyalása túlmutat a jelenlegi dolgozat témakörén. Meg lehet próbálni a tenzoros formalizmuson belül hozzáadni a $L^{\oplus q}$ -hoz egy konstans c_q -t, mely megfelelően lassan tart 0-hoz. Egy másik lehetőség a $L^{\oplus q}$ módosítása $(L^{\oplus q})^{a_q}$ -ra, ahol $a_q \rightarrow 0$. Egy további ötlet az eredeti multifraktálok alapuló formalizmusnál a multifraktál elforgatása az egységnyígyzeten belül, amit az [S13] publikáció mutat be részletesebben.

D FÜGGELÉK: *A multifraktál alapú gráfgeneráló termodinamikai határeset*

Saját publikációk

A tézispontok alapjául szolgáló 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.

A dolgozathoz kapcsolódó további saját publikációk

- [S1] I. Farkas – I. Derényi – G. Palla – T. Vicsek: Equilibrium statistical mechanics of network structures. *Lect. Notes Phys.*, 650. évf. (2004), 163–187. p.
- [S2] G. Palla – D. Ábel – I. J. Farkas – P. Pollner – I. Derényi – T. Vicsek: k-clique Percolation and Clustering. In B. Bollobás – R. Kozma – D. Miklós (szerk.): *Handbook of Large-scale Random Networks*. 9. fejezet. 2009, Springer, 369–408. p.

- [S3] I. J. Farkas – D. Ábel – G. Palla – T. Vicsek: Weighted network modules. *New J. Phys.*, 9. évf. (2007), 180. p.
- [S4] P. Pollner – G. Palla – T. Vicsek: Preferential attachment of communities: The same principle, but a higher level. *Europhys. Lett.*, 73. évf. (2006), 478–484. p.
- [S5] G. Palla – T. Vicsek – A.-L. Barabási: Community dynamics in social networks. *Fluct. Noise Lett.*, 7. évf. (2007), L273–L287. p.
- [S6] G. Palla – T. Vicsek – A.-L. Barabási: Statistical Properties of Community Dynamics in Large Social Networks. *International Journal of Agent Technologies & Systems*, 1. évf. (2009), 1–16. p.
- [S7] G. Palla – P. Pollner – A.-L. Barabási – T. Vicsek: Social Group Dynamics in Networks. In T. Gross – H. Sayama (szerk.): *Adaptive Networks*. 2. fejezet. Berlin/Heidelberg, 2009, Springer, 11–38. p.
- [S8] Palla G. – Barabási A.-L. – Vicsek T.: Társas kapcsolatok hálózata. *Természet Világa*, 139. évf. (2008), 108–110. p.
- [S9] Palla G. – Kertész J.: Szociofizika: humán kapcsolatok hálózata nagy skálán. *Fizikai Szemle*, 2008/6. évf. (2008), 217. p.
- [S10] G. Tibély – P. Pollner – T. Vicsek – G. Palla: Extracting Tag Hierarchies. *PLoS ONE* 8. évf. (2013), e84133. p.
- [S11] G. Palla – G. Tibély – E. Mones – P. Pollner – T. Vicsek: Hierarchical networks of scientific journals. *Palgrave Communications* 1. évf. (2015), 15016. p.
- [S12] D. Czégel – G. Palla: Random walk hierarchy measure: What is more hierarchical, a chain, a tree or a star? *Scientific Reports* 5. évf. (2015), 17994. p.
- [S13] G. Palla – P. Pollner – T. Vicsek: Rotated multifractal network generator. *J. Stat. Mech.*, (2011), P02003. p.
- [S14] B. Adamcsek – G. Palla – I. J. Farkas – I. Derényi – T. Vicsek: CFinder: Locating cliques and overlapping modules in biological networks. *Bionformatics*, 22. évf. (2006), 1021–1023. p.
- [S15] P. Pollner – G. Palla – D. Ábel – A. Vicsek – I. J. Farkas – I. Derényi – T. Vicsek: Centrality properties of directed module members in social networks. *Physica A*, 387. évf. (2008), 4959–4966. p.

