

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Lokális tulajdonságok véletlen gráfokban

Szakdolgozat

Nagy Gábor

Matematika BSc

Alkalmazott matematikus szakirány

Témavezető:

Backhausz Ágnes

tanársegéd

Valószínűségelméleti és Statisztikai Tanszék



Budapest, 2012

Köszönetnyilvánítás

Sok kisebb és nagyobb dolog közrejátszott abban, hogy elkészülhetett a szakdolgozatom, és sokaknak kell köszönetet mondanom, úgyszólván hely hiányában csak a legfontosabbakat említeném.

Köszönöm a szüleimnek, hogy biztosították számomra a háttérrel és köszönöm a lakótársaimnak, hogy hagytak nyugodtan dolgozni.

Köszönöm azoknak, akik kinyomtatták és bekötötték a dolgozatomat, illetve köszönöm azoknak a szaktársaimnak, akik segítségükkel közelebb vittek a célhoz.

Legfőképp azonban témavezetőmnek, a mindig kedves és segítőkész Backhausz Ágnesnek köszönöm, hogy mindig számíthattam rá. Magyarázataival és útmutatásával nagy szerepet vállalt a dolgozatom elkészítésében.

Köszönöm Nektek! Nélkületek nem jöhetett volna létre a szakdolgozatom.

Tartalomjegyzék

1. Komplex hálózatok	8
1.1. Skálafüggetlen és kis világ-sorozatok	9
1.2. Az Erdős–Rényi véletlen gráf	11
1.2.1. Az Erdős–Rényi véletlen gráf egy érdekes tulajdonsága	12
1.2.2. Az Erdős–Rényi gráf komponensei	12
1.2.3. Nagy csúcsszámú Erdős–Rényi véletlen gráfok	15
1.3. Komplex hálózatokat modellező véletlen gráfok	16
2. A Preferential Attachment Modell	19
2.1. A modell bemutatása	19
2.1.1. A preferential attachment modell $m > 1$ -re	20
2.1.2. Preferential attachment egy kicsit máshogy	21
2.1.3. Hurokélmentes modell	22
2.2. A modell bemutatása - Barabási–Albert-modell	23
2.2.1. Az általánosított Barabási–Albert-fa felépítésének módjai	24
3. Fokszámok vizsgálata véletlen gráfokban	26
3.1. Rögzített csúcsok fokszámainak vizsgálata	26

3.2.	A preferential attachment modellből adódó fokszámsorozatok	28
3.3.	A Barabási–Albert-fa fokszámeloszlása	29
3.4.	A Barabási–Albert-fa alsó szintjei	30
3.4.1.	A Barabási–Albert-fa alsó szintjeinek nagysága	31
3.4.2.	A legnagyobb szinteken	32
3.4.3.	Lokális fokszámeloszlás az általánosított Albert–Barabási-fában	32
3.5.	Valós hálózatok	33
3.5.1.	Hat lépés távolság	33
3.5.2.	Az Erdős-számok	34

Előszó

A hálózatok tudománya egy viszonylag új ága a matematikának. Ez a tudományág a valós és virtuális világban fellelhető hálózatok létrejöttét, növekedését, változását és tulajdonságait kutatja.

Egy hálózatot azonban nem könnyű leírni, mert a legtöbb valós hálózat amivel találkozunk, rettentően nagy. Gondoljunk csak az internetes hálózatra, vagy az emberi ismeretségek hálózatára. Mivel ilyen nagyok ezek a hálózatok, pontos, számszerű leírásuk egyáltalán nem lehetséges. Ezért az egész hálózat leírása helyett annak lokális jellemzésével próbálunk képet alkotni az egész rendszerről. Ezek a helyi tulajdonságok azonban általában valószínűségeken alapulnak, ezért a valós hálózatokat a legjobban véletlen gráfokkal tudjuk leírni.

A véletlen gráfok elmélete a múlt század közepén jött létre Erdős Pál és Rényi Alfréd munkássága nyomán és azóta igen sokat fejlődött. Erdős és Rényi a matematikusok közül elsőként kezdtek foglalkozni a véletlen gráfokkal. Jelentős szerepük volt a később róluk elnevezett Erdős–Rényi véletlen gráf megalkotásában, bár azt még nem hálózatok modellezésére, hanem diszkrét matematikai állítások bizonyítására használták. A dolgozat elején az Erdős–Rényi véletlen gráfról és annak tulajdonságairól lesz szó. Ki fog azonban derülni, hogy ennek a modellnek nincsenek olyan jellemzői, mint egy valós hálózatnak, így nem modellezi jól azokat.

Az Erdős–Rényi véletlen gráf után ezért az úgynevezett preferential attachment modellel foglalkozunk majd, ami olyan tulajdonságokkal rendelkezik, mint a való életben előforduló hálózatok. Például kis világok illetve skálafüggetlenek, vagyis a gráf méretéhez képest nagy fokszámú csúcsok is előfordulnak bennük. Ez annak köszönhető, hogy ebben a mo-

dellben az alapján kapnak a csúcsok éleket, ahány éllel már rendelkeznek. A több éllel rendelkező csúcsok nagyobb, míg a kevesebb éllel rendelkező csúcsok kisebb valószínűséggel kapnak új éleket. A preferential attachment modell nagy előnye, hogy a paramétereit változtathatjuk úgy, hogy jól közelítsen valós hálózatokat, ezáltal a valós hálózatok tulajdonságait is megismerhetjük.

A dolgozat végén pedig a Barabási–Albert-fáról illetve az általánosított Barabási–Albert-fáról és annak lokális tulajdonságairól lesz szó. A Barabási–Albert-fa is egy preferential attachment modell, amit úgy kapunk meg, hogy a paramétereit megfelelő módon rögzítjük. Érdekes tulajdonságokat fedezhetünk majd fel ennek a fának az alsóbb szintjein, ugyanis a gyökérhez közeli szinteken más lesz a fokszámok eloszlása, mint a legnagyobb szinteken, vagy általában véve az egész fában.

1. fejezet

Komplex hálózatok

Két alapvető tulajdonságban szinte minden hálózat osztozik. Az egyik ilyen tulajdonság a kis világ, a másik, hogy az azonos fokszámú csúcsok száma fordítottan arányos a fokszám egy bizonyos hatványával.

A kis világ azt jelenti, hogy a csúcsok közti átlagos távolság kicsi. Tehát egy viszonylag nagy gráfban is könnyen, kevés lépéssel el tudunk jutni egy tetszőleges csúcsból egy másik tetszőleges csúcsba. Igaz ez a tulajdonság az internetes hálózatra, ennek az egyik következménye az, hogy e-mailjeinket gyorsan kézhez kapjuk. Ha az internet nem rendelkezne ezzel a tulajdonságával, az elektronikus levélkézbesítés egyszerűen nem működne, vagy ha mégis, akkor nem sokkal lenne gyorsabb mint a hagyományos, postai úton történő levelezés. Ilyen tulajdonsággal rendelkezik az emberek közti ismeretségek hálózata is, amiről később lesz még szó.

A fokszámok és a hatványaik közti fordított arányosság pedig a következőt jelenti. Legyen N_k a k fokszámú csúcsok száma, n a csúcsok száma, c_n egy konstans, γ pedig valamilyen kitevő, ekkor

$$N_k \sim c_n k^{-\gamma}$$

teljesül.

Az fenti arányt legjobban log–log skálán tudjuk ábrázolni, mert ilyenkor az arány egy egyenes vonalhoz közelít. Vegyük tehát az arány mindkét

oldalának logaritmusát, ekkor azt kapjuk, hogy

$$\log N_k \sim \log c_n - \gamma \log k.$$

Vagyis a $\log k \mapsto \log N_k$ hozzárendelés grafikonjában a $\log c_n$ azt adja meg, hogy hol metszi az egyenes az y tengelyt, a γ pedig az egyenes meredekségét adja meg. Emellett az is igaz, hogy

$$\sum_k N_k = n,$$

ezért feltesszük, hogy $\gamma > 1$. γ nem lehet 1, hiszen akkor $\sum_{k=1}^{\infty} \frac{c_n}{k}$ végtelent adna, ami véges hálózat eset azt jelentené, hogy végtelen sok csúcsunk van és ez nyilván nem lehetséges.

Ahhoz, hogy megfelelően tudjuk modellezni a skálafüggetlen gráfokat, olyan gráfokat veszünk, amik előre megadott fokszámsorozattal rendelkeznek. Ezért legyen F_X az X valószínűségi változó eloszlásfüggvénye, eloszlását pedig jelölje $\{f_k\}_{k=1}^{\infty}$, tehát

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{k \leq x} f_k.$$

Olyan gráfmodellt szeretnénk elérni, ahol N_k , a k fokszámú csúcsok száma nagyjából egyenlő nf_k -val, ahol n a gráf nagysága, vagyis csúcsainak a száma. Hatványrendű kapcsolatnál az

$$N_k \sim nf_k$$

aránynak kell teljesülnie és így

$$f_k \propto k^{-\gamma}$$

relációnak, ahol a bal oldal a jobb oldalhoz aránylik. Ilyen fokszámsorozattal rendelkező véletlen gráf előállításának egyik módja Bollobás Béla konfigurációs modellje, amelyet később ismertetünk.

