

EÖTVÖS LORÁND TUDOMÁNYEGYETEM  
TERMÉSZETTUDOMÁNYI KAR

---

Rokob Sándor

## Véletlen gráfok duplikációs modelljei

Matematika Bsc szakdolgozat

Témavezető:

Móri Tamás

Valószínűségelméleti és Statisztika Tanszék



Budapest, 2016



# Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani témavezetőmnek, Móri Tamásnak a szakdolgozat megírása során nyújtott számos tanácsért és észrevételért, melyekkel a munkámat segítette.



# Tartalomjegyzék

<b>1. Bevezetés</b>	<b>7</b>
<b>2. Duplikáció</b>	<b>9</b>
2.1. Szükséges alapismeretek . . . . .	10
2.2. Duplikációs modellek . . . . .	13
<b>3. Parciális duplikáció</b>	<b>15</b>
3.1. A modell és változatai . . . . .	15
3.2. Megcáfolt eredmények . . . . .	18
3.3. Eredmények . . . . .	21
<b>4. Thörnblad modell</b>	<b>28</b>
4.1. A modell és változatai . . . . .	28
4.2. Aszimptotikus fokszámeloszlás . . . . .	30
4.3. Maximális fokszám . . . . .	38
4.4. Domináns klikk szuperkritikus esetben . . . . .	49
<b>5. Összefoglalás</b>	<b>61</b>
<b>Irodalomjegyzék</b>	<b>62</b>



# 1. Bevezetés

A technológia fejlettségének köszönhetően az elmúlt években rengeteg figyelem irányult a nagyméretű, való életben megtalálható hálózatokra és ezek véletlen gráfokkal való modellezésére, habár ez a fajta megfeleltetés egyáltalán nem egy ismeretlen gondolat. Ezek között is legnépszerűbbek az úgynevezett skálafüggetlen változatok (a hatványrendben lecsengő fokszámeloszlással rendelkezők). A kutatók számára hamar kiderült, hogy az olyan klasszikusnak mondható véletlen gráf modellek, mint például az Erdős és Rényi nevéhez fűződő [17], itt nem használhatóak.

Talán a legtöbbet emlegetett ilyen hálózat a Barabási és Albert [5] által javasolt Barabási-Albert modell. Ennek legnagyobb előnye az, hogy viszonylag egyszerű fejlődési dinamika segítségével éri el a hatványrendben lecsengő fokszámeloszlást. Ez nem más, mint a fokszámárányos (preferential attachment) választás, ami annyit jelent, hogy annak valószínűsége, hogy a hálózat egy új eleme csatlakozik-e egy korábbihoz, csak ennek fokszámától függ.

Ennek a modellnek a fő célja világháló a jobb megismerése volt, ugyanis az interneten található honlapok és a közöttük futó linkek is hasonlóan viselkednek. Ahhoz, hogy egy új honlap minél több látogatót vonzzon, a legfelkapottabb honlapokkal próbál kapcsolatot létesíteni, ahol a népszerűség az adott oldal linkjeinek számával egyezik meg. Ezáltal pontosan az a "gazdag gazdagabb lesz" elv figyelhető meg, amelyet a Barabási-Albert modell is reprezentál.

Természetesen nem csak ez az egy hálózat létezik a világban, ahogy nem csak ezt az egy fajta növekedési mechanizmust lehet használni modellezés esetén. Sok más területtel (marketing, közlekedés, emberi kapcsolatok) megegyezően az orvostudomány és a biológia is rengeteg érdekes hálózatot tartalmaz. Ezek külön is izgalmasak, mert az egyre precízebb mérőműszerek által szolgáltatott adatok mellett a biológiai folyamatok alapvető formáló erőinek ismerete is rendelkezésünkre áll.

Az egyik legfontosabb ilyen formáló erő, a duplikáción, azaz másoláson keresztüli fejlődés. Ez jelen esetben abban nyilvánul meg, hogy a hálózat új elemei valamilyen formában a korábbiak másolatai lesznek. A dolgozat során végig az effajta dinamikával ellátott biológiai hálózatok véletlen gráfokkal való modellezésével foglalkozunk.

A dolgozat felépítése a következő. A második fejezetben részletesebben elemezzük a problémát és azt, hogy ez hogyan is modellezhető véletlen gráfok segítségével. Majd miután felvázoltuk a szükséges matematikai ismereteket, megemlítünk pár a szakirodalomban megtalálható modellt és a rájuk vonatkozó eredményeket. Ezután a harmadik fejezetben ismertetjük a duplikáció egyik legáltalánosabb matematikai formáját és az erre vonatkozó megcáfolt, illetve bizonyított állításokat. Itt a cáfolatok hangsúlyozása azért is lényeges, mert ahogyan a Barabási-Albert modellnél, ezen biológiai hálózatok esetén is sokszor előfordul, hogy a heurisztikus gondolkodás

átveszi a szigorú matematikai érvelés szerepét. Végül a negyedik fejezetben részletezzük az egyik legtöbb, szigorú matematikával bizonyított állítást magának tudható modellt.



## 2. Duplikáció

A kölcsönhatások szerepe minden biológiai rendszerben alapvető, legyen szó akár a különböző fajok közti kapcsolatról, vagy egy sejtet felépítő molekulák interakciójáról. Az aktuális kutatások jelentős része inkább ezen utóbbival foglalkozik, köszönhetően az adatmennyiségnek, mely a modern technológiák segítségével hozzáférhető. Természetesen a sejten belüli rendszereknek is több fajtája ismert (például anyagcsere-hálózatok, génhálózatok és fehérje-fehérje interakció hálózatok), habár ezek kialakításánál mind a genomban rejlő információ megduplázásának mechanizmusa húzódik meg a háttérben.

Egy adott rendszer elemeinek és a közöttük fennálló kölcsönhatásoknak, kapcsolatoknak matematikai modellezésére gráfok használata legkézenfekvőbb. Ekkor az egyes elemeket a gráf csúcsai, míg a közöttük fellépő interakciókat a gráfban található élek reprezentálják. Ebben az esetben az információ kettőzése nem jelent mást, mint korábbi csúcsokkal megegyező szomszédsággal ellátott csúcsoknak a rendszerhez való hozzávétele, vagyis egyszerűbben megfogalmazva, egy, már a hálózatban található csúcs másolása. Tehát ilyenkor az az információ, amit duplikálunk, pont az adott elem kölcsönhatásaira vonatkozik.

Ezek mellett az is igaz, hogy akárcsak a legtöbb természetben előforduló folyamatban, egy-egy duplikáció során is felléphetnek hibák az új elem keletkezésekor. Ez a fajta eltérés, avagy mutáció úgy írható le az előbbi modellben, hogy a másolat szomszédsága nem esik egybe a másolt csúcs szomszédjainak halmazával. Például előfordulhat, hogy a másolat a régi csúcs szomszédjai mellett továbbiakkal is interakcióba lép, ezzel bővítve az előbb említett halmazt. Habár a valóságban ezek a mutációk nem feltétlenül teljesen véletlenszerűek, gondolva itt a biológiai folyamatokban gyakran feltűnő adaptációra, az előbbi modellhez ezt mégis kevésbé determinisztikus formájában szokás hozzáadni.

Összesítve, a sejten belüli molekulák kölcsönhatásainak matematikai modellje nem más, mint egy dinamikusan fejlődő véletlen gráf, melynek növekedési mechanizmusa egy másolási, illetve egy mutációs fázisra bontható. Az persze egyáltalán nem magától értetődő, hogy ezt a két lépést pontosan hogyan definiáljuk, mi lesz az a kikötés, mely a valóságnak valamilyen értelemben a legjobban megfelel, így a matematikailag előállítható modellek tárháza végtelen. Természetesen ezek közül főleg a valóságot minél jobban visszaadó variációkat tanulmányozták eddig.

A fejezet célja ezeknek a különböző változatoknak az ismertetése, a teljességre való törekvés nélkül. Ezt azért is érdemes kihangsúlyozni, ugyanis ez a kutatási terület túl friss és nagy népszerűségnek örvendő ahhoz, hogy átfogó képet adhasunk az eddig vizsgált modellekről. Emellett egy külön szakaszban ismertetjük a későbbi vizsgálatainkhoz felhasznált matematikai háttérrel, illetve a leghatékonyabb hálózatokat leíró fogalmakat.

## 2.1. Szükséges alapismeretek

Mielőtt a duplikáció dinamikájával fejlődő véletlen gráfok részletes kifejtésébe kezdenénk, érdemes megtenni az alapvető megköötéseket, megemlíteni a későbbiekben használatos nevezetes valószínűségi számítások állításait és kitérni arra, melyek azok a paraméterek, amiket a leggyakrabban vizsgálunk véletlen hálózatok elemzése esetén. A következőben ezeket fogjuk taglalni, azzal a megjegyzéssel, hogy az egyes tételek részletes kifejtését, illetve bizonyítását mellőzzük, ugyanis jelen esetben számunkra csak ezek ismerete szükséges. A legtöbbjük a megfelelő szakkönyvekben megtalálható, melyek közül kifejtésünk során néhányat meg is említünk.

A további vizsgálataink közben megfigyelt gráfokra a következő, gráfelméleti kikötéseket tesszük. Habár sok hálózat esetén (például az internet általánosabb modellje, melyben a linkekhez irányokat is adunk) fontos az egyes csúcsok között fennálló kapcsolatok iránya is, a későbbiekben végig irányítatlan gráfokra szorítunk. Ezen felül, bár a párhuzamos és hurokélek jelenléte további információt adhat az adott gráf struktúrájáról, ezek a soron következő modellek mindegyikénél, vagy a dinamikából adódóan, vagy pedig külön szabályok által mellőzve maradnak. Tehát összesítve, az elkövetkezőkben végig csak irányítatlan, egyszerű gráfokat fogunk tekinteni.

Ahogy a legtöbb véletlen hálózat esetén, a legfontosabb formáló erő itt is az lesz, hogy a gráf aktuális állapota egyedül a közvetlenül előtte lévő állapottól függ. Ezen sajátosságot a valószínűségi számításban Markov-tulajdonságként tartjuk számon. Tekintettel ennek fontosságára, ezt precízen is definiáljuk.

**2.1. Definíció.** *Legyen adott az  $(X_n)_{n \in \mathbb{N}}$  folyamat, ahol  $X_n : \Omega \rightarrow I$  valószínűségi változók az  $(\Omega, \mathcal{F}, P)$  valószínűségi mezőn,  $I$  pedig megszámlálható halmaz. Ekkor a folyamat Markov-tulajdonságú, ha  $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$  jelöléssel minden  $B \subseteq I$  és minden  $m \geq n$  esetén*

$$P(X_m \in B | \mathcal{F}_n) = P(X_m \in B | X_n).$$

Minden olyan sztochasztikus folyamatot, melyben ez a fajta emlékeztethiány jelen van, Markov-folyamatnak, vagy Markov-láncnak nevezünk. Tehát az előbbieket alapján levonható a következtetés, melyet a későbbiek során leginkább közvetetten fogunk felhasználni, miszerint a dinamikus fejlődő véletlen gráfok Markov-láncok segítségével modellezhetőek. Mivel ezen diszkrét folyamat számunkra fontos alaptulajdonságainak ismertetése könnyen megtöltene egy ehhez hasonló dolgozatot, így ezt nem tesszük meg. Ezek elsajátításához és elmélyítéséhez a [8] könyvet, vagy a [24] könyv 2. és 3. fejezetét ajánljuk.

Természetesen a szakirodalomban előforduló legtöbb modellt diszkrét időben definiálják ennek jól kezelhetősége miatt, viszont néhány változat esetén, ahogyan azt majd egy későbbi fejezet során látni fogjuk, a folytonos idejű megfogalmazás is előkerül. Ezeknél szintén jelen van a Markov tulajdonság, de természetesen ennek a folytonos időben definiált változata. Ennek következtében érdemes az úgynevezett folytonos idejű Markov-láncokra vonatkozó, számunkra fontos állításokat összefoglaló irodalmat is megemlíteni. Erre a célra megfelelő lesz a [25] könyv 14. fejezete.

Emellett az is igaz, hogy nem csak általános tételekre lesz szükségünk a folytonos idejű Markov-láncokkal kapcsolatban, hanem azok speciálisabb eseteinek ismeretére

is. Ezek közé tartoznak a születési-halálozási és Yule folyamatok, melyekről a [31] 6., vagy a [24] 4. fejezetében található részletesebb összefoglalást.

Vizsgálataink során egy másik sztochasztikus folyamat jó tulajdonságait is ki fogjuk használni. Ez nem más, mint az úgynevezett martingál, melyet a következőképp definiálhatunk diszkrét paraméter mellett.

**2.2. Definíció.** *Legyenek  $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{A}$  rész- $\sigma$ -algebrák és  $X_1, X_2, \dots$  valószínűségi változók. Ekkor az  $(X_n, \mathcal{F}_n)$  sorozatot martingálnak nevezzük, ha teljesíti a következő feltételeket.*

1. Minden  $n \in \mathbb{N}$  esetén  $E|X_n| < \infty$ .
2. Minden  $n \in \mathbb{N}$  esetén  $X_n$   $\mathcal{F}_n$ -mérhető.
3. Minden  $m \geq n \geq 1$  esetén  $E[X_m | \mathcal{F}_n] = X_n$ .

A későbbiekben gyakran előfordul, hogy egy martingál megadásánál a  $\sigma$ -algebrát nem definiáljuk. Ebben az esetben a megfelelő  $\sigma$ -algebrának a  $\sigma(X_1, X_2, \dots, X_n)$  természetes  $\sigma$ -algebrát tekintjük. A számunkra legfontosabb martingálelméleti tétel a következő.

**2.1. Tétel. (Martingál-konvergenciatétel)** *Legyen az  $(X_n, \mathcal{F}_n)$  martingál. Ekkor a következők teljesülnek:*

1. Ha  $\sup_n E|X_n| < \infty$ , akkor  $X_n$  majdnem biztosan konvergál.
2.  $\{X_n : n \in \mathbb{N}\}$  pontosan akkor egyenletesen integrálható, ha  $X_n$  (majdnem biztosan és)  $\mathcal{L}^1$ -ben (is) konvergál.
3. Legyen  $r > 1$ . Ekkor  $\sup_n E|X_n|^r < \infty$  (azaz  $X_n$   $\mathcal{L}^r$ -ben korlátos) pontosan akkor ha  $X_n$   $\mathcal{L}^r$ -ben konvergens.

A továbbiak során sok alapvető valószínűségszámítási eredmény (például Borel-Cantelli lemmák, karakterisztikus függvények folytonossági tétele, Dirichlet eloszlás) és fogalom is elő fog kerülni, mely a legtöbb felsőbb valószínűségszámítási bevezető könyvben megtalálható. Ilyen könyvek például a [18], vagy magyar nyelven a [15]. Természetesen ezek mellett olyan általunk gyakran használt kifejezések is szerepelni fognak, melyek habár nem mondhatók újnak, mégsem feltétlenül találhatók meg az előbbi forrásokban. Ezek alatt az alábbiakat értjük.

Már az előző fejezetben is szerepelt a későbbiekben gyakran előforduló hatványrendben lecsengő eloszlás elnevezés. Ezt a véletlen hálózatok témakörében rendszeresen felbukkanó fogalmat a következőképp definiálhatjuk.

**2.3. Definíció.** *Hatványrendben lecsengő eloszláson olyan diszkrét  $(p_0, p_1, \dots)$  eloszlást értünk, melyre*

$$p_k \sim Ck^{-\alpha} \quad (k \rightarrow \infty),$$

ahol  $\alpha > 1$ ,  $C > 0$  konstans.

Érdeemes ennek pár hasznos tulajdonságát is megjegyezni. Például, hogy ellentétben a haranggörbét követő eloszlások azon tulajdonságával, hogy a várható érték egy környezetére koncentrálnak, a hatványrendben lecsengő eloszlás nagyobb kilengést is megenged. Továbbá, hogy skálafüggetlen, abban az értelemben, hogy konstans szorzó nem változtatja meg az eloszlás ezen jellemzőjét. Emellett fontos lesz még számunkra ennek egy másik változata, melyet exponenciális levágással ellátott hatványrendben lecsengő eloszlásnak nevezünk.

**2.4. Definíció.** *Exponenciális levágással ellátott hatványrendben lecsengő eloszláson olyan diszkrét  $(p_0, p_1, \dots)$  eloszlást értünk, melyre*

$$p_k \sim Ck^{-\alpha}a^{-k} \quad (k \rightarrow \infty),$$

ahol  $\alpha, a > 1, C > 0$  konstans.

Mindezen alapok mellett, melyeket mostantól ismertnek feltételezünk, az is számunkra válhat a későbbiekben, ha a leggyakrabban vizsgált hálózatokat leíró paramétereket is definiáljuk. Természetesen, ahogyan azt az eddigiekben tettük, itt sem törekszünk teljességre, csak a számunkra lényeges mérőszámokat említjük meg.

Már a bevezetőben is szerepelt az úgynevezett fokszámeloszlás, mely kifejezés alatt nem meglepően, a véletlen gráf fokszámainak tapasztalati eloszlását értjük. Tekintettel arra, hogy egy ilyen folyamat olykor sokkal összetettebb, mint azt elsőre gondolnánk, így előfordulhat, hogy a konkrét fokszámeloszlásról nem tudunk semmit sem mondani. Ekkor ehelyett a várható fokszámeloszlást szokták vizsgálni. Abban az esetben pedig, ha a hálózat időben vett határértékének fokszámain értelmezett eloszlását nézzük, akkor aszimptotikus fokszámeloszlásról beszélünk.

Ahhoz, hogy a következő mérőszámot be tudjuk vezetni, szükségünk lesz az alábbi definícióra.

**2.5. Definíció.** *Legyen  $\mathbb{G}$  egyszerű gráf és  $v$  ennek egy olyan csúcsa, melynek legalább két szomszédja van, ekkor a  $v$ -beli lokális klaszterezettségi együtthatón a*

$$C_v(\mathbb{G}) = \frac{a \text{ } v \text{ szomszédjai közti élek száma}}{a \text{ } v \text{ szomszédjainak számával megegyező csúcsszámú teljes gráf élszáma}}$$

*mennyiséget értjük.*

Ebből, ha a gráf összes csúcsán végigmenve megsúlyozzuk a lokális klaszterezettségi együtthatót az adott csúcs szomszédjai között behúzható összes élek számával, majd ezt átlagoljuk, akkor a kapott értéket globális klaszterezettségi együtthatónak nevezzük. Erről könnyen belátható az alábbi állítás, melynek bizonyítását az olvasóra hagyjuk.

**2.1. Állítás.** *Jelölje a  $\mathbb{G}$  gráf globális klaszterezettségi együtthatóját  $C(\mathbb{G})$ . Ekkor erre az alábbi teljesül:*

$$C(\mathbb{G}) = \frac{a \text{ gráfban található háromszögek száma}}{a \text{ gráfban található szomszédos élek száma}}.$$

Habár a továbbiakban a globális jelzőt elhagyjuk, ezt a globális klaszterezettségi együtthatót fogjuk tekinteni, mely azt adja meg, hogy a gráf az éleit nézve mennyire sűrű. Szóval, ha ez az érték nagy, akkor magas klaszterezettségről beszélhetünk.

Vizsgálataink során elő fog kerülni a robusztusság és a gócpont fogalma is. Az előbbi egyáltalán nem ismeretlen kifejezés a statisztikában, de jelen esetben pontosan úgy fogalmazható meg, hogy a hálózat mennyire érzékeny az apró változtatásokra. Míg az utóbbi alatt egy olyan csúcsot értünk, melyből a legtöbb másik csúcs közvetlenül elérhető. Ezek mellett olyan gyakran vizsgált tulajdonságok is léteznek, melyekre nem adunk külön definíciót. Ilyen például a maximális fokszám, vagy, hogy a kiindulási gráf szerkezete mennyire befolyásolja, határozza meg a későbbi struktúrát.

Hasonlóan a matematikai alap megteremtéséhez, ezen véletlen gráfokkal kapcsolatos, előbbieken felsorolt fogalmakat is ismertnek feltételezzük a továbbiakban. Ha esetleg ez a rövid és szemléletes összefoglaló nem lenne elég a megértéshez, akkor ezek jobb elsajátítása érdekében a [37] jegyzetet ajánljuk.

## 2.2. Duplikációs modellek

Ahhoz, hogy átfogó képet kaphassunk a biológiai hálózatok dinamikusan fejlődő véletlen gráfokkal való modellezéséről, mindenképp hasznos néhány már vizsgált modell és az ezekkel kapcsolatban elért, vagy elvárt eredmények felvázolása. Itt az utóbbi alatt azt értjük, hogy habár egy adott állítást statisztikai, vagy heurisztikus úton alátámasztottak a modellt vizsgáló kutatók, az nincs bizonyítva szigorú matematikai eszközökkel. Ez a fajta kutatási módszertan egyáltalán nem meglepő, tekintettel arra, hogy ezeket a modelleket főként biológusok és fizikusok alkotják meg. Sőt, hasonlóan a Barabási-Albert modellhez, ennek következményeképp a duplikációs modelleknél is előfordul korábbi „eredmények“ megcáfolása, melyet majd egy későbbi szakaszban részletesebben fejtünk ki.

Mielőtt rátérnénk a konkrét modellek felsorolására, azt is érdemes megemlíteni, hogy lévén az összes mögött ugyanaz a korábban említett másolás és mutáció mechanizmus van, így legtöbbjüket egy kalap alá lehetne venni, ahogyan azt a következő fejezetben tesszük. Jelen szakaszban viszont a kategorizálás helyett inkább arra törekszünk, hogy minél több ötletet, modell koncepciót bemutassunk.

Az első lépés egy ilyen duplikációs modellnél azt meghatározni, hogy a duplikálandó csúcsot hogyan válasszuk ki. A véletlenszerűség természetesen adott, de az, hogy ez milyen eloszlásnak felel meg, már nem. A legtöbb modellben a legkézenfekvőbb lehetőséget, az egyenletes eloszlást, azaz valamilyen értelemben a legvéletlenszerűbb választást preferálják, de persze van ettől eltérő próbálkozás is. Cohen, Jordan and Voliotis [13] modelljében például a másolandó csúcs kiválasztása a Barabási-Albert modellből ismert fokszámárányos módon történik. Ezután pedig megtörténik a csúcs lemásolása, mely a cikk java részében arra a speciális esetre koncentrálódik mikor a másolat megegyezik az eredetivel. Az erre belátott fő eredmény, hogy tetszőleges kiindulási gráf esetén, ha már elég hosszú ideje megy a növekedés, akkor található olyan részgráfja a inicializáló gráfnak, hogy a legtöbb születő csúcs az ebben található másolata.

A továbbiakban már csak olyan változatok lesznek, ahol az egyes időpillanatokban mindegyik csúcs ugyanakkora eséllyel duplikálódik. Ennek következtében a modellek definiálásához elegendő lesz csak a konkrét duplikáció leírása. Az egyik

legtöbbet vizsgált variáció az, melyben a másolás nem adja vissza az eredeti csúcsot, egyes kapcsolatok hiányoznak az újonnan született elemből. Ilyen modell szerepel például a [12, 7, 27, 23] cikkekben. Ezt a későbbiekben sokkal részletesebben fogjuk kifejteni, így előljáróban csak annyit érdemes megjegyezni, hogy sokáig azt feltételezték, az internet modellezésére használt véletlen gráfokhoz hasonló, hatványrendben lecsengő fokszámeloszlással rendelkeznek, melyről később kiderült, hogy ilyen általánosságban nem teljesülhet.

Persze rengeteg olyan modell is létezik, amely kibővíti ezen másolás és duplikáció korlátjait. Az egyik közülük a [30] és [26] cikkekben szereplő, mely a korábbiak mellett egy olyan plusz lépést is tartalmaz, hogy az új csúcs kicsi valószínűséggel tetszőleges korábbival is összeköthető. Amellett, hogy ez jobban reprezentálja a mutációról alkotott elvárásainkat, sokkal nagyobb teret enged a hálózat struktúrájának megváltozására. Habár a [30] cikkben statisztikai módszereket felhasználva megbizonyosodtak arról, hogy itt is hatványrendben lecsengő fokszámeloszlás van, a későbbi [7] cikk szerzői heurisztikus bizonyítékot adtak arra, hogy ez csak akkor lehetséges, ha egy alsó korlátot adunk a keletkező csúcsok fokszámára.

Ezek mellett az információ elvesztését, pontosabban törlését megengedő variációkat is felvetettek. Ilyen például a Backhausz, Móri és Thörnblad [3, 4, 35, 36] által vizsgált modell is, melyben vagy teljesen lemásolunk egy korábbi csúcsot, vagy minden élt törölve izolálttá tesszük. Ez a változat jól visszaadja a valóságban megfigyelt magas klaszterezettséget. Továbbá azért is érdekes, mert könnyebb kezelhetősége miatt sokkal több tulajdonságát tudták bebizonyítani precíz matematikai eszközökkel, mint a többi biológiai hálózatnak. Ezeket az eredményeket a későbbiekben bővebben ismertetjük.

Kevésbé drasztikus fokszámcsökkenés szerepel a Vázquez, Flamminia, Maritana és Vespignani [38] szerzőnégyes által felvázolt dinamikában. Itt a mutációval ellátott duplikáció után, ha egy csúcs a másolatnak és a eredetinek is szomszédja, akkor valamilyen valószínűséggel ezen két él közül csak egyet tartunk meg. Külön érdekesség, hogy a valóságban megfigyelt fehérje-fehérje hálózatokkal való, statisztikai módszereket használó összehasonlítás mellett a modell robusztusságát is megnézték. Erről, főként heurisztikus gondolatmenetet használva, azt sikerült belátni, hogy jól tükrözi a való életben jelenlévő tűrőképességet.

Olyan megfontolások is léteznek, hogy az aktuális duplikációt egy a háttérben megbújó sztochasztikus folyamat irányítja. Ilyen például a Hamdi, Krishnamurthy és Yin [20] szerzők által tanulmányozott modell, melyben egy él lemásolásának a valószínűsége egy rejtett Markov lánc állapotától függ. Sajnos ezt teljes általánosságban nem vizsgálták, csak olyan speciális esetben, amikor a gráf mérete nem változik. A cikkben szereplő állítások szerint ekkor a fokszámok várható értékének határeloszlása hatványrendben lecsengő.