Irodalomjegyzék

- [1] D. Achlioptas–R. M. D’Souza–J. Spencer: Explosive percolation in random networks. *Science*, 323. évf. (2009), 1453–1455. p.
- [2] L. A. Adamic–B. A. Huberman: Internet: growth dynamics of the world-wide web. *Nature*, 401. évf. (1999), 131. p.
- [3] Y.-Y. Ahn–J. P. Bagrow–S. Lehmann: Link communities reveal multiscale complexity in networks. *Nature*, 466. évf. (2010), 761–764. p.
- [4] C. Aicher–A. Z. Jacobs–A. Clauset: Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3. évf. (2015), 221–248. p.
- [5] W. Aiello–F. Chung–L. Lu: A random graph model for massive graphs. In *Proc. 32nd ACM Symp. on the Theory Comput.* (konferenciaanyag). New York, 2000, ACM, 171–180. p.
- [6] T. Aittokallio–B. Schwikowski: Graph-based methods for analysing networks in cell biology. *Briefings in Bioinformatics*, 7. évf. (2006), 243–255. p.
- [7] R. Albert–A.-L. Barabási: Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74. évf. (2002), 47–97. p.
- [8] R. Albert–H. Jeong–A.-L. Barabási: Diameter of the world wide web. *Nature*, 401. évf. (1999), 130–131. p.
- [9] N. A. M. Araújo–H. J. Herrmann: Explosive percolation via control of the largest cluster. *Phys. Rev. Lett.*, 105. évf. (2010), 035701. p.
- [10] 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.
- [11] M. Baiesi–S. S. Manna: Scale free networks from a hamiltonian dynamics. *Phys. Rev. E*, 68. évf. (2003), 047103. p.
- [12] A.-L. Barabási–Z. N. Oltvai: Network Biology: Understanding the Cells’s Functional Organization. *Nat. Rev. Genet.*, 5. évf. (2004), 101–113. p.

- [13] A.-L. Barabási: *Network Science*. 2015, Cambridge University Press.
URL <http://barabasi.com/networksciencebook/>.
- [14] A.-L. Barabási–R. Albert: Emergence of scaling in random networks. *Science*, 286. évf. (1999), 509–512. p.
- [15] 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.
- [16] A. L. Barabási–E. Ravasz–T. Vicsek: Deterministic scale-free networks. *Physica A*, 299. évf. (2001), 559–564. p.
- [17] E. R. Barnes: An algorithm for partitioning the nodes of a graph. *SIAM J. Algebr. Discrete Methods*, 3. évf. (1982), 541–550. p.
- [18] D. Ben-Avraham–S. Havlin: *Diffusion and Reactions in Fractals and Disordered Systems*. Cambridge, 2000, Cambridge University Press.
- [19] D. Bennett-Wood–I. G. Enting–D. S. Gaunt–A. J. Guttmann–J. L. Leask–A. L. Owczarek–S. G. Whittington: Exact enumeration study of free energies of interacting polygons and walks in two dimensions. *J. Phys. A*, 31. évf. (1998), 4725. p.
- [20] J. Berg–M. Lässig: Correlated random networks. *Phys. Rev. Lett.*, 89. évf. (2002), 228701. p.
- [21] A. Beyer–S. Bandyopadhyay–T. Ideker: Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.*, 9. évf. (2007), 699–710. p.
- [22] G. Bianconi: The entropy of randomized network ensembles. *Europhys. Lett.*, 81. évf. (2008), 28005. p.
- [23] G. Bianconi–A.-L. Barabási: Bose-einstein condensation in complex networks. *Phys. Rev. Lett.*, 86. évf. (2001), 5632. p.
- [24] G. Bianconi–M. Marsili: Emergence of large cliques in random scale-free networks. *Europhys. Lett.*, 74. évf. (2006), 740–746. p.
- [25] G. Bianconi–M. Marsili: Number of cliques in random scale-free network ensembles. *Physica D*, 224. évf. (2006), 1–6. p.
- [26] M. Blatt–S. Wiseman–E. Domany: Superparamagnetic clustering of data. *Phys. Rev. Lett.*, 76. évf. (1996), 3251–3254. p.
- [27] V. D. Blondel–J.-L. Guillaume–R. Lambiotte–E. Lefebvre: Fast unfolding of communities in large networks. *J. Stat. Mech.*, 2008., P10008. p.
- [28] M. Boguñá–R. Pastor-Satorras: Class of correlated random networks with hidden variables. *Phys. Rev. E*, 68. évf. (2003), 036112. p.