1.1. Skálafüggetlen és kis világ-sorozatok

A valóságban a legtöbb hálózat hatalmas, ennek ellenére jópár hasonló tulajdonsággal rendelkeznek, többek között azzal, hogy ritkák. Vagyis

viszonylag alacsony fokszámú csúcsokból állnak, ha azt nézzük, hogy egy n csúcsú gráfban a maximális fokszám $n - 1$ lehet. Ezen kívül még általában skálafüggetlenek és kis világok is.

Sok hálózat az idő előrehaladtával méretben és kiterjedésben változik, ezért célszerű gráfsorozatokat bevezetnünk és ezekkel modelleznünk a valós hálózatokat. Egy gráfsorozatot jelöljünk $\{G_n\}_{n=1}^\infty$ -nel, ahol n a G_n gráf nagyságát mutatja, azaz a csúcsainak számát adja meg. Legyen $P_k(n)$ a k fokszámú csúcsok aránya a G_n gráfban, azaz

$$P_k(n) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{D_i(n)=k\}},$$

ahol $D_i(n)$ az i csúcs fokszámát jelöli a G_n gráfban. Így a $\{P_k(n)\}_{n=0}^\infty$ a G_n gráf fokszámsorozatát adja meg.

Most definiáljuk, mi az, hogy skálafüggetlenség.

Ehhez először is nézzük meg, hogy pontosan mikor nevezünk egy gráfsorozatot ritkának. Egy $\{G_n\}_{n=1}^\infty$ gráfsorozatot ritkának nevezünk, ha

$$\lim_{n \rightarrow \infty} P_k(n) = p_k$$

teljesül egy adott $\{p_k\}_{k=1}^\infty$ eloszlásra. A p_k ebben a képletben determinisztikus, ezért a konvergenciát nézhetjük valószínűségben vagy eloszlásban. És mivel $\sum_{k=0}^\infty p_k$ egyet ad, ezért a csúcsok nagy részének fokszáma felülről korlátos, innen ered a ritka gráf elnevezés. Ezt felhasználva most definiáljuk a skálafüggetlen gráfsorozatokat.

Egy $\{G_n\}_{n=1}^\infty$ véletlen gráfsorozatot skálafüggetlennek hívunk, ha ritka és ha

$$\lim_{k \rightarrow \infty} \frac{\log p_k}{\log \frac{1}{k}} = \gamma$$

létezik.

És most határozzuk meg azt is, mit jelent az, hogy egy gráfsorozat kis világ.

Legyen H_n két egyenletesen választott csúcs távolsága, amik között létezik út a gráfban. Tehát kiválasztunk az összeköttetésben lévő csúcspárok halmazából egy párt, és a H_n e két csúcs közötti távolság, azaz

a köztük futó legrövidebb út élszáma. Nevezzük H_n -t a G_n gráf tipikus távolságának. Ekkor azt mondjuk, hogy a $\{G_n\}_{n=1}^\infty$ gráfsorozat kis világ, ha létezik egy K konstans úgy, hogy

$$\lim_{n \rightarrow \infty} \mathbb{P}(H_n \leq K \log n) = 1.$$

Néhány modellben a tipikus távolság sokkal kisebb is lehet, mint $\log n$, az ilyen gráfokat angolul 'ultra small world'-nek nevezzük, ez magyarra lefordítva körülbelül annyit jelent, hogy nagyon kis világ. Egy $\{G_n\}_{n=1}^\infty$ gráfsorozatot nagyon kicsi világnak nevezünk, ha létezik olyan K konstans, hogy

$$\lim_{n \rightarrow \infty} \mathbb{P}(H_n \leq K \log \log n) = 1.$$

Az egyik legelső véletlen gráfmodellt Erdős Pál, Rényi Alfréd és Edgar Gilbert munkássága nyomán jött létre 1959-ben. Lássuk hát ezt a modellt [6] alapján.

1.2. Az Erdős–Rényi véletlen gráf

A véletlen gráfok területe az '50-es évek végén, illetve a '60-as évek elején alakult ki a XX. században. Akadtak korábban is a témával foglalkozó cikkek, például Yule [10, 1925] munkája, de annak, amit általában ennek a területnek az alapkövének neveznek, Erdős és Rényi voltak a szerzői. Nézzük meg, hogyan is néz ki ez a modell. Vesszünk a véletlen gráfban csúcspárokat és mindegyikről eldöntjük egymástól függetlenül, hogy összekötjük-e őket, vagy sem. Erdős és Rényi rögzítették a kiválasztott élek számát, a terület másik úttörője, Edgar Gilbert azonban nem. Gilbert modellje közelebb áll a nemsokára tárgyalt preferential attachment modellhez, ezért most nézzük ezt a modellt.

Az eljárás nagyon egyszerű: kiválasztunk két csúcst, majd p valószínűséggel összekötjük őket egymással, $1 - p$ valószínűséggel pedig nem kötjük össze őket. Azt az esélyt, amellyel két csúcs össze lesz kötve egymással, élvalószínűségnek is szokták nevezni. Az így kapott gráfot a

megalkotóik tiszteletére Erdős–Rényi véletlen gráfnak nevezzük és $ER_n(p)$ -vel jelöljük. Erdős és Rényi 1959-ben mutatták be ennek a gráfnak egy változatát [4], Gilbert pedig 1960-ban írta le ezt a modellt [5].

1.2.1. Az Erdős–Rényi véletlen gráf egy érdekes tulajdonsága

Annak ellenére, hogy az $ER_n(p)$ az egyik legegyszerűbb véletlen gráf, ami véletlen hálózatokat modellez, már ez is érdekes állapotváltozással rendelkezik p változtatása közben. Az állapotváltozást hasonlóan kell érteni, mint a víz halmazállapotjának változását. Ha 0 °C alá hűtjük a vizet, akkor megfagy, ha pedig 100 °C fölé melegítjük, akkor gáz halmazállapotúvá válik.

Körülbelül ez történik az Erdős–Rényi véletlen gráfban is. Ha ugyanis $p = \frac{\lambda}{n}$ -et választunk élvalószínűségnek, ahol $\lambda < 1$, akkor a véletlen gráfunk több kicsi komponensből fog állni, amik mérete $\Theta(\log n)$. Ha viszont olyan $p = \frac{\lambda}{n}$ élvalószínűséggel készítjük el a gráfot, ahol $\lambda > 1$, akkor a gráfunk egy nagy komponensből és több kis komponensből fog állni, ahol a nagy komponens $\Theta(n)$ méretű, a kicsik pedig $\Theta(\log n)$ nagyságúak.

Az, hogy egy komponens mérete $\Theta(n)$ azt jelenti, hogy a komponens csúcsainak a száma és az egész gráf csúcsainak a számának aránya két pozitív szám közé esik minden n -re. Tehát az arány nullánál nagyobb és végtelennél kisebb lesz $n \rightarrow \infty$ esetén. Vagyis, ha n_c -vel jelöljük a komponens méretét, akkor

$$0 < c_1 < \frac{n_c}{n} < c_2 < \infty, \quad n \rightarrow \infty.$$

1.2.2. Az Erdős–Rényi gráf komponensei

Határozzuk most meg az Erdős–Rényi véletlen gráf csúcsai által alkotott komponenseket. Azt mondjuk, hogy u csúcsból elérhető v csúcs, ha létezik legalább egy olyan élsorozat, amin végighaladva u -ból v -be ju-

tunk. Ezt így jelöljük: $u \leftrightarrow v$. Ezek után definiáljuk egy tetszőleges v csúcs komponensét a következő képpen:

$$\mathcal{C}(v) = \{u : v \leftrightarrow u\}.$$

Tehát minden olyan csúcs beletartozik $\mathcal{C}(v)$ -be, ami össze van kötve v -vel. Ebbe a komponensbe megegyezés szerint maga a v csúcs is beletartozik, mert úgy tekintjük, hogy v össze van kötve saját magával. Az Erdős–Rényi véletlen gráfban előforduló legnagyobb komponens pedig jelöljük így:

$$|\mathcal{C}_{max}| = \max_{v \in \{1, \dots, n\}} |\mathcal{C}(v)|,$$

ami, mint ahogy azt az előző fejezetben láttuk, nagyban függ a p érték nagyságától.

Ahhoz, hogy a legnagyobb komponenssel foglalkozni tudjunk, először az összes komponens fel kell térképeznünk. Ezt úgy tesszük, hogy először is megszámozzuk tetszőleges módon a csúcsokat, majd vesszük az 1-gyel jelölt csúcsot és megnézzük, hogy mi tartozik az ő komponensébe. Ezt úgy érjük el, hogy az összes 1-ből induló élen végigmegyünk és megnézzük, hogy milyen csúcsokba jutunk el. Ezek a csúcsok a v közvetlen szomszédai és ezek a csúcsok persze benne vannak $\mathcal{C}(1)$ -ben, vagyis az 1-es csúcs komponensében, mert létezik köztük út. Azt mondjuk, hogy ezek a csúcsok 1 távolságra vannak az 1-es csúcstól a vizsgált véletlen gráfban. Ezeknek a csúcsoknak a számát jelöljük X_1 -gyel. X_1 csak a véletlentől függ, ezért valószínűségi változóként kezeljük. Az X_1 eloszlása, vagyis az 1-es csúcs közvetlen szomszédainak a számának az eloszlása megegyezik egy $n - 1$ rendű és p paraméterű binomiális valószínűségi változó eloszlásával, vagyis $\text{Bin}(n - 1, p)$ -vel.