Tehát összefoglalva az előbbiekben felsorolt modellek mindegyike az alap másolás és mutáció mechanizmushoz ad hozzá valami pluszt, ami a valóságot jobban reprezentáló dinamika születését eredményezi. Ennek következményeképp a következő fejezetben ezt a hozzáadott műveleteket nem tartalmazó változatot és pár általánosítását vizsgáljuk meg részletesebben.

## 3. Parciális duplikáció

Az eddigiek alapján megfigyelhető tendencia, hogy a biológiai hálózatok során legtöbbször a következő mechanizmus volt a háttérben. Adott pillanatban a hálózat egy csúcsa a mutáció miatt előforduló módosítások mellett lemásolódik. Matematikailag ez egy olyan dinamikusan fejlődő véletlen gráffal fogható meg, hogy egy, az egyenletes eloszlás szerint választott csúcs másolása után a másolat éleit valamilyen valószínűséggel ritkítjuk. Ezen általános metódust szokás részleges, avagy parciális duplikációnak nevezni.

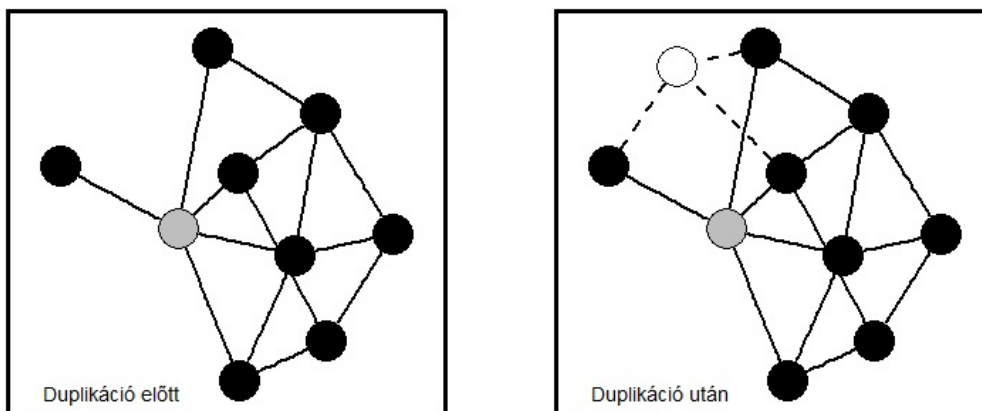
Tekintettel arra a nyilvánvaló tényre, hogy a parciális duplikációs modell is empirikus megfigyelésekből fogant, az erre vonatkozó eredmények java része szintén vagy statisztikai megfontolásokon, vagy számítógépes szimulációkon, azaz kevésbé szigorú matematikán alapul. Ennek következtében a szakirodalomban nemcsak a biológiai hálózatokat leíró tulajdonságok újabb és újabb bizonyítása [21], hanem korábbi „eredmények“ cáfolatai is megtalálhatóak [7].

A fejezet célja a parciális modellre vonatkozó aktuális eredmények ismertetése, aminek következtében az alábbiaknak megfelelően fogunk haladni. Először definiáljuk a parciális duplikációs modellt és felvázoljuk ennek pár további általánosítását, illetve speciális esetét. Ezután egy kisebb kitérőt téve példát mutatunk nem túl precízen belátott korábbi állítások cáfolatára. Mindezek után az utolsó szakaszban a teljességre való törekvés igénye nélkül felvázolunk néhányat a modellre vonatkozó eredmények közül.

### 3.1. A modell és változatai

A szakasz elsődleges céljaként definiáljuk precízen a parciális duplikációs véletlen gráf modelljét, majd ismertetünk néhányat ennek további általánosításai közül.

**Parciális duplikációs modell.** Legyen  $0 \leq p \leq 1$  rögzített érték. A folyamatot az  $n_0$  időben indítjuk egy olyan összefüggő  $\mathbb{G}_{n_0}$  gráfból, mely csúcsainak száma pont  $n_0$ . Ezután diszkrét időben előrehaladva, az  $n > n_0$  pillanatban  $\mathbb{G}_n$  a következő dinamika szerint képződik  $\mathbb{G}_{n-1}$ -ből. Az egyenletes eloszlásnak megfelelően választunk egy  $v$  csúcsot a  $\mathbb{G}_{n-1}$  gráfból. Majd ezen  $\mathbb{G}_{n-1}$  gráfhoz felvesszünk egy  $w$  csúcsot, méghozzá úgy, hogy az előbb választott  $v$  összes szomszédjával egymástól függetlenül  $p$  valószínűséggel összekötjük.



1. ábra. A parciális duplikáció egy lépése. A bal oldali ábrán a duplikáció előtti állapot látható egy adott időpontban. A jobb oldali ábrán fehérrel jelölt csúcs a kiválasztott szürke csúcs egy parciális duplikációs lépés utáni másolata.

Ha feltesszük, hogy  $p < 1$ , akkor előfordulhat, hogy a duplikáció izolált csúcsot hoz létre a gráfban. Tehát tekintettel arra feltételre, hogy a kiindulási gráf összefüggő, látható, hogy ebben az esetben  $\mathbb{G}_n$  végig nem lesz más, mint egy összefüggő komponens és izolált csúcsok összességéből álló gráf. Ahogyan az majd a későbbiekben ki fog derülni, ezen izolált csúcsok arányának ismerete nagy szerepet fog kapni a gráf aszimptotikus viselkedésének leírásában.

Ezen felül érdemes azt is megvizsgálni, hogy modellünk milyen viszonyban áll a népszerű fokszámárányos növekedési dinamikával ellátott társaival. Ehhez az [22] cikk heurisztikus gondolatmenetét írhatjuk fel. Tegyük fel, hogy a  $p$  paraméter annyira kicsi, hogy az új csúcs várhatóan egy éllel fog kapcsolódni a korábbi gráfhoz. Ekkor ha egy  $k$  fokszámú csúcsot duplikálunk, akkor annak a valószínűsége, hogy az új csúcs az összefüggő komponenshez csatlakozik, megegyezik  $1 - (1 - p)^k$  értékkel, ami közelítőleg  $pk$ . Tehát annak a valószínűsége, hogy az új csúcs egy  $k$  fokszámú csúcs valamely szomszédjához kapcsolódik, arányos lesz  $k$ -val. Azaz, ha feltesszük, hogy a modell nem generál izolált csúcsokat, akkor a  $p$  paraméterrel a nullához tartva egyfajta fokszámárányos választás lesz megfigyelhető.

A parciális duplikáció egy fontos speciális eseteként lehet értelmezni az úgynevezett teljes duplikációs modellt. Ez úgy kapható meg, hogy a részleges másolás  $p$  paraméterét 1-nek választjuk. Ekkor ajánlott megvizsgálni, hogy a kiindulás gráf által leírt struktúra hogyan változik ezen teljes duplikációs lépések után. Ehhez az eredeti gráfból definiáljunk egy újat a következőképp. Csoportosítsuk a gráf csúcsait úgy, hogy egy halmazba azon csúcsok kerüljenek, melyeknek pontosan ugyanazok a szomszédjaik. Ezen halmazokat szokás a gráf orbitjainak is nevezni. Az új gráf csúcsai legyenek ezen orbitok, továbbá két orbitot kössünk össze, ha az eredeti gráfban csúcsaik között megy él. Ebben az esetben látható, hogy teljes duplikáció során, mivel a születő másolatok nem változnak a korábbi verzióhoz képest, a folyamat során fejlődő gráfok orbitjainak előbbieken definiált gráfja végig megegyezik a kiindulási gráféval. Tehát, habár a csúcsok száma megnő, strukturailag a gráf nagyon nem változik.



Most térjünk rá, hogy a parciális duplikációs modellt milyen természetes, azaz a valóságot még jobban reprezentáló irányokban lehet általánosítani. Ahogyan azt már korábban is láthattuk, a parciális duplikáció rengeteg teret hagy a dinamika bővítésére (például egyenletestől eltérő eloszlás szerinti választás), a teljességre való törekvés nélkül csak kettő, a szakirodalomban legtöbbször előforduló, valamilyen formában továbbgondolt esetet fogunk vizsgálni.

A parciális modell egyik nagy hiányossága, hogy csak az újonnan született csúcson enged meg változtatást a gráf struktúrájában, holott az úgynevezett mutáció korábbi csúcs esetén is előfordulhat. Ezt a gondolatot próbálja megfogni a következő, alábbiakban definiált Pastor-Satorras [30] modell.

**Pastor-Satorras modell.** Legyen  $0 \leq p \leq 1$  és  $r \geq 0$  rögzített valós számok.

A folyamatot az  $n_0$  időben indítjuk egy olyan összefüggő  $\mathbb{G}_{n_0}$  gráfból, mely csúcsainak száma pont  $n_0$ . Ezután diszkrét időben előrehaladva, az  $n > n_0$  pillanatban  $\mathbb{G}_n$  a következő kétlépéses dinamika szerint képződik  $\mathbb{G}_{n-1}$ -ből.

- (1) Először az egyenletes eloszlásnak megfelelően választunk egy  $v$  csúcsot a  $\mathbb{G}_{n-1}$  gráfból. Majd ezen  $\mathbb{G}_{n-1}$  gráfhoz felvesszünk egy  $w$  csúcsot, méghozzá úgy, hogy az előbb választott  $v$  összes szomszédjával egymástól függetlenül  $p$  valószínűséggel összekötjük.
- (2) Majd a  $\mathbb{G}_{n-1}$  minden  $u$  csúcsát  $\frac{r}{n-1}$  valószínűséggel összekötjük az új csúccsal, úgy, hogy az esetlegesen keletkező párhuzamos éleket összeolvasztjuk.

Látható, hogy ha itt  $r$  értékének a nullát választjuk, akkor a korábbi parciális duplikációt kapjuk vissza.

Emellett az eredeti modell másik gyenge pontja az, hogy a dinamika egyedül a gráf fejlődését engedi. Ez pedig nem túl célszerű, ugyanis a természetben előforduló folyamatok többségénél jelen van egy kihalási tényező is. Ennek értelmében a hálózatban jelenlévő csúcsok idővel szerepüket veszíthetik és törlődhetnek. Ezt a jelenséget próbálja a modellbe foglalni a következő fejlődési mechanizmus.

**Duplikációs törléses modell.** Legyen  $0 \leq p, q \leq 1$  rögzített értékek. A folyamatot az  $n_0$  időben indítjuk egy olyan összefüggő  $\mathbb{G}_{n_0}$  gráfból, mely csúcsainak száma pont  $n_0$ . Ezután diszkrét időben előrehaladva, az  $n > n_0$  pillanatban  $\mathbb{G}_n$  a következő dinamika szerint képződik  $\mathbb{G}_{n-1}$ -ből.

- (1) Vagy  $q$  valószínűséggel végrehajtunk egy parciális duplikációt, vagyis egyenletes eloszlás szerint választunk egy tetszőleges  $v$  csúcsot a  $\mathbb{G}_{n-1}$ -ből, ezután a  $\mathbb{G}_{n-1}$  gráfhoz felvesszünk egy  $w$  pontot, melyet az előbb választott  $v$  összes szomszédjával és magával  $v$ -vel egymástól függetlenül  $p$  valószínűséggel összekötünk.
- (2) Vagy  $1 - q$  valószínűséggel végrehajtunk egy törlést, azaz a  $\mathbb{G}_{n-1}$  egy egyenletes eloszlás szerint választott csúcsát a belőle kiinduló élek törlésével izolálttá tesszük.

Fontos megjegyezni, hogy ezt a modellt, ilyen általánosságban még nem vizsgálták, így a szakirodalomban sem szerepel. Viszont speciálisabb változatairól már

több eredmény is született. A számunka legfontosabb variáció, melyet a következő fejezetben fogunk részletesen ismertetni, az úgynevezett Thörnblad modell [35]. Ez abban speciális az általános duplikáció-törlés modellhez képest, hogy részleges duplikáció helyett egy olyan teljes másolást végzünk, amiben a másolt és a másolat csúcsok között is keletkezik egy új él. A duplikáció-törlés modell egy másfajta speciális megfogalmazása található a [20] cikkben, amiben megengedjük, hogy egy lépés során akár egymás után történjen másolás és törlés is.

## 3.2. Megcáfolt eredmények

Mint ahogyan az már a korábbiakban említésre került, a biológiai hálózatokra vonatkozó eredmények jó része, lévén a kutatási terület empirikus tudományágaknak köszönhetően született, statisztikai megfontolásokból és szimulációkból származik. Jelen szakasz célja nem csak ezeknek az állításoknak a matematikai megcáfolása, hanem olyan korábbi eredmények ismertetése, melyek a kelleténél több heurisztikus gondolatot és nem szigorú matematikát tartalmaznak. A szakasz során természetesen felvázoljuk az ezeknek ellentmondó állításokat is.

Hasonlóan a többi véletlen gráf modellhez, az elsődleges kérdés szintén az aszimptotikus fokszámeloszlásra vonatkozik. Így a jelen szakaszban felsoroltatott hibás elképzelések is mind ezt taglalják. A [7] cikk alapján fogjuk vizsgálni, hogy a parciális duplikációs és a Pastor-Satorras modelleknél felvetett fokszámeloszlások láthatóan nem teljesülnek.

A [12] cikk második felében a parciális duplikáció által generált gráf fokszámeloszlására, egy korábbi [9], számítógépes szimulációkat alkalmazó érvelés alapján feltételeztek hatványrendben lecsengő eloszlást. Ezt a feltételezést kihasználva a cikk írói azt vizsgálták, hogy milyen összefüggést kell kielégítenie az eloszlást leíró paraméternek. Sajnos ehhez közvetetten az a hibás kikötés is használatra kerül, miszerint a  $k$  fokú csúcsok hányada az időtől független. Ezzel szemben itt csak azal foglalkozunk, hogy miért nem lehet a fokszámeloszlás hatványrendben lecsengő. Ahhoz, hogy ezt megmutassuk, vezessük be az alábbi jelölést.

**Jelölés.** Legyen  $k \geq 1$  természetes szám. Ekkor a folyamat  $n$ -edik pillanatában a  $\mathbb{G}_n$  gráfban található  $k$  fokú csúcsok számát jelöljük  $D_{k,n}$ -nel.

Tegyük fel, hogy a  $\frac{D_{k,n}}{n}$  mennyiségnek, amint  $n$ -nel a végtelenbe tartunk, minden  $k$  esetén létezik határértéke, azaz beszélhetünk az adott fokszámú csúcsok arányának határeloszlásáról. Ekkor, ha ezen határértékeket rendre  $c_k$ -val jelöljük, két eset lehetséges, vagy  $\sum_{k=0}^{\infty} c_k = 1$ , vagy  $\sum_{k=0}^{\infty} c_k < 1$ . Ez utóbbi, amit nevezünk a továbbiakban defektív esetnek, akkor lehetséges, ha  $c_{\infty} > 0$ , vagyis a gráfban létezik egy óriás komponens. Például ez a helyzet áll fenn teljes duplikáció esetén. Nézzük meg mikor lehet a határeloszlás nem defektív.

A modell dinamikáját felhasználva könnyen belátható, hogy az  $n$ -edik lépésben, a gráfban található élek várható száma aszimptotikusan  $e_{n_0} n^{2p}$ , ahol  $e_{n_0}$  a kiindulási gráf éleinek száma. Tehát a parciális modell várható átlagos fokszáma  $n^{2p-1}$  rendben változik. Így pedig nyilvánvaló, hogy a  $p < \frac{1}{2}$  esetben ha létezik határeloszlás, akkor ez nem lehet defektív. Érdeemes megjegyezni, habár tévesen úgy tűnhet, hogy ha

$p \geq \frac{1}{2}$ , akkor a kapott határeloszlás már defektív, ahogyan a későbbiekben látni fogjuk, ez nem igaz.

Ahhoz, hogy a hatványrendben lecsengő fokszámeloszlást megcáfoljuk, arra lesz szükségünk, hogy hogyan alakul az izolált csúcsok száma a gráf fejlődése során. Már korábban is szerepelt, a folyamat közben a gráf végig egy összefüggő komponens és izolált pontok összessége. Így azt a kérdést szeretnénk megválaszolni, hogy ha időben a végtelenbe tartunk, akkor előfordulhat-e, hogy az izolált csúcsok aránya 1-hez tart, azaz tetszőleges  $k \geq 1$  esetén a  $k$  fokú csúcsok aránya a 0-hoz konvergál. Ezt viszont a [7] cikkben található következő lemma garantálja.

**3.1. Lemma.** *Legyen  $0 < p < 1$  és tekintsük a parciális duplikációs modellt. Ekkor a korábbi jelölésünk mellett, a  $\frac{D_{0,n}}{n}$   $n$ -nek monoton növekvő függvénye, melyre*

$$\lim_{n \rightarrow \infty} \frac{D_{0,n}}{n} \leq 1$$

*teljesül. Továbbá, minden  $k \geq 1$  esetén*

$$\lim_{n \rightarrow \infty} \frac{D_{k,n}}{n} = 0.$$

A lemma mindkét állítása egy, az adott pillanatban a gráfban található izolált csúcsok számára felírt rekurzió alapján könnyen bizonyítható, így a bizonyítást az olvasóra hagyjuk.

Tehát a 3.1. Lemma alapján látható, hogy ha kizárjuk a  $p = 1$  szélsőséges esetet, akkor a parciális duplikációs modell esetén nem lehetséges olyan aszimptotikus fokszámeloszlásról beszélni, melyre a  $k \geq 1$  fokú csúcsok aránya nem 0-val egyenlő. Például  $p < \frac{1}{2}$  esetén, mint láttuk, az összefüggő komponens átlagos fokszáma időben a 0-hoz konvergál, tehát a lemma alapján  $c_0 = 1$ .

Természetesen ez sem teljesen korrekt, ugyanis a kimondás során közvetetten feltettük azt, hogy valamilyen határeloszlás létezik, ami egyáltalán nem garantált. Például előfordulhat, hogy egy  $p > \frac{1}{2}$  paraméter esetén a gráfban található összefüggő komponens egy olyan óriáskomponens, melyben a minimális fokszám a végtelenhez tart.

Érdeemes megjegyezni azt is, hogy a 3.1. Lemma csak azt cáfolja, hogy az egész gráfban van aszimptotikusan hatványrendben lecsengő fokszámeloszlás. Ettől függetlenül nagy a valószínűsége, hogy a [12] cikkben szereplő eredmény és feltételek a gráf összefüggő komponensén belül igazak, habár ezt eddig nem sikerült bebizonyítani.

Térjünk rá most az előbbiekben bevezetett Pastor-Satorras modell aszimptotikus fokszámeloszlására vonatkozó állításokra. Ehhez vizsgáljuk meg, hogy hogyan alakulhat az  $l$  pillanatban született csúcs fokszáma az  $n > l$  időpontban. Egyszerűség kedvéért ehhez tegyük fel, hogy a Pastor-Satorras modellben szereplő duplikáció utáni mutációt leíró  $r$  paraméter 0-val egyenlő. Ez azért is tehető meg, mert ha már ebben a szélsőséges állapotban is sikerül belátni, hogy a fokszám viszonylag nagy, akkor ha további plusz éleket húzunk, azzal nem ronthatunk.

Jelölje tehát az  $l$  pillanatban született csúcs fokszámát az  $n \geq l$  időpontban  $d_l(n)$ . Ekkor, mivel a csúcsok fokszáma csak akkor tud növekedni, ha valamely

szomszédját duplikáljuk, így a  $d_l(n)$  várható értékére a következő összefüggés írható fel:

$$E[d_l(n+1)] = E[d_l(n)] \left(1 + \frac{p}{n}\right) = d_l(l) \left(1 + \frac{p}{l}\right) \left(1 + \frac{p}{l+1}\right) \dots \left(1 + \frac{p}{n}\right).$$

Ebből viszont felhasználva azt a jól ismert tényt, miszerint

$$\log(1+x) = x - O(x^2),$$

azt kapjuk, hogy létezik olyan  $C_1$  konstans, melyre

$$E[d_l(n+1)] > C_1 E[d_l(l)] \left(\frac{n}{l}\right)^p$$

teljesül. Ezt kihasználva, a [7] cikkben szereplő további heurisztikus ötletekkel belátható, hogy a  $k_{\max}$ -al jelölt maximális fokszám várható értékére a következő alsó becslés érvényes

$$E[k_{\max}] > C_2 n^p,$$

ahol  $C_2 > 0$  egy alkalmas konstans.

Nyilvánvalóan az ellentmondás nem ebben a gondolatmenetben van, hanem egy korábbi, a modellről íródott legelső [30] cikkben található fejtegetésben. Itt az aszimptotikus fokszámeloszlást taglaló végső konklúzióként az szerepelt, hogy ez exponenciális levágással ellátott hatványrendben lecsengő eloszlás. Ez viszont az előbbieket miatt nem lehetséges, ugyanis ha ilyen lenne a fokszámeloszlás, akkor léteznének olyan  $a > 1$  és  $b, c > 0$  konstansok, hogy a  $k$  fokú csúcsok száma az  $n$  pillanatban  $nck^{-b}a^{-k}$  értékkel egyezik aszimptotikusan. Így pedig ha a  $\frac{\log n}{\log a}$  számnál nagyobb  $k$ -ra összegezzük a  $k$ -fokú csúcsok számait, akkor az összeg 0-hoz fog konvergálni, tehát a legnagyobb fokszám, azaz  $k_{\max}$ , nagyságrendje  $O\left(\frac{\log n}{\log a}\right)$ -val fog meg egyezni. Ez viszont a korábbi alsó becslés miatt nem lehetséges. Szóval összegezve a következő állítást bizonyítottuk be.

**3.1. Állítás.** *A Pastor-Satorras modell aszimptotikus fokszámeloszlása nem lehet exponenciális levágással ellátott hatványtörvény-eloszlás.*

Mielőtt továbbmennénk a következő szakaszra és a parciális duplikációra vonatkozó eredmények tárgyalására, mindenképp hasznos lehet megemlíteni, mi az, amit a Pastor-Satorras-féle általánosításról elmondhatunk. Könnyen belátható, hogy ha feltesszük, hogy  $r > 0$ , akkor itt nem fordulhat elő olyan fajta degenerált határeloszlás, mint ami a parciális duplikáció esetén volt. Persze sajnos ekkor is megtörténhet, hogy egyáltalán nem létezik aszimptotikus fokszámeloszlás. Azt, hogy ez mikor áll fenn és mikor létezik határeloszlás, még nem sikerül formálisan belátni. Így a [7, 26] cikkekben a Pastor-Satorras modellre kimondott (de nem bizonyított) következő tétel ezt az erős feltételt is tartalmazza.

**3.1. Tétel.** *Legyen  $0 < p < 1$  és  $r > 0$  rögzítettek. Ekkor ha feltesszük, hogy a Pastor-Satorras modellben létezik aszimptotikus fokszámeloszlás, akkor a  $k$  fokú csúcsok arányának határértéke  $(1 + O\left(\frac{1}{k}\right))ck^{-b}$ , ahol  $c$  egy alkalmas konstans és  $b$  az*

$$1 = pb - p + p^{b-1}$$

*egyenlet legnagyobb megoldása.*

Ezen felül érdekességként megjegyezzük, hogy a [7] cikkben nemcsak a fokszámoszlás kérdését, hanem az élek várható számát is vizsgálták. Ezt az eredményt foglalja össze következő lemma, melyet szintén nem bizonyítunk.

**3.2. Lemma.** *Legyen  $0 < p < 1$  és  $r > 0$  rögzítettek. Ekkor a Pastor-Satorras modellben az  $n$ -edik lépésében a gráfban található élek számának  $e(n)$  várható értékére a következő összefüggés írható fel:*

$$e(n) \sim \begin{cases} \frac{r}{1-2p}n, & \text{ha } p < \frac{1}{2}; \\ rn \log n, & \text{ha } p = \frac{1}{2}; \\ \left(e(0) + \frac{r}{2p-1}\right)n^{2p}, & \text{ha } p > \frac{1}{2}. \end{cases}$$

### 3.3. Eredmények

Az előbbi, inkább módszertani jellegű kitérő után térjünk rá a parciális duplikációs modellre vonatkozó eredmények összefoglalására. Ehhez a témában íródott legaktuálisabb [21] cikkben taglaltakat fogjuk vázlatosan áttekinteni. Fő eredménynek a korábbi 3.1. Lemmából következő defektív határeloszlás létezésének korlátjára adott kritikus érték tekinthető. Ezt nem bizonyítjuk, ugyanis a bizonyításához felhasznált módszerek nagy része túlmutat a dolgozat témáján. Ezzel szemben vázlatosan belátjuk, hogy hogyan alakul a kezdeti gráf csúcsainak fokszáma a folyamat során. Sőt, említés szintjén szó lesz a modell klaszterezettségi együtthatójáról is.

Vizsgálataink megkezdése előtt vezessük be a továbbiakban felhasználandó legfontosabb jelöléseket.

**Jelölés.** Legyen  $k \geq 1$  természetes szám. Jelölje a folyamat  $n$ -edik pillanatában a  $\mathbb{G}_n$  gráfban található  $k$ -csillagok számát  $S_{k,n}$ , míg a  $k$ -klikkek számát  $C_{k,n}$ .

Fontos megjegyezni, hogy itt a csillagok esetén a csúcsok számozottak, míg a klikkek esetén pedig nem.

Ezek nyilvánvalóan összekapcsolhatóak az adott foksámú csúcsok számával. Például könnyen felírható a csillagok és foksámok közti

$$S_{k,n} = \sum_{l=k}^{\infty} l \dots (l-k+1) D_{l,n}$$

összefüggés. Ez azért teljesül, mivel minden  $l \geq k$  foksámú csúcshoz található  $\frac{l!}{(l-k)!}$  darab, az adott csúccsal megegyező középpontú  $k$ -csillag.