-
- [29] B. Bollobás: *Graph Theory, An Introductory Course*. Berlin, 1979, Springer-Verlag.
 - [30] B. Bollobás: *Random graphs*. 2. kiad. Cambridge, 2001, Cambridge University Press.
 - [31] B. Bollobás–S. Janson–O. Riordan: The phase transition in inhomogeneous random graphs. *Random Struct. Algor.*, 31. évf. (2007), 3–122. p.
 - [32] B. Bollobás–O. Riordan: Clique percolation. *Random Struct. Algor.*, 35. évf. (2009), 294–322. p.
 - [33] 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.
 - [34] C. Borgs–J. Chayes–L. Lovász–V. T. Sós–K. Vesztergombi: Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219. évf. (2008), 1801–1851. p.
 - [35] E. I. Boyle–S. A. Weng–J. Gollub–H. Jin–D. Botstein–J. M. Cherry–G. Sherlock: GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, 20. évf. (2004), 3710–3715. p.
 - [36] U. Brandes–D. Delling–M. Gaertler–R. Görke–M. Hoefer–Z. Nikoloski–D. Wagner: On modularity clustering. *IEEE T. Knowl. Data En.*, 20. évf. (2008), 172–188. p.
 - [37] S. R. Broadbent–J. M. Hammersley: Percolation processes i. crystals and mazes. *Proc. Camb. Phil. Soc.*, 53. évf. (1957), 629–641. p.
 - [38] C. Bron–J. Kerbosch: Algorithm 457: Finding all cliques of an undirected graph. *Commun. ACM*, 16. évf. (1973), 575–577. p.
 - [39] S. V. Buldyrev–R. Parshani–G. Paul–H. E. Stanley–S. Havlin: Catastrophic cascade of failures in interdependent networks. *Nature*, 464. évf. (2010), 1025–1028. p.
 - [40] Z. Burda–J. D. Correia–A. Krzywicki: Statistical ensemble of scale-free random graphs. *Phys. Rev. E*, 64. évf. (2001), 046118. p.
 - [41] Z. Burda–A. Krzywicki: Uncorrelated random networks. *Phys. Rev. E*, 67. évf. (2003), 046118. p.
 - [42] 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.
 - [43] R. F. Cancho–R. V. Solé: The small world of human language. *Proc. R. Soc. Lond. B*, 268. évf. (2001), 2261–2265. p.
 - [44] X. Castelló–V. M. Eguíluz–M. San Miguel: Ordering dynamics with two non-excluding options: bilingualism in language competition. *New J. Phys.*, 8. évf. (2006), 308. p.

-
- [45] X. Castelló–R. Toivonen–V. M. Eguíluz–K. Kaski–M. San Miguel: Anomalous lifetime distributions and topological traps in ordering dynamics. *Europhys. Lett.*, 79. évf. (2007), 66006. p.
 - [46] D. Chen–M. Shang–Z. Lv–Y. Fu: Detecting overlapping communities of weighted networks via a local algorithm. *Physica A*, 389. évf. (2010), 4177–4187. p.
 - [47] J. M. Cherry–C. Ball–S. Weng–G. Juvik–R. Schmidt–C. Adler–B. Dunn–S. Dwight–L. Riles–R. K. Mortimer–D. Botstein: Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature*, 387(6632 Suppl). évf. (1997), 67–73. p.
 - [48] Y. S. Cho–J. S. Kim–J. Park–B. Kahng–D. Kim: Percolation transitions in scale-free networks under the achlioptas process. *Phys. Rev. Lett.*, 103. évf. (2009), 135702. p.
 - [49] A. Clauset–C. Moore–M. E. J. Newman: Hierarchical structure and the prediction of missing links in networks. *Nature*, 453. évf. (2008), 98–101. p.
 - [50] A. Clauset–M. E. J. Newman–C. Moore: Finding community structure in very large networks. *Phys. Rev. E*, 70. évf. (2004), 066111. p.
 - [51] B. Corominas-Murtra–J. Goñi–R. V. Solé–C. Rodríguez-Caso: On the origins of hierarchy in complex networks. *Proc. Natl. Acad. Sci. USA*, 110. évf. (2013), 13316–13321. p.
 - [52] B. Corominas-Murtra–C. Rodríguez-Caso–J. Goñi–R. Solé: Measuring the hierarchy of feedforward networks. *Chaos*, 21. évf. (2011), 016108. p.
 - [53] P. Csermely: *Weak Links*. Berlin Heidelberg, 2006, Springer-Verlag.
 - [54] R. A. da Costa–S. N. Dorogovtsev–A. V. Goltsev–J. F. F. Mendes: Explosive percolation transition is actually continuous. *Phys. Rev. Lett.*, 105. évf. (2010), 255701. p.
 - [55] P. Devillard–H. E. Stanley: Exact enumeration approach to the directed polymer problem. *Phys. Rev. A*, 41. évf. (1989), 2942. p.
 - [56] P. S. Dodds–R. Muhamad–D. J. Watts: An experimental study of search in global social networks. *Science*, 301. évf. (2003), 827–829. p.
 - [57] R. M. D’Souza–M. Mitzenmacher: Local cluster aggregation models of explosive percolation. *Phys. Rev. Lett.*, 104. évf. (2010), 195702. p.
 - [58] J. Duch–A. Arenas: Community detection in complex networks using extremal optimization. *Phys. Rev. E*, 72. évf. (2005), 027104. p.
 - [59] H. Ebel–J. Davidsen–S. Bornholdt: Dynamics of social networks. *Complexity*, 8. évf. (2002), 24–27. p.