Ha $X_1 = 0$, akkor az 1-es csúcs komponense saját maga, ha viszont $X_1 \geq 1$, akkor van legalább egy szomszédja és ennek a szomszédos csúcsnak is tudjuk majd nézni a szomszédait. Ezt is fogjuk tenni, de először jelöljük el az 1-gyel szomszédos csúcsokat i_1, i_2, \dots, i_{X_1} -gyel, úgy, hogy $i_1 < i_2 < i_3 < \dots < i_{X_1}$ teljesüljön. Így könnyebb lesz a csúcsokat számon tartani. Akkor fedezzük most fel az i_1 csúcs szomszédait. Ennek

a csúcsnak azonban csak azokra a szomszédaira vagyunk kíváncsiak, amelyekről még nem tudjuk, hogy az 1-es csúcs komponensébe tartoznak-e. Tehát ha például i_1 -ből fut él i_j -be, $1 < j \leq X_1$, akkor i_j -t nem tekintjük új, most felfedezett csúcsnak, mert már ismerjük, tudjuk, hogy benne van az 1-es csúcs komponensében. i_1 újonnan felfedezett szomszédainak számát jelöljük X_2 -vel. Ezen szomszédok számának is binomiális eloszlása lesz, de nem $n - 1$ renddel, hanem $n - 1 - X_1$ renddel. Tehát az X_2 feltételes eloszlása X_1 -re nézve $\text{Bin}(n - 1 - X_1, p)$.

Ha $X_1 \geq 2$, akkor i_2 azon közvetlen szomszédait is feltérképezhetjük, amik még nincsenek $C(1)$ -ben. Legyen X_3 az i_2 -vel szomszédos, újonnan felfedezett csúcsok száma. Ezeknek a csúcsoknak a száma is binomiális eloszlású, de ez már nem csak X_1 -től, hanem X_2 -től is függ. Azaz X_3 feltételes eloszlása X_1 -re és X_2 -re nézve $\text{Bin}(n - 1 - X_1 - X_2, p)$. Tehát sorban minden olyan csúcsnak felfedezzük a szomszédait, amik 1 távolságra vannak az 1-es csúcstól. Majd ha ezek a csúcsok elfogytak, azok a csúcsok következnek, amik 2 távol vannak az 1-es csúcstól, azután azok, amik 3 távol vannak, és így tovább, amíg még van felfedezetlen csúcs a komponensében. A csúcsok ilyen módon történő felfedezését szélességi keresésnek nevezzük.

Általánosságban, amikor a $C(1)$ $(i + 1)$ -dik csúcsának szomszédait nézzük meg, akkor ezeknek a csúcsoknak a számát jelöljük X_{i+1} -gyel. A kapott szomszédok mind $C(1)$ elemei és számuk szintén binomiális eloszlású p paraméterrel és $n - 1 - X_1 - X_2 - \dots - X_i$ renddel. Feltéve persze, hogy $1 \leq k \leq i$ -re X_k -t ismerjük.

Mielőtt az i -dik csúcsot felfedoznénk, azon csúcsok száma, amelyek szomszédait még nem térképeztük fel, a következővel egyenlő:

$$1 + X_1 + X_2 + \dots + X_i - i.$$

Azért, mert az 1-es csúcs komponensében lévő csúcsok számából, vagyis $X_1 + X_2 + \dots + X_i$ -ből levonunk $i - 1$ -et, azon csúcsok számát, amelyek szomszédait már teljesen feltérképeztük. Ebből következik, hogy a $C(1)$ felfedezésének folyamata addig tart, amíg ez a különbség nagyobb, mint

nulla, vagyis amíg:

$$1 + X_1 + \cdots + X_i - i > 0$$

teljesül. Amikor pedig már felfedeztünk minden csúcsot $C(1)$ -ben, a $C(1)$ csúcsainak a számára a következőt kapjuk:

$$|C(1)| = \min\{i : X_1 + \cdots + X_i = i - 1\}. \quad (1.1)$$

Hasonlóan meg tudjuk vizsgálni, hogy a többi, $C(1)$ -be nem tartozó csúcsok mely egyéb komponensekbe tartoznak. Legyen j a legkisebb számú csúcs, ami már nem tartozik az 1-es csúcs komponensébe. Ebből a csúcsból kiindulva felfedezhetjük ennek a csúcsnak a komponensét. Persze ez előtt el kell távolítanunk a $C(1)$ -be tartozó csúcsokat, mivel azokra már nem lesz szükség. Miután feltérképeztük a j csúcs komponensét, megnézzük, hogy melyik számú csúcs a legkisebb olyan, ami még nem került be az előző komponensekbe. Ebből a csúcsból is kiindulva egy újabb komponenst kapunk. Majd újra kiválasztjuk a legkisebb csúcsot és ezeket a lépéseket addig ismételjük, amíg már nem marad csúcs, amit ne térképeztünk volna fel. Ekkor befejeződött az eljárás, végigértünk a gráfon és megkaptuk az összes komponenst.

1.2.3. Nagy csúcsszámú Erdős–Rényi véletlen gráfok

Vizsgáljuk meg, hogy mi történik, ha n nagy. Ha n nagy, akkor tudjuk, hogy az n rendű és $\frac{\lambda}{n}$ paraméterű binomiális eloszlás közel van a λ paraméterű Poisson-eloszláshoz. Más szóval:

$$\mathbb{P}(\text{Bin}(n, \lambda/n) = k) = e^{-\lambda} \frac{\lambda^k}{k!} + o(1), \quad k = 0, 1, 2, \dots$$

teljesül. Azt is mondhatjuk, hogy a k fokszámú csúcsok aránya konvergál a Poisson-eloszláshoz valószínűségben.

Minden adott i -re ha X_1, X_2, \dots, X_i nem túl nagy, akkor

$$N_i = n - 1 - X_1 - X_2 - \cdots - X_i \approx n.$$

Így a $p = \frac{\lambda}{n}$. Vagyis X_i -ket is közelíthetjük Poisson-eloszlású valószínűségi változókkal. Ha ezeket X_i^* -gal jelöljük, akkor azt kapjuk, hogy S_i^* , vagyis a még felfedezettlen csúcsok száma

$$S^* \sim 1 + X_1^* + \cdots + X_i^* - i.$$

A folyamat természetesen itt is addig tart, amíg S_i^* először el nem éri a nullát, azaz

$$\min\{i : S_i^* = 0\} = T^* = \min\{i : X_1^* + \cdots + X_i^* = i - 1\}$$

teljesül. Ebben az egyszerűsített modellben T^* lehet végtelen is, ami (1.1)-ben nem lehetséges.

Bár az Erdős–Rényi véletlen gráf érdekesen viselkedik ha változtatjuk az élek behúzásának valószínűségét, a fokszámai nem skálafüggetlenek, ezáltal nem használható megfelelően véletlen hálózatok modellezésére. A tipikus fokszám ebben a modellben a fokszámok átlaga, vagy ettől csak kicsit tér el. Bebizonyítható azonban az, hogy a skálafüggetlen gráfokhoz hasonlóan a $p = \frac{\lambda}{n}$ élvalószínűségű Erdős–Rényi véletlen gráf ritka.

1.3. Komplex hálózatokat modellező véletlen gráfok

Tehát már tudjuk, hogy az Erdős–Rényi véletlen gráf nem modellezi jól a valós hálózatokat a fokszámeloszlás szempontjából, mivel nem skálafüggetlen. A további fejezetekben olyan modellekkkel fogunk foglalkozni, illetve úgy alakítjuk ki ezeket a modelleket, hogy skálafüggetlenek legyenek és ezáltal jobban jellemezzék a valóságban kialakult hálózatokat. Nézzük először meg a skálafüggetlen véletlen gráfok általános modelljét [6].

Minden i csúcshoz rendelünk egy W_i valószínűségi változót, ami megadja az adott csúcs súlyát. Így egy általános gráfban annak a valószínűsége, hogy két csúcs, s és t össze vannak kötve:

$$p_{st} = \frac{W_s W_t}{W_s W_t + L_n}.$$

$S_n = \sum_i W_i$ jelöli a gráf csúcsainak összsúlyát és ezek a súlyok függetlenek egymástól, feltéve, hogy ismerjük az összes W_i -t. Belátható, hogy ha az Erdős–Rényi véletlen gráfot így véletlenítjük, skálafüggetlen gráfot kapunk, abban az esetben, ha W_i -k független és azonos eloszlású valószínűségi változók.