Térjünk rá tehát a szakasz első tételére. Eddigi ismereteink szerint, ha a  $p$  paraméter kisebb, mint  $\frac{1}{2}$ , akkor az aszimptotikus foksámoszlás nem lehet defektív. Ahogyan a [7] cikkben is szerepel, ez egyáltalán nem garantálja azt hogy az  $\frac{1}{2}$  felett már biztosan defektív határeloszlást kapunk. A következő eredmény, amit a korábban taglaltak miatt nem bizonyítunk, ezt a kritikus értéket adja meg.

**3.2. Tétel.** *Legyen  $0 \leq p \leq 1$  rögzített érték és tekintsük a parciális duplikációs modellt. Továbbá jelölje a  $pe^p = 1$  egyenlet egyértelmű megoldását  $p^*$  és a foksámok eloszlásának generátorfüggvényét az  $n$ -edik pillanatban  $H_n(q) = \sum_{k=0}^{\infty} \frac{D_{k,n}}{n} q^k$ . Ezek mellett a következők írhatók fel:*

1. Ha  $p \leq p^*$ , akkor  $H_n(q) \rightarrow 1$  sztochasztikusan, amint  $n \rightarrow \infty$ , minden  $q \in [0, 1)$  esetén.
2. Ha  $p > p^*$ , akkor  $E[H_n(q)] \rightarrow q_\infty < 1$ , amint  $n \rightarrow \infty$ , minden  $q \in [0, 1)$  esetén, ahol

$$q_\infty = 1 - \left(1 - \frac{1}{p} \log\left(\frac{1}{p}\right)\right) \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \frac{S_{k,n_0}}{n_0} \prod_{l=1}^{k-1} \left(1 - \frac{1-p^l}{pl}\right).$$

Látható, hogy a  $q = 0$  helyettesítés mellett  $H_n(0) = \frac{D_{0,n}}{n}$  teljesül. Ezáltal a tétel első pontja éppen az általunk említett kritikus értéket adja, ugyanis behelyettesítve azt kapjuk, hogy  $\frac{D_{0,n}}{n} \rightarrow 1$ , azaz az izolált csúcsok hányada 1-hez konvergál. Ezáltal megkaptuk, hogy  $p^*$  lesz az a kritikus érték, mely a nem defektív és defektív aszimptotikus fokszámoszlásokat elválasztja egymástól. Megjegyezendő érdekesség, hogy a tételben szereplő  $p^*$  érték, ami megközelítőleg 0.567143, a korábbi [12] cikk parciális duplikációra vonatkozó eredményei között is feltűnik, habár a jelenlegitől eltérő kontextusban.

A Lévy-féle folytonossági tétel generátorfüggvényekre vonatkozó változata értelmében az is elmondható, hogy  $p \leq p^*$  mellett a  $\frac{D_{k,n}}{n}$  sorozat eloszlásban a konstans 1 generátorfüggvényű eloszláshoz konvergál. Tehát az is igaz, hogy az aszimptotikus fokszámoszlás megegyezik  $\delta_0$ -lal, ha  $p \leq p^*$ . Természetesen a tétel értelmében az is igaz, hogy ha  $p > p^*$ , akkor ilyen aszimptotikus fokszámoszlás nem létezik.

Ajánlott újból megemlíteni, hogy a 3.2. Tétel nem zárja ki annak lehetőségét, hogy a gráf összefüggő komponensében hatványrendben lecsengő fokszámoszlás figyelhető meg. Azaz lehetséges, hogy léteznek  $b, c > 0$  értékek, hogy

$$\lim_{k \rightarrow \infty} k^b \lim_{n \rightarrow \infty} \frac{D_{k,n}}{\sum_{l=1}^{\infty} D_{l,n}} = c$$

teljesül. Ezt a sejtést a korábban hangoztatott fokszámarányos, illetve a parciális duplikációs dinamikával fejlődő véletlen gráfok közti kapcsolat is megerősíti. Habár a [22] cikkben a problémával bővebben foglalkoztak, a felsorakoztatott eredmények bizonyításához sokszor inkább a heurisztikus utat választották.

Ahhoz, hogy a parciális duplikáció minél jobban reprezentálja a valóságot, fontos, hogy a klaszterezettségi együttható magas legyen. A definíció alapján ahhoz, hogy ezt az értéket meghatározzuk, szükség lesz a gráfban található háromszögek és kétágú csillagok várható számára. Ezekre a következő két tétel mondható ki, melyek bizonyítását, mivel azok jórészt technikai lépéseken alapulnak, szintén nem részletezzük.

**3.3. Tétel.** *Legyen  $0 \leq p \leq 1$  rögzített érték és tekintsük a parciális duplikációs modellt. Tegyük fel, hogy  $C_{k,n_0} > 0$  teljesül. Ekkor minden  $k$  esetén majdnem biztosan létezik*

$$\lim_{n \rightarrow \infty} n^{-kp^{k-1}} C_{k,n} = C_{k,\infty}$$

és a  $k = 2$  esetben ez  $\mathcal{L}^2$ -ben is fennáll. Továbbá a  $C_{k,n}$  várható értékére

$$E[C_{k,n}] = C_{k,n_0} \prod_{m=n_0}^{n-1} \frac{m + kp^{k-1}}{m} \sim \frac{C_{k,n_0} \Gamma(n_0)}{\Gamma(n_0 + kp^{k-1})} n^{kp^{k-1}} \quad (n \rightarrow \infty)$$

teljesül.

**3.4. Tétel.** Legyen  $0 \leq p \leq 1$  rögzített érték és tekintsük a parciális duplikációs modellt. Ekkor minden  $k$  esetén majdnem biztosan létezik

$$\lim_{n \rightarrow \infty} n^{-(kp+p^k)} S_{k,n} = S_{k,\infty}$$

és a  $k = 1$  esetben pedig  $\mathcal{L}^2$ -ben. Továbbá  $S_{2,n}$  várható értékére

$$\begin{aligned} E[S_{2,n}] &= (S_2(n_0) + \frac{2}{p} S_1(n_0)) \prod_{k=n_0}^{n-1} \frac{k+2p+p^2}{k} - S_1(n_0) \frac{2}{p} \prod_{k=n_0}^{n-1} \frac{k+2p}{k} \\ &\sim (S_2(n_0) + \frac{2}{p} S_1(n_0)) \frac{\Gamma(n_0)}{\Gamma(n_0+2p+p^2)} n^{2p+p^2} \quad (n \rightarrow \infty) \end{aligned}$$

teljesül.

Az első észrevétel a 3.4. Tétellel kapcsolatos, ugyanis az állítást elég lenne  $k \geq 2$  feltétellel kimondani, amivel az 1-csillagokat kizárva azt a téves gondolatot ébresztenénk, hogy esetleg ebben az esetben nem tudunk semmit sem mondani. Természetesen ez nem igaz, az  $S_{1,n}$  várható értékére a 3.3. Tételből is tudunk következtetéseket levonni, ugyanis az 1-csillagok és 2 csúcsú teljes gráfok között az  $S_{1,n} = 2C_{2,n}$  összefüggés írható fel.

Az előbbi két tétel segítségével a 3.2. Tételben szereplőhöz hasonló eredményt kaphatunk a  $k$ -klikkek, valamint a  $k$ -csillagok relatív gyakoriságának, azaz a  $\frac{C_{k,n}}{n}$  és  $\frac{S_{k,n}}{n}$  hányadosok aszimptotikus viselkedésének leírására. Nyilvánvalóan konvergenciartományuk a parciális duplikáció  $p$  paraméterének függvényében két részre osztható. Ha  $p$  elég nagy, akkor ezen arányok minden határon túl nőnek. Viszont ha  $p$  viszonylag kicsi, akkor a gráfban található  $k$  csúcsú klikkek és  $k$ -csillagok száma eltöprel a csúcsok számához képest. Precízebben rögzített  $k$  mellett a következők írhatók fel. A  $k$ -klikkek arányának határértékére,

$$\lim_{n \rightarrow \infty} \frac{C_{k,n}}{n} = \begin{cases} \infty, & \text{ha } p > k^{-\frac{1}{k-1}}; \\ C_{k,\infty}, & \text{ha } p = k^{-\frac{1}{k-1}}; \\ 0, & \text{ha } p < k^{-\frac{1}{k-1}}; \end{cases}$$

teljesül. Ehhez hasonlóan a  $k$ -csillagok arányának határértékére pedig a

$$\lim_{n \rightarrow \infty} \frac{S_{k,n}}{n} = \begin{cases} \infty, & \text{ha } p > p_k; \\ S_{k,\infty}, & \text{ha } p = p_k; \\ 0, & \text{ha } p < p_k; \end{cases}$$

összefüggést kapjuk, ahol  $p_k$  a  $pk + p^k = 1$  egyenlet  $[0, 1]$  intervallumbéli megoldása.

Érdekességként megemlíthető, hogy ezen kritikus értékek semmilyen  $k$  értékre nem lesznek egyenlők  $p^*$ -gal, viszont egy egyszerű kapcsolat létesíthető a  $p_k$ -k és a  $p^*$  között. Ha feltesszük, hogy  $p \geq p^*$ , akkor a

$$px + p^x \geq pxe^{-px} > 1$$

egyenlőtlenség minden  $x > 0$  esetén fennáll. Azaz  $p^*$  a legkisebb olyan  $[0, 1]$ -beli paraméter, melyre a  $px + p^x = 1$  egyenletnek nem létezik pozitív megoldása  $x$ -ben.

Emlékezzünk vissza arra, hogy egy gráf klaszterezettségi együtthatóján a gráfban található háromszögek háromszorosának és a 2-utak számának hányadosát értjük. Mivel a háromszögek nem mások, mint a 3-klikkek, továbbá minden 2-út kétféleképp fogható fel, mint 2-csillag, így az eddigi jelöléseink mellett az  $n$ -edik pillanatban a gráf klaszterezettségi együtthatója a

$$\frac{6C_{3,n}}{S_{2,n}}$$

érték. Tehát az előző két tétel alapján, a keresett paraméterre, némi egyszerűsítés követően felírható, hogy

$$\frac{6E[C_{3,n}]}{E[S_{2,n}]} \sim \frac{6C_{3,n_0}}{S_{2,n_0} + \frac{2}{p}S_{1,n_0}} \frac{\Gamma(n_0 + 2p + p^2)}{\Gamma(n_0 + 3p^2)} n^{-2p(1-p)},$$

amint  $n$  a  $\infty$ -hez tart.

Így mivel az első két szorzótényező konstans, elegendő megnézni, hogy az  $n$  kitevőjében milyen értékek szerepelhetnek. Tekintettel arra, hogy ez a kitevő minden nem szélsőséges  $p$  értékre negatív, azt a következtetést vonhatjuk le, hogy a modell klaszterezettségében nem reprezentálja jól a valóságban megfigyelt biológiai hálózatokat.

Az utolsó, e szakaszban taglalt eredmény a kiindulási gráfban található csúcsok fokszámainak folyamat közbeni fejlődésére vonatkozik. Ahhoz, hogy ezt vázlatosan bizonyítsuk, szükségünk lesz az alábbi martingáleméleti állításra.

**3.2. Állítás.** *Legyen  $(X_n)_{n=n_0, n_0+1, \dots}$  egy nemnegatív, integrálható sztochasztikus folyamat, mely az  $(\mathcal{F}_n)_{n=n_0, n_0+1, \dots}$  filtrációra nézve adaptált,  $\mathcal{F}_\infty = \sigma(\bigcup_{n=n_0}^\infty \mathcal{F}_n)$  és  $x_0 = E[X_{n_0}]$ . Továbbá tegyük fel, hogy*

$$E[X_{n+1} | \mathcal{F}_n] = \left(1 + \frac{a}{n}\right) X_n$$

*minden  $n = n_0, n_0+1, \dots$  esetén, ahol  $a > -n_0$  tetszőleges szám. Ekkor a következők teljesülnek:*

1. *Az alábbi  $\mathcal{M} = (M_n)_{n=n_0, n_0+1, \dots}$  sztochasztikus folyamat, martingál*

$$M_{n_0} = X_{n_0}; \quad M_n = X_n \prod_{k=n_0}^{n-1} \frac{k}{k+a}; \quad n \geq n_0$$

*Továbbá*

$$E[X_n] = x_0 \prod_{k=n_0}^{n-1} \frac{k+a}{k} \sim \frac{x_0 \Gamma(n_0)}{\Gamma(n_0+a)} n^a$$

*amint  $n$  a  $\infty$ -hez tart.*

2. *Majdnem biztosan létezik*

$$\lim_{n \rightarrow \infty} n^{-a} X_n = X_\infty,$$

*ahol*

$$E[X_\infty] \leq \frac{x_0 \Gamma(n_0)}{\Gamma(n_0+a)}$$

*teljesül.*



3. Ha az előző pontban még az is igaz, hogy léteznek  $r > 1$  és  $c > 0$  konstansok úgy, hogy

$$E[X_n^r] \leq cn^{ar},$$

akkor  $\mathcal{L}^r$ -beli konvergencia is teljesül.

**Bizonyítás.** A feltételek miatt az  $\mathcal{M}$  adaptáltsága és integrálhatósága nyilvánvaló. A martingál tulajdonság pedig az alábbiak szerint látható:

$$E[M_{n+1}|\mathcal{F}_n] = E[X_{n+1}|\mathcal{F}_n] \prod_{k=n_0}^n \frac{k}{k+a} = X_n \prod_{k=n_0}^{n-1} \frac{k}{k+a} = M_n.$$

Az  $E[X_n]$ -re vonatkozó összefüggés teljes indukció segítségével könnyedén ellenőrizhető, míg az aszimptotikához a gamma-függvény következő nevezetes tulajdonságát használva juthatunk el:

$$\prod_{k=n_0}^{n-1} \frac{k+a}{k} = \frac{\Gamma(n+a)}{\Gamma(n)} \frac{\Gamma(n_0)}{\Gamma(n_0+a)} \sim \frac{\Gamma(n_0)}{\Gamma(n_0+a)} n^a \quad (n \rightarrow \infty).$$

A második ponthoz alkalmazzuk a martingál-konvergenciatétel azon következményét, hogy egy nemnegatív martingálnak majdnem biztosan létezik határértéke. Így  $\mathcal{M}$  egy  $M_\infty$  valószínűségi változóhoz konvergál majdnem biztosan. Tehát a korábbiak mellett felírható, hogy

$$n^{-a} X_n = n^{-a} M_n \prod_{k=n_0}^{n-1} \frac{k+a}{k} \rightarrow M_\infty \frac{\Gamma(n_0)}{\Gamma(n_0+a)} = X_\infty,$$

amiből már az állítás konvergencia része következik. A limesz várható értékére vonatkozó egyenlőtlenség a Fatou-lemma következménye. Végül használjuk fel a feltétel által adott  $E[X_n^r] \leq cn^{ar}$  korlátot és a gamma-függvényre korábban már alkalmazott aszimptotikát. Összesítve a következőt kapjuk:

$$\sup_n E[M_n^r] = \sup_n E[X_n^r] \left( \prod_{k=n_0}^{n-1} \frac{k}{k+a} \right)^r \leq \sup_n \frac{\Gamma(n_0+a)^r cn^{ar}}{\Gamma(n_0)^r c' n^{ar}} < \infty,$$

tehát az  $\mathcal{M}$  martingál  $\mathcal{L}^r$ -ben korlátos, azért  $\mathcal{L}^r$ -ben konvergens. Azaz az állítást beláttuk. ■

Mondjuk ki tehát a kezdeti gráf fokszámainra vonatkozó eredményt.

**3.5. Tétel.** Legyen  $0 \leq p \leq 1$  rögzített érték és tekintsük a parciális duplikációs modellt. Címkezzük a kiindulási gráf csúcsait az  $1, 2, \dots, n_0$  számokkal, illetve jelölje  $d_k(n) > 0$  a  $k \leq n_0$  címkéjű csúcs foksámát az  $n$ -edik pillanatban. Ekkor  $n \geq n_0$  és  $l \geq a$  esetén

$$P(d_k(n) = l | d_k(n_0) = a) = \sum_{m=a}^l (-1)^{m-a} \binom{l-1}{m-1} \binom{m-1}{a-1} \prod_{j=n_0}^{n-1} \left(1 - \frac{pm}{j}\right)$$

teljesül. Továbbá

$$n^{-p}d_k(n) \rightarrow d_k(\infty) \quad (n \rightarrow \infty)$$

majdnem biztosan és  $\mathcal{L}^r$ -ben tetszőleges  $r \in \mathbb{N}$  esetén. Jelölje az úgynevezett eltolt faktoriális  $(l)_r = l(l+1)(l+2)\dots(l+r-1)$ , ahol  $r > 0$  és  $(l)_0 = 1$ . Ekkor az

$$E[n^{-rp}(d_k(n))_r] = \frac{(d_k(n_0))_r}{n^{rp}} \prod_{k=n_0}^{n-1} \frac{k+rp}{k} \rightarrow \frac{\Gamma(n_0)}{\Gamma(n_0+rp)} (d_k(n_0))_r = E[d_k(\infty)^r]$$

összefüggés áll fenn, amint  $n$  tart végtelenbe.

Amint az látható, három eredményt kell belátnunk a tétel bizonyításához. Ezek közül a legelsőnél néhány mélyebb kombinatorikai előismeretet használó lépést nem részletezünk, továbbá a korábbi tisztán martingáleméleti állításba való behelyettesítéseket is mellőzük, tehát a bizonyítás vázlatosnak tekinthető.

**Bizonyítás.** A feltételes valószínűségekre vonatkozó eredményhez elegendő belátni, hogy

$$P(d_k(n) \leq l | d_k(n_0) = a) = \sum_{m=a}^l (-1)^{m-a} \binom{l}{m-1} \binom{m-1}{a-1} \prod_{j=n_0}^{n-1} \left(1 - \frac{pm}{j}\right)$$

teljesül, ugyanis ekkor kihasználva a binomiális együtthatók rekurzív megadási képletét, az állítást kapjuk. Ezt  $n$  szerinti teljes indukcióval fogjuk megmutatni. Ehhez rögzítsük az  $n_0$ ,  $k$  és  $a$  értékeket és vezessük be a

$$\Phi_l(n) = P(d_k(n) \leq l | d_k(n_0) = a)$$

jelölést. Ha  $n = n_0$ , akkor nyilvánvalóan  $\Phi_l(n_0) = \mathbb{I}_{\{l \geq a\}}$  igaz, sőt különböző, itt nem részletezett leszámplálási módszerekkel megmutatható, hogy ez tetszőleges  $l$  választása mellett is fennáll.

Mivel egy rögzített csúcs fokszáma tetszőleges lépés során legfeljebb 1-gyel nőhet, így az alábbi rekurziót tudjuk felírni:

$$\Phi_l(n+1) = \Phi_l(n) - \frac{pl}{n} P(d_k(n) = l | d_k(n_0) = a) = \Phi_l(n) - \frac{pl}{n} (\Phi_l(n) - \Phi_{l-1}(n)).$$

Tegyük fel tehát, hogy állításunk  $n$ -re és minden  $l$ -re teljesül, ekkor az előbbi rekurziót használva a következő egyenlőséget kapjuk,

$$\begin{aligned} \Phi_l(n+1) &= \sum_{m=a}^l (-1)^{m-a} \binom{l}{m} \binom{m-1}{a-1} \prod_{j=n_0}^{n-1} \left(1 - \frac{pm}{j}\right) \\ &\quad - \sum_{m=a}^l (-1)^{m-a} \frac{pl}{n} \left( \binom{l}{m} - \binom{l-1}{m} \right) \binom{m-1}{a-1} \prod_{j=n_0}^{n-1} \left(1 - \frac{pm}{j}\right) \\ &= \sum_{m=a}^l (-1)^{m-a} \binom{l}{m} \binom{m-1}{a-1} \prod_{j=n_0}^{n-1} \left(1 - \frac{pm}{j}\right) \\ &\quad - \frac{pm}{n} \sum_{m=a}^l (-1)^{m-a} \binom{l}{m} \binom{m-1}{a-1} \prod_{j=n_0}^{n-1} \left(1 - \frac{pm}{j}\right) \\ &= \sum_{m=a}^l (-1)^{m-a} \binom{l}{m} \binom{m-1}{a-1} \prod_{j=n_0}^{n-1} \left(1 - \frac{pm}{j}\right). \end{aligned}$$

amiből pedig az első állítás már következik.

A tételben megmaradt két további állításhoz a 3.2. Állítást használhatjuk. Mint ahogyan már az előbbiekben is szerepelt, egy adott csúcs fokszáma egy lépés során legfeljebb csak eggyel nőhet. Mivel ez a növekedés akkor következik be, ha az adott csúcs egyik szomszédját duplikáljuk, így az egy lépést követően bekövetkező változás:

$$d_k(n+1) - d_k(n) = \begin{cases} 1, & p \frac{d_k(n)}{n} \text{ valószínűséggel;} \\ 0, & 1 - p \frac{d_k(n)}{n} \text{ valószínűséggel.} \end{cases}$$

Az eltolt faktoriálisra nyilvánvalóan az alábbi összefüggés is teljesül:

$$(l+1)_r = l_r + r(l+1)_{r-1}.$$

Tehát ezeket összesítve azt kapjuk, hogy

$$E[(d_k(n+1))_r | \mathcal{F}_n] = (d_k(n))_r + \frac{p d_k(n)}{n} r (d_k(n) + 1)_{r-1} = (d_k(n))_r \left(1 + \frac{rp}{n}\right).$$

Itt ha az  $r$  értékét 1-nek választjuk, akkor a 3.2. Állítás 2. pontja alapján a kívánt majdnem biztos konvergencia, illetve a várható értékre felírt összefüggés következik. Ha  $r \geq 2$ , akkor ugyanezen állítás 1. eredménye alapján megkaphatjuk, hogy

$$E[d_k(n)^r] \leq E[(d_k(n))_r] \geq cn^{rp}$$

valamilyen alkalmas  $c > 0$  konstans mellett. Tehát használhatjuk az előbbi állítás 3. pontját, ami miatt az  $\mathcal{L}^r$ -beli konvergencia is teljesül. Sőt, mivel  $d_k(n) \rightarrow \infty$  ( $n \rightarrow \infty$ ), így a szükséges utolsó állítást is megkapjuk, azaz az

$$E[d_k(\infty)^r] = \lim_{n \rightarrow \infty} E[(n^{-rp} d_k(n))^r] = \lim_{n \rightarrow \infty} E[n^{-rp} (d_k(n))_r]$$

összefüggés teljesül, így a bizonyítást befejeztük. ■

Természetesen előfordulhat, hogy számunkra nem a kiindulási gráf egyik csúcsának fokszáma érdekes, hanem a folyamat során születőé. Ebben az esetben egyszerűen tekintjük azt az állapotot inicializálónak, melyben a vizsgálandó csúcs született. Így az előbbi tétel tetszőleges csúcs fokszámának leírására alkalmas.

## 4. Thörnblad modell

A való életben megfigyelt biológiai hálózatok egyik legszembetűnőbb sajátossága az, hogy olyan nagy és sűrűn összekapcsolt blokkokból állnak, melyeket csak egy-két él kapcsol össze. Ahogyan az már korábban említésre került, ezt a tulajdonságot magas klaszterezettségnek nevezzük, melynek fennállására kézenfekvő magyarázat az ilyen rendszerek evolucionáris - azaz másolatokon keresztül - fejlődése. Az előző fejezetben taglalt általános parciális duplikációs modellek klaszterezettségi együtthatójára kapott korábbi becslés miatt levonhattuk azt a következtetést, miszerint ez a tulajdonság ezeknél nem figyelhető meg. Ezzel ellentétben a most következő speciális dinamikus fejlődő gráfban ez a jellemző lesz, amelyet az egyes eredmények bizonyítása közben leginkább ki fogunk használni.

A továbbiakban elemzett diszkrét modell egy kevesebb véletlent tartalmazó változatát elsőként Backhausz és Móri ismertette a [3] cikkben. Később ezt Thörnblad általánosította [35], a fejlődési dinamikát tovább randomizálva. Tekintettel arra, hogy a kérdéskör biológiai környezetből származik, így sokkal természetesebb lenne ezen általánosított változat egy folytonos idejű megfelelőjét elemezni, ahogyan azt Champagnat, Lambert és Richard tették [10, 11] pár évvel korábban, egy mutációt tartalmazó populációs modell vizsgálatakor. Viszont itt mégis a diszkrét esettel foglalkozunk, ugyanis a folytonossá tétel könnyen megtehető egy később ismertetett módszerrel, így az itt szereplő összes eredmény kimondható a diszkrét idejűségből fakadó megszorítás elhagyásával is.

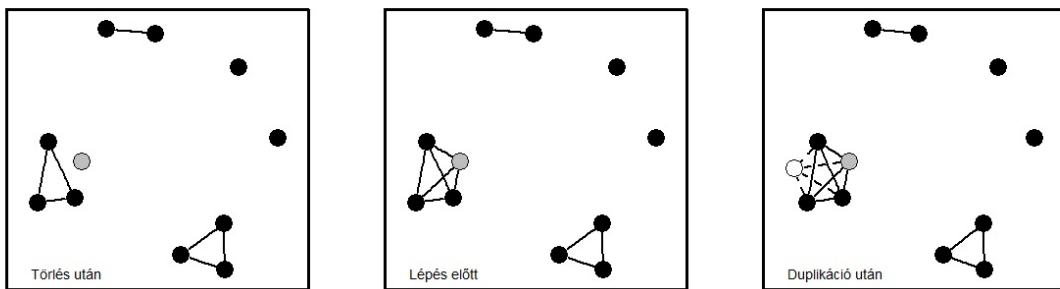
Az alábbi fejezetben a következőképpen fogunk haladni. Először ismertetjük a modellt és annak szakirodalomban gyakran előforduló verzióit. Ezután rátérünk a hálózat szerkezetét leíró tulajdonságainak vizsgálatára. A második szakaszban taglaljuk a gráf aszimptotikus fokszámeloszlásra vonatkozó eredményeket, illetve a coupling módszer segítségével egy speciális esetben átemeljük ezt egy korábbi dinamikára is. A harmadik alfejezetben áttekintjük a maximális fokszámú csúcs kérdéskörét, és majdnem biztos aszimptotikus korlátot adunk ezen fokszámra, mely egy kitüntetett esettől eltekintve tovább nem gyengíthető. Az utolsó fejezetben az előbbi korlátot élesítjük ezen megmaradt esetben, mely következtében a modell kapcsolata a valószínűségi számítás más területével is előkerül.