-
- [60] H. Ebel–L.-I. Mielsch–S. Bornholdt: Scale-free topology of e-mail networks. *Phys. Rev. E*, 66. évf. (2002), 35103(R). p.
 - [61] J.-P. Eckmann–E. Moses–D. Sergi: Entropy of dialogues creates coherent structures in e-mail traffic. *Proc. Natl. Acad. Sci. USA*, 101. évf. (2004), 14333–14337. p.
 - [62] F. Eggenberger–G. Pólya: Über die statistik verketteter vorgänge. *Zeitschrift für Angewandte Mathematik und Mechanik*, 3. évf. (1923), 279–289. p.
 - [63] P. Erdős–A. Rényi: On random graphs I. *Publ. Math. Debrecen*, 6. évf. (1959), 290–297. p.
 - [64] P. Erdős–A. Rényi: On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5. évf. (1960), 17–61. p.
 - [65] A. V. Esquivel–M. Rosvall: Compression of flow can reveal overlapping-module organization in networks. *Phys. Rev. X*, 1. évf. (2011), 021025. p.
 - [66] L. Euler: Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8. évf. (1741), 128–140. p.
 - [67] M. G. Everett–S. P. Borgatti: Analyzing Clique Overlap. *Connections*, 21. évf. (1998), 49–61. p.
 - [68] B. S. Everitt: *Cluster Analysis*. 3. kiad. London, 1993, Edward Arnold.
 - [69] M. Faloutsos–P. Faloutsos–C. Faloutsos: On power-law relationships of the internet topology. *Comput. Commun. Rev.*, 29. évf. (1999), 251. p.
 - [70] Mérei Ferenc: *Közösségek rejtett hálózata*. Budapest, 1971, Közgazdasági és Jogi Könyvkiadó.
 - [71] G. Finocchiaro–F. M. Mancuso–D. Cittaro–H. Muller: Graph-based identification of cancer signaling pathways from published gene expression signatures using PubLiME. *Nucl. Ac. Res.*, 35. évf. (2007), 2343–2355. p.
 - [72] S. Fortunato: Community detection in graphs. *Phys. Rep.*, 486. évf. (2010), 75–174. p.
 - [73] S. Fortunato–M. Barthélemy: Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA*, 104. évf. (2007), 36–41. p.
 - [74] O. Frank–D. Strauss: Markov graphs. *J. Am. Stat. Assoc.*, 81. évf. (1986), 832–842. p.
 - [75] J. Gao–S. V. Buldyrev–S. Havlin–H. E. Stanley: Robustness of a network of networks. *Phys. Rev. Lett.*, 107. évf. (2011), 195701. p.
 - [76] J. Gao–S. V. Buldyrev–H. E. Stanley–S. Havlin: Networks formed from interdependent networks. *Nat. Phys.*, 8. évf. (2012), 40–48. p.

- [77] J. Gao–S. V. Buldyrev–H. E. Stanley–X. Xu–S. Havlin: Percolation of a general network of networks. *Phys. Rev. E*, 88. évf. (2013), 062816. p.
- [78] W. Gao–K.-F. Wong: Natural document clustering by clique percolation in random graphs. *Lect. Notes in Comp. Sci.*, 4182. évf. (2006), 119–131. p.
- [79] R. Gibrat: *Les Inégalités économiques*. Paris, France, 1931, Librairie du Recueil Sirey.
- [80] E. N. Gilbert: Random graphs. *Ann. Math. Stat.*, 30. évf. (1959), 1141–1144. p.
- [81] M. Girvan–M. E. J. Newman: Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99. évf. (2002), 7821–7826. p.
- [82] A. V. Goltsev–S. N. Dorogovtsev–J. F. F. Mendes: Critical phenomena in networks. *Phys. Rev. E*, 67. évf. (2003), 026123. p.
- [83] M. C. González–H. J. Herrmann–J. Kertész–T. Vicsek: Community structure and ethnic preferences in school friendship networks. *Physica A*, 379. évf. (2007), 307–316. p.
- [84] M. C. González–P. G. Lind–H. J. Herrmann: System of mobile agents to model social networks. *Phys. Rev. Lett.*, 96. évf. (2006), 088702. p.
- [85] M. Granovetter: *Decision Making: Alternatives to Rational Choice Models Economic Action and Social Structure: The Problem of Embeddedness*. Newbury Park, 1992, CA:SAGE.
- [86] G. R. Grimmett: *Percolation*. 2. kiad. Berlin Heidelberg, 1999, Springer-Verlag.
- [87] J. W. Grossman: The evolution of the mathematical research collaboration graph. *Congressus Numerantium*, 158. évf. (2002), 202–212. p.
- [88] J. W. Grossman–P. D. F. Ion: On a Portion of the Well-Known Collaboration Graph. *Congressus Numerantium*, 108. évf. (1995), 129–131. p.
- [89] 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.
- [90] 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.
- [91] R. Guimerà–L. A. N. Amaral: Functional cartography of complex metabolic networks. *Nature*, 433. évf. (2005), 895–900. p.
- [92] 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.