Egy másik modell úgy alkot véletlen gráfot, hogy a foksámokból indul ki. Minden i csúcshoz rendelünk egy D_i foksámot úgy, hogy az összegük, vagyis $L_n = \sum D_i$ páros legyen. Ezután minden i csúcshoz illesztünk D_i darab élt úgy, hogy ezeknek az éleknek a másik élvégiük szabad legyen. Ezután kiválasztunk egyenletesen két szabad élvéget és összeillesztjük őket, így a két élből egy él lesz. A többi szabad élvéggel is elvégezzük ezt az összeillesztést. A folyamat során bármely élvég párt egyenlő valószínűséggel választunk ki. Az így kapott modellt konfigurációs modellnek nevezzük, ezt a modellt Bollobás Béla találta ki [2].

Persze az is előfordulhat, sőt elő is fordul, hogy a kapott gráf nem egyszerű. Ez azért történhet meg, mert a párhuzamos éleket és hurokéleket ez a fajta modell nem zárja ki. Ha foksámok véges szórással rendelkeznek, akkor az így kapott gráf pozitív valószínűséggel egyszerű lesz. Tehát ha többször megismételjük a konfigurációs eljárást, vagyis ha többször egymás után megrajzoljuk a gráfot, akkor előbb-utóbb egyszerű gráfot kapunk. Ezt ismétléses konfigurációs modellnek (angolul: repeated configuration model) hívjuk. Van egy másik módja is annak, hogy a gráfot egyszerűvé alakítsuk, mégpedig az, hogy a párhuzamos éleket és a hurokéleket kitöröljük. Ezek a műveletek nem változtatnak sokat a gráf foksámain, úgyhogy nyugodtan elvégezhetjük őket. Az ilyen modellt törléses konfigurációs modellnek (angolul: erased configuration model) hívjuk. A konfigurációs modellek és a véletlen gráfok jól írják le a valódi hálózatokat, sőt olyan modelleket teremtenek, amik összehasonlíthatók a foksám szempontjából valódi hálózatokkal. Sajnos azt viszont nem magyarázzák el, hogy hogyan jöttek létre ezek hálózatok. Sajnos, hiszen a hálózatok vizsgálatánál nagyon fontos szempont a kialakulásuk és növekedésük. A világháló sem egy pillanat alatt jött létre, hanem az idő elteltével folyamatosan növekedett és a mai napig is tart ez a folyamat.

A világháló mellett egy másik hálózat, ami szintén növekszik, az a saját ismeretségi hálózatunk. Ez a rendszer sem egyik pillanatról a másikra épül ki, hanem szép fokozatosan. Minden nap új emberekkel találkozunk, akikkel kialakíthatunk kapcsolatokat, így bővíthetjük a "saját hálózatunkat". Érdekes tehát egy hálózat növekedését szemügyre venni. A skálafüggetlen gráfok kialakulásának az egyik lehetséges magyarázatát Albert Réka és Barabás Albert-László adta, akik a valódi hálózatok kialakulását a 'preferential attachment'-tel magyarázták. Ez azt takarja, hogy azok a csúcsok, amik újonnan kerülnek bele a gráfba, nagyobb valószínűséggel csatlakoznak olyan csúcsokhoz, amelyek fokszáma nagyobb, mint azokhoz, amelyeké kisebb. Ez logikusnak tűnik, hiszen a világhálón is hasonlóan működnek a weblapok közti kapcsolatok, az ismertebb weblapokra (pl.: Wikipedia, Google, Youtube) sok hivatkozást találhatunk a weblapokon, míg egy olyan oldalra, amit havonta tíz személy látogat, sokkal kevesebbszer hivatkoznak. Az ismeretségi körünk kialakulása során is nagyobb az esélye annak, hogy olyan valakit ismerünk meg, aki sokakkal van kapcsolatban, mint hogy olyat, akinek kevés ismerőse van. A következőkben a 'preferential attachment' modellek általánosításával fogunk foglalkozni.

2. fejezet

A Preferential Attachment Modell

Véletlen gráfokat modellezni kétféleképpen lehet: statikus vagy dinamikus modellekkel. A statikus modellek nem veszik számításba a gráf bővülését, illetve csökkenését, mivel a gráf tulajdonságait egy kiragadott időpontban mutatják be. A dinamikus modelleket azért nevezzük dinamikusnak, mert a véletlen gráfok kialakulását, változását modellezik. Az ilyen modellek hatványrendű fokszámsorozatokhoz vezettek minket, így megmagyarázzák az ilyen fokszámsorozatok véletlen gráfokban való megjelenését. Az egyik lehetséges magyarázatát annak, hogy miért jelennek meg hatványrendű fokszámsorozatok a valódi hálózatokban, a 'preferential attachment' minta adja meg. A 'preferential attachment' modellnek a lényege az, hogy az újdonsült csúcsok éleit nem azonos valószínűséggel kötjük össze bármelyik csúccsal, hanem nagyobb eséllyel kötjük össze olyan csúcsokkal, amiknek nagyobb a fokszáma, mint azokkal, amiknek kisebb. Lássuk hát, hogyan is néz ki általánosan ez a modell.

2.1. A modell bemutatása

A 'preferential attachment' modell egy gráfsorozatot állít elő, amit $\{PA_t(m, \delta)\}_{t=1}^{\infty}$ -gel jelölünk. Ez a sorozat minden t -re egy t csúcsú gráfot ad, aminek mt éle van, ahol $m \in \{1, 2, \dots\}$ és $\delta \geq -1$ teljesül.

Nézzük meg a modellt $m = 1$ -re. $PA_t(1, \delta)$ csúcsait jelöljük a következőképpen: $\{v_1(1), v_2(1), \dots, v_t(1)\}$, a $v_i(1)$ fokszámát pedig $D_i(t)$ -vel. A definíció szerint $PA_1(1, \delta)$ egy csúcsból áll, ami egy hurokéllel rendelkezik. Vizsgáljuk meg, hogyan néz ki a gráf növekedése, vagyis a $(t + 1)$ -edik csúcsnak a gráfba vétele. Hozzáadunk először is egy csúcsot és egy élt a gráfhoz úgy, hogy az él egyik vége rögzítve legyen az új csúcshoz, a másik vége pedig egyelőre szabad legyen. Ezt a szabad véget most kössük össze a gráf egy tetszőleges csúcsával, méghozzá a fokszámmal arányos valószínűség szerint. Vagyis a szabad élvéget $\frac{1+\delta}{t(2+\delta)+(1+\delta)}$ valószínűséggel saját magához, $\frac{D_i(t)+\delta}{t(2+\delta)+(1+\delta)}$ valószínűséggel a gráf egy tetszőleges $v_i(t) \in PA_t(1, \delta)$ csúcsához kötjük. Tehát annak a valószínűsége, hogy az új csúcsot egy adott csúcshoz kötjük:

$$\mathbb{P}(v_{t+1}^{(1)} \rightarrow v_i^{(1)} | PA_t(1, \delta)) = \begin{cases} \frac{1+\delta}{t(2+\delta)+(1+\delta)} & \text{ha } i = t + 1, \\ \frac{D_i(t)+\delta}{t(2+\delta)+(1+\delta)} & \text{ha } i \in [t], \end{cases} \quad (2.1)$$

ahol $[t] = \{1, 2, \dots, t\}$ és δ pedig a modell paramétere. Ez a paraméter mindig olyan, hogy $\delta \geq -1$, ezért könnyű azt belátni, hogy a (2.1)-ben szereplő valószínűségek összege egyet ad. Az is igaz, hogy $D_i(t) + \delta \geq 0$, mivel $D_i(t) \geq 1$ minden t -re és minden i -re.

Említsünk meg néhány szót a δ -ról. Általában véve minden valós hálózatnak van egy jellemző paramétere, ezt a paramétert jelöltük γ -val. A hálózat ezen γ -ját meg tudjuk határozni, majd miután ezt megtettük, a valós hálózatra egy megfelelő δ -jú preferential attachment modellt tudunk illeszteni. A δ -t a $\gamma = 3 + \frac{\delta}{m}$ összefüggésből kapjuk meg. Tehát egy γ paraméterű valós hálózathoz tudunk találni egy δ paraméterű preferential attachment modellt, ami jól közelíti azt.