### 4.1. A modell és változatai

A következőkben definiáljuk a Thörnblad modellt, illetve ennek különböző, bizonyításokban használatos változatait. Ennek keretein belül kitérünk egy, a diszkrét idő folytonosítására vonatkozó általános módszerre, melynek bővebb kifejtése a [2] könyv 5. fejezetében megtalálható.

**Thörnblad modell.** Legyen  $0 < p < 1$  fixált érték. A folyamatot az  $n = 0$  időben indítjuk egy olyan  $\mathbb{G}_0$  gráfból, mely egyetlen izolált csúcsból áll. Majd a diszkrét időben előrehaladva, az  $n \geq 1$  pillanatban  $\mathbb{G}_n$  a következő dinamika szerint képződik  $\mathbb{G}_{n-1}$ -ből.

- (1) Vagy  $p$  valószínűséggel végrehajtunk egy duplikációt. Ez jelen esetben azt takarja, hogy egyenletes eloszlás szerint választunk egy tetszőleges  $v$  csúcsot a  $\mathbb{G}_{n-1}$ -ből. Ezután a  $\mathbb{G}_{n-1}$  gráfhoz felveszünk egy  $w$  pontot, melyet az előbb választott  $v$  összes szomszédjával és magával  $v$ -vel is összekötünk.
- (2) Vagy  $1 - p$  valószínűséggel végrehajtunk egy törlést, mely annyit jelent, hogy a  $\mathbb{G}_{n-1}$  egy egyenletes eloszlás szerint választott csúcsát a belőle kiinduló élek törlésével izolálttá tesszük.



2. ábra. Egy lépés implementálása. A középső ábrán szerepel a gráf egy adott tetszőleges időpillanatban, ahol a szürkével színezett csúcs lesz az amelyiket duplikáljuk vagy töröljük. A jobb oldali ábra mutatja a lépés utáni állapotot, amennyiben kiválasztott csúcsot duplikáltuk, míg a bal oldali azt, amikor töröltük. Ezek rendre  $p$  és  $1 - p$  valószínűségekkel következhetnek be.

Könnyen látható, hogy a definíció elején rögzített  $p$  értéke szerint a folyamat hosszútávú viselkedése három különböző csoportra osztható. Ha  $0 < p < \frac{1}{2}$ , akkor a törlés dominál, ha  $p = \frac{1}{2}$ , akkor átlagosan ugyanannyi törlés és duplikálás történik, míg ha  $\frac{1}{2} < p < 1$ , akkor a duplikációs lépések dominálnak. Ezen kategóriákra a továbbiakban mint szubkritikus ( $0 < p < \frac{1}{2}$ ), kritikus ( $p = \frac{1}{2}$ ) és szuperkritikus ( $\frac{1}{2} < p < 1$ ) esetre fogunk utalni.

A Thörnblad modell alapvető tulajdonsága, hogy a gráf a fejlődés tetszőleges pillanatában nem lesz más, mint egymástól diszjunkt teljes gráfok összessége. Ezt következőképp kell érteni. Egy ilyen összefüggő komponensen belül minden csúcspár össze van kötve, viszont a különböző klikkek között sosem alakul ki kapcsolat. Mint ahogyan már korábban is előkerült, ez az egyszerűbb struktúra nagyban hozzájárul a dinamika kezelhetőségéhez. Viszont megvan a maga hátulütője, ugyanis azon egyszerűsítés, hogy az egyes komponensek között ritkán előforduló éleket véglegesen elhagyjuk a modellből, a természetben megfigyelt biológiai hálózatoktól való szembevető eltérést eredményez. Így érdemes szót ejteni a folyamatot motiváló, Backhausz és Móri által megfogalmazott dinamikáról, mely nem él az előbbi egyszerűsítéssel.

**Backhausz-Móri modell.** Induljunk ki az egy izolált csúcsot tartalmazó,  $\mathbb{G}_0$ -lal jelölt gráfból. A folyamat során a  $\mathbb{G}_{n-1}$ -ből a  $\mathbb{G}_n$ -et ( $n \geq 1$ ) a következőképp

kapjuk. Egymástól függetlenül válasszunk két tetszőleges, de nem feltétlenül különböző csúcsot  $(v_1, v_2) \in \mathbb{G}_{n-1}$ -ből. Ezután vegyünk egy új csúcsot  $(w)$  a gráfhoz, méghozzá úgy, hogy ezt összekötjük  $v_1$ -gyel és annak minden szomszédjával (duplikáció). Majd a másodikként választott  $v_2$  csúcs minden olyan élét töröljük, mely nem illeszkedik  $w$ -re.

Nem nehéz észrevenni, hogy ha az előbbi dinamikából kivesszük az új csúcsból induló élek megtartására vonatkozó megkötést, akkor a hálózat a fejlődés során végig diszjunkt teljes gráfok rendszere marad. Sőt, az is igaz, hogy az így kapott folyamat aszimptotikusan hasonló tulajdonságokat produkál, mint a Thörnblad modell kritikus esetben. Ezen egyezés azért léphet fel, mivel ekkor a Thörnblad modellben ugyanolyan valószínűséggel duplikálunk vagy törölünk, így átlagosan ugyanannyi duplikációs lépés lesz, mint törölő. Míg a Backhausz és Móri által eredetileg vizsgált folyamatban ezek száma a mechanizmusból adódóan azonos.

A Thörnblad modell valóságba való beültetéséhez a diszkrét idő folytonosításán keresztül is közelebb kerülhetünk. Ezt a következő, valószínűségi számításban gyakran előforduló gondolatmenettel érhetjük el. A gráfban lévő összes csúcshoz rendeljünk két exponenciális órát, azaz exponenciális eloszlással rendelkező valószínűségi változót. Ez az elnevezés azért is célszerű, mivel az exponenciális eloszlásra tekinthetünk úgy, mint egy véletlenszerű esemény bekövetkezéséhez szükséges időre, mely jelen megfogalmazásban lehet egy óra csörgése. Visszatérve eredeti mondanivalónkra, az első paramétere legyen azonos a modellben szereplő  $0 < p < 1$  paraméterrel, míg a másodiké  $(1 - p)$ -vel. Ha ez előbbi óra csörög, akkor duplikáljunk egy csúcsot, viszont ha az utóbbi, akkor töröljünk, a modell dinamikájának megfelelően. Az, hogy a kapott folytonos folyamat, azon időpontokban, amikor történik valami, megegyezik az eredetivel, nyilvánvaló. Viszont, mivel itt a diszkrét idő folytonosba való beágyazásáról van szó, a két folyamat nem tekinthető ekvivalensnek. Ezt nem elfelejtve mégis elmondható, hogy a következőkben megemlített, illetve belátott eredmények elmondhatóak folytonos esetben is, esetleges technikai módosítások mellett.

## 4.2. Aszimptotikus fokszámeloszlás

Az első kérdés, ami felmerülhet, hogy lehet-e beszélni aszimptotikus fokszámeloszlásról, és ha igen, akkor tudunk-e erre explicit képletet adni. Ezen alfejezetben ezt a problémakört járjuk végig, továbbá a coupling módszer alkalmazásával eredményünket kritikus esetben kiterjesztjük a Backhausz-Móri modellre. Az alábbi tétel, mely a [3], illetve [35] cikkekben a szerzők által bizonyított eredmények összefűzése, kielégítő választ ad a fokszámeloszlásra vonatkozó kérdéseinkre.

**4.1. Tétel.** *Tekintsük a Thörnblad modellt valamely  $0 < p < 1$  paraméter mellett. Továbbá vezessük be az alábbi jelöléseket. Legyen  $D_{k,n}$  a  $k$  fokú csúcsok száma  $n$  lépés után,  $N_n$  az összes csúcs száma  $n$  lépés után,  $\beta = \frac{p}{2p-1}$  és  $\gamma = \frac{1-p}{p}$ . Ekkor*

$$\frac{D_{k,n}}{N_n} \rightarrow c_k \quad (n \rightarrow \infty) \quad \text{majdnem biztosan,}$$

ahol  $(c_k)_{k=0}^{\infty}$  pozitív számok olyan sorozata, hogy  $\sum_{k=0}^{\infty} c_k = 1$  és tagjaira az alábbi összefüggés teljesül:

$$c_0 = \frac{1-p}{1+p}(1+c_1); \quad c_k = \frac{k+1}{k+1+p}(pc_{k-1} + (1-p)c_{k+1}) \quad (k \geq 1).$$

Ezenfelül a sorozat tagjai explicit módon is megadhatóak, a különböző esetekben különféleképp, a következők szerint.

- Szubkritikus eset. Ha  $0 < p < \frac{1}{2}$ , akkor

$$c_k = \gamma^{-k-1} \int_0^1 \frac{t^{k+1}(1-t)^{-1-\beta}}{(1-\gamma^{-1}t)^{1-\beta}} dt \quad (k \geq 0), \text{ és}$$

$$c_k \sim (-\beta)^{-1}(1-\beta)^{-\beta} \Gamma(1-\beta) \gamma^{-k} k^{\beta} \quad (k \rightarrow \infty).$$

- Kritikus eset. Ha  $p = \frac{1}{2}$ , akkor

$$c_k = (k+1) \int_0^{\infty} \frac{t^k e^{-t}}{(1+t)^{k+2}} dt \quad (k \geq 0), \text{ és}$$

$$c_k \sim (e\pi)^{\frac{1}{2}} k^{\frac{1}{4}} e^{-2\sqrt{k}} \quad (k \rightarrow \infty).$$

- Szuperkritikus eset. Ha  $\frac{1}{2} < p < 1$ , akkor

$$c_k = \gamma \int_0^1 \frac{t^{k+1}(1-t)^{\beta-1}}{(1-\gamma t)^{1+\beta}} dt \quad (k \geq 0), \text{ és}$$

$$c_k \sim \gamma \beta^{\beta} \Gamma(\beta+1) k^{-\beta} \quad (k \rightarrow \infty).$$

Ahogy az látható, a tétel a Thörnblad modellre vonatkozó három eredményt foglalja össze. Az első tartalmazza az aszimptotikus foksámeloszlás létezését és egyértelműségét, a második az eloszlás tagjainak explicit meghatározását, míg az utolsó ezen tagok aszimptotikus viselkedését. Habár a tételt precízen nem bizonyítjuk, de a bizonyítást felvázoljuk, ugyanis a benne szereplő ötletek és állítások önmagukban is érdekesek, sőt kiindulási alapot adhatnak hasonló modellek vizsgálatánál.

Először nézzük, hogyan juthatunk el az egyértelmű aszimptotikus foksámeloszlás létezéséig. Látható, hogy a modell vizsgálatát a törlések előfordulása teszi nehezebbé a szokásos dinamikusan fejlődő véletlen gráfokénál, mivel ezek erős ingadozást okoznak a hálózat viselkedésében. Például előfordulhat, hogy a véletlenszerűen törlésre választott csúcs fokszáma viszonylag nagy, így a lépés után a foksámokban jelentős változás történik. A probléma kiküszöbölésére felhozott természetes gondolat, hogy a foksámok megfigyelése helyett inkább meghatározott elemszámú klikkek számának arányát tekintsük, ahol az arány szintén, az összes csúcsok számához vett arányt takarja. Ennek következtében egy olyan mérőszámot kapunk, mely az egymást követő lépések során nem módosul olyan intenzíven.

A tétel első részének belátásához szintén fontos összetevő a következő lemma, melyet Backhausz és Móri a [3] cikkben ismertetett. Ennek bizonyítása martingáleméleti módszerek segítségével történik, amit szintén nem részletezünk. A kimondás előtt érdemes még megjegyezni, hogy a lemma általánosabb formában is igaz, sőt hasonlóan megfogalmazható a limesz szuperiorra is, viszont ezekre a jelen esetben nem lesz szükségünk, így csak a [35] cikkbeli alakot ismertetjük.

**4.1. Lemma.** Legyen  $(\mathcal{F}_n)_{n=0}^\infty$  növekvő  $\sigma$ -algebrák sorozata, illetve  $(\xi_n)_{n=0}^\infty$  ehhez adaptált nemnegatív folyamat. Legyenek  $(u_n)_{n=1}^\infty$  és  $(v_n)_{n=1}^\infty$  nemnegatív előrejelezhető folyamatok úgy, hogy  $u_n < n$  teljesül minden  $n \geq 1$  esetén, és  $\lim_{n \rightarrow \infty} u_n = u > 0$  létezik majdnem biztosan. Továbbá tegyük fel, hogy létezik  $\delta > 0$  konstans, melyre

$$E[(\xi_n - \xi_{n-1})^2 | \mathcal{F}_n] = O(n^{1-\delta}) \text{ és } E[\xi_n | \mathcal{F}_{n-1}] \geq (1 - \frac{u_n}{n})\xi_{n-1} + v_n.$$

Ekkor

$$\liminf_{n \rightarrow \infty} v_n \geq v$$

esetén

$$\liminf_{n \rightarrow \infty} \frac{\xi_n}{n} \geq \frac{v}{u+1} \text{ majdnem biztosan.}$$

Így megsejthető, hogy az aszimptotikus fokszámeloszlás bizonyítása azzal folytatódik, hogy a folyamat dinamikáját felhasználva megadjuk a  $k$  csúcsot tartalmazó klikkek várható számát az  $n$ -edik lépésben, feltételezve, hogy a hálózatot ismerjük az  $(n-1)$ -edik időpontban. Ezután a klikkekre átfogalmazott fokszámeloszlást leíró sorozat segítségével rekurzíve definiálunk megszámlálható sok sorozatot, melyekről az előbbi lemma segítségével belátható, hogy mindnek létezik felső korlátja, mely legfeljebb 1 lehet. Ezen felső korlátok természetesen nem mások, mint a megfelelő csúcsszámú klikkek számának arányának időben vett limesz inferiorjai. Így viszont a rekurzív sorozataink mind konvergálnak, és a határértékek szintén legfeljebb az előbbi, megfelelő limesz inferiorok lehetnek, továbbá, a konstrukcióból adódóan megfelelnek a klikkszám eloszlást leíró sorozat tagjaival. Tehát, használva, hogy egy  $k$  darab csúcsot tartalmazó klikkben minden csúcs foka  $k-1$ , a kapott eredmény visszafogalmazható fokszámokra, szóval azt kapjuk, hogy a  $k$  fokú csúcsok arányának időben vett limesz inferiora alulról becsülhető a tételben szereplő fokszámeloszlást leíró sorozattal. Azt, hogy ebből miért következik a határérték megegyezése is a következő, könnyen bizonyítható segédállítás garantálja [35].

**4.1. Állítás.** Tegyük fel, hogy  $(Y_{n,k})_{k=1}^\infty$  nemnegatív valószínűségi változók egy, diszkrét idő  $(n)$  által indexelt, megszámlálható családja, úgy, hogy  $\sum_{k=1}^\infty Y_{n,k} = 1$  teljesül minden  $n \in \mathbb{N}$  esetén. Továbbá legyen  $(b_k)_{k=1}^\infty$  olyan nemnegatív sorozat, melyre

$$\sum_{k=1}^\infty b_k = 1, \text{ és } \liminf_{n \rightarrow \infty} Y_{n,k} \geq b_k \text{ minden } k \in \mathbb{N} \text{ esetén, majdnem biztosan.}$$

Ekkor  $\lim_{n \rightarrow \infty} Y_{n,k} = b_k$  majdnem biztosan.

**Bizonyítás.** Elegendő belátni a  $\limsup_{n \rightarrow \infty} Y_{n,k} \leq b_k$  egyenlőtlenséget minden  $k \geq 1$  esetén, ugyanis ekkor  $b_k \leq \liminf_{n \rightarrow \infty} Y_{n,k} \leq \limsup_{n \rightarrow \infty} Y_{n,k} \leq b_k$  miatt következik az állítás. Ehhez viszont nincs másra szükség, csak a limesz inferior és szuperior általános tulajdonságaira, illetve a Fatou lemmára. Legyen  $l \in \mathbb{N}$  tetszőleges, így

$$\begin{aligned} \limsup_{n \rightarrow \infty} Y_{n,l} &= \limsup_{n \rightarrow \infty} \left( 1 - \sum_{k \geq 1, k \neq l} Y_{n,k} \right) = 1 + \limsup_{n \rightarrow \infty} \left( - \sum_{k \geq 1, k \neq l} Y_{n,k} \right) \\ &= 1 - \liminf_{n \rightarrow \infty} \sum_{k \geq 1, k \neq l} Y_{n,k} \leq 1 - \sum_{k \geq 1, k \neq l} \liminf_{n \rightarrow \infty} Y_{n,k} \\ &\leq 1 - \sum_{k \geq 1, k \neq l} b_k = 1 - (1 - b_l) = b_l. \end{aligned}$$



Ez pedig pont az, amit bizonyítani szerettünk volna. ■

Ennek következtében az első eredmény igazolásához csak két további lépésre van szükség. Az, hogy a kapott sorozat tényleg egy eloszlást definiál, azaz a tagok összege 1, majd az aszimptotikák segítségével lesz bizonyítva. Így a létezést beláttuk, már csak az egyértelműség hiányzik. Ez pedig indirekt úton, teljes indukciót és a sorozat különbségek sorozatát felhasználva könnyen belátható.

Térjünk rá a 4.1. Tételben kapott fokszámeloszlás explicit megadásáról szóló eredmények bizonyításának felvázolására. Itt a szub- és szuperkritikus, illetve a kritikus eseteket külön-külön kezeljük, méghozzá az előbbieket a [35] szerinti gondolatmenettel, míg az utóbbit a [3] cikkben szereplő eljárás segítségével. Előbb viszont érdemes visszaemlékezni azon korábbi megjegyzésünkre, hogy habár a fejlődés lefolyásában a Backhausz-Móri, illetve Thörnblad modell a kritikus esetben eltér, átlagosan ugyanannyi duplikációs és törlési lépést produkálnak. Így, ahogyan az a fentiekből is látható, az előbbieken taglalt aszimptotikus fokszámeloszlást leíró sorozat megegyezik. Ennek következtében a Backhausz és Móri által az egzakt képlet megadásra használt eljárás a Thörnblad modellben szereplő dinamikára is alkalmazható.

A szub- és szuperkritikus esetekben a bizonyítás kiindulási ötlete megszokott a dinamikus fejlődő véletlen gráfok témakörében [32, 14]. Tegyük fel, hogy a  $k$ -fokú csúcsok aránya előáll a következő kifejezésként:

$$c_k = \int_{t_0}^{t_1} t^k v(t) dt = \left[ \frac{t^{k+1}}{k+1} v(t) \right]_{t_0}^{t_1} - \int_{t_0}^{t_1} \frac{t^{k+1}}{k+1} v'(t) dt \text{ minden } k \geq 0 \text{ esetén,}$$

ahol  $t_0 < t_1$  és  $v : [t_0, t_1] \rightarrow \mathbb{R}$  teljesülnek. Ekkor ezt behelyettesítve a fokszámeloszlást megadó formulába azt kapjuk, hogy ennek érvényben maradásához a következő két feltétel teljesülése szükséges:

$$\frac{v'(t)}{v(t)} = \frac{-(1-p)t^2 - pt + p}{t(t-1)((1-p)t - p)} \quad \text{és} \quad \left[ t^{k+1} v(t) (t-1) ((1-p)t - p) \right]_{t_0}^{t_1} = 0.$$

Használva a 4.1. Tételben bevezetett  $\beta = \frac{p}{2p-1}$  és  $\gamma = \frac{1-p}{p}$  jelöléseket, továbbá az előbbi logaritmikus deriváltra vonatkozó feltételt, integrálás segítségével kapjuk:

$$v(t) = Rt(1-t)^{\beta-1}(1-\gamma t)^{-\beta},$$

ahol  $R$  egy pozitív konstans. Ebben a pontban a szubkritikus és a szuperkritikus eset bizonyítása elválik, ugyanis a megfelelő  $t_0$  és  $t_1$  intervallum végpontok megválasztása  $\beta$ -n és  $\gamma$ -n keresztül függ  $p$  értékétől. Itt azt kell szem előtt tartani, hogy teljesüljön a megváltozásra kiírt feltétel, illetve, hogy a logaritmikus derivált nevezője ne legyen nulla. A kapott intervallumok tudatában, az eredeti feltevésbe való visszahelyettesítés után már csak az  $R$  konstans meghatározására van szükségünk. Ehhez a következő, tisztán függvénytan definíciót, illetve állítást kell feleleveníteni, mely habár komplex esetben is megfogalmazható ([1] 2. fejezet), itt csak a valós megfelelőjére szorítkozunk.

**4.1. Definíció.** Jelölje ismét az eltolt faktoriális  $(a)_n = a(a+1)(a+2) \dots (a+n-1)$ , ahol  $n > 0$  és  $(a)_0 = 1$ . Ha  $a, b$  tetszőleges valós számok, illetve  $c$  olyan, hogy  $(c)_n \neq 0$ , akkor az

$${}_2F_1(a, b, c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n n!} x^n \text{ ha } |x| < 1,$$

függvényt hipergeometrikus függvénynek nevezzük.

**4.2. Állítás.** Legyenek  $a, b$  és  $c$  olyan valós számok úgy, hogy  $b > 0$  és  $c > 0$ . Ekkor az előbb definiált hipergeometrikus függvényre

$${}_2F_1(a, b, c; x) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-xt)^{-a} dt \text{ ha } |x| < 1,$$

teljesül.

Így például szuperkritikus esetben, az előző állítás alapján:

$$\begin{aligned} c_k &= \int_0^1 t^k v(t) dt = R \int_0^1 \frac{t^{(k+1)}(1-t)^{\beta-1}}{(1-\gamma t)^{\beta+1}} dt \\ &= R \frac{\Gamma(k+2)\Gamma(\beta)}{\Gamma(\beta+k+2)} {}_2F_1(\beta+1, k+2, \beta+k+2; \gamma), \end{aligned}$$

minden  $k \geq 0$  esetén. Ennek következtében, használva a hipergeometrikus függvény definícióját az  $R = c_0 = \frac{1-p}{p} = \gamma$  összefüggést kapjuk, amiből következik az állítás. Hasonlóan járhatunk el a szubkritikus esetben, ahol a 4.2. Állítás használatához még egy plusz integráltranszformációra lesz szükségünk.

A kritikus esetben a generátorfüggvény használata a fő ötlet, mely megtehető, ugyanis a fokszámoszlás nemnegatív egészeken értelmezett. Jelöljük ezt  $G(z)$ -vel egy  $|z| \leq 1$  helyen. Ekkor ha az eloszlást megadó sorozat tagjait egytől végtelenig összegezzük, akkor a következő differenciálegyenletet kapjuk:

$$(1-z)^2 G'(z) = (3-2z)G(z) - 1, \quad G(0) = c_0.$$

Ennek megoldása a következő kifejezés:

$$G(z) = \frac{c(z)}{(1-z)^2} \exp\left(\frac{z}{1-z}\right), \text{ ahol } c(z) = c_0 - \int_0^z \exp\left(-\frac{y}{1-y}\right) dy.$$

Felhasználva, hogy  $c_k$  a  $k$  fokú csúcsok arányának majdnem biztos határértéke, a Fatou lemma alkalmazásával azt kapjuk, hogy  $G(1) \leq 1$ . Tekintettel arra, hogy az előbbi differenciálegyenletből  $G(z)$ -re kapott összefüggés  $c(z)$ -n kívüli összes tagja végtelenhez tart, ha  $z$ -vel az egyhez tartunk, a következőnek kell teljesülnie:

$$c_0 = \int_0^1 \exp\left(-\frac{y}{1-y}\right) dy.$$

Ha ezt visszahelyettesítjük az előbbi képletbe, akkor megfelelő behelyettesítések és megszokott sorösszegzési trükkök után az alábbi összefüggés írható fel:

$$\begin{aligned} \sum_{k=0}^{\infty} c_k z^k &= G(z) = \int_0^{\infty} \frac{e^{-y}}{(1+(1-z)y)^2} dy = \int_0^{\infty} \sum_{k=0}^{\infty} (k+1) \frac{z^k y^k e^{-y}}{(1+y)^{k+2}} dy \\ &= \sum_{k=0}^{\infty} \left[ \int_0^{\infty} (k+1) \frac{y^k e^{-y}}{(1+y)^{k+2}} dy \right] z^k, \end{aligned}$$

ahol az utolsó egyenlőségnél a Beppo Levi tételt használtuk. Ebből viszont már következik az állítás.

Mielőtt felvázolnánk a 4.1. Tételben szereplő utolsó eredmény bizonyítását, érdemes megnézni, hogy a szub- és szuperkritikus esetekben kapott explicit kifejezések hogyan viselkednek a kritikus érték közelében. Heurisztikusan gondolkozva a következő kapcsolatoknak kell teljesülniük:

$$\begin{aligned} \lim_{p \uparrow \frac{1}{2}} \gamma^{-k-1} \int_0^1 \frac{t^{k+1}(1-t)^{-1-\beta}}{(1-\gamma^{-1}t)^{1-\beta}} dt &= (k+1) \int_0^\infty \frac{t^k e^{-t}}{(1+t)^{k+2}} dt \\ &= \lim_{p \downarrow \frac{1}{2}} \gamma \int_0^1 \frac{t^{k+1}(1-t)^{\beta-1}}{(1-\gamma t)^{1+\beta}} dt, \end{aligned}$$

minden  $k \geq 0$  esetén. Ez viszont teljesülni is fog, ugyanis a szélső integrálokhoz könnyen megadható integrálható majoráns (például  $t^{k+2}$  jó lesz), továbbá ezen integrandusok  $p \rightarrow \frac{1}{2}$  esetén pontonként konvergálnak a  $\frac{t^k}{(1-t)^2} e^{-\frac{t}{1-t}}$  kifejezéshez. Így a Lebesgue-féle majorált konvergenciatételt használva a határértékvételt és az integrálást fel tudjuk cserélni, ahonnan helyettesítéssel és parciális integrálással a kívánt összefüggést kapjuk.