-
- [93] L. H. Hartwell–J. J. Hopfield–S. Leibler–A. W. Murray: From molecular to modular cell Biology. *Nature*, 402. évf. (1999), 6761, supplement C47–C52. p.
- [94] T. Hastie–R. Tibshirani–J. H. Friedman: *The Elements of Statistical Learning*. Berlin, Germany, 2001, Springer.
- [95] W. K. Hastings: Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57. évf. (1970), 97–109. p.
- [96] 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.
- [97] P. W. Holland–S. Leinhardt: Transitivity in structural models of small groups. *Comp. Group Stud.*, 2. évf. (1971), 107–124. p.
- [98] R. A. Holley–T. M. Liggett: Ergodic theorems for weakly interacting infinite systems and voter model. *Ann. Probab.*, 3. évf. (1975), 643–663. p.
- [99] P. Holme–Ch. R. Edling–F. Liljeros: Structure and time-evolution of an internet dating community. *Soc. Networks*, 26. évf. (2004), 155–174. p.
- [100] D. C. Hong–S. Havlin–H. J. Herrmann–H. E. Stanley: Breakdown of alexander-orbach conjecture for percolation: Exact enumeration of random walks on percolation backbones. *Phys. Rev. B*, 30. évf. (1984), 4083. p.
- [101] 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.
- [102] K. Hukushima: Extended ensemble monte carlo approach to hardly relaxing problems. *Comp. Phys. Comm.*, 147. évf. (2002), 77–82. p.
- [103] I. Ispolatov–I. Mazo–A. Yuryev: Finding mesoscopic communities in sparse networks. *J. Stat. Mech.*, 2006., P09014. p.
- [104] H. Jeong–Z. Neda–A.-L. Barabási: Measuring preferential attachment for evolving networks. *Europhys. Lett.*, 61. évf. (2003), 567–572. p.
- [105] 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.
- [106] P. F. Jonsson–P. A. Bates: Global topological features of cancer proteins in the human interactome. *Bioinformatics*, 22. évf. (2006), 2291–2297. p.
- [107] P. F. Jonsson–T. Cavanna–D. Zicha–P. A. Bates: Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis. *BMC Bioinformatics*, 7. évf. (2006), 2. p.

- [108] F. Karinthy: Láncszemek. In *Minden másképpen van*. Budapest, 1929, Atheneum Irodai és Nyomdai R.-T., 85–90. p.
- [109] B. W. Kernighan–S. Lin: An efficient heuristic procedure for partitioning graphs. *Bell Syst. Tech. J.*, 49. évf. (1970), 291–307. p.
- [110] K. Klemm–V. M. Eguíluz–R. Toral–M. S. Miguel: Nonequilibrium transitions in complex networks: A model of social interaction. *Phys. Rev. E*, 67. évf. (2003), 026120. p.
- [111] S. Knudsen: *A Guide to Analysis of DNA Microarray Data*. 2. kiad. 2004, Wiley-Liss.
- [112] D. König: *Theorie der endlichen und unendlichen Graphen*. Providence, Rhode Island, 1936, AMS Chelsea Publishing.
- [113] G. Kossinets–D. J. Watts: Empirical analysis of an evolving social network. *Science*, 311. évf. (2006), 88–90. p.
- [114] I. A. Kovács–R. Palotai–M. Szalay-Bekő–P. Csermely: Community landscapes: a novel, integrative approach for the determination of overlapping network modules. *PLoS ONE*, 7. évf. (2010), e12528. p.
- [115] 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.
- [116] J. M. Kumpula–M. Kivelä–K. Kaski–J. Saramäki: Sequential algorithm for fast clique percolation. *Phys. Rev. E*, 78. évf. (2008), 026109. p.
- [117] J. M. Kumpula–J.-P. Onnela–J. Saramäki–K. Kaski–J. Kertész: Emergence of communities in weighted networks. *Phys. Rev. Lett.*, 99. évf. (2007), 228701. p.
- [118] A. Lancichinetti–F. Radicchi–J. J. Ramasco–S. Fortunato: Finding statistically significant communities in networks. *PLoS ONE*, 6. évf. (2011), e18961. p.
- [119] A. Lázár–D. Ábel–T. Vicsek: Modularity measure of networks with overlapping communities. *Europhys. Lett.*, 90. évf. (2010), 18001. p.
- [120] 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.
- [121] 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.
- [122] Ch. Li–Ph. K. Maini: An evolving network model with community structure. *J. Phys. A*, 38. évf. (2005), 9741–9749. p.