2.1.1. A preferential attachment modell $m > 1$ -re

Láttuk, hogy $m = 1$ -re így működik a modell, nézzük meg, hogy $m > 1$ -re hogyan vihető át mindez. Egy $PA_{mt}(1, \frac{\delta}{m})$ gráfsorozattal kezdünk, aminek a csúcsai legyenek $v_1(1), \dots, v_{mt}(1)$. Miután ezt a gráfsorozatot

elkészítettük, a $PA_{mt}(1, \frac{\delta}{m})$ -beli $v_1(1), \dots, v_m(1)$ csúcsok legyenek a $PA_t(m, \delta)$ gráfban a $v_1(m)$ csúcs. A $PA_t(1, \frac{\delta}{m})$ -ben szereplő következő m darab csúcs legyen a $PA_t(m, \delta)$ -ban a $v_2(m)$ csúcs, és így tovább, vagyis általánosan a $v_{(j-1)m+1}(1), \dots, v_{jm}(1)$ legyen a $v_j(m)$ csúcs. Az új $PA_t(m, \delta)$ gráfban két tetszőleges csúcs, $v_k(m)$ és $v_l(m)$ össze van kötve, ha a régi $PA_{mt}(1, \frac{\delta}{m})$ gráfban ennek a két csúcsnak bármelyik két őse össze volt kötve. Egy $v_i(m)$ csúcs őseinek azt az m darab csúcsot tekintjük, amik összevonásával ő maga keletkezett. Így természetesen hurokélek és párhuzamos élek is létrejöhetnek a gráfban. Hurokélek akkor alakulnak ki, ha egy $PA_t(m, \delta)$ -beli csúcs m darab őse közül bármelyik kettő össze volt kötve, párhuzamos élek pedig akkor keletkeznek két csúcs között az új gráfban, ha e két csúcs ősei között a $PA_{mt}(1, \frac{\delta}{m})$ gráfban több él is futott.

Ezzel definiáltuk a modellt $m > 1$ -re is. Ezt az eljárást szokták 'collapsing of vertices'-nek (csúcsok összevonása) hívni. Az így keletkezett $PA_t(m, \delta)$ gráf egy multigráf, aminek t csúcsa, mt éle van és így az összfokszámja $2mt$. Multigráfnak nevezzük az olyan gráfokat, amiknek vannak hurokéleik vagy párhuzamos éleik.

2.1.2. Preferential attachment egy kicsit máshogy

A preferential attachment modell működésének egy másik megközelítése is lehetséges. Ebben a megközelítésben a $PA_t(m, \delta)$ gráfsorozatot egy kicsit másképpen alkotjuk meg. Nem a $PA_t(1, \frac{\delta}{m})$ gráfsorozatból indulunk ki, hanem vesszük a $PA_t(m, \delta)$ -t és a következőképpen kötünk hozzá új csúcsot és új éleket. Vesszünk egy csúcsot és m darab élt, amiknek az egyik vége a csúcshoz csatlakozik, a másik végük pedig szabad. Majd az élek közül kiválasztunk egyet és ennek a szabad végét a már meglévő gráf egy csúcsához kötjük olyan valószínűségekkel, ahogy az a (2.1)-ben szerepel. Természetesen saját magához is köthetjük. Annak a csúcsnak a fokszámát, amihez az új élt kötöttük, megnöveljük eggyel. Ezután választunk egy másik szabad végű élt, ezt is hozzákötjük a gráfhoz, majd

megint megnöveljük annak a csúcsnak a fokszámát, amihez a szabad élt kötöttük. Ezt addig folytatjuk, amíg az összes szabad végű él el nem fogy. Ezt a módszert 'intermediate updating of the degrees'-nek nevezzük.

A fent leírt modell és mindkét változata úgy működik, hogy azok a csúcsok, amiknek nagyobb a fokszámuk, nagyobb eséllyel kapnak új éleket, mint egyéb csúcsok, aminek következtében még nagyobb valószínűséggel kapcsolódnak hozzájuk újabb élek, vagyis a folyamat öngerjesztő. Szokták néha ezt a modellt 'Rich-get-Richer' modellnek is nevezni, mivel megmagyarázza, hogy miért találhatóak nagyon nagy fokszámú csúcsok is a véletlen gráfokban.

2.1.3. Hurokélmentes modell

A modellt megváltoztathatjuk úgy, hogy hurokélek ne forduljanak elő a gráfokban. Ezt úgy tesszük, hogy módosítjuk a kiindulási feltételeket és azt is, hogyan köthető egy új, szabad végű él a csúcsokhoz. Ezt a módosított gráfsorozatot $\{PA_t^{(b)}(m, \delta)\}_{t \geq 2}$ -vel jelöljük. A legelső gráf $m = 1$ -re, vagyis $PA_2(1, \delta)$ álljon két csúcsból, $v_1(1)$ és $v_2(1)$ -ből és fússon köztük két él. Az új él hozzáadásának szabályát pedig módosítsuk így:

$$\mathbb{P}(v_{t+1}^{(1)} \rightarrow v_i^{(1)} | PA_t^{(b)}(1, \delta)) = \frac{D_i(t) + \delta}{t(2 + \delta)}.$$

Ezáltal mindig egy olyan gráfot kapunk, ami hurokélmentes és összefüggő. Ezek után definiálhatjuk a gráfsorozatot $m \geq 2$ -re a 'collapsing of vertices' módszerrel.

Nem kell a párhuzamos élek eltüntetésével bajlódnunk, ha $m = 1$, hiszen ebben az esetben ezek nem alakulhatnak ki. És nemcsak egyszerű gráfot kapunk az $m = 1$ esetben, hanem ez a gráf még fa is lesz. Ez könnyen látható, hiszen minden új csúcs csak egy korábbi csúcscsal lesz összeköttetésben, így nem alakulhat ki a gráfban kör. A továbbiakban az $m = 1$ eset fog minket érdekelni, mivel az Albert–Barabási-fa tulajdonságait fogjuk a következőkben megvizsgálni.

2.2. A modell bemutatása - Barabási–Albert-modell

Ezt a modellt Barabási Albert-László és tanítványa, Albert Réka dolgozták ki 1999-ben [1]. Barabási Albert-László fizikus és hálózatkutató Csíkszeredától 23 km-re északra, Karcfalván született Romániában. Az Indiana állambeli Notre Dame Egyetem professzora volt egészen 2007-ig, jelenleg Bostonban a Northeastern Egyetemen, illetve a Harvardon dolgozik. A hálózatok területén elért eredményeit a szociológia és a természettudományok különböző területein hasznosítják. Eredményei nem csak publikációk, hanem könyvek formájában is elérhetőek, a legismertebb ezek közül a 'Behálózva - A hálózatok új tudománya' címet viseli. Lássuk hát az említett modellt.

A Barabási–Albert-modell egy olyan preferential attachment modell, ahol $\delta = 0$. A Barabási–Albert-modellben az egy lépésben hozzáadott élek száma bármilyen pozitív egész szám lehet, azaz $m = \{1, 2, \dots\}$. Ha csak $m = 1$ élt adunk hozzá a gráfhoz minden egyes lépésben, akkor Barabási–Albert-fáról beszélünk. A Barabási–Albert-fa tehát a preferential attachment modellnek azon speciális esete, amikor $\delta = 0$ és $m = 1$. Ezt a Barabási–Albert-fát tudjuk általánosítani azzal, hogy megengedjük azt, hogy a δ nem csak nulla, hanem más is lehet. Ekkor általánosított Barabási–Albert-fáról beszélünk. A következőkben az általánosított Barabási–Albert-fáról lesz szó, de előtte nézzük még meg, hogy a Barabási–Albert-fát hogy hozzuk létre. Ez a módszer adja majd az általánosított Barabási–Albert-fa elkészítésének alapját.

A folyamat kezdetén egyetlen csúcsunk van, amit gyökérnek nevezünk. Egy általános lépés úgy néz ki, hogy választunk egy már meglévő csúcsot a gráfból, ehhez rögzítjük az új élünk egyik végét, majd a másik végét az újonnan hozzáadott csúcshoz kötjük. A bekötött csúcs annak a csúcsnak a gyereke, amihez bekötötték. Egy gráfban lévő csúcsot akkora valószínűséggel választunk, amekkora a fokszámának és a gráf összfokszámának aránya és az új csúcs bekötése minden korábbi lépéstől függetlenül tör-

ténik.

2.2.1. Az általánosított Barabási–Albert-fa felépítésének módjai

Az általánosított Barabási–Albert-fát kétféleképpen is felépíthetjük. Következzen most az egyik módszer.

Úgy kezdődik egy új csúcs hozzáadása a gráfhoz, hogy kiválasztunk a már meglévő csúcsokból egy csúcsot a súlyával arányosan, ahol egy k fokszámú csúcs súlya $k + \delta$, $\delta \geq -1$ pedig a modellre jellemző paraméter. Tehát egy konkrét csúcsot $\frac{k+\delta}{S_n}$ valószínűséggel választunk. S_n -nel jelöljük az n élű gráf csúcsainak összsúlyát, ami $S_n = 2n + (n + 1)\delta$ -val egyenlő. Majd a kiválasztott csúcsot hozzákötjük az új csúcshoz. Az új csúcs bekötése mindig az előző csúcsok bekötésétől függetlenül történik.

Van egy másik módja a Barabási–Albert-fa felépítésének, ami a következőképpen néz ki. Magát a gyökeret $\frac{\delta}{S_n}$ valószínűséggel válasszuk ki. Egyébként pedig válasszunk egy élt a gráfbeli élek közül egyenletesen. Ezután válasszuk ki ennek az élnek a felső végpontját $\frac{1+\delta}{2+\delta}$ valószínűséggel, az alsó végpontját $\frac{1}{2+\delta}$ valószínűséggel. Beszélhetünk az él alsó és felső végpontjáról, mivel a gráfunk fa, és így minden csúcsnak egy és csakis egy távolsága van a gyökértől számítva. Ez alapján egy él alsó végpontja az a végpont, amelyik közelebb van a gyökérhez, a felső végpont pedig az, amelyik távolabb van a gyökértől. Nézzük meg, hogy ezzel a módszerrel ugyanazokat a súlyokat kapják-e a csúcsok, mint amiket az előző eljárás során.