Mindezek után foglaljuk össze a 4.1. Tétel utolsó, aszimptotikákat taglaló állítás belátásához alkalmazott ötleteket. Mielőtt ebbe belekezdenénk, célszerű rápillantani az előbbi explicit formulákra. Látható, hogy a szub- és szuperkritikus esetek formailag jobban hasonlítanak, mint a kritikus bármelyikre. Így, mivel ezen eredményeket fogjuk felhasználni, megint érdemes lesz az egyes eseteket kritikus és nem kritikus csoportokban vizsgálni.

A nem kritikus esetekben az aszimptotikus viselkedés meghatározása az egzakt formulából egy kétoldali becslés segítségével történik, kihasználva, hogy a megadott képletek Béta integrálhoz hasonló alakúak. A felső korlát könnyedén megkapható a nevező minimumával történő becslés segítségével. Ezzel szemben az alsó korlát meghatározásánál egyre kisebb tartományon vett integrálás segítségével lehet eljárni. Így persze, azért, hogy ekkor is a Béta-függvény kellemes tulajdonságait tudjuk használni, minimális korrekcióra van szükség, ami, ha egyre nagyobb fokszámokat vizsgálunk (azaz  $k \rightarrow \infty$ ), akkor elenyésző lesz. A kapott két korlát megegyezik, így szub- és szuperkritikus esetekben beláttuk az állítást.

A hátralévő kritikus esetben az előbbi Béta integrálos megfeleltetés nem működik, így máshogyan kell eljárjunk. Jelölje  $f(y)$  az explicit képletben szereplő integrandus logaritmusát,  $y_k$  ezen függvény maximumhelyét, továbbá legyen

$$t = t(y) = y_k^{-\frac{1}{2}}(y - y_k)$$

segédváltozó és

$$g(t) = f(y) - f(y_k)$$

segédfüggvény. Ekkor a maximumhelyről analízisbeli gondolatmenettel belátható a következő:

$$y_k = -\frac{3}{2} + \sqrt{k + \frac{9}{4}} = \sqrt{k} - \frac{3}{2} + o(1).$$

Sőt, felhasználva az  $f(y)$  függvény második és harmadik deriváltját, a  $g(t)$ -re a következő két összefüggés látható:

$$g(t) \leq \frac{y_k}{2} f''(y_k) t^2 = a_k t^2 \text{ ha } y_k^{\frac{1}{2}} \leq t \leq 0, \text{ azaz } y \leq y_k,$$

ahol  $a_k \rightarrow -1$ , amint  $k \rightarrow \infty$ , illetve,

$$g(t) \leq \left( \frac{y_k}{2} f''(y_k) + \frac{y_k^2}{12} f'''(y_k) \right) t^2 = b_k t^2 \text{ ha } 0 \geq t \geq \frac{1}{2} y_k^{\frac{1}{2}}, \text{ azaz } y_k \geq y \geq \frac{3}{2} y_k,$$

ahol  $b_k \rightarrow -\frac{1}{2}$ , amint  $k \rightarrow \infty$ . Így integrálbehelyettesítés után használható lesz a Lebesgue-féle majorált konvergenciatétel, amivel a következőt kapjuk:

$$\int_0^{\frac{3}{2} y_k} e^{f(y)} dy \sim y_k^{\frac{1}{2}} \exp(f(y_k)) \int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi} y_k^{\frac{1}{2}} \exp(f(y_k)).$$

Természetesen szükségünk van a  $[\frac{3}{2} y_k, \infty]$  intervallumon vett integrál becslésére is. Ehhez először jegyezzük meg, hogy minimális átalakítás után  $f(y)$  az alábbi alakban írható fel:

$$f(y) = -2 \log y - (d+2) \log \left( \frac{y+1}{y} \right) - y.$$

Ennek és a számtani és mértani közepek közti egyenlőtlenségnek segítségével a kívánt becslés megadható, méghozzá a következőképp:

$$\int_{\frac{3}{2} y_k}^{\infty} e^{f(y)} dy \leq (2y_k)^{-2} \exp \left( -\sqrt{2k} + \frac{1}{2} - \frac{3}{4} y_k \right) = o(y_k^{-2} \exp(-2\sqrt{d})).$$

Emellett ha felhasználjuk a  $k$  és az  $y_k$  közti kapcsolatot, akkor az  $f(y_k)$  maximum érték is megkapható, amelyet az eddigiekbe visszahelyettesítve pont a kívánt eredményt kapjuk, befejezve evvel az aszimptotikus viselkedésre vonatkozó eredmények bizonyításának vázlatát.

Tehát ezen aszimptotikákra kapott eredmények azt mutatják, hogy az aszimptotikus fokszámeloszlás szubkritikus esetben exponenciálisan, míg a superkritikusban pedig polinomiálisan csökken. A kettő közti kritikus esetben az elvárásainknak megfelelően az exponenciálisnál lassabb, a polinomiálisnál gyorsabb lecsengésű fokszámeloszlás figyelhető meg.

Ahhoz, hogy a bizonyítás vázlatát teljessé tehesük, még szükségünk van egy korábbi állításra, melyet a következő lemmában fogalmazunk meg újra.

**4.2. Lemma.** *Legyen  $0 < p < 1$ . Ekkor a*

$$c_0 = \frac{1-p}{1+p}(1+c_1); \quad c_k = \frac{k+1}{k+1+p}(pc_{k-1} + (1-p)c_{k+1}) \quad (k \geq 1),$$

*segítségével megadott  $(c_k)_{k=1}^{\infty}$  sorozat eloszlást definiál, azaz  $\sum_{k=0}^{\infty} c_k = 1$  teljesül.*

**Bizonyítás.** A tagokra vonatkozó összefüggést átrendezve és 0-tól  $n$ -ig összegezve a következőt kapjuk:

$$\begin{aligned}
\sum_{k=0}^n (k+1+p)c_k &= (1+p)c_0 + \sum_{k=1}^n (k+1+p)c_k \\
&= (1-p)(1+c_1) + \sum_{k=1}^n (k+1)[pc_{k-1} + (1-p)c_{k+1}] \\
&= (1-p)(1+c_1) + p \sum_{k=0}^n (k+2)c_k + (1-p) \sum_{k=0}^n kc_k - (1-p)c_1 \\
&\quad - p(n+2)c_n + (1-p)(n+1)c_{n+1} \\
&= (1-p) + \sum_{k=0}^n kc_k + 2p \sum_{k=0}^n c_k - p(n+2)c_n + (1-p)(n+1)c_{n+1}.
\end{aligned}$$

Így egyszerűsítés és rendezés után a

$$\sum_{k=0}^n c_k = 1 + \frac{1}{1-p} \left[ -p(n+2)c_n + (1-p)(n+1)c_{n+1} \right]$$

egyenlőség írható fel. Ekkor viszont felhasználva az aszimptotikák által adott tényt, miszerint tetszőleges  $0 < p < 1$  esetén  $c_k = o(k^{-1})$  teljesül, az előbbi egyenlőségben  $n$ -nel végtelenhez tartva pont az állítást kapjuk. ■

Habár ezzel az alfejezet fő tételét vázlatosan beláttuk, a korábbiak szerint kritikus esetben a coupling módszer segítségével ez kiterjeszhető a valóságot jobban reprezentáló Backhausz-Móri modellre is. Ez a hazai szakirodalomban párosításként vagy csatolásként emlegetett eljárás egy megszokott és elegáns bizonyítási technika a valószínűségi számításban, mely különböző valószínűségi változók, illetve véletlen folyamatok, valamilyen értelemben vett összekapcsolásán alapul. Elterjedtségéből adódóan a módszert, melynek alapjai elsajátíthatóak a [28, 34] könyvekből, itt nem ismertetjük, csak a konstrukciót adjuk meg. Ezen felül megjegyezzük, hogy habár a kritikus jelzőt elhagyjuk, a következőkben végig erről az esetről lesz szó.

Felelevenítve, hogy ha a Backhausz-Móri féle modelltől kivesszük az új csúcsból induló élek megtartására vonatkozó megkötést, akkor az így kapott verzió aszimptotikusan azonos tulajdonságokat produkál mint, a Thörnblad modell. Ennek következtében elegendő a plusz megkötések nélküli Backhausz-Móri dinamikával fejlődő gráfot összekapcsolni az eredeti modellel. Ezen előbbit a továbbiakban nevezzük második verziónak.

A két különböző verzió párosítása a módosításra kiválasztott csúcsok segítségével történik. Ezt a következőképp értjük. Ha a két eltérő variációhoz tartozó folyamatokat egyidejűleg vizsgáljuk, akkor feltehető, hogy a duplikációra és a törlésre választott csúcsok az összes lépés során azonosak. Ez azért lehetséges, mivel ezen pontokat mindkét variációban egymástól függetlenül, az egyenletes eloszlás szerint választjuk.

Ebben az esetben a második változatban elhagyott, korábban említett megkötés miatt a két folyamat lefolyásában csak a törülő lépések esetén lesz különbség, ugyanis az eredeti dinamikában egy újonnan behúzott élet ugyanabban a lépésben nem

törölhetünk. Így a következő ötletet használhatjuk. Ha egy él a második verzióban törölnék, de az eredetiben a megkötés miatt nem lehet, akkor azt színezzük pirosra. Duplikáció során ezen piros éleket szintén piros élként másoljuk, ami alatt azt értjük, hogy ha valamely él piros a duplikált csúcs és szomszédja között, akkor az új csúcsból ebbe a szomszédba menő él is legyen az. A többi él duplikáció során nem tud átszíneződni.

Ezen élszínezést felhasználva a csúcsokra is bevezetünk egy hasonló színezési sémát. Először színezzünk pirosra a piros élekre illeszkedő csúcsokat. Ezután járjunk el a következőképp. Egy duplikáció során az új csúcs csak akkor legyen piros, ha a másolt is az volt. Ha színezetlen csúcsot duplikáltunk, akkor a kapott új csúcs is az marad egyelőre. Viszont ez a törlés során megváltozhat, ugyanis ha egy újonnan behúzott él is törölni kellene akkor a korábbiak szerint azt pirosra színezzük. Ebben az esetben ennek mindkét végpontját átszínezzük. Természetesen ha ez a helyzet nem áll elő, azaz a törlés során a kiválasztott csúcs minden élet töröljük, akkor ez szintelen lesz. Azonban előfordulhat, hogy egy olyan piros csúcs szomszédját választjuk törlés során, melyből egy darab piros él indul ki, ebben az esetben ezt nem színezzük át. Tehát előfordulhat, hogy egy olyan csúcs lesz piros, amelynek nincs piros éle. Mindent összevetve az alapfelállítás, miszerint színezett él nem indulhat nem piros csúcsból, megmarad.

A kapott színezés következménye, hogy a nem színezett él pont a második verzió realizációját adják az eredeti folyamatban, ugyanis a piros él megsemmisítés nélkül mind törölnének, így a másolatuk sem szerepelne a gráfban. Tehát a két különböző változatban a piros csúcsok fokszáma lesz egyedül, ami eltér, aminek következtében elegendő megmutatni azt, hogy ezek hányada majdnem biztosan nullához tart a lépésszám növelése mellett. Ennek bizonyítását nem részletezzük, de martingálelméleti módszerek és a Kronecker lemma segítségével belátható.

Érdemes megjegyezni, hogy annak következményeképp, hogy a kettőt összekapcsoltuk, az eredeti modell további jó tulajdonságai aszimptotikusan érvényesek lesznek a Backhausz-Móri modell esetén. Például, mivel a Thörnblad modell klaszterezettségi együtthatója konstans 1, megkapható, hogy az utóbbié az 1-hez konvergál. Így, a Backhausz-Móri modell ezen mérőszám alapján közelebb áll a valóságban megfigyelhető biológiai hálózatokhoz, mint a korábban taglalt parciális duplikációval ellátott társai.

Tehát összegezve, az szakaszban végigvettük a Thörnblad modell aszimptotikus fokszámeloszlására vonatkozó 4.1. Tétel összes állításának belátására felhasznált ötletet, továbbá a kritikus esetben ezen eredményt a valóságot jobban reprezentáló Backhausz-Móri modellre is átvittük. Ezek tudatában térhetünk rá a maximális fokszerű csúcsról szóló eredmények vizsgálatára.

### 4.3. Maximális fokszerű

A következő hálózatok esetén gyakran vizsgált probléma, hogy mit lehet mondani a maximális fokszerűről egy adott időpillanatban. Jelen alfejezetben ezt és az ehhez kapcsolódó kérdéskört vizsgáljuk. A [4] cikkben szereplő úton haladva először megmutatjuk, hogy miként alakul egy rögzített csúcs fokszerűje az idő múlásával,

majd ezt felhasználva majdnem biztos korlátot adunk a legnagyobb fokszámmra. Hasonlóan a korábbiakhoz, a most következő tételeket és állításokat sem bizonyítjuk precízen.

A továbbiakhoz szükségünk lesz a Thörnblad modell korábban már említett folytonosított, illetve valamilyen értelemben vett lelassított változataira. Az előbbit, melyet mostantól nevezzük folytonos Thörnblad modellnek, már korábban ismertettük, míg utóbbit a következőképp definiálhatjuk.

**Lassított Thörnblad modell.** Legyen  $0 < p < 1$  fixált érték és jelölje  $N_{n-1}$  a csúcsok számát az  $(n - 1)$ -edik lépés után. A folyamatot az  $n = 0$  pillanatban a  $\mathbb{G}_0$ -lal jelölt izolált pontból indítva, diszkrét időben előrehaladva az  $n \geq 1$  időpontban a  $\mathbb{G}_n$  gráfot a következő szabály szerint képezzük  $\mathbb{G}_{n-1}$ -ből.

- (1) Vagy  $1 - \frac{N_{n-1}}{n}$  valószínűséggel  $\mathbb{G}_n$  azonos lesz a  $\mathbb{G}_{n-1}$  gráffal, azaz nem módosítunk semmit.
- (2) Vagy  $\frac{N_{n-1}}{n}$  valószínűséggel végrehajtunk egy lépést az eredeti Thörnblad modell dinamikája szerint.

Itt érdemes visszaautalni arra a korábban említett tényre, hogy azon pillanatokban, ahol valami történik a folytonos és az eredeti folyamat azonosak. Látható, hogy ugyanez elmondható az előbb megadott lassított modell esetén is. Ennek következménye, hogy az egyes bizonyítások során használhatjuk az adott körülmények között legkézenfekvőbb verziót. Az alfejezetben fontos lesz számunkra a következő, általunk nem bizonyított lemma.

**4.3. Lemma.** *Jelölje  $N_n$  a csúcsok számát  $n$  lépés után a lassított és az eredeti, továbbá  $N(t)$  a gráf méretét a folytonos Thörnblad modellben. Ekkor a következők teljesülnek:*

- (1)  $N_n \sim pn$  majdnem biztosan az eredeti változatban,
- (2)  $N_n \sim \zeta n^p$  majdnem biztosan a lassított változatban, ahol  $\zeta$  egy pozitív valószínűségi változó,
- (3)  $N(t) \sim \eta e^{pt}$  majdnem biztosan a folytonos változatban, ahol  $\eta$  egy pozitív valószínűségi változó.

Ennek ereje abban rejlik, hogy segítségével egyes eredményeinket a bizonyításukban, illetve kimondásukban használt verziótól függetlenül az eredeti Thörnblad modellre is át tudjuk vinni. Ennek következményeképp a soron lévő tételeket és állításokat majd minden esetben a bizonyításukban használt változatnak megfelelően mondjuk ki, az eredeti modellre való visszatranszformálást az olvasóra hagyjuk.

Ahhoz, hogy az adott pillanatbeli maximális fokszámot becsülni tudjuk, először azt kell megvizsgálnunk, hogy egy rögzített csúcs fokszáma hogyan változik a folyamat során. Ennél a kérdésnél talán legkézenfekvőbb a folytonos változatot használni, úgyhogy a következő állítások esetén eszerint járunk el. Ezen felül szükségünk lesz a csúcsok egy címkézésére, amit a születési sorrendjük szerinti számozással érünk el. Vezessük be a következő jelölést.

**Jelölés.** Az előbbi címkézés mellett jelölje  $d_i(t)$  az  $i = 1, 2, \dots$  csúcs fokszámát a  $t$  pillanatban a folytonos modellben.

Ezt felhasználva kimondhatjuk a problémakörbe tartozó első állításunkat, amely megadja egy fixált csúcs aszimptotikus fokszámeloszlását.

**4.3. Állítás.** *Tekintsük a folytonos Thörnblad modellt és jelölje  $\gamma$  az  $\frac{1-p}{p}$  értéket  $0 < p < 1$  rögzített paraméter mellett. Ekkor*

$$\lim_{t \rightarrow \infty} P(d_i(t) = k) = q_k, \quad k = 0, 1, 2, \dots,$$

ahol

$$q_0 = \gamma(1 - c_0), \quad q_k = c_{k-1} - \gamma c_k, \quad (k \geq 1),$$

és a  $(c_k)_{k=0}^\infty$  a 4.1 Tételben definiált sorozat. Továbbá  $q_k > 0$  minden  $k \geq 0$  esetén és  $\sum_{k=0}^\infty q_k = 1$ .

**Bizonyítás.** Először belátjuk a rekurziót. Ehhez rögzítsünk egy csúcsot. Ekkor ez eddigiek mellett világos, hogy fokszámának változása egy folytonos idejű Markov-lánccal írható le, amely infinitezimális generátora a folyamat dinamikájából adódóan a következő:

$$\mathbf{Q} = \begin{bmatrix} 1-p & 0 & \cdots \\ 1-p & 0 & \cdots \\ 1-p & 0 & \cdots \\ 1-p & 0 & \cdots \\ \vdots & \vdots & \end{bmatrix} + \begin{bmatrix} -1 & p & 0 & 0 & 0 & \cdots \\ 1-p & -2 & 2p & 0 & 0 & \cdots \\ 0 & 2(1-p) & -3 & 3p & 0 & \cdots \\ 0 & 0 & 3(1-p) & -4 & 3p & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}.$$

Itt a mátrix összegre bontása csak áttekinthetőség érdekében történt. Mivel a folyamat pozitív rekurrens, tehát létezik stacionárius eloszlás  $q = (q_0, q_1, \dots)$ , melyről tudjuk, hogy a  $q\mathbf{Q} = 0$  egyenletrendszer egyértelmű megoldása. Ezt felírva azt kapjuk, hogy

$$q_0 = (1-p)(1+q_1); \quad q_k = \frac{kpq_{k-1} + (k+1)(1-p)q_{k+1}}{k+1} \quad (k \geq 1).$$

Ennek megoldása viszont a 4.1. Tétel miatt pont az állításban szereplő összefüggés, azaz a rekurziót beláttuk.

Emellett szükségünk van még a pozitivitásra, illetve arra, hogy a sorozatot összegezve 1-et kapjunk. Az utóbbi könnyen látható, ugyanis, szintén a 4.1. Tételt használva,

$$\sum_{k=0}^{\infty} q_k = \gamma(1 - c_0) + \sum_{k=1}^{\infty} c_{k-1} - \gamma c_k = \gamma + (1 - \gamma) \sum_{k=0}^{\infty} c_k = \gamma + 1 - \gamma = 1$$

teljesül. Míg az előbbihez két dolgot kell észrevennünk. Ha  $k = 0$ , akkor felhasználva megint a  $c_k$  együtthatók összegére vonatkozó állítást, felírható, hogy

$$q_0 = \gamma(1 - c_0) = \gamma \left( \sum_{n=0}^{\infty} c_n - c_0 \right) = \gamma \sum_{n=1}^{\infty} c_n > 0,$$



ahol az egyenlőtlenség megint csak a 4.1. Tétel miatt teljesül. Viszont ha  $k \geq 1$ , akkor a  $c_k$  együtthatók integrálalakjaiból az alábbiak szerint levezethető egy explicit képlet a  $q_k$ -kra a különböző esetekben, amelyből az állítás következik.

Szuperkritikus esetben:

$$\begin{aligned} q_k &= \gamma \int_0^1 \frac{t^k(1-t)^{\beta-1}}{(1-\gamma t)^{1+\beta}} dt - \gamma^2 \int_0^1 \frac{t^{k+1}(1-t)^{\beta-1}}{(1-\gamma t)^{1+\beta}} dt \\ &= \gamma \int_0^1 \frac{t^k(1-t)^{\beta-1}}{(1-\gamma t)^\beta} \left( \frac{1-\gamma t}{1-\gamma t} \right) dt = \gamma \int_0^1 \frac{t^k(1-t)^{\beta-1}}{(1-\gamma t)^\beta} dt. \end{aligned}$$

Kritikus esetben:

$$\begin{aligned} q_k &= k \int_0^\infty \frac{t^{k-1}e^{-t}}{(1+t)^{k+1}} dt - (k+1) \int_0^\infty \frac{t^k e^{-t}}{(1+t)^{k+2}} dt \\ &= k \int_0^\infty \frac{t^{k-1}e^{-t}}{(1+t)^{k+1}} dt + \left[ \frac{t^k e^{-t}}{(1+t)^{k+1}} \right]_0^\infty - \int_0^\infty \frac{kt^{k-1}e^{-t} - t^k e^{-t}}{(1+t)^{k+1}} dt \\ &= \int_0^\infty \frac{t^k e^{-t}}{(1+t)^{k+1}} dt. \end{aligned}$$

Szubkritikus esetben:

$$\begin{aligned} q_k &= \gamma^{-k} \int_0^1 \frac{t^k(1-t)^{-1-\beta}}{(1-\gamma^{-1}t)^{1-\beta}} dt - \gamma^{-k} \int_0^1 \frac{t^{k+1}(1-t)^{-1-\beta}}{(1-\gamma^{-1}t)^{1-\beta}} dt \\ &= \gamma^{-k} \int_0^1 \frac{t^k(1-t)^{-\beta}}{(1-\gamma^{-1}t)^{1-\beta}} \left( \frac{1-t}{1-t} \right) dt = \gamma^{-k} \int_0^1 \frac{t^k(1-t)^{-\beta}}{(1-\gamma^{-1}t)^{1-\beta}} dt \end{aligned}$$

Így az állítást beláttuk. ■

Ha összevetjük az előbbi eredményt a korábban szerepelt 4.1. Tétellel, akkor azt az érdekes tényt vehetjük észre, hogy habár a foksámok határeloszlása minden csúcs esetén azonos, a gráf aszimptotikus foksámeloszlása mégsem egyezik meg ezzel. Ugyan ez elsöre ellentmondásnak tűnhet, természetesen nem az. Ahhoz, hogy a gráf foksámeloszlása egybeessen az előbbi állításban szereplőkkel, az kell, hogy a foksámok egymástól függetlenek és azonos eloszlásúak legyenek, amely feltételek közül modellünkben egyik sem teljesül. Például, jelen esetben, ha egy csúcsból  $k$  darab él indul, akkor ugyanez mondható el mind a  $k$  szomszédjára is. Tehát egy magas foksámú csúcs létezése megannyi ugyanekkora foksámmal rendelkező csúcs jelenlétét vonja maga után. Ezzel szemben, ha feltesszük a függetlenséget, akkor a sok szomszédal rendelkező csúcsok száma Poisson eloszlást kell kövessen [6].

Ezen tulajdonság jobb megértéséhez érdemes megfigyelni egy tetszőleges csúcs foksámának változását a folyamat során. Ehhez emlékezzünk vissza arra a korábban említett tényre, hogy egy csúcs egyenletes eloszlás szerinti választása ugyanaz, mint ha a gráfból méretarányosan kijelölnénk egy klikket, majd annak valamelyik csúcsát vennénk. Ennek következménye, hogy minél magasabb egy csúcs foksáma, annál gyorsabban kerül kiválasztásra egy lépés során, azaz törlés esetén hamarabb lesz izolált. Ezáltal elmondható, hogy annak valószínűsége, hogy a vizsgált csúcs foksáma megegyezik  $k$ -val az adott pillanatban, kisebb mint az ekkor a gráfban található  $k$  fokú csúcsok aránya.

Visszatérve a maximális foksámra vonatkozó eredmények tárgyalására, könnyen látható, hogy a folyamat során minden csúcsot végtelen sokszor kiválasztunk a

lépések során. Ennek következtében a törlések jelenléte miatt egy-egy csúcs végtelen sokszor válik izolálttá. A kérdés az, hogy mi mondható el a kimagaslóan nagy fokszámról. Ezt válaszolja meg a következő tétel, melyet szintén csak vázlatosan bizonyítunk.

**4.2. Tétel.** *Legyen  $0 < p < 1$  rögzített érték és használjuk a korábbi  $\beta = \frac{p}{2p-1}$ , illetve  $\gamma = \frac{1-p}{p}$  jelöléseket. Ekkor egy fixált csúcs fokszámról az alábbi mondható el a különböző esetekben.*

- *Szubkritikus eset.*  $\limsup_{t \rightarrow \infty} \frac{d_i(t)}{\log \log N(t)} = \frac{1}{\log \gamma}.$
- *Kritikus eset.*  $\limsup_{t \rightarrow \infty} \frac{d_i(t)}{(\log \log N(t))^2} = 1.$
- *Szuperkritikus eset.*  $\limsup_{t \rightarrow \infty} \frac{\log d_i(t)}{\log \log N(t)} = \frac{1}{1-\beta}.$

A tétel levezetése két lépésből áll. Először megmutatjuk, hogy két egymást követő törlés között a foksám mekkorára tud megnőni, majd ezt felhasználva, a jól ismert Borel-Cantelli lemmák segítségével kiterjesztjük az egész folyamatra. Ehhez vezessük be az  $s_i(r)$  jelölést, ami annak a valószínűségét takarja, hogy egy  $i$  fokú csúcs valamikor a legközelebbi törlési lépés előtt lesz-e legalább  $r$  foksámú. Ekkor erre nyilvánvalóan teljesül a következő összefüggés:

$$\begin{aligned} s_0(r) &= ps_1(r); \\ s_i(r) &= \frac{(i+1)ps_{i+1}(r) + i(1-p)s_{i-1}(r)}{i+1}, \quad i = 1, 2, \dots, r-1; \\ s_r(r) &= 1. \end{aligned}$$

Ebből viszont az is jól látható, hogy az  $i < r$  feltétel mellett az  $\frac{s_i(r)}{s_0(r)}$  hányados nem függ a választott  $r$  értékétől, tehát ha ezt a hányadost  $a_i$ -vel jelöljük, akkor a korábbiak szerint, az átrendezéssel kapott  $s_0(r) = \frac{1}{a_r}$  kifejezésre vagyunk kíváncsiak. Ehhez használjuk fel az alábbi,  $(a_r)_{r \geq 0}$  sorozatra vonatkozó rekurziót, amely az eddigiek alapján könnyen ellenőrizhető.

$$a_0 = 1; \quad a_1 = \frac{1}{p}; \quad a_i = pa_{i+1} + \frac{i}{i+1}(1-p)a_{i-1}, \quad i = 1, 2, \dots$$

Ettől a ponttól a bizonyítás menete, ahogyan az elvárható, a különböző esetekben más és más.