-
- [123] 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.
- [124] L. Lovász: *Large Networks and Graph Limits*. American Mathematical Society Colloquium Publications sorozat, 60. köt. Providence, RI, 2012, American Mathematical Society.
- [125] L. Lovász – B. Szegedy: Limits of dense graph sequences. *J. Comb. Theory B*, 96. évf. (2006), 933–957. p.
- [126] 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.
- [127] 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.
- [128] B. B. Mandelbrot: *The Fractal Geometry of Nature*. San Francisco, 1982, Freeman.
- [129] O. Mason – M. Verwoerd: Graph theory and networks in Biology. *IET Syst. Biol.*, 1. évf. (2007), 89–119. p.
- [130] J. F. F. Mendes – S. N. Dorogovtsev: *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford, 2003, Oxford University Press.
- [131] R. K. Merton: The matthew effect in science. *Science*, 159. évf. (1968), 56–63. p.
- [132] N. Metropolis – A. W. Rosenbluth – M. N. Rosenbluth – A. H. Teller – E. Teller: Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21. évf. (1953), 1087–1092. p.
- [133] M. Mezard – G. Parisi – M. Virasoro: *Spin glass theory and beyond*. Singapore, 1987, World Scientific.
- [134] S. Milgram: The small world problem. *Psychol. Today*, 2. évf. (1967), 60–67. p.
- [135] M. Molloy – B. Reed: A critical point for random graphs with a given degree sequence. *Random Struct. Algor.*, 6. évf. (1995), 161–179. p.
- [136] 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.
- [137] E. Mones – L. Vicsek – T. Vicsek: Hierarchy measure for complex networks. *PLoS ONE*, 7. évf. (2012), e33799. p.
- [138] J. M. Montoya – R. V. Solé: Small world patterns in food webs. *J. Theor. Biol.*, 214. évf. (2002), 405–412. p.

- [139] G. Mukherjee–S. S. Manna: Quasistatic scale-free networks. *Phys. Rev. E*, 67. évf. (2003), 012101. p.
- [140] D. L. Nelson–C. L. McEvoy–T. A. Schreiber: The university of south florida word association, rhyme, and word fragment norms. 1998.
URL <http://www.usf.edu/FreeAssociation/>.
- [141] T. Nepusz–F. Bazsó: Likelihood-based clustering of directed graphs. In *IEEE Proceedings of the 3rd International Symposium on Computational Intelligence and Intelligent Informatics* (konferenciaanyag). 2007, 189–194. p.
- [142] T. Nepusz–L. Négyessy–G. Tusnády–F. Bazsó: Reconstructing cortical networks: case of directed graphs with high level of reciprocity. In B. Bollobás–R. Kozma–D. Miklós (szerk.): *Handbook of Large-scale Random Networks*. Berlin, 2008, Springer, 325–368. p.
- [143] M. E. J. Newman: Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64. évf. (2001), 025102. p.
- [144] M. E. J. Newman: From the cover: The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98. évf. (2001), 404–409. p.
- [145] M. E. J. Newman: Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Phys. Rev. E*, 2001., 016132. p.
- [146] M. E. J. Newman: The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98. évf. (2001), 404–409. p.
- [147] M. E. J. Newman: Detecting community structure in networks. *Eur. Phys. J. B*, 38. évf. (2004), 321–330. p.
- [148] M. E. J. Newman: Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69. évf. (2004), 066133. p.
- [149] M. E. J. Newman: Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74. évf. (2006), 036104. p.
- [150] M. E. J. Newman: From the cover: Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, 103. évf. (2006), 8577–8582. p.
- [151] M. E. J. Newman: *Networks: An Introduction*. New York, 2010, Oxford Univ. Press.
- [152] M. E. J. Newman–A.-L. Barabási–D. J. Watts (szerk.): *The Structure and Dynamics of Networks*. Princeton and Oxford, 2006, Princeton University Press.
- [153] M. E. J. Newman–M. Girvan: Finding and evaluating community structure in networks. *Phys. Rev. E*, 69. évf. (2004), 026113. p.

-
- [154] M. E. J. Newman–J. Park: Why social networks are different from other types of networks. *Phys. Rev. E*, 68. évf. (2003), 036122. p.
 - [155] M. E. J. Newman–S. H. Strogatz–D. J. Watts: Random graphs with arbitrary degree distribution and their applications. *Phys. Rev. E*, 64. évf. (2001), 026118. p.
 - [156] V. Nicosia–G. Mangioni–V. Carchiolo–M. Malgeri: Extending the definition of modularity to directed graphs with overlapping communities. *J. Stat. Mech.*, 2009., P03024. p.
 - [157] 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.
 - [158] 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.
 - [159] 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.
 - [160] J. Park–M. E. J. Newman: Solution of the two-star model of a network. *Phys. Rev. E*, 70. évf. (2004), 066146. p.
 - [161] Y. Park–C. Moore–J. S. Bader: Dynamic networks from hierarchical bayesian graph clustering. *PLoS ONE*, 5. évf. (2010), e8118. p.
 - [162] R. Parshani–S. V. Buldyrev–S. Havlin: Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition. *Phys. Rev. Lett.*, 105. évf. (2010), 048701. p.
 - [163] R. Pastor-Satorras–A. Vázquez–A. Vespignani: Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87. évf. (2001), 258701. p.
 - [164] T. P. Peixoto: Parsimonious module inference in large networks. *Phys. Rev. Lett.*, 110. évf. (2013), 148701. p.
 - [165] T. P. Peixoto: Efficient monte carlo and greedy heuristic for the inference of stochastic block models. *Phys. Rev. E*, 89. évf. (2014), 012804. p.
 - [166] T. P. Peixoto: Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X*, 4. évf. (2014), 011047. p.
 - [167] D. J. De Sola Price: Networks of scientific papers. *Science*, 14. évf. (1965), 510–515. p.
 - [168] D. J. De Solla Price: A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inform. Sci.*, 27. évf. (1976), 292–306. p.