Ehhez először alakítsuk át az összfokszámra kapott képletünket:

$$S_n = 2n + (n + 1)\delta = (2 + \delta)n + \delta.$$

Vizsgáljuk meg a módszert egy 1 fokszámú csúcstra. Ekkor a csúcsnak csak egy éle van és ahhoz a csúcshoz kapcsolódik, amelyiknek ő a gyereke. Tehát az él felső végpontjánál helyezkedik el, ez pedig azt jelenti, hogy ha ezt az élt választjuk, akkor $\frac{1+\delta}{2+\delta}$ valószínűséggel választjuk ezt a csúcsot.

Az élt $\frac{2+\delta}{S_n}$ valószínűséggel választjuk, hiszen egy él $2 + \delta$ -nyi súlyt ad hozzá a gráfhoz. Mivel az él és a csúcs választásának is meg kell történnie és ezek egymástól függetlenek, ezért az él választásának valószínűségét összeszorozva a csúcs választásának valószínűségével azt kapjuk, hogy az egy fokszámú csúcsunkat

$$\frac{1 + \delta}{2 + \delta} \cdot \frac{2 + \delta}{n(2 + \delta) + \delta} = \frac{1 + \delta}{n(2 + \delta) + \delta}$$

valószínűséggel választjuk, tehát a csúcs súlya $1 + \delta$. Épp ennek kellett kijönnie a modell alapján.

Egy $k > 1$ fokszámú csúcsra pedig a következő a helyzet. A gyökéren kívül minden csúcsnak egy és csakis egy szülője van, vagyis egyetlen olyan éle van, aminek a felső végpontjánál helyezkedik el. Ezért az előzőek alapján a csúcs súlya eddig $1 + \delta$. A szülőjéhez futó élen kívül a csúcsnak van még $k - 1$ darab éle, amik az ő gyerekeihez futnak, tehát ezeknél az éleknél ő van az alsó végpontoknál. Ezek az élek összesen $k - 1$ nagyságú súlyt jelentenek. Vagyis a csúcs összsúlya $1 + \delta + (k - 1) = k + \delta$, úgy, ahogy azt vártuk. Ez a gyökér súlyánál is ennyi, mivel van k darab éle és kap hozzá még δ -nyi súlyt, így a kettő összege szintén $k + \delta$.

Ezek tehát az általánosított Barabási–Albert-fa felépítésének módja. Ezek a módszerek a korábban látottak alapján egy egyszerű gráfot, azaz hurokélektől és párhuzamos élektől mentes gráfot eredményeznek.

3. fejezet

Fokszámok vizsgálata véletlen gráfokban

3.1. Rögzített csúcsok fokszámainak vizsgálata

Ahhoz, hogy adott csúcsok fokszámai vizsgálhassuk, be kell vezetnünk a Γ -függvényt:

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx,$$

ahol $t > 0$. Azt fogjuk belátni, hogy $D_i(t)$ arányos $t^{\frac{1}{2+\delta}}$ -val és ez az arány egy pozitív ξ valószínűségi változóhoz konvergál. Pontosabban megfogalmazva [6, 8.1 tétel]:

1. tétel. *Adott $m = 1$ és $\delta \geq -1$. Ekkor $D_i(t)t^{-\frac{1}{2+\delta}}$ egy pozitív ξ_i valószínűségi változóhoz konvergál majdnem mindenütt, és*

$$\mathbb{E}(d_i(t) + \delta) = (1 + \delta) \cdot \frac{\Gamma(t+1)\Gamma(i - \frac{1}{2+\delta})}{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(i)}$$

teljesül, ha $t \rightarrow \infty$.

Ez a tétel kiterjeszthető a maximális fokszám konvergálására is.

Bizonyítás: Vegyük az egyenlet bal oldalán található várható értéket és

vegyük feltételesen az i csúcs fokszámára, azaz $D_i(t)$ -re. Így ez adódik:

$$\begin{aligned}\mathbb{E}(D_i(t+1) + \delta | D_i(t)) &= D_i(t) + \delta + \mathbb{E}(D_i(t+1) - D_i(t) | D_i(t)) = \\ &= D_i(t) + \delta + \frac{D_i(t) + \delta}{t(2+\delta) + 1 + \delta} = \\ &= (D_i(t) + \delta) \frac{t(2+\delta) + 2 + \delta}{t(2+\delta) + 1 + \delta} = \\ &= (D_i(t) + \delta) \frac{(2+\delta)(t+1)}{t(2+\delta) + 1 + \delta}\end{aligned}$$

Kiszámoljuk az i csúcs fokszámának várható értékét is:

$$\begin{aligned}\mathbb{E}(D_i(i) + \delta) &= 1 + \delta + \frac{1 + \delta}{(2+\delta)(i-1) + 1 + \delta} = \\ &= (1 + \delta) \frac{(2+\delta)(i-1) + 2 + \delta}{(2+\delta)(i-1) + 1 + \delta} = \\ &= (1 + \delta) \frac{(2+\delta)i}{(2+\delta)(i-1) + 1 + \delta}.\end{aligned}$$

Ezek segítségével a következő martingált határozzuk meg:

$$M_i(t) = \frac{D_i(t) + \delta}{1 + \delta} \prod_{s=i-1}^{t-1} \frac{(2+\delta)s + 1 + \delta}{(2+\delta)(s+1)}.$$

Ez az $M_i(t)$ egy nemnegatív martingál, aminek a várható értéke 1. A konvergencia-tétel miatt $M_i(t)$ egy ξ_i valószínűségi változóhoz konvergál 1 valószínűséggel, ha $t \rightarrow \infty$.

Az $M_i(t)$ martingál determinisztikus részét tovább alakíthatjuk a következő módon:

$$\prod_{s=i-1}^{t-1} \frac{(2+\delta)s + 1 + \delta}{(2+\delta)s + 2 + \delta} = \prod_{s=i-1}^{t-1} \frac{s + \frac{1+\delta}{2+\delta}}{s+1} = \frac{\Gamma(t + \frac{1+\delta}{2+\delta})\Gamma(i)}{\Gamma(t+1)\Gamma(i - \frac{1}{2+\delta})}.$$

Belátható, hogy $D_i(t)/t^{\frac{1}{2+\delta}}$ egy valószínűségi változóhoz konvergál eloszlásban, és ennek a valószínűségi változónak $(1 + \delta) \frac{\Gamma(i - \frac{1}{2+\delta})}{\Gamma(i)}$ a várható értéke.

Ez az eredmény azért lesz fontos, mert a későbbiekben a Barabási–Albert-fa lokális fokszámeloszlását fogjuk vizsgálni. Pontosabban a gyökér szomszédságában számítjuk majd ki a k fokú csúcsok arányát, a tétel pedig pont ennek a halmaznak a méretéről szól.

3.2. A preferential attachment modellből adódó fokszámsorozatok

Jelöljük $P_k(t)$ -vel a t időpontban k fokkal rendelkező csúcsok arányát, azaz $P_k(t)$ legyen

$$P_k(t) = \frac{1}{t} \sum_{i=1}^t \mathbb{I}_{\{D_i(t)=k\}}.$$

Jelöljük $\{p_k\}_{k=0}^{\infty}$ -nel azt az eloszlást, amit a

$$p_k = \left(2 + \frac{\delta}{m}\right) \frac{\Gamma(k + \delta)\Gamma(m + 2 + \delta + \frac{\delta}{m})}{\Gamma(m + \delta)\Gamma(k + 3 + \delta + \frac{\delta}{m})}.$$

képlet határoz meg $m \geq k$ -ra, $k \in \{0, 1, 2, \dots, m - 1\}$ -re pedig $p_k = 0$. Mielőtt összefüggést találnánk a k fokszámú csúcsok aránya és a $\{p_k\}_{k=1}^{\infty}$ eloszlás között, említsük meg a p_k eloszlás alakját $m = 1$ -re. Ekkor a következőre egyszerűsödik az eloszlás:

$$p_k = (2 + \delta) \frac{\Gamma(k + \delta)\Gamma(3 + 2\delta)}{\Gamma(k + 3 + 2\delta)\Gamma(1 + \delta)}.$$

Ez a képlet fog előjönni, amikor a Barabási–Albert-fa fokszámeloszlását tárgyaljuk. Térjünk vissza a k fokú csúcsok arányához: belátható, hogy $\{p_k\}_{k=1}^{\infty}$ valóban eloszlás és ez az eloszlás a $PA_t(m, \delta)$ fokszámeloszlása. Ezt a következő tétel mondja [6, 8.2 tétel].

2. tétel. *Adott $\delta > -m$ és $m \geq 1$. Ekkor létezik olyan m -től és δ -tól függő $C = C(m, \delta) > 0$ konstans, hogy $t \rightarrow \infty$ esetén*

$$\mathbb{P} \left(\max_k |P_k(t) - p_k| \geq C \sqrt{\frac{\log t}{t}} \right) = o(1)$$

teljesül.