Kritikus esetben az előbbi  $(a_r)_{r \geq 0}$  sorozatra adott összefüggés leegyszerűsödik. Sőt, ez olyan módon történik, hogy összekapcsolhatóvá válik a korábban említett hipergeometrikus függvények egy másfajta speciális esetével, az úgynevezett Laguerre polinomokkal, melyekről több információ szerezhető a [33] könyvből. Ezt a kapcsolatot és egy Laguerre polinomokról ismert állítást felhasználva, kis számolás után megkapható, hogy ezen esetben

$$s_0(r) = \frac{1}{a_r} \sim 2\sqrt{e\pi} r^{\frac{1}{4}} e^{-2\sqrt{r}}$$

teljesül.

A megmaradt két esetből elegendő lesz csak a szuperkritikus vizsgálni, ugyanis a szubkritikus könnyen visszavezethető erre. Vezessük be az  $a'_r = \left(\frac{p}{1-p}\right)^r a_r = \gamma^{-r} a_r$

jelölést. Ekkor  $a'_r$  szintén kielégíti az  $(a_r)_{r \geq 0}$  rekurziót, abban az esetben, ha  $p$ -t  $(1-p)$ -vel helyettesítjük. Tehát ezen behelyettesítés ténylegesen visszavezeti a szubkritikus esetet a szuperkritikusra, nem elfelejtve azt, hogy ekkor a  $\beta$  is  $(1-\beta)$ -ra változik.

Vegyük tehát a szuperkritikus esetet. Ekkor az  $(a_r)_{r \geq 0}$  sorozat generátorfüggvényéről, közönséges differenciálegyenletes fejtegetéssel belátható, hogy

$$G(z) = \sum_{r=0}^{\infty} a_r z^r = (1-z)^{-\beta} (1-\gamma z)^{\beta-1}.$$

Ugyanakkor használva a binomiális sorokra ismert összefüggéseket, az előbbi továbbírható az alábbi formában

$$G(z) = \sum_{r=0}^{\infty} a_r z^r = \sum_{r=0}^{\infty} z^r (-1)^r \sum_{i=0}^r \binom{\beta-1}{i} \binom{-\beta}{r-i} \gamma^i,$$

amiből pedig az egyenlő együtthatók módszere miatt

$$a_r = (-1)^r \sum_{i=0}^r \binom{\beta-1}{i} \binom{-\beta}{r-i} \gamma^i$$

következik. Ekkor viszont használva az általánosított binomiális együtthatókra vonatkozó ismert tulajdonságokat, azt kapjuk, hogy

$$(-1)^r r^{1-\beta} \binom{-\beta}{r} = \frac{r^{1-\beta} \Gamma(r+\beta)}{\Gamma(r+1) \Gamma(\beta)},$$

ahol a jobb oldal az  $\frac{1}{\Gamma(\beta)}$  értékhez tart, ha  $r$ -rel a végtelenhez tartunk. Ezáltal az  $r^{1-\beta} a_r$  együtthatókat leíró sort dominálni tudjuk egy konvergens sorral, aminek következtében az alábbi írható fel, használva a Lebesgue-féle majorált konvergencia tételt és a binomiális sorok korábban már alkalmazott összegzési képletét:

$$\begin{aligned} \lim_{r \rightarrow \infty} r^{1-\beta} a_r &= \sum_{i=0}^{\infty} \lim_{r \rightarrow \infty} (-1)^r r^{1-\beta} \binom{\beta-1}{i} \binom{-\beta}{r-i} \gamma^i = \frac{1}{\Gamma(\beta)} \sum_{i=1}^{\infty} \binom{\beta-1}{i} (-\gamma)^i \\ &= \frac{(1-\gamma)^{\beta-1}}{\Gamma(\beta)} = \frac{\beta^{1-\beta}}{\Gamma(\beta)}. \end{aligned}$$

Ebből pedig visszahelyettesítéssel a számunkra szükséges,

$$s_0(r) = \frac{1}{a_r} \sim \beta^{\beta-1} \Gamma(\beta) r^{1-\beta}$$

összefüggést kapjuk a szuperkritikus esetben. Továbbá alkalmazva a korábban taglalt visszavezetési ötletet, a megfelelő behelyettesítés után a szubkritikus eset is felírható az alábbiak szerint:

$$s_0(r) = \frac{1}{\gamma^r a'_r} \sim (1-\beta)^{-\beta} \Gamma(1-\beta) r^{\beta} \gamma^{-r}.$$

Ezek után rátérhetünk a bizonyítás második részének áttekintésére. Tekintettel a folytonosított modellben szereplő exponenciális óras megfogalmazásra, a  $t$

időpillanatig bekövetkező, két törlés által közrefogott intervallumok száma megegyezik  $(1-p)t(1+o(1))$ -gyel, ami, felhasználva a korábbi 4.3. Lemmában magába foglalt eredményt, aszimptotikusan  $\gamma \log N(t)$ -vel egyenlő. Ennek következtében az általunk keresett  $\max\{d_i(u)|u \leq t\}$  mennyiség aszimptotikusan meg fog egyezni  $(1+o(1))\gamma \log N(t)$  darab egymástól független,  $P(\xi \geq r) = s_0(r)$  eloszlású valószínűségi változó maximumával. Ez viszont a Borel-Cantelli lemmák alkalmazásával a következő eredményeket adja az egyes esetekben.

$$\max\{d_i(u)|u \leq t\} \sim \frac{\log \log N(t)}{\log \gamma}$$

a szubkritikus,

$$\max\{d_i(u)|u \leq t\} \sim \log^2 t \sim (\log \log N(t))^2$$

a szuperkritikus és

$$\log \max\{d_i(u)|u \leq t\} \sim \frac{\log \log N(t)}{\beta - 1}$$

a kritikus esetben, melyek pontosan a tételben jelzett értékeknek felelnek meg.

Habár alapvetően érdekes tény, hogy egy-egy rögzített csúcs fokszáma mennyire tud megnőni a folyamat során, a szakasz fő célja, ahogyan az már korábban említésre került, az, hogy egy adott pillanatban mekkora az egész gráfon vett maximális fokszám. Mielőtt viszont az ezen eredményeket összefoglaló fő tételt kimondanánk, érdemes megjegyezni, egy olyan, az elkövetkezőkben felhasználandó állítást, mely a 4.3. Állításban szereplő  $q_k$  ( $k \geq 0$ ) határeloszlást leíró együtthatók farokeloszlásának aszimptotikáját adja meg. Ennek bizonyítása az eddigiekből könnyedén megkapható, így nem részletezzük.

**4.4. Állítás.** *Legyen  $0 < p < 1$  rögzített érték és használjuk a korábbi  $\beta = \frac{p}{2p-1}$ , illetve  $\gamma = \frac{1-p}{p}$  jelöléseket. Ekkor egy fixált csúcs fokszámának farokeloszlásának aszimptotikus viselkedéséről az egyes esetekben következők mondhatóak el.*

- *Szubkritikus eset.*  $q_k + q_{k+1} + \dots \sim (1-\beta)^{2-\beta} \Gamma(1-\beta) k^{\beta-1} \gamma^{-k}$ , ha  $k \rightarrow \infty$ .
- *Kritikus eset.*  $q_k + q_{k+1} + \dots \sim (e\pi)^{\frac{1}{2}} k^{\frac{1}{4}} e^{-2\sqrt{k}}$ , ha  $k \rightarrow \infty$ .
- *Szuperkritikus eset.*  $q_k + q_{k+1} + \dots \sim \gamma \beta^\beta \Gamma(\beta-1) k^{1-\beta}$ , ha  $k \rightarrow \infty$ .

ahol a  $q_k$  ( $k \geq 0$ ) együtthatók az előbbi 4.3. Állításban szereplő fokszámeloszlást leírókkal azonosak.

A legmagasabb fokszám kérdéskörét taglaló tétel kimondásához és bizonyításának felvázolásához érdemes bevezetni az alábbi jelöléseket.

**Jelölés.** Jelölje  $M_n$  a gráf maximális fokszámú csúcsának fokszámát  $n \geq 0$  lépés után az eredeti Thörnblad modellben.

**Jelölés.** Használva a korábban ismertetett  $N_n$  jelölést az  $n$  lépés után a gráfban található csúcsok számára, jelölje  $d_i(n)$  a születési sorrendben  $i$ -edik ( $i = 1, 2, \dots, N_n$ ) fokszámát az  $n$ -edik lépés után.

**Jelölés.** Az előbbieket mellett legyen

$$S_n(r) = \sum_{i=1}^{N_n} \binom{d_i(n)}{r},$$

ahol  $r$  tetszőleges nemnegatív egész szám.

Térjünk rá tehát az alfejezet legfontosabb eredményére, melyet bár nem az eredeti modell segítségével bizonyítunk, mégis ebben az alakban írunk fel, tekintettel a korábbi, illetve soron lévő tételekkel való könnyű összeegyeztethetőségre.

**4.3. Tétel.** *Legyen  $0 < p < 1$  rögzített érték és használjuk az előbbi, illetve korábbi  $\beta = \frac{p}{2^{p-1}}$  és  $\gamma = \frac{1-p}{p}$  jelöléseinket. Ekkor az  $n$ -edik lépés utáni maximális fokszámról az alábbi mondható el a különböző esetekben.*

- *Szubkritikus eset.*  $\frac{1-p}{\log \gamma} \leq \liminf_{n \rightarrow \infty} \frac{M_n}{\log N_n} \leq \limsup_{n \rightarrow \infty} \frac{M_n}{\log N_n} \leq \frac{1+p}{p \log \gamma}.$
- *Kritikus eset.*  $\frac{1}{16} \leq \liminf_{n \rightarrow \infty} \frac{M_n}{\log^2 N_n} \leq \limsup_{n \rightarrow \infty} \frac{M_n}{\log^2 N_n} \leq \frac{9}{4}.$
- *Szuperkritikus eset.*  $\frac{p}{\beta} \leq \liminf_{n \rightarrow \infty} \frac{\log M_n}{\log N_n} \leq \limsup_{n \rightarrow \infty} \frac{\log M_n}{\log N_n} \leq \frac{1}{\beta}.$

Természetesen az, hogy ezen korlátok gyengíthetőek-e tovább vagy sem, nem feltétlenül derül ki a tétel állításából. Ehhez érdemes végiggondolni, mégis mennyi az a maximális fokszám, amit még a modell megengedhet. A 4.1. Tételben szereplők miatt nyilvánvaló, hogy  $M_n$  tart a végtelenhez, amint  $n$  tart végtelenhez. Sőt, az ugyanezen tételben bizonyított  $(c_k)_{k=1}^{\infty}$  aszimptotikus fokszámeloszlást felhasználva  $M_n$  a következőképp karakterizálható. Tekintettel arra, hogy a  $c_k$  érték a  $k$ -ad fokú csúcsok arányának majdnem biztos határértéke, így nagyságrendben csak olyan érték jöhet szóba, mint azon maximális  $k$  egész szám, melyre  $N_n c_k \geq 1$  teljesül, azaz  $M_n \asymp \max\{k | N_n c_k \geq 1\}$ . Ez alapján azt kapjuk, hogy  $M_n \asymp \log N_n$  szubkritikus,  $M_n \asymp \log^2 N_n$  kritikus és  $M_n \asymp N_n^{1/\beta}$  szuperkritikus esetben. Tehát ha ezen becsléseinket összevetjük a tételben szereplő eredményekkel, akkor látható, hogy a szubkritikus és kritikus esetben ezek egybeesnek, viszont a megmaradt szuperkritikus esetben csak kevesebbet állítunk. Ez persze tovább javítható az elvárásainknak megfelelően, amelyet a következő alfejezetben fogunk bővebben taglalni. Jelen esetben viszont a teljességre való törekvés miatt csak ezt vizsgáljuk.

A tétel bizonyítását szintén nem teljes részletességgel ismertetjük, csak a legfontosabb lépéseket és ötleteket vázoljuk fel. A jobb strukturáltság érdekében először nézzük a felső korlátokra vonatkozó eredményeket. Ezek belátásához a lassított Thörnblad modellt használjuk, továbbá a következő lemmát, melynek igazolását, figyelembe véve, hogy indukción múlik, nem prezentáljuk.

**4.4. Lemma.** *Legyen  $0 < p < 1$  rögzített és használjuk a megszokott  $\beta = \frac{p}{2^{p-1}}$  jelölést. Ekkor a korábban bevezetett  $S_n(r)$  változó várható értékére az egyes esetekben az alábbi felső becsléseket adhatjuk.*

- *Szubkritikus eset.*  $ES_n(r) \leq 2(r+1)(-\beta)^r n^p.$
- *Kritikus eset.*  $ES_n(r) \leq 2(r+1)! \sqrt{n}.$

- *Szuperkritikus eset.*  $ES_n(r) \leq \begin{cases} C_r(p)n^p, & \text{ha } r < \beta - 1; \\ C_r(p)n^p(1 + \log n), & \text{ha } r = \beta - 1; \\ C_r(p)n^{(r+1)(2p-1)}, & \text{ha } r > \beta - 1, \end{cases}$   
ahol  $C_r(p)$  egy  $n$ -től független,  $r$  és  $p$  értékektől függő konstans.

Ennek megfelelően nézzük hogyan is gondolkozhatunk az egyes esetekben.

A szubkritikus esethez rögzítsünk egy  $a$  és egy  $z$  értéket, hogy  $0 < z < -\frac{1}{\beta}$  és  $a \log(1+z) > 1+p$  megkötéseink teljesüljenek. Ekkor alkalmazva a 4.4. Lemmát, illetve a mértani sorok jól ismert tulajdonságait azt kapjuk, hogy:

$$\begin{aligned} E\left(\sum_{i=1}^{N_n} (1+z)^{d_i(n)}\right) &= E\left(\sum_{i=1}^{N_n} \sum_{r=0}^n \binom{d_i(n)}{r} z^r\right) = E\left(\sum_{r=0}^n \sum_{i=1}^{N_n} \binom{d_i(n)}{r} z^r\right) \\ &= \sum_{r=0}^n ES_n(r) z^r \leq 2 \sum_{r=0}^{\infty} (r+1)(-\beta z)^r n^p = \frac{2}{(1+\beta z)^2} n^p = Kn^p. \end{aligned}$$

Ebből viszont a Markov egyenlőtlenség segítségével majdnem biztos felső korlátot adhatunk a maximális fokszámmra, mégpedig a következőképp:

$$\begin{aligned} P(M_n \geq a \log n) &= P((1+z)^{M_n} \geq (1+z)^{a \log n}) \\ &\leq n^{-a \log(1+z)} E\left(\sum_{i=1}^{N_n} (1+z)^{d_i(n)}\right) \leq Kn^{p-a \log(1+z)}. \end{aligned}$$

Ekkor az előbbi egyenlőtlenséget a pozitív  $n$ -ekre összegezve, tekintettel a jobb oldal konvergenciájára, azt kapjuk, hogy a bal oldalon szereplő sor is konvergál, így a Borel-Cantelli lemma alkalmazásával azt kapjuk, hogy elég nagy  $n$  esetén  $M_n < a \log n$  teljesül, egy nullmértékű halmazt leszámítva. Tehát az  $a$ -ra és  $z$ -re adott feltételek mellett, a

$$\limsup_{n \rightarrow \infty} \frac{M_n}{\log n} \leq \frac{1+p}{\log(1-\beta^{-1})} = \frac{1+p}{\log \gamma}$$

felső becslés teljesül, amiből a 4.3. Lemma segítségével látható  $\log N_n \sim p \log n$  ( $n \rightarrow \infty$ ) összefüggés miatt a kívánt állítás következik.

Hasonlóan járhatunk el a kritikus esetben is. Itt először azt érdemes észrevenni, hogy a struktúra miatt ha van a gráfban  $k$  fokú csúcs, akkor legalább  $k+1$  darab ilyen kell legyen. Ezáltal minimum  $M_n+1$  darab maximális foksámú csúcs csúcs van az  $n$ -edik pillanatban, amiből

$$S_n(r-1) = \sum_{i=1}^{N_n} \binom{d_i(n)}{r-1} \geq (M_n+1) \binom{M_n}{r-1} = r \binom{M_n+1}{r}$$

következik. Így, felhasználva a 4.4. Lemmát, a

$$E\left(\binom{M_n}{r}\right) \leq E\left(\binom{M_n+1}{r}\right) \leq \frac{1}{r} ES_n(r-1) \leq 2r! \sqrt{n}, \quad r = 1, 2, \dots$$

becslést kapjuk, melyből behelyettesítés után az alábbi egyenlőtlenség írható fel:

$$E\left(\sum_{r=0}^n \binom{M_n}{r} \frac{y^r}{r!}\right) \leq 1 + 2\sqrt{n} \sum_{r=1}^n y^r \leq \frac{2\sqrt{n}}{1-y},$$

ha  $0 < y < 1$  teljesül. Ekkor ha fixálunk egy olyan  $a$  értéket, melyre  $ya > \frac{9}{16}$ , akkor a szubkritikus esetben felhasznált Markov egyenlőtlenséget és Borel-Cantelli lemmát tartalmazó gondolatmenettel megkapható, hogy majdnem biztosan minden elég nagy  $n$  esetén  $M_n < a \log^2 n$  teljesül. Tehát

$$\limsup_{n \rightarrow \infty} \frac{M_n}{\log^2 n} \leq a$$

teljesül minden  $a > \frac{9}{16}$  esetén, ahonnan, a megfelelően módosított 4.3. Lemmabeli  $\log n \sim 2 \log N_n$  állítást alkalmazva a kívánt eredményt kapjuk.

A megmaradt szuperkritikus esetben az eddigiekkel megegyező ötleteket használhatjuk. Ennek megfelelően legyen  $a > 2p - 1$  és  $r$  olyan nagy, hogy  $r > 1 - \beta$  és  $r(2p - 1 - a) < -1$  teljesüljenek. Tehát felhasználva a Markov egyenlőtlenséget, a 4.4. Lemmát és a korábban adott becslést az  $\binom{M_n}{r}$  várható értékére az alábbi egyenlőtlenséget kapjuk:

$$P(M_n \geq n^a) = P\left(\binom{M_n}{r} \geq \binom{n^a}{r}\right) \leq \frac{E\binom{M_n}{r}}{\binom{n^a}{r}} \leq \frac{ES_n(r-1)}{r\binom{n^a}{r}} = O(n^{r(2p-1-a)}).$$

Ebből a Borel-Cantelli lemmát alkalmazva látható, hogy majdnem biztosan, elég nagy  $n$  esetén  $\log M_n < a \log n$  fennáll, amiből következik, hogy

$$\limsup_{n \rightarrow \infty} \frac{\log M_n}{\log n} \leq 2p - 1.$$

Ekkor a 4.3. Lemma segítségével elvégezhető  $\log N_n \sim p \log n$  transzformáció után készen vagyunk.

Térjünk rá tehát az alsó korlátok belátásához használt ötletek összefoglalására. Ezekhez a folytonosított Thörnblad modellt vesszük igénybe, aminek következtében a maximális fokszámot egy adott pillanatban értelemszerűen  $M(t)$ -vel fogjuk jelölni, a korábbi  $M_n$  helyett.

Először azt mutatjuk meg, hogy az  $(1-p)\sqrt{n}$  pillanatban az izolált csúcsok száma viszonylag sok. Vezessük be az  $N_n = N((1-p)\sqrt{n} - 1)$  jelölést. Ekkor az adott pillanatban, mind az  $N_n$  darab csúcs esetén annak a valószínűsége, hogy egységnyi idő elteltéig (azaz  $(1-p)\sqrt{n}$ -ig) törölve lesz és utána nem duplikálódik, megegyezik  $\sigma = (1-p)(1 - e^{-1})$ -nel. Ennek következtében az  $(1-p)\sqrt{n}$  pillanatban a gráfban található izolált csúcsok száma legalább akkora, mint egy  $N_n$  rendű és  $\sigma$  paraméterű binomiális eloszlású valószínűségi változó, melyet jelöljük  $Q_n$ -nel. Mivel a folytonosított modell definíciójából következően  $N_n \sim Geo(p_n)$ , ahol  $p_n = \exp\{-p((1-p)\sqrt{n} - 1)\}$ , így  $Q_n$  eloszlása a következő keverék eloszlás lesz:

$$Q_n = \begin{cases} Geo\left(\frac{p_n}{\sigma + (1-\sigma)p_n}\right), & \frac{\sigma}{\sigma + (1-\sigma)p_n} \text{ valószínűséggel;} \\ 0, & \frac{(1-\sigma)p_n}{\sigma + (1-\sigma)p_n} \text{ valószínűséggel.} \end{cases}$$

Ezután coupling segítségével belátható, hogy annak a valószínűsége, hogy egy, az  $(1-p)\sqrt{n}$  időpontban izolált csúcs fokszáma  $p\sqrt{n}$  idő után nagyobb lesz mint egy  $k_n$  érték, legalább

$$\pi_n = \sum_{k=k_n}^{\infty} q_k - \exp\{-p(1-p)\sqrt{n}\},$$

ahol  $q_k$  az 4.3. Állításban szereplő stacionárius eloszlás. Tehát ha  $k_n$  számot úgy választjuk, hogy

$$\sum_{n=1}^{\infty} \frac{p_n}{\pi_n} < \infty$$

teljesüljön, akkor  $p_n = o(\pi_n)$ , amiből a  $Q_n$  változóra kapott korábbi eredményeinket és a geometriai eloszlás generátorfüggvényét felhasználva a

$$P(M(\sqrt{n}) \leq k_n) \leq E((1 - \pi_n)^{Q_n}) \sim \frac{p_n}{\sigma \pi_n}$$

aszimptotikát kapjuk. Így a feltétel miatt ezen valószínűségeket összegezve egy konvergens sort kapunk. Ennek következtében a Borel-Cantelli lemma ismételt alkalmazásával következik, hogy majdnem biztosan minden elég nagy  $n$  esetén  $M(\sqrt{n}) > k_n$  teljesül.

Mindezek mellett az is igaz, hogy a maximális foksám nem tud lényegesen csökkenni  $\sqrt{n}$  és  $\sqrt{n+1}$  időpontok között. Ugyanis az előbbiek alapján feltéve, hogy  $M(\sqrt{n}) > k_n$ , egy rögzített maximális foksámú csúcs esetén ki tudunk választani  $k_n$  darabot annak szomszédjai közül. Ekkor annak a valószínűsége, hogy ezek közül legalább  $\varepsilon k_n$  darab izolálttá válik a kijelölt intervallumban, egy  $k_n$  rendű és

$$1 - \exp\{(1-p)(\sqrt{n} - \sqrt{n+1})\} \leq (1-p)(\sqrt{n} - \sqrt{n+1}) \leq 2(1-p)n^{-\frac{1}{2}}$$

paraméterű  $Z$  binomiális eloszlású valószínűségi változóval adható meg. A Hoeffding egyenlőtlenséget használva azt kapjuk, hogy

$$P(Z > \varepsilon k_n) \leq \exp\{-2(\varepsilon - 2(1-p)n^{-\frac{1}{2}})^2 k_n\}.$$

Tehát, ha a  $k_n$  értékek sorozatától azt is megköveteljük, hogy a

$$\sum_{n=1}^{\infty} \exp\{-\varepsilon^2 k_n\} < \infty$$

feltételt teljesítse, akkor szintén a Borel-Cantelli lemmát alkalmazva, nagy  $n$  esetén a

$$\min\{M(t) | \sqrt{n} \leq t \leq \sqrt{n+1}\} > (1-\varepsilon)k_n$$

egyenlőtlenséget kapjuk. Következésképp  $n = \lfloor t^2 \rfloor$  választás mellett, ahol  $t$  elég nagy,  $M(t) \geq (1-\varepsilon)k_n$  teljesül.

Összegezve az eddigieket, arra van szükségünk, hogy a három különböző esetben megadjuk a feltételeket kielégítő  $k_n$  és az általa számolható  $\pi_n$  sorozatokat. Ezeket a következő állításban foglaljuk össze.

#### 4.5. Állítás. A korábbi

- I)  $\sum_{n=1}^{\infty} \frac{p_n}{\pi_n} < \infty$
- II)  $\sum_{n=1}^{\infty} \exp\{-\varepsilon^2 k_n\} < \infty$



feltételeket kielégítő  $k_n$  és az általa meghatározott  $\pi_n$  sorozatok a következőképpen alakulnak az egyes esetekben:

- Szubkritikus eset.

$$k_n = \frac{p(1-p)}{\log \gamma}(1-\varepsilon); \quad \pi_n = \exp\{-(1+o(1))k_n \log \gamma\}.$$

- Kritikus eset.

$$k_n = \frac{(1-\varepsilon)^2}{64}n; \quad \pi_n = \exp\{-(1+o(1))\frac{1-\varepsilon}{4}\sqrt{n}\}.$$

- Szuperkritikus eset.

$$k_n = \exp\{(1-\varepsilon)p(2p-1)\sqrt{n}\}; \quad \pi_n = \exp\{-(1+o(1))(1-\varepsilon)p(1-p)\sqrt{n}\}.$$

Ezen állítás bizonyítása az 4.4. Állítás és további behelyettesítés segítségével történik, melyet itt nem részletezünk. Viszont ekkor az is látható, hogy ha az előbbi  $k_n$  sorozatokat visszahelyettesítjük az  $M(t) \geq (1-\varepsilon)k_n$  egyenlőtlenségbe és használjuk a folytonosított modellre vonatkozó 4.3. Lemmabeli eredményt, akkor pont a kívánt alsó korlátokat kapjuk, azaz a tételt vázlatosan beláttuk.