- [169] D. Pumain (szerk.): *Hierarchy in Natural and Social Sciences*. Methodos Series sorozat, 3. köt. Dodrecht, The Netherlands, 2006, Springer Netherlands.
- [170] F. Radicchi–C. Castellano–F. Cecconi–V. Loreto–D. Parisi: Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. USA*, 101. évf. (2004), 2658–2663. p.
- [171] F. Radicchi–S. Fortunato: Explosive percolation in scale-free networks. *Phys. Rev. Lett.*, 103. évf. (2009), 168701. p.
- [172] F. Radicchi–S. Fortunato: Explosive percolation: A numerical analysis. *Phys. Rev. E*, 81. évf. (2010), 036110. p.
- [173] A. Ramezanpour–V. Karimipour–A. Mashaghi: Generating correlated networks from uncorrelated ones. *Phys. Rev. E*, 67. évf. (2003), 046107. p.
- [174] B. Ráth–B. Tóth: Triangle percolation in mean field random graphs—with pde. *J. Stat. Phys.*, 131. évf. (2008), 385–391. p.
- [175] 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.
- [176] S. Redner: How popular is your paper? an empirical study of the citation distribution. *Eur. Phys. J. B*, 4. évf. (1998), 131–134. p.
- [177] J. Reichardt–S. Bornholdt: Detecting fuzzy community structures in complex networks with a potts model. *Phys. Rev. Lett.*, 93. évf. (2004), 218701. p.
- [178] J. Reichardt–S. Bornholdt: Statistical mechanics of community detection. *Phys. Rev. E*, 74. évf. (2006), 016110. p.
- [179] T. Richardson–P. J. Mucha–M. A. Porter: Spectral tripartitioning of networks. *Phys. Rev. E*, 80. évf. (2009), 036111. p.
- [180] O. Riordan–L. Warnke: Explosive percolation is continuous. *Science*, 333. évf. (2011), 322–324. p.
- [181] O. Riordan–L. Warnke: L. Achlioptas process phase transitions are continuous. *Ann. Appl. Probab.*, 22. évf. (2012), 1450–1464. p.
- [182] A. W. Rives–T. Galitski: Modular organization of cellular networks. *Proc. Natl. Acad. Sci. USA*, 100. évf. (2003), 1128–1133. p.
- [183] 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.

-
- [184] V. Rosato–L. Issacharoff–F. Tiriticco–S. Meloni–S. De Porcellinis–R. Setola: Modelling interdependent infrastructures using interacting dynamical models. *Int. J. Crit. Infrastruct.*, 4. évf. (2008), 63–79. p.
 - [185] M. Rosvall–C. T. Bergstrom: Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA*, 105. évf. (2008), 1118–1123. p.
 - [186] A. D. Sánchez–J. M. López–M. A. Rodríguez: Nonequilibrium phase transitions in directed small-world networks. *Phys. Rev. Lett.*, 88. évf. (2002), 048701. p.
 - [187] J. Scott: *Social Network Analysis: A Handbook*. 2. kiad. London, 2000, Sage Publications.
 - [188] M. Á. Serrano–M. Boguñá: Tuning clustering in random networks with arbitrary degree distributions. *Phys. Rev. E*, 72. évf. (2005), 036133. p.
 - [189] H.-W. Shen–X.-Q. Cheng–J.-F. Guo: Quantifying and identifying the overlapping community structure in net works. *J. Stat. Mech.*, (2009). évf. (2009), P07042. p.
 - [190] R. M. Shiffrin–K. Börner: Mapping knowledge domains. *Proc. Natl. Acad. Sci. USA*, 101. évf. (2004), 5183–5185. p.
 - [191] H. A. Simon: On a class of skew distribution functions. *Biometrika*, 42. évf. (1955), 425–440. p.
 - [192] R. D. Smith: The network of collaboration among rappers and its community structure. *J. Stat. Mech.*, 2006., P02006. p.
 - [193] S.-W. Son–H. Jeong–J. D. Noh: Random field ising model and community structure in complex networks. *Eur. Phys. J. B*, 50. évf. (2006), 431–437. p.
 - [194] C. M. Song–S. Havlin–H. A. Makse: Self-similarity of complex networks. *Nature*, 433. évf. (2005), 392–395. p.
 - [195] V. Spirin–K. A. Mirny: Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. USA*, 100. évf. (2003), 12123–12128. p.
 - [196] 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.
 - [197] 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.
 - [198] J. J. Sylvester: Chemistry and algebra. *Nature*, 17. évf. (1878), 284–284. p.
 - [199] G. Szabó–G. Fáth: Evolutionary games on graphs. *Phys. Rep.*, 446. évf. (2007), 97–216. p.