A tétel szerint annak a valószínűsége, hogy $P_k(t)$ és p_k különbségének maximuma nagyobb, mint egy nullához tartó függvény értéke, tart a nullához $t \rightarrow \infty$ esetén. Vagyis a fokszámok aránya, azaz $P_k(t)$ a $\{p_k\}_{k=1}^{\infty}$ fokszámeloszláshoz tart, ahogy a gráf a preferential attachment szabályai

szerint növekszik.

Az az előnye ennek a képletnek, hogy kapcsolatot teremt a véletlen és a determinisztikus között. A k fokú csúcsok arányát, azaz $P_k(t)$ -t, teljes egészében a véletlen irányítja, akár teljesen el is térhet a várt értékektől. A p_k eloszlás ezzel ellentétben egy determinisztikus, előre meghatározható eloszlás, és ehhez konvergál hosszú távon a véletlen szabályai alapján működő $P_k(t)$.

A p_k eloszlás ezen kívül még egy hatványrendű eloszláshoz is közel van. Ezt onnan tudjuk, hogy $k \rightarrow \infty$ esetén

$$p_k = c_{m,\delta} k^{-\gamma} (1 + O(1/k))$$

igaz, ahol $\gamma = 3 + \frac{\delta}{m} > 2$. Vagyis a 'preferential attachment' modellel létrehozott gráfok fokszámsorozata hatványrendű, $\gamma = 3 + \frac{\delta}{m}$ kitevővel. Bármilyen $\gamma > 2$ elérhető, ha δ -t és m -et megfelelően választjuk.

Ezért van az, hogy a preferential attachment modell jó bizonyítékot ad arra, hogy miért is jelennek meg a hatványrendű fokszámsorozatok.

3.3. A Barabási–Albert-fa fokszámeloszlása

A Barabási–Albert-fa fokszámeloszlása valóban az előző pontban említett p_k eloszlás $m = 1$ esete, ezt többek között Móri Tamás [8] és Bollobás Béla [3] bizonyította be. Belátták, hogy a k fokszámú csúcsok aránya a Barabási–Albert-fában egy q_i eloszláshoz konvergál, és

$$q_i = \frac{2 + \delta}{i + \delta} \prod_{j=1}^i \frac{j + \delta}{j + 2 + 2\delta} \sim (2 + \delta) \frac{\Gamma(2\delta + 3)}{\Gamma(\delta + 1)} \cdot \frac{1}{i^{\delta+3}} \quad (i \rightarrow \infty).$$

Ez a p_k fokszámeloszlás $m = 1$ esete, hiszen $\frac{\Gamma(k+\delta)}{\Gamma(k+3+2\delta)} t^{-\frac{1}{\delta+3}}$. Katona Zsolt [7] pedig azt bizonyította be, hogy ugyanez az eloszlás figyelhető meg a legnagyobb méretű szinteken is. A gyökértől azonos távolságra lévő csúcsok alkotnak egy szintet.

A $\{p_k\}_{k=1}^{\infty}$ fokszámeloszlás egy másik egyszerűsített alakja $k \geq m$ és

$\delta = 0$ esetén érhető el. Ekkor:

$$p_k = \frac{2\Gamma(k)\Gamma(m+2)}{\Gamma(k+3)\Gamma(m)} = \frac{2m(m+1)}{k(k+1)(k+2)}$$

Ezt és az előző képletet összerakva kapjuk a Barabási–Albert-fa fokszám-eloszlását. Azaz, ha $\delta = 0$ és $m = 1$, akkor:

$$p_k = 2 \frac{\Gamma(k)\Gamma(3)}{\Gamma(k+3)\Gamma(1)} = 2 \frac{2}{(k+2)(k+1)k} = \frac{4}{(k+2)(k+1)k}.$$

A levezetéshez felhasználtuk a

$$\Gamma(k+1) = k\Gamma(k)$$

összefüggést, illetve az ebből következő

$$\Gamma(k) = (k-1)!$$

egyenlőséget.

Tehát a Barabási–Albert-fa fokszám-eloszlása $\frac{4}{k(k+1)(k+2)}$ -höz tart, vagyis konyhanyelven fogalmazva: nagyon sok lépés után $\frac{4}{k^3}$ valószínűséggel választunk ki egy k fokszámú csúcsot, ha egyenletesen választunk a csúcsok közül. Vagyis a Barabási–Albert-fa skálafüggetlen $\gamma = 3$ kitevővel.

3.4. A Barabási–Albert-fa alsó szintjei

Ahhoz, hogy a Barabási–Albert-fa szintjeinek nagyságáról beszélhessünk, először is be kell vezetnünk a szint fogalmát. Már tudjuk, hogy ha két csúcs össze van kötve, akkor a régebbi csúcs az újabbnak a szülője, az újabb csúcs pedig a régebbinek a gyereke. Beszélhetünk régebbi és újabb csúcsokról, ha azt nézzük, hogy mikor került a csúcs a gráfba. Azt mondjuk, hogy a gráfban néhány csúcs egy szintet alkot, ha ugyanolyan távol, vagyis ugyanannyi lépésre vannak a gyökércsúcstól. A nulladik szint csak a gyökeret tartalmazza. Az első szintet a gyökér gyerekei alkotják. A második szintet a gyökércsúcs gyerekeinek a gyerekei, vagyis az első szint

csúcsainak a gyerekei alkotják. A harmadik szintet a második szint gyerekei alkotják, és így tovább. Fában vagyunk, ahol bármely két csúcsot pontosan egy út köt össze, ezért minden csúcsnak csak egy távolsága lehet a gyökértől. Jelöljük $X[n, k]$ -val az n lépés után kialakult k -edik szint nagyságát. A fa szélessége pedig legyen a legnagyobb szint nagysága és jelöljük $W_n = \max\{X[n, k] : 1 \leq k \leq n\}$ -nel.

A szint fogalmán kívül be kell még vezetnünk egy szint összsúlyának fogalmát is. Legyen $W[n, k]$ a k -edik szintet alkotó csúcsok összsúlya. Ez a gyökér esetén $W[n, 0] = X[n, 1] + \delta$, mivel a gyökérből kilépő összes élnek 1 súlya van és ehhez jön még magának a gyökérnek a δ súlya. Egy szint súlya $k > 0$ -ra pedig: $W[n, k] = X[n, k + 1] + (1 + \delta)X[n, k]$, mert a k -edik szintből $X[n, k + 1]$ számú él lép ki, itt mindegyik él súlya 1, és a k -edik szinten $X[n, k]$ darab csúcs található, ezek pedig a $(k - 1)$ -edik szintről befutó élek felső végpontjánál helyezkednek el, ezért ezek $1 + \delta$ súllyal szerepelnek. Most már meghatározhatjuk, hogy mekkora eséllyel fog nőni a k -edik szint mérete a Barabási–Albert fában.

3.4.1. A Barabási–Albert-fa alsó szintjeinek nagysága

Legyen \mathcal{F}_n az első n lépés által generált σ -algebra. Ha erre a σ -algebrára nézzük a k -edik szint növekedésének valószínűségét, akkor azt kapjuk, hogy ez:

$$\begin{aligned} P([n + 1, k] = X[n, k] + 1 | \mathcal{F}_n) &= 1 - P(X[n + 1, k] = X[n, k] | \mathcal{F}_n) = \\ &= \frac{W[n, k - 1]}{S_n}, \end{aligned}$$

vagyis a $(k - 1)$ -edik szint csúcsainak az összsúlya és a gráf csúcsainak az összsúlyának az aránya jön ki. Ez logikus is, hiszen egy új csúcs csak akkor lesz a k -edik szinten, ha egy $(k - 1)$ -edik szintű csúcshoz csatlakozik. Annak a valószínűsége pedig, hogy egy ilyen csúcshoz csatlakozik, az pont az előbb kijött valószínűség. Lássuk akkor a k -edik szint nagyságáról szóló tételt [9, 2.1. tétel].

3. tétel. *Rögzített k -ra*

$$X[n, k] \frac{\zeta}{(k-1)!} \mu(n)^{k-1} n^{\frac{1}{2+\delta}} = n\zeta \text{Poi}_{k-1}[\mu(n)]$$

1 valószínűséggel, ha $n \rightarrow \infty$.

ζ egy pozitív valószínűségi változó, $\mu(n) = \frac{1+\delta}{2+\delta} \log n$, a $\text{Poi}_{k-1}[\mu(n)]$ pedig a $\mu(n)$ paraméterű Poisson-eloszlás $(k-1)$ -edik tagja. Ez a tétel is hasonló dolgot állít, mint a rögzített csúcsok fokszámáról szóló tétel. Ott egy adott csúcs fokszáma volt arányos $n^{\frac{1}{2+\delta}}$ -val, itt pedig az egész k -adik szint mérete arányos $n^{\frac{1}{2+\delta}}$ -val.