Ahogy az a tétel bizonyítása előtt megjegyeztük, a szubkritikus, illetve kritikus esetekben a sejtéseinknek megfelelő eredményeket kapjuk. Viszont a szuperkritikus tartományban ennél kevesebbet tudtunk belátni az előbbi módszerek segítségével. Ezzel szemben a fejtegetésben található felvetés teljesül ebben az esetben is, melynek részletes kifejtése a következő szakasz célja.

## 4.4. Domináns klikk szuperkritikus esetben

Mint ahogyan az előbbieken beharangozásra került, ezen szakasz célja, hogy a szubkritikus esetben felvetett  $M_n \asymp N_n^{1/\beta}$  sejtést igazoljuk. Kitűzött célunk eléréséhez a következő sorrendben fogunk haladni. Először belátunk egy alapvető lemmát a születési-halálozási folyamatok aszimptotikus viselkedéséről. Ezután a Thörnblad modellt leíró struktúrát és dinamikát átfogalmazzuk urnamodellre, melyre egy már korábbról ismert, a [19] cikkben leírt módszert felhasználva belátjuk a sejtést. Ennek következtében nemcsak a kívánt bizonyítást kapjuk meg, hanem az urna- és véletlen gráf modellek közti megfeleltetés ötletét is, mely habár régóta ismert módszer, esetlegesen a további kutatások alapjaként szolgálhat.

Tekintettel arra, hogy az alfejezetben végig a szuperkritikus tartományra szorítkozunk, a duplikációt leíró  $p$  paraméterről, hacsak külön nem mondunk mást, végig feltesszük, hogy az  $\frac{1}{2}$  és 1 közti nyílt intervallumból választott, rögzített érték. Ezen felül használjuk a korábban már alkalmazott  $\beta = \frac{p}{2p-1}$  és  $\gamma = \frac{1-p}{p}$  jelöléseket.

Ahhoz, hogy a legelső állítást beláthassuk, szükségünk lesz az alábbi tételre, mely az úgynevezett Hamburger-momentumproblémára ad megoldást.

**4.4. Tétel.** Legyen  $\mu$  egy mérték az  $\mathbb{R}$ -en és jelölje az ez által definiált momentumokat

$$m_n = \int_{-\infty}^{+\infty} x^n d\mu(x) \quad (n = 0, 1, 2, \dots).$$

Ha most

$$\sum_{n=1}^{\infty} m_{2n}^{-\frac{1}{2n}} = +\infty,$$

akkor a momentumok egyértelműen meghatározzák a  $\mu$  mértéket.

Ennek a tételnek, amelyet a szakirodalomban Carleman feltételként emlegetnek, következménye, hogy annak eldöntésére, hogy egy eloszlást a momentumai meghatároznak, elegendő egy sor divergenciájáról megbizonyosodni. Természetesen a tételnek létezik erősebb változata is, mely megtalálható a [16] könyv 3. fejezetében, de számunkra ez is elegendő lesz.

Lássuk tehát a születési-halálozási folyamatokra vonatkozó lemmát, mely fő eredményünk belátásának alappillére. Mielőtt kimondanánk, érdemes megjegyezni, hogy habár a bizonyítás máshogyan is megtehető, itt a folyamat rekurzív tulajdonságát kihasználó, úgynevezett kontrakciós eljárást fogunk alkalmazni.

**4.5. Lemma.** Legyen  $(X(t))_{t \geq 0}$  olyan lineáris növekedésű folytonos idejű születési-halálozási folyamat, amelynek születési intenzitása  $p$ , halálozási intenzitása  $1 - p$  és  $X(0) = 1$ , ekkor

$$\frac{X(t)}{e^{(2p-1)t}} \rightarrow U \quad (t \rightarrow \infty) \quad \text{majdnem biztosan,}$$

ahol az  $U$  valószínűségi változó eloszlása:  $(1 - \gamma)Exp(\frac{1}{\beta}) + \gamma\delta_0$  (keverék).

**Bizonyítás.** Vizsgáljuk először a lemma állításában szereplő  $X(t)$  folyamatot és jelöljük  $\tau$ -val az első esemény időpontját. Ez a feltételek miatt  $p$  valószínűséggel egy születés és  $1 - p$  valószínűséggel halálozás (ez azt jelentené, hogy minden  $t > \tau$  esetén  $X(t) = 0$  teljesül). Sőt, az is igaz, hogy  $\tau \sim Exp(1)$ . Ezek tudatában az alábbi eloszlásbeli egyenlőség írható fel:

$$X(t) \stackrel{d}{=} \mathbb{I}_{\{\tau \leq t\}}(X'(t - \tau) + X''(t - \tau))Y + \mathbb{I}_{\{\tau > t\}}$$

ahol  $X'(t)$  és  $X''(t)$  az eredeti  $X(t)$ -vel azonos eloszlású független valószínűségi változók, továbbá  $Y \sim Ind(p)$ , független  $\tau$ -tól. Normalizálás után felírható, hogy

$$\frac{X(t)}{e^{(2p-1)t}} \stackrel{d}{=} e^{-(2p-1)\tau} \mathbb{I}_{\{\tau \leq t\}} \left( \frac{X'(t - \tau)}{e^{(2p-1)(t-\tau)}} + \frac{X''(t - \tau)}{e^{(2p-1)(t-\tau)}} \right) Y + \frac{1}{e^{(2p-1)t}} \mathbb{I}_{\{\tau > t\}}.$$

Nyilvánvaló, hogy a képletben szereplő indikátorokra majdnem biztosan teljesül, hogy  $\mathbb{I}_{\{\tau \leq t\}} \rightarrow 1$  és  $\mathbb{I}_{\{\tau > t\}} \rightarrow 0$ , amint  $t \rightarrow \infty$ . Ezen felül, hasonlóan mint diszkrét paraméterű martingálok esetén, belátható, hogy  $X(t)/e^{(2p-1)t}$  egy nemnegatív martingál. Így felhasználva a martingál-konvergenciatételt, ez konvergálni fog egy  $U$  valószínűségi változóhoz, míg az egyenlet jobb oldalán található kifejezés pedig a következő keverékhez:

$$e^{-(2p-1)\tau}(U' + U'')Y,$$

ahol  $U$ ,  $U'$  és  $U''$  megint csak független azonos eloszlású valószínűségi változók,  $Y \sim \text{Ind}(p)$  és  $\tau \sim \text{Exp}(1)$  függetlenek egymástól és az  $U$ -któl. Ezekre ekkor nyilvánvalóan teljesül, hogy

$$U \stackrel{d}{=} e^{-(2p-1)\tau}(U' + U'')Y.$$

A továbbiakban definiálunk egy eloszlásokon értelmezett metrikus teret és ezen egy operátort, mely az egyes térbeli eloszlásokhoz az előbbi egyenlőség jobb oldalán álló transzformációt rendeli. Erről belátjuk, hogy létezik egyértelmű fixpontja, tehát ha az  $U$  határérték benne van a metrikus térben, akkor eloszlása meg kell egyezzen a fixponttal, amiből az állítás következik.

Nézzük először a fixpontot. Jelölje a továbbiakban  $\mathcal{M}$  azon véges szórású eloszlások terét, melyek várható értéke 1-gyel egyezik. Tekintsük itt a

$$d(\lambda_1, \lambda_2) = \inf \|Z_1 - Z_2\|_2$$

metrikát, melyet szokás Wasserstein metrikának is nevezni, ahol az infimum az összes olyan  $Z_1, Z_2$  valószínűségi változóra értendő, melyekre  $Z_1 \sim \lambda_1$  és  $Z_2 \sim \lambda_2$  teljesül és  $\|\cdot\|_2$  az  $L_2$ -norma. Legyen  $T : \mathcal{M} \rightarrow \mathcal{M}$  egy olyan eloszlásoperátor, amely tetszőleges  $Z \in \mathcal{M}$  valószínűségi változóhoz a

$$TZ \stackrel{d}{=} e^{-(2p-1)\tau}(Z' + Z'')Y$$

valószínűségi változót rendeli, ahol  $Y \sim \text{Ind}(p)$ ,  $\tau \sim \text{Exp}(1)$  a korábbiaknak megfelelően,  $Z'$  és  $Z''$  pedig  $Z$ -vel megegyező eloszlású független valószínűségi változók. Belátható, hogy minden  $Z \in \mathcal{M}$  esetén  $TZ \in \mathcal{M}$  teljesül, tehát a  $T$  operátor jól definiált.

Vizsgáljuk most a  $T$  kontrakciós állandóját. Legyen  $Z_1, Z_2 \in \mathcal{M}$  tetszőlegesek, válasszuk  $TZ_1$  és  $TZ_2$  megadásához ugyanazt a  $\tau$  és  $Y$  változót, továbbá legyen  $(Z'_1, Z'_2)$  pár független a  $(Z''_1, Z''_2)$  pártól, és a megfelelő  $Z$ -k azonos eloszlásúak. Ekkor

$$\begin{aligned} \|TZ_1 - TZ_2\|_2 &= \|e^{-(2p-1)\tau}(Z'_1 - Z'_2 + Z''_1 - Z''_2)Y\|_2 \\ &= \|e^{-(2p-1)\tau}Y\|_2 \cdot \|Z'_1 - Z'_2 + Z''_1 - Z''_2\|_2 \end{aligned}$$

teljesül. Itt második tényező a függetlenség miatt  $\sqrt{2}\|Z'_1 - Z'_2\|_2$ , míg az első pedig  $\sqrt{\frac{p}{4p-1}}$ . Így felírható, hogy

$$d(TZ_1, TZ_2) \leq \sqrt{\frac{2p}{4p-1}}d(Z_1, Z_2).$$

Tehát az operátor kontrakciós állandója kisebb, mint 1.

Ennek következtében a Banach-féle fixpont tétel értelmében a

$$Z \stackrel{d}{=} e^{-(2p-1)\tau}(Z' + Z'')Y$$

egyenlőségnek egyértelmű megoldása van az  $\mathcal{M}$  metrikus térben. A továbbiakban megmutatjuk, hogy a megoldás nem más, mint

$$V^* = (1 - \gamma)\text{Exp}\left(\frac{1}{\beta}\right) + \gamma\delta_0.$$

Ahhoz, hogy belássuk, ez ténylegesen teljesíti az egyenlőséget, az előbb megemlített 4.4. Tételt fogjuk felhasználni. Ehhez két lépésen keresztül juthatunk el. Először belátjuk, hogy a  $V^*$  változó eloszlását momentumai egyértelműen meghatározzák, majd megmutatjuk, hogy a két oldal momentumai  $V^*$  eloszlása mellett azonosak, amikből a 4.4. Tétel miatt az állítás következik.

Tekintettel arra, hogy a 0-ra koncentrált eloszlás nem befolyásolja a  $V^*$  kevert valószínűségi változó momentumait, így tetszőleges  $k = 1, 2, \dots$  esetén felírható:

$$E[(V^*)^k] = (1 - \gamma)E[(Exp(\frac{1}{\beta}))^k] = k!\beta^{k-1},$$

ahol az utolsó egyenlőségénél az exponenciális eloszlás momentumaira ismert azonosságot, illetve a  $\gamma$  és  $\beta$  definícióit használtuk. Ezekre viszont a Stirling-formula alkalmazásával

$$\sum_{k=1}^{\infty} (E[(V^*)^{2k}])^{-\frac{1}{2k}} = \sum_{k=1}^{\infty} ((2k)!\beta^{2k-1})^{-\frac{1}{2k}} = \sum_{k=1}^{\infty} \frac{a_k}{k} = +\infty$$

adódik, ahol  $a_k \rightarrow \frac{e}{2\beta}$ , ha  $k \rightarrow \infty$ . Így felhasználva a 4.4. Tételt, azt kapjuk, hogy  $V^*$  eloszlását a momentumai egyértelműen meghatározzák.

Emellett szükségünk van még a jobb oldalon található kevert eloszlás  $k$ -adik ( $k = 1, 2, \dots$ ) momentumára, amihez rögzítsük  $k$ -t. Mivel ez független valószínűségi változók szorzata, így igaz, hogy

$$E[(e^{-(2p-1)\tau}(V' + V'')Y)^k] = E[(e^{-k(2p-1)\tau})]E[Y^k]E[(V' + V'')^k]$$

tejesül, ahol  $V'$  és  $V''$  egymástól független,  $V^*$ -gal megegyező eloszlású valószínűségi változók. Könnyű számolással belátható, hogy

$$E[(e^{-k(2p-1)\tau})]E[Y^k] = \frac{p}{(2p-1)k+1} = \frac{1}{k(1-\gamma) + \gamma + 1}.$$

A második egyenlőségénél szintén a  $\gamma$  definícióját használtuk. A megmaradt tényezőhöz azt kell észrevenni, hogy

$$V' + V'' = \gamma^2\delta_0 + 2\gamma(1-\gamma)Exp\left(\frac{1}{\beta}\right) + (1-\gamma)^2\Gamma\left(2, \frac{1}{\beta}\right),$$

azaz az összeg szintén egy keverék eloszlás. Ennek következtében pedig a  $k$ -adik momentumra az alábbi írható fel:

$$\begin{aligned} E[(V' + V'')^k] &= E\left[\left(\gamma^2\delta_0 + 2\gamma(1-\gamma)Exp\left(\frac{1}{\beta}\right) + (1-\gamma)^2\Gamma\left(2, \frac{1}{\beta}\right)\right)^k\right] \\ &= \gamma^2E\delta_0^k + 2\gamma(1-\gamma)E\left[Exp^k\left(\frac{1}{\beta}\right)\right] + (1-\gamma)^2E\left[\Gamma^2\left(2, \frac{1}{\beta}\right)\right] \\ &= 2\gamma(1-\gamma)k!\beta^k + (1-\gamma)^2(k+1)!\beta^k = k!\beta^{k-1}(k(1-\gamma) + \gamma + 1). \end{aligned}$$

Így összegezve ezeket, azt kapjuk, hogy

$$E[(e^{-(2p-1)\tau}(V' + V'')Y)^k] = \frac{k(\gamma-1) + \gamma + 1}{k(1-\gamma) + \gamma + 1} k!\beta^{k-1} = k!\beta^{k-1} = E[(V^*)^k]$$

teljesül tetszőleges  $k$  esetén, amiből következik, hogy  $V^*$  a korábbi  $T$  operátor fixpontja.

A lemma bizonyításához már csak arra van szükségünk, hogy belássuk, az  $U$  eleme  $\mathcal{M}$ -nek, azaz azt kell megmutatni, hogy  $E[U] = 1$  és  $E[U^2] < \infty$ . Felhasználva a születési-halálozási folyamatok ismert tulajdonságát ([2] 3. fejezet)

$$D^2[X(t)] = \frac{e^{2(2p-1)t}(1 - e^{-(2p-1)t})}{2p - 1},$$

kapható, amiből következik, hogy

$$D^2\left[\frac{X(t)}{e^{(2p-1)t}}\right] \leq \frac{1}{2p - 1},$$

minden  $t \geq 0$  esetén, tehát az  $U$  határérték második momentuma is véges. Így az  $U$  változó várható értéke is véges kell legyen, sőt, mivel a vizsgált martingál várható értéke 1, az  $E[U] = 1$  feltétel is teljesül. Tehát a lemmát beláttuk. ■

Ezt szem előtt tartva térhetünk rá a duplikációs modell urnamodellel való reprezentálására.

Habár a soron lévő tételek és állítások átfogalmazhatóak, sőt bizonyíthatóak is a véletlen gráfos megfogalmazásban, mégis a következőkben definiálandó urnamodellt alkalmazzuk. Ugyanis ez rámutat arra a korábban megemlített tényre, hogy a valószínűségi számítás két különböző területe összekapcsolható.

**Thörnblad-féle urnamodell.** Legyen  $0 < p < 1$  rögzített érték. A folyamatot az  $n = 0$  időben indítjuk egy olyan urnával, amely egyetlen darab tetszőleges színű golyót tartalmaz. Ezután a diszkrét időben előrehaladva, az  $n \geq 1$  pillanatban véletlenszerűen választunk egy golyót az urnából és a következő két lépés egyikét alkalmazzuk rá:

- (1) Vagy  $p$  valószínűséggel visszarakjuk a golyót az urnába egy másik, vele azonos színű golyóval,
- (2) vagy  $1 - p$  valószínűséggel újraszínezzük a golyót egy, az urnában még nem fellelhető színnel, és ezt az átszínezett golyót rakjuk vissza.

Az urnamodell és a Thörnblad modell ekvivalenciája nyilvánvaló, ugyanis a véletlen gráf egyes klikkjei ugyanúgy viselkednek, mint az urnában található azonos színű golyók. Így a korábban használt jelöléseink gond nélkül átvihetőek az urnás megfogalmazásra is.

Az előbbi 4.5. Lemma jelenléte természetesen azt is sejteti, hogy bizonyításaink során szükségünk lesz a Thörnblad-féle urnamodell folytonos időbe való beágyazására is. Ez pont úgy tehető meg, ahogyan a folytonosított Thörnblad modell esetén történt. Azaz az urnában lévő minden golyóhoz hozzárendelünk két exponenciális órát, ahol az egyik  $0 < p < 1$ , míg a másik  $1 - p$  rátával csörög. Ha az első csörög, akkor berakunk az urnába egy azonos színű golyót, míg ha a második, akkor átszínezzük. Ekkor megint teljesül, hogy a két folyamat átmenetvalószínűségei megegyeznek, sőt ha egy új szín születik, akkor az ezen színhez tartozó golyók úgy viselkednek, mint egy  $p$  születési és  $1 - p$  halálozási intenzitással rendelkező folytonos idejű születési-halálozási folyamat. Ezt, és a 4.5. Lemma eredményét felhasználva bizonyíthatjuk az alábbi tételt.

**4.5. Tétel.** Jelölje  $X_n$  egy rögzített színhez tartozó golyók számát az  $n$  pillanatban. Ekkor feltéve, hogy ez a szín nem hal ki az urnából,

$$X_n \sim \nu N_n^{\frac{1}{\beta}}$$

teljesül, ahol  $\nu$  pozitív valószínűségi változó.

**Bizonyítás.** Tekintsük tehát a folytonosított urnamodellt, és ennek megfelelően jelölje  $X(t)$  a fixált színű golyók számát a  $t$  pillanatban. Tegyük fel, hogy ezen szín a  $t_0$  időpillanatban született. Ekkor, mivel feltettük a szín túlélését folyamat során, alkalmazva a 4.5. Lemmát azt kapjuk, hogy

$$X(t) \sim e^{(2p-1)(t-t_0)}U$$

majdnem biztosan, ahol  $U \sim \text{Exp}(\frac{1}{\beta})$ . Ezen felül, ha alkalmazzuk a korábbi  $N(t)$  jelölést az urnában a  $t$  időpillanatban található golyók számára, akkor a 4.3. Lemma miatt tudjuk, hogy

$$N(t) \sim e^{pt}V$$

majdnem biztosan. Itt viszont, mivel  $N(t)$  szintén felfogható születési-halálozási folyamatként, megint használva 4.5. Lemmát, azt is megkapjuk, hogy  $V \sim \text{Exp}(1)$ . Így ezen két eredmény egybefésülésével a ki nem halás feltétele miatt az

$$X(t) \sim \nu N(t)^{\frac{2p-1}{2}}$$

összefüggés teljesül majdnem biztosan. Itt  $\nu$  olyan pozitív valószínűségi változó, mely  $t_0$ -tól,  $U$ -tól és  $V$ -tól függ. Megint csak alkalmazva a 4.3. Lemmát, ezen eredmény egy-az-egyben átemelhető a diszkrét idejű modellre is, azaz a bizonyítás kész. ■

Érdemes itt megemlíteni azt a jól ismert tényt, hogy egy  $p$  születési és  $1-p$  halálozási intenzitással rendelkező lineáris növekedésű születési-halálozási folyamat esetén a kihalás valószínűsége éppen  $\gamma = \frac{1-p}{p}$ . Ezt alkalmazva az előbbi tételre, a következőt írhatjuk fel.

**4.1. Következmény.** Jelölje  $X_n$  egy rögzített színhez tartozó golyók számát az  $n$  pillanatban. Ekkor létezik egy pozitív  $\nu$  valószínűségi változó úgy, hogy

$$\lim_{n \rightarrow \infty} N_n^{-\frac{1}{\beta}} X_n = \begin{cases} 0, & \gamma \text{ valószínűséggel;} \\ \nu, & 1 - \gamma \text{ valószínűséggel.} \end{cases}$$

Mindezek tudatában rátérhetünk a szakasz fő eredményének bizonyítására. Ehhez először vizsgáljunk a Thörnblad-féle urnamodellünknél egyszerűbbet, mely valamilyen értelemben az eredeti projekciója. Tegyük fel tehát, hogy urnánk csak fekete és fehér színű golyókat tartalmaz, méghozzá úgy, hogy ezekből kezdetben rendre  $B_0 = b$  és  $W_0 = w$  darab van. A folyamat során minden diszkrét pillanatban véletlenszerűen húzunk egy golyót az urnából és következő lépések egyikét hajtjuk végre. Vagy  $\frac{1}{2} < p < 1$  valószínűséggel visszarakjuk egy vele megegyező színű golyóval együtt, vagy  $1-p$  valószínűséggel végleg kivesszük az urnából. Ahhoz, hogy az esetleges degenerációt elkerüljük, a dinamikát leállítjuk, ha valamelyik szín

eltűnik az urnából. Ennek következtében, ha feltesszük, hogy  $B_n \neq 0$  és  $W_n \neq 0$  teljesülnek, akkor az átmenetvalószínűségek az alábbiak szerint adhatóak meg:

$$(B_{n+1}, W_{n+1}) = \begin{cases} (B_n + 1, W_n), & p \frac{B_n}{B_n + W_n} \text{ valószínűséggel;} \\ (B_n, W_n + 1), & p \frac{W_n}{B_n + W_n} \text{ valószínűséggel;} \\ (B_n - 1, W_n), & (1 - p) \frac{B_n}{B_n + W_n} \text{ valószínűséggel;} \\ (B_n, W_n - 1), & (1 - p) \frac{W_n}{B_n + W_n} \text{ valószínűséggel.} \end{cases}$$

A célunk eléréséhez vizsgáljuk meg ezen két szín együttes viselkedésének alakulását, azaz az

$$f_{b,w}(n) = \frac{B_n}{W_n + B_n}$$

hányadost. Ahogyan az említésre került,  $f_{b,w}(n) = 0$  vagy  $f_{b,w}(n) = 1$  fennállása esetén megállunk, szóval ezek az  $(f_{b,w}(n))_{n=0}^{\infty}$  folyamat elnyelő állapotai. Viszont ahelyett, hogy efféle megállításokra szorítkoznánk, érdemes feltenni, hogy egyik szín sem tűnik el az urnából. Ezen pozitív valószínűségi esemény, habár csak közvetett módon, de elő fog még kerülni a továbbiakban.

Szükségünk lesz az előbb definiált hányados időben nézett határértékének meghatározására. Ehhez megint a folytonos időbe való beágyazást fogjuk tekinteni, mégpedig a következőképp. Az urnában kezdetben jelenlévő  $b$  darab fekete és  $w$  darab fehér golyó mindegyikéhez rendeljünk hozzá egy-egy egymástól független születési-halálozási folyamatot,  $p$  születési és  $1 - p$  halálozási intenzitással. Ezeket jelölje rendre  $B_i(t)$  ( $i = 1 \dots b$ ) és  $W_j(t)$  ( $j = 1 \dots w$ ). Ha a

$$\left( \sum_{i=1}^b B_i(t), \sum_{j=1}^w W_j(t) \right)_{t \geq 0}$$

folyamatot csak azon pillanatokban nézzük, mikor történik valami, akkor természetesen visszakapjuk az eredeti  $(B_n, W_n)_{n=0}^{\infty}$  modellt. Így az átmenetvalószínűségek sem változnak, tehát  $f_{b,w}(n)$  helyett elegendő a

$$g_{b,w}(t) = \frac{\sum_{i=1}^b B_i(t)}{\sum_{i=1}^b B_i(t) + \sum_{j=1}^w W_j(t)}$$

folytonos változatot tekinteni. Ha a  $g_{b,w}(t) = 0$  vagy  $g_{b,w}(t) = 1$  események bármelyike teljesül, akkor a folyamatot szintén megállítjuk. Szóval, ha valami tudunk mondani a  $g_{b,w}(t)$  hányadosról, akkor az rögtön átvihető  $f_{b,w}(n)$ -re is.

Ennek tudatában vizsgáljuk a  $g_{b,w}(t)$  időben vett határértékét, amelyről a következő igazolható.

**4.6. Állítás.** *Az előbbi jelöléseink mellett a  $\lim_{t \rightarrow \infty} g_{b,w}(t)$  határérték majdnem biztosan létezik és eloszlása az alábbi:*

$$\gamma^b \left(1 - \frac{b}{b+w} \gamma^w\right) \delta_0 + (1 - \gamma^b)(1 - \gamma^w) \text{Beta}(G_b, H_w) + \gamma^w \left(1 - \frac{w}{b+w} \gamma^b\right) \delta_1,$$

ahol  $G_b$  és  $H_w$  olyan egymástól független binomiális valószínűségi változók, melyek nem vehetnek fel nullát, azaz eloszlásukra,

$$P(G_b = k) = \frac{1}{1 - \gamma^b} \binom{b}{k} (1 - \gamma)^k \gamma^{b-k}, \quad k = 1, \dots, b,$$

$$P(H_w = k) = \frac{1}{1 - \gamma^w} \binom{w}{k} (1 - \gamma)^k \gamma^{w-k}, \quad k = 1, \dots, w,$$

teljesül.

**Bizonyítás.** Jelölje  $\tau_b = \inf\{t \mid \sum_{i=1}^b B_i(t) = 0\}$  és  $\sigma_w = \inf\{t \mid \sum_{i=1}^w W_i(t) = 0\}$  az egyes színek kihalási idejét. Ekkor a  $\{\sigma_w < \infty\}$  és  $\{\tau_b < \infty\}$  események rendre a fehér, illetve fekete szín urnából való kihalásával egyenértékűek. Természetesen a feltételek miatt az is igaz, hogy ezen két esemény egymástól független.