- [200] 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.
- [201] G. Szabó–M. Alava–J. Kertész: Structural transitions in scale-free networks. *Phys. Rev. E*, 67. évf. (2003), 056102. p.
- [202] M. Szalay-Bekő–R. Palotai–B. Szappanos–I. A. Kovács–B. Papp–P. Csermely: Moduland plug-in for cytoscape: extensively overlapping network modules, community centrality and their use in biological networks. *Bioinformatics*, 28. évf. (2012), 2202–2204. p.
- [203] E. Szemerédi: Regular partitions of graphs. In *Problèmes combinatoires et théories des graphes*. Orsay, 1976, Colloq. Internat. CNRS, Univ. Orsay, 399–401. p.
- [204] R. Toivonen–J.-P. Onnela–J. Saramäki–J. Hyvönen–K. Kaski: A model for social networks. *Physica A*, 370. évf. (2006), 851–860. p.
- [205] A. Trusina–S. Maslov–P. Minnhagen–K. Sneppen: Hierarchy measures in complex networks. *Phys. Rev. Lett.*, 92. évf. (2004), 178702. p.
- [206] S. Valverde–R. Ferrer Cancho–R. V. Solé: Scale-free networks from optimal design. *Europhys. Lett.*, 60. évf. (2002), 512. p.
- [207] S. Valverde–R. V. Solé: Self-organization versus hierarchy in open-source social networks. *Phys. Rev. E*, 76. évf. (2007), 046118. p.
- [208] T. Vicsek: *Fractal Growth Phenomena*. 2. kiad. Singapore, New Jersey, London, Hong Kong, 1992, World Scientific.
- [209] 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.
- [210] A. Wagner–D. A. Fell: The small world inside large metabolic networks. 2000. Tech. Rep. 00-07-041, Santa Fe Institute.
- [211] 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.
- [212] 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.
- [213] S. Warner: E-prints and the Open Archives Initiative. *Library Hi Tech*, 21. évf. (2003), 151–158. p.
- [214] S. Wasserman–K. Faust: *Social network analysis: methods and applications Structural Analysis in the Social Sciences*. Cambridge, 1994, Cambridge University Press.

-
- [215] 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.
 - [216] D. J. Watts: A twenty-first century science. *Nature*, 445. évf. (2007), 489. p.
 - [217] D. J. Watts–P. S. Dodds–M. E. J. Newman: Identity and search in social networks. *Science*, 296. évf. (2002), 1302–1305. p.
 - [218] D. J. Watts–S. H. Strogatz: Collective dynamics of 'small-world' networks. *Nature*, 393. évf. (1998), 440–442. p.
 - [219] 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.
 - [220] D. M. Wilkinson–B. A. Huberman: A method for finding communities of related genes. *Proc. Natl. Acad. Sci. USA*, 101. évf. (2004), 5241–5248. p.
 - [221] F. Y. Wu: The potts model. *Rev. Mod. Phys.*, 54. évf. (1982), 235–268. p.
 - [222] I. Xenarios–D. W. Rice–L. Salwinski–M. K. Baron: DIP: the Database of Interacting Proteins. *Nucl. Ac. Res.*, 28. évf. (2000), 289–291. p.
 - [223] J. Xie–S. Kelley–B. K. Szymanski: Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Comput. Surv.*, 45. évf. (2013), 43:1–43:35. p.
 - [224] 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.
 - [225] G. U. Yule: A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 213. évf. (1925), 21–87. p.
 - [226] S. Zhang–X. Ning–X.-S. Zhang: Identification of functional modules in a PPI network by clique percolation clustering. *Comput. Biol. Chem.*, 30. évf. (2006), 445–451. p.
 - [227] X. Zhu–M. Gerstein–M. Snyder: Getting connected: analysis and principles of biological networks. *Gene. Dev.*, 21. évf. (2007), 1010–1024. p.
 - [228] R. M. Ziff: Explosive growth in biased dynamic percolation on two-dimensional regular lattice networks. *Phys. Rev. Lett.*, 103. évf. (2009), 045701. p.
 - [229] G. K. Zipf: *Human behavior and the principle of least resort*. Oxford, England, 1949, Addison-Wesley Press.