3.4.2. A legnagyobb szinteken

Katona Zsolt bebizonyította [7], hogy nem csak rögzített k -ra érvényes az előző tételben kimondott becslés, hanem k -ban egyenletesen teljesül. Arra az eredményre jutott, hogy

$$X[n, k] \sim e^{-\nu}$$

k -ban egyenletesen, ahol $\nu = \frac{(k-\mu(n))^2}{2\mu(n)}$. Meghatározta ezáltal a legnagyobb szint súlyát is, arra jutott, hogy $W_n = \frac{n}{\sqrt{2\pi\mu(n)}} \sim 1$ valószínűséggel.

A 3. tételben megjelenő ζ valószínűségi változó azonban érdekes módon eltűnik, ha a legnagyobb szinteket vizsgáljuk. Ez a következő tételből derül majd ki.

Legyen $X[n, k, i]$ az i fokszámú csúcsok száma n lépés után a k -adik szinten. Katona Zsolt bizonyította be azt is, hogy a legnagyobb szinteken az aszimptotikus fokszámeloszlás ugyanaz, mint az egész fában. Vagyis

$$X[n, k, i] = q_i X[n, k] + \mathcal{O}\left(\frac{n}{\log n}\right) \quad \text{minden } i \in 1, 2, \dots\text{-re.}$$

3.4.3. Lokális fokszámeloszlás az általánosított Albert–Barabási-fában

Az alacsonyabb szintek ennek ellenére máshogy viselkednek, ezt a következő tétel mondja ki [9, 3.1. tétel].

4. tétel. Minden $k \geq 1$, $i \geq 1$ -re

$$\lim_{n \rightarrow \infty} \frac{X[n, k, i]}{X[n, k]} = \frac{1 + \delta}{(i + \delta)(i + 1 + \delta)} = (1 + \delta) \left(\frac{1}{i + \delta} - \frac{1}{i + 1 + \delta} \right)$$

1 valószínűséggel.

Ez az arány nem azonos a q_i aránnyal, ami az egész fa i fokú csúcsainak arányát mutatja meg. A q_i $\frac{1}{i^{\delta+3}}$ -mal arányos, tehát hatványrendű és ez a hatvány függ a modell paraméterétől. Ezzel szemben az alacsonyabb szinteken a foksámeloszlás hatványrendű ugyan, de a kitevője mindig -2 , azaz nem függ a modell paraméterétől, δ -tól.

3.5. Valós hálózatok

3.5.1. Hat lépés távolság

Egy bizonyos Stanley Milgram, amerikai pszichológus érdekes kísérletet bonyolított le 1967-ben. Kansas állam különböző részeibe összesen 60 levelet küldött ki azzal a kéréssel, hogy próbálják a kapott levelet eljuttatni egy Massachusetts állambeli nőhöz. A levelet viszont mindenki csak olyan személynek adhatta tovább, akit személyesen is ismer. Ezzel próbálta meg Milgram felmérni, hogy hány személyes ismeretségen keresztül lehet kapcsolatot teremteni két tetszőleges egyén között.

Az előbbi kísérlet még nem járt nagy sikerrel, de Milgram megismételte ezt a kísérletet többször is, például úgy, hogy több információt adott meg a célszemélyről, vagy fontosnak tüntette fel a kézbesítendő levelet. Ilyen és hasonló módszerekkel sikerült elérnie, hogy a levelek 95%-a megérkezzen a kijelölt személyhez. A kísérletből Milgram azt a következtetést vonta le, hogy valószínűleg bármely két ember a világon hat személyes ismeretségen keresztül ismeri egymást. Angolul ezt a hat lépés távolságot 'six degrees of separation'-nek nevezik.

Érdekes módon nem a tudományban jelent meg először a fenti gondolat, hanem az irodalomban. Ismert hazai írónk, Karinthy Frigyes vetette papírra először ezt az ötletet 1929-ben a *Láncszemek* című novellájában.

”Annak bizonyításául, hogy a Földgolyó lakossága sokkal közelebb van egymáshoz, mindenféle tekintetben, mint ahogy valaha is volt, próbát ajánlott fel a társaság egyik tagja. Tessék egy akármilyen meghatározható egyént kijelölni a Föld másfél milliárd lakója közül, bármelyik pontján a Földnek — ő fogadást ajánl, hogy legfölbjebb öt más egyéneken keresztül, kik közül az egyik neki személyes ismerőse, kapcsolatot tud létesíteni az illetővel, csupa közvetlen — ismeretség — alapon. . . ”

Habár most már nem csak másfél milliárdan élünk ezen a bolygón, hanem majdnem ötször annyian, ez az állítás minden bizonnyal továbbra is igaz, hiszen a technika fejlődése révén a Föld lakói sokkal közelebb kerültek egymáshoz, mint amilyen közel valaha is voltak. Például a Facebook nevű közösségi hálózat, ami 900 millió felhasználót kapcsol össze a világ minden tájáról, szintén egy kis világ. Állítólag bármely két felhasználó között legfeljebb négy lépés a távolság. Ez azt jelenti, hogy az emberek ismeretségének hálózata egy olyan hálózat, ami kis világ, sőt nagyon kis világ. Tehát tényleg jogos az a felkiáltás, amit nap mint nap hallunk: ”De kicsi a világ!”.

3.5.2. Az Erdős-számok

Tudjuk, hogy híres matematikusunk, Erdős Pál rengeteg cikknek volt a szerzője illetve társszerzője. Olyan sok matematikussal dolgozott együtt, hogy barátai kitalálták az Erdős-szám fogalmát. Ez a következőképpen működik. Erdős Pál Erdős-száma 0. Akik közös cikket publikáltak Erdőssel, azoknak az Erdős-száma 1. Akik olyan személlyel publikáltak cikket, aki publikált előtte Erdőssel cikket, de közvetlenül Erdőssel nem írtak közös cikket, azoknak az Erdős-száma 2, és így tovább. A legnagyobb Erdős-szám a 13 a <http://www.oakland.edu/enp/trivia/> weboldal szerint. Ha ebben a hálózatban a matematikusokat csúcsoknak tekintjük, akkor két csúcs akkor van összekötve, ha a nekik megfelelő személyek írtak közös cikket egymással. Egy ilyen gráfban Erdős egyértelműen az egyik legnagyobb fokszámú csúcs. Ebben a gráfban a matematikusok

általában 4, 7 távolságra vannak Erdőstől.

Hasonló gráfot lehet felépíteni a színészek között is. A színészek legyenek a gráf csúcsai és két csúcs között fusson él, ha a nekik megfelelő színészek játszottak egy filmben.

Utószó

Megtudtuk tehát, hogy mik is azok a véletlen gráfok. Volt szó az egyik legelső véletlen gráfról, az Erdős–Rényi gráfról, annak fázisátmenetéről és komponenseinek nagyságáról. Kiderült azonban, hogy mivel nem skálafüggetlen ez a gráf, így nem lehet jól használni valós hálózatok modellezésére.

Ezután megismerkedtünk a preferential attachment modellel. Észre vettük, hogy az ez alapján létrehozott véletlen gráfok skálafüggetlenek és ezáltal hasonlítanak a valós hálózatokra. Definiáltuk ezt a dinamikus modellt $m = 1$ -re illetve $m > 1$ -re is, ahol m az egy lépésben a gráfhoz adott új élek száma volt. Megtudtuk azt is, hogy egy rögzített csúcs fokszáma egy pozitív ξ valószínűségi változóhoz tart, ahogy a gráf méretét növeljük.

Majd rögzítettük a preferential attachment modell bizonyos paramétereit, így jutottunk el a Barabási–Albert-fához és annak általánosításához. Megvizsgáltuk a k fokszámú csúcsok arányát a gráfban, valamint az alsó szintek nagyságát is. Így jutottunk arra az eredményre, hogy bár a legnagyobb szinteken a fokszámeloszlás ugyanaz, mint az egész gráfban, az alsó szinteken ez nem így van. Ott a fokszámeloszlás hatványrendű, de nem függ a modell paraméterétől.

Irodalomjegyzék

- [1] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286** (1999), no. 5439, 509–512.
- [2] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Combin.* **1** (1980), no. 4, 311–316.
- [3] B. Bollobás et al., The degree sequence of a scale-free random graph process, *Random Structures Algorithms* **18** (2001), no. 3, 279–290.
- [4] P. Erdős and A. Rényi, On random graphs. I, *Publ. Math. Debrecen* **6** (1959), 290–297.
- [5] E. N. Gilbert, Random graphs, *Ann. Math. Statist.* **30** (1959), 1141–1144.
- [6] R. van der Hofstad, *Random Graphs and Complex Networks*. Előkészületben.
- [7] Zs. Katona, Levels of a scale-free tree, *Random Struct. Algorithms*, **29**(2), 194–207.
- [8] T. F. Móri, On random trees, *Studia Sci. Math. Hungar.* **39** (2002), no. 1-2, 143–155.
- [9] T. F. Móri, A surprising property of the Barabási-Albert random tree, *Studia Sci. Math. Hungar.* **43** (2006), no. 2, 265–273.

- [10] G. U. Yule. A mathematical theory of evolution, based on the conclusions of dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London*, Ser. B **213** (402-410):21-87, 1925.