Értelemszerűen a  $\{\tau_b < \sigma_w < \infty\} \cup \{\tau_b < \sigma_w = \infty\}$  esemény fennállása esetén  $g_{b,w}(\tau_b) = 0$  teljesül. Mivel viszont a két esemény kizáró, így elég egyesével nézni ezek valószínűségét. Az első eseménynél annak a valószínűsége kell, hogy mindkét szín kihál, de a fekete előbb. Ehhez újra megemlítjük, hogy egy  $p$  születési és  $1-p$  halálozási intenzitással rendelkező születési-halálozási folyamat kihalásának valószínűsége  $\frac{1-p}{p}$ , ami jelen esetben pont  $\gamma$ -val egyezik. Így tehát, felhasználva a függetlenséget, annak valószínűsége, hogy az összes folyamat kihál, megegyezik  $\gamma^{b+w}$ -nel. Kellene még továbbá, hogy az utolsóként kiháló folyamat fehér. Mivel feltettük, hogy mindegyik folyamat kihál, ennek valószínűsége pont  $\frac{w}{b+w}$ . A születési-halálozási folyamat kihalási valószínűségére vonatkozó, előbb felhasznált állítás segítségével az unióban szereplő második esemény valószínűségét nyilvánvalóan a  $\gamma^b(1 - \gamma^w)$  érték adja. Ezek összege a vizsgált esemény valószínűsége, azaz  $\gamma^b(1 - \frac{b}{b+w}\gamma^w)$ . Hasonlóan belátható, hogy a  $\{\sigma_w < \tau_b < \infty\} \cup \{\sigma_w < \tau_b = \infty\}$  esemény valószínűségére pedig  $\gamma^w(1 - \frac{w}{b+w}\gamma^b)$  teljesül.

Mint ahogyan az már az előbbieken említésre került,  $g_{b,w}(\tau_b) = 0$  fennáll, amint az  $\{\tau_b < \sigma_w < \infty\} \cup \{\tau_b < \sigma_w = \infty\}$  esemény bekövetkezik. Mivel 0 a folyamat egy elnyelő állapota, így ez a korábbiak miatt egy  $\gamma^b(1 - \frac{b}{b+w}\gamma^w)$  aszimptotikus valószínűségű pont lesz. Ugyanígy, a  $\{\sigma_w < \tau_b < \infty\} \cup \{\sigma_w < \tau_b = \infty\}$  esemény fennállásával együtt  $g_{b,w}(\tau_b) = 1$  is igaz, ami szintén egy elnyelő állapot. Így az eddigiek alapján az 1 pont aszimptotikus valószínűsége  $\gamma^w(1 - \frac{w}{b+w}\gamma^b)$  lesz a folyamat során.

Vizsgáljuk a megmaradt esetet, ahol egyik szín sem hal ki a folyamat során, azaz a  $\{\tau_b = \infty; \sigma_w = \infty\}$  eseményt. A függetlenség fennállása miatt ennek valószínűségére  $(1 - \gamma^b)(1 - \gamma^w)$  adódik. Ezenfelül az is látható, hogy a színek túlélésének feltétele mellett a nem kiháló fekete, illetve fehér folyamatok számának eloszlása rendre a  $G_b$  és  $H_w$  eloszlásokkal egyezik meg. Viszont az összes ilyen túlélő folyamatra egyesével alkalmazva a 4.5. Lemmát a következőt kapjuk:

$$g_{b,w}(t) = \frac{e^{-(2p-1)t} \sum_{k=1}^{G_b} B_{k_i}(t)}{e^{-(2p-1)t} \sum_{k=1}^{G_b} B_{k_i}(t) + e^{-(2p-1)t} \sum_{k=1}^{H_w} W_{k_i}(t)} \rightarrow \frac{U}{U + V}$$

majdnem biztosan, ha  $t \rightarrow \infty$ , ahol  $U \sim \Gamma(G_b, \frac{1}{\beta})$  és  $V \sim \Gamma(H_w, \frac{1}{\beta})$ , függetlenek. Erről viszont tudjuk, hogy  $\frac{U}{U+V} \sim \text{Beta}(G_b, H_w)$ , a gamma eloszlás ismert tulajdonságai alapján. Így összegezve pont az állításban szereplő eloszlást kaptuk. ■



Habár a konkrét átvitelt nem végezzük el, de az eredmény fontossága miatt az előbbi állítást diszkrét modellre vonatkozó formában is kimondjuk.

**4.6. Tétel.** *Az előbbi jelöléseink mellett, a  $\lim_{n \rightarrow \infty} f_{b,w}(n)$  határérték majdnem biztosan létezik és eloszlása az alábbi:*

$$\gamma^b \left(1 - \frac{b}{b+w} \gamma^w\right) \delta_0 + (1 - \gamma^b)(1 - \gamma^w) \text{Beta}(G_b, H_w) + \gamma^w \left(1 - \frac{w}{b+w} \gamma^b\right) \delta_1,$$

ahol  $G_b$  és  $H_w$  olyan egymástól független binomiális valószínűségi változók, melyek nem vehetnek fel nullát, nevezetesen,

$$P(G_b = k) = \frac{1}{1 - \gamma^b} \binom{b}{k} (1 - \gamma)^k \gamma^{b-k}, \quad k = 1, \dots, b,$$

$$P(H_w = k) = \frac{1}{1 - \gamma^w} \binom{w}{k} (1 - \gamma)^k \gamma^{w-k}, \quad k = 1, \dots, w.$$

Érdeemes megjegyezni, hogy abban az általunk nem vizsgált esetben, amikor  $p$  megegyezik 1-gyel, tehát csak duplikáló lépések történnek, visszakapjuk a már sokat tanulmányozott Pólya-féle urnamodellt. Ezen felül  $p = 1$  esetén  $\gamma = 0$  teljesül, amiből azt kapjuk, hogy a 4.6. Tételben kapott eloszlás  $\text{Beta}(b, w)$ -ra redukálódik, mely klasszikus eredmény a szakirodalomban [29].

Természetes gondolat ezt az eredményt több szín együttes viselkedésének leírására is kiterjeszteni. Például ha fixálunk egy  $k > 2$  egész számot, akkor hasonlóan levetíthetjük az eredeti urnamodellünket egy  $k$  szín tartalmazóra a következőképp. Tekintsünk egy urnát melyben  $k$  különféle színű golyó lehet és jelölje  $X_i(n)$  a  $i$ -edik színhez tartozó golyók számát az  $n$  időpillanatban, továbbá tegyük fel, hogy kezdetben minden színre  $X_i(0) > 0$  teljesül. Minden lépés során húzzunk egy golyót véletlenszerűen, majd vagy  $p$  valószínűséggel egy vele azonos színűvel együtt visszarakjuk, vagy  $1 - p$  valószínűséggel nem rakjuk vissza többé. Egy adott  $n$  pillanatban az urnában lévő golyók számát jelölje  $S(n) = X_1(n) + \dots + X_k(n)$ . Ekkor ez előbbi állításban szereplő módszerrel belátható, hogy ha feltesszük, hogy egyik szín sem tűnik el az urnából, akkor

$$\left( \frac{X_1(n)}{S(n)}, \dots, \frac{X_k(n)}{S(n)} \right) \xrightarrow{d} (Y_1, \dots, Y_k),$$

ahol nem meglepően, a határeloszlás egy olyan keverék Dirichlet eloszlás, melynek paraméterei binomiális eloszlást követnek a túlélés feltétele mellett. Tekintettel arra, hogy ezt nem fogjuk felhasználni a továbbiakban, az állítás helyességének meggondolása az olvasóra marad.

Mindezek után térjünk vissza a két szín tartalmazó egyszerűsített modellhez, és vizsgáljuk azon esemény gyakoriságát, hogy urnában azonos számú fekete és fehér golyó van. Könnyen belátható, hogy ezen állapot a folyamat lefolyása alatt csak véges sokszor fordulhat elő.

**4.7. Állítás.** *Az az esemény, hogy a két színből azonos számú golyó van az urnában, majdnem biztosan csak véges sokszor fordulhat elő.*

**Bizonyítás.** A 4.6. Tételben szereplő  $G_b$  és  $H_w$  összes lehetséges realizációja mellett a  $Beta(G_b, H_w)$  eloszlás abszolút folytonos a  $(0, 1)$  intervallumon. Ennek következményeképp az ugyanezen tételben szereplő kevert eloszlás is abszolút folytonos lesz a  $(0, 1)$  nyílt intervallumon. Így tehát a  $\frac{B_n}{B_n+W_n}$  hányados majdnem biztosan konvergál egy  $\varphi \neq \frac{1}{2}$  (véletlen) határértékhez. Ez viszont pont azt jelenti, hogy a vizsgált arány csak véges sokszor lehet  $\frac{1}{2}$ , azaz a fekete és fehér golyókból csak véges sokszor lehet azonos számú az urnában. ■

Korábbi jelöléseinket megtartva az előbbi állítást úgy is felírható, hogy a  $B_n$  és a  $W_n$  egyezősége legfeljebb csak véges sok  $n$  esetén teljesül. Ennek ismeretében érdemes azt is megnézni, hogy ezen esemény milyen valószínűséggel következik be. Erről szól a következő állítás.

**4.8. Állítás.** *Tegyük fel, hogy  $b > w$  teljesül, ahol a korábbiak alapján  $B_0 = b$  és  $W_0 = w$ . Jelölje a 4.6. Tételben szereplő*

$$\gamma^b \left(1 - \frac{b}{b+w} \gamma^w\right) \delta_0 + (1 - \gamma^b)(1 - \gamma^w) Beta(G_b, H_w) + \gamma^w \left(1 - \frac{w}{b+w} \gamma^b\right) \delta_1$$

*keverék eloszlás eloszlásfüggvényét  $F_{b,w}$ , továbbá  $P(b, w)$  annak valószínűségét, hogy valamilyen  $n$  esetén  $\frac{B_n}{B_n+W_n} = \frac{1}{2}$  fennáll. Ekkor erre*

$$P(b, w) = 2F_{b,w}\left(\frac{1}{2}\right)$$

*teljesül.*

**Bizonyítás.** Jelöljük  $\varphi$ -vel az  $f_{b,w}(n) = \frac{B_n}{B_n+W_n}$  hányados határértékét, illetve legyen  $\mathcal{E} = \inf\{n \geq 0 \mid f_{b,w}(n) = \frac{1}{2}\}$ . Mivel a korábbiak alapján már tudjuk, hogy  $P(\varphi = \frac{1}{2}) = 0$ , így

$$\{\mathcal{E} < \infty\} = \left\{ \mathcal{E} < \infty; \varphi < \frac{1}{2} \right\} \cup \left\{ \mathcal{E} < \infty; \varphi > \frac{1}{2} \right\}$$

felírható. Az is igaz, hogy  $(B_n, W_n)_{n=\varepsilon+1}^\infty$  Markov-folyamat, tehát egyedül a  $(B_\varepsilon, W_\varepsilon)$  állapottól függ. Ezért az  $\{\mathcal{E} < \infty\}$  esemény fennállása esetén a  $\varphi$  határérték a  $F_{B_\varepsilon, W_\varepsilon}$  szerint választódik. Ez pedig az  $\mathcal{E}$  definíciója miatt  $\frac{1}{2}$  körül szimmetrikus, aminek következtében

$$P\left(\mathcal{E} < \infty; \varphi < \frac{1}{2}\right) = P\left(\mathcal{E} < \infty; \varphi > \frac{1}{2}\right)$$

teljesül. Viszont mivel feltettük, hogy  $b > w$ , ezért  $\{\varphi < \frac{1}{2}\} \subseteq \{\mathcal{E} < \infty\}$ , amiből  $P(\mathcal{E} < \infty; \varphi < \frac{1}{2}) = P(\varphi < \frac{1}{2})$  felírható. Így pedig

$$P(b, w) = P(\mathcal{E} < \infty) = 2P\left(\varphi < \frac{1}{2}\right) = 2F_{b,w}\left(\frac{1}{2}\right)$$

igaz. ■

A szakasz fő állításának bizonyításához szükséges lesz annak ismerete is, hogy milyen valószínűséggel történhet meg az, hogy egy új színhez tartozó golyók száma az idő múlásával beéri egy korábbi szín golyóinak számát. A leegyszerűsített modellben az előbbi jelölések megtartásával ezen keresett valószínűség nem más, mint  $P(b, 1)$ . A következő lemmában megmutatjuk, hogy ez az érték  $b$  függvényeként legalább exponenciálisan csökken.

**4.6. Lemma.** *A kiegyenlítődéés  $P(b, 1)$  valószínűségére a*

$$P(b, 1) \leq 2 \left( \frac{1}{2p} \right)^b$$

*egyenlőtlenség teljesül.*

**Bizonyítás.** Használva az 4.8. Állítást és a 4.6. Tételt az alábbi írható fel:

$$\begin{aligned} P(b, 1) &= 2F_{b,1} \left( \frac{1}{2} \right) \\ &= 2\gamma^b \left( 1 - \frac{b}{b+1}\gamma \right) + \\ &\quad 2(1-\gamma^b)(1-\gamma) \sum_{k=1}^b k \int_0^{\frac{1}{2}} x^{k-1}(1-x)^{1-1} dx \binom{b}{k} \frac{(1-\gamma)^k \gamma^{b-k}}{1-\gamma^b} \\ &\leq 2\gamma^b + 2 \sum_{k=1}^b \frac{1}{2^k} \binom{b}{k} (1-\gamma)^k \gamma^{b-k} = 2 \sum_{k=0}^b \binom{b}{k} \left( \frac{1-\gamma}{2} \right)^k \gamma^{b-k} \\ &= 2 \left( \frac{1-\gamma}{2} + \gamma \right)^b = 2 \left( \frac{1}{2p} \right)^b. \end{aligned}$$

■

Térjünk vissza a megszámlálhatóan végtelen sok színt tartalmazó eredeti modellre. A  $P(b, 1)$ -re adott előbbi korlát egy jól használható becslést ad annak valószínűségére, hogy egy kis elemszámú szín túlnő egy nála határozottan nagyobbát. A következőkben ezt és a Borel-Cantelli lemmát felhasználva megmutatjuk, hogy a folyamat során legfeljebb véges sok szín lehet a legnagyobb méretű.

**4.7. Tétel.** *A folyamat során majdnem biztosan legfeljebb véges sok különböző szín lehet maximális elemszámú.*

**Bizonyítás.** Jelölje  $(T_n)_{n=1}^\infty$  az egyes új színek születési idejét. Ezen felül jelöljük  $\mathcal{H}_n$ -nel azt az eseményt, hogy a  $T_n$  pillanatban születő színhez tartozó golyók száma akkora lesz, mint az abban a pillanatban maximális elemszámú szín. Ekkor ezen új és az addigi legnagyobb méretű szín együttes viselkedése pont olyan, mint a korábbi egyszerűsített 2-dimenziós urna az  $(M_n, 1)$  kezdeti állapotból indítva. Legyen  $\mathcal{C}_r$  ( $r \in \mathbb{N}$ ) az az esemény, hogy  $M_n \geq n^{\frac{1}{2\beta}}$  teljesül, minden  $n > r$  esetén. Érdeemes megjegyezni, hogy ekkor  $M_n \geq r^{-\frac{1}{2\beta}} n^{\frac{1}{2\beta}}$  fennáll minden  $n \geq 1$  mellett. Ennek értelmében, ha fixálunk egy  $r \in \mathbb{N}$  számot, akkor

$$P(\mathcal{H}_n \cap \mathcal{C}_r) \leq \sup \left\{ P(A, 1) \mid A \geq r^{-\frac{1}{2\beta}} n^{\frac{1}{2\beta}} \right\} \leq 2 \left( \frac{1}{2p} \right)^{r^{-\frac{1}{2\beta}} n^{\frac{1}{2\beta}}}$$

felírható. Ha most  $n$ -re összegzünk, akkor azt kapjuk, hogy

$$\sum_{n=1}^{\infty} P(\mathcal{H}_n \cap \mathcal{C}_r) \leq 2 \sum_{n=1}^{\infty} \left( \frac{1}{2p} \right)^{r^{-\frac{1}{2\beta}} n^{\frac{1}{2\beta}}} < \infty.$$

Ekkor a Borel-Cantelli lemmát felhasználva következik, hogy a  $\mathcal{H}_n \cap \mathcal{C}_r$  esemény 0 valószínűséggel teljesül végtelen sok  $n$ -re. Elevenítsük fel a 4.5. Tételt, miszerint egy

rögzített szín golyóinak száma, a túlélés feltevése mellett, úgy nő, mint  $\nu N_n^{1/\beta}$ , ahol  $\nu$  egy pozitív valószínűségi változó. Így tehát  $P(\mathcal{C}_r) \rightarrow 1$ , amint  $r$  tart a végtelenbe, aminek következménye, hogy  $\mathcal{H}_n$  0 valószínűséggel áll fenn végtelen sok  $n$  esetén. Ez pedig pont azt jelenti, hogy majdnem biztosan csak véges sok szín lehet maximális elemszámú a folyamat során. ■

Érdekes visszaemlékezni a korábbi 4.7. Állításra, mely azt mondta ki, hogy csak véges sokszor fordulhat elő, hogy két színből azonos számú van. Így ezt összekapcsolva az előbbi tétellel, azt kapjuk, hogy 1 valószínűséggel létezik a folyamat során egy szín, mely véges idő után a legnagyobb elemszámú marad örökre. Ha ezt a szín domináló színnek nevezzük, akkor erre ezzel a gondolatmenettel beláttuk az alábbi tételt.

**4.8. Tétel.** *A folyamat során lesz egy domináló szín majdnem biztosan.*

Így az eddigieket összefoglalva megkaphatjuk a szakasz céljaként kitűzött eredményt, az eredeti Thörnblad-féle duplikációs véletlen gráf modell szuperkritikus esetére.

**4.9. Tétel.** *Az eddigi jelöléseink mellett, a szuperkritikus esetben az  $n$ -edik lépés utáni maximális fokszám eloszlása*

$$M_n \sim \mu N_n^{\frac{1}{\beta}}$$

*majdnem biztosan, ahol  $\mu$  egy pozitív valószínűségi változó.*

A szakasz lezárása előtt érdemes még megjegyezni, hogy a törlések jelenléte miatt a modellben nem létezhet gócpont a korábban bevezetett értelemben. Ezzel szemben ebben az alfejezetben azt is beláttuk, hogy ha a gócpont fogalmát nem csúcsra leszűkítve, hanem klikkekre nézve vizsgáljuk, akkor szuperkritikus esetben ilyen majdnem biztosan létezik.

## 5. Összefoglalás

A dolgozat során bemutatásra került, milyen alapvető problémák, illetve gondolatok játszhatnak szerepet a biológiai hálózatok véletlen gráfokkal való modellezése esetén. Szó esett az ilyen folyamatok háttérében meghúzódó mechanizmus matematikai megfelelőjéről, továbbá ennek egy általános és egy speciális formájáról. Az előbbi esetén pár aktuális eredmény mellett említésre került a túlzott heurisztika használatának problémája, amely nem ritka az ilyen empirikus tudományokból fakadó kérdések esetén. Az utóbbi tulajdonságairól részletes áttekintést adtunk.

Természetesen, mivel a véletlen gráfok ezen területe viszonylag friss és rohamosan fejlődik, így sajnos teljes antológia még nem adható a témában. Ennek következménye, hogy habár igyekeztünk a lehető legátfogóbb képet adni a kérdéskörrel, ez sem feltétlenül fedi az összes aktuális eredményt. Viszont így is elmondható, hogy a dolgozat irodalomjegyzéke jó támpontot adhat a továbbiakhoz.

Persze a terület fiatalsága olyan erényekkel is szolgál, hogy az elérhető szakirodalom rengeteg további kérdést, illetve sejtést vet fel, melyek lehet, hogy heurisztikus elvárásaink szerint nyilvánvalóak, de szigorú matematikai bizonyításuk nehéz. Az egyik ilyen kérdés például, melyet a dolgozatban is kihangsúlyoztunk, hogy mi mondható el a parciális duplikációs modell összefüggő komponensének aszimptotikus fokszámeloszlásáról.

Összegezve, mindenképp érdemes a biológiai hálózatok matematikai modellezésével foglalkozni, ugyanis rengeteg nyitott probléma érhető el, melyek megoldása nem csak matematikai, hanem akár a biológia és az orvostudomány fejlődését is elősegítheti.

# Irodalomjegyzék

- [1] ANDREWS, G. E.; ASKEY, R.; ROY, R. *Special Functions*. Cambridge University Press, 1999.
- [2] ATHREYA, K. B.; NEY, P. E. *Branching Processes*. Springer, Berlin Heidelberg, 1972.
- [3] BACKHAUSZ, Á.; MÓRI, T. F. Asymptotic properties of a random graph with duplications. *Journal of Applied Probability* **52** (2015), 375–390.
- [4] BACKHAUSZ, Á.; MÓRI, T. F. Further properties of a random graph with duplications and deletions. *Stochastic Models* **32** (2016), 99–120.
- [5] BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *Science* **286** (1999), 509–512.
- [6] BARBOUR, A. D.; HOLST, L.; JANSON, S. *Poisson Approximation*. Clarendon Press, Oxford, 1992.
- [7] BEBEK, G.; P., B.; COOPER, C.; FRIEDETZKY, T.; NADEAU, J.; SAHINALP, S. C. The degree distribution of the generalized duplication model. *Theoretical Computer Science* **369** (2006), 239–249.
- [8] BEHRENDTS, E. *Introduction to Markov Chains*. Springer, 2000.
- [9] BHAN, A.; GALAS, D. J.; DEWEY, T. G. A duplication growth model of gene expression networks. *Bioinformatics* **18** (2002), 1486–1493.
- [10] CHAMPAGNAT, N.; LAMBERT, A. Splitting trees with neutral Poissonian mutations i: Small families. *Stochastic Processes and their Applications* **122** (2012), 1003–1033.
- [11] CHAMPAGNAT, N.; LAMBERT, A.; RICHARD, M. Birth and death processes with neutral mutations. *International Journal of Stochastic Analysis* **2012** (2012).
- [12] CHUNG, F.; LU, L.; DEWEY, T. G.; GALAS, D. J. Duplication models for biological networks. *Journal of Computational Biology* **10** (2003), 677–687.
- [13] COHEN, N.; JORDAN, J.; VOLIOTIS, M. Preferential duplication graphs. *Journal of Applied Probability* **47** (2010), 572–585.
- [14] COOPER, C.; FRIEZE, A.; VERA, J. Random deletion in a scale-free random graph process. *Internet Mathematics* **1** (2004), 463–483.
- [15] CSÖRGŐ, S. *Fejezetek a Valószínűségelméletből*. Polygon Kiadó, Szeged, 2010.

- [16] DURRETT, R. *Probability: Theory and Examples*. Cambridge University Press, 2010.
- [17] ERDŐS, P.; RÉNYI, A. On random graphs. *Publicationes Mathematicae Debrecen* **6** (1959), 290–297.
- [18] FELLER, W. *An Introduction to Probability and its Applications, Vol. II*. Wiley, New York, 1971.
- [19] GALASHIN, P. Existence of a persistent hub in the convex preferential attachment model. *arXiv:1310.7513* (2013).
- [20] HAMDI, M.; KRISHNAMURTHY, V.; YIN, G. Tracking the empirical distribution of a Markov-modulated duplication-deletion random graph. *arXiv:1303.0050* (2013).
- [21] HERMANN, F.; PFAFFELHUBER, P. Large-scale behavior of the partial duplication random graph. *arXiv:1408.0904* (2014).
- [22] ISPOLATOV, I.; KRAPIVSKY, P. L.; YURYEV, A. Duplication-divergence model of protein interaction network. *Physical Review E* **71** (2005), 061911.
- [23] JASRA, A.; PERSING, A.; BESKOS, A.; HEINE, K.; DE IORIO, M. Bayesian inference for duplication–mutation with complementarity network models. *Journal of Computational Biology* **22** (2015), 1025–1033.
- [24] KARLIN, S.; TAYLOR, H. E. *A First Course in Stochastic Processes*. Academic Press, 1975.
- [25] KARLIN, S.; TAYLOR, H. E. *A Second Course in Stochastic Processes*. Elsevier, 1981.
- [26] KIM, J.; KRAPIVSKY, P. L.; KAHNG, B.; REDNER, S. Infinite-order percolation and giant fluctuations in a protein interaction network. *Physical Review E* **66** (2002), 055101.
- [27] LI, S.; CHOI, K. P.; WU, T. Degree distribution of large networks generated by the partial duplication model. *Theoretical Computer Science* **476** (2013), 94–108.
- [28] LINDVALL, T. *Lectures on The Coupling Method*. Courier Corporation, 2002.
- [29] MAHMOUD, H. *Pólya Urn Models*. CRC Press, 2008.
- [30] PASTOR-SATORRAS, R.; SMITH, E.; SOLÉ, R. V. Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology* **222** (2003), 199–210.
- [31] PINSKY, M.; KARLIN, S. *An Introduction to Stochastic Modeling*. Academic Press, 2010.
- [32] PRAŁAT, P.; WANG, C. An edge deletion model for complex networks. *Theoretical Computer Science* **412** (2011), 5111–5120.
- [33] SZEGŐ, G. Orthogonal polynomials. *Amer. Math. Soc, Providence, RI* (1959), 100–284.

- [34] THORISSON, H. *Coupling, Stationarity, and Segeneration*. Springer, New York, 2000.
- [35] THÖRNBLAD, E. Asymptotic degree distribution of a duplication–deletion random graph model. *Internet Mathematics* **11** (2015), 289–305.
- [36] THÖRNBLAD, E. The dominating colour of an infinite pólya urn model. *arXiv:1506.05862* (2015).
- [37] VAN DER HOFSTAD, R. Random graphs and complex networks. <http://www.win.tue.nl/rhofstad/NotesRGCN>.
- [38] VÁZQUEZ, A.; FLAMMINI, A.; MARITANA, A.; VESPIGNANI, A. Modeling of protein interaction networks. *Complexus* **1** (2003), 38–44.