

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

TERMÉSZETTUDOMÁNYI KAR

Kovács Bálint

**GRÁFLIMESZEK ÉS A SZTOCHASZTIKUS
BLOKKMODELL KAPCSOLATA**

Szakdolgozat

Matematika BSc

matematikus specializáció

Témavezető:

dr. Backhausz Ágnes

egyetemi adjunktus

Valószínűségelméleti és Statisztika Tanszék



Budapest, 2022

Köszönetnyilvánítás

Szeretnék ezúton is köszönetet mondani a *témavezetőmnek*, dr. Backhausz Ágnesnek, hogy megismertette velem ezt a témát, hasznos észrevételekkel, tanácsokkal látott el munkám során, és hogy amikor kérdésem volt, mindig számíthattam gyors, alapos és körültekintő válaszára. Köszönöm, hogy a Valószínűségszámítás előadások során megszerettette velem ezt a területet.

Köszönöm továbbá a *családomnak* a folyamatos támogatást, amit az alapképzés három éve alatt nyújtottak számomra, és a nehéz időszakokban is mindig biztattak a folytatásra.

Köszönettel tartozom végül, de nem utolsó sorban az általános és középiskolai *matematikatanáraitmnak*, akik elindítottak ezen a hosszú úton. Kiemelném közülük *Édesapámat*, akinek a matematika iránti – olykor fanatikus – szeretete az egyetemi éveim alatt is sokszor ösztönzőleg hatott rám.

Köszönöm.

Minden, ami emberi, akár rossz, akár
jó, előbb-utóbb véget ér... kivéve a
matematikát.

Erdős Pál

Tartalomjegyzék

Jelölések	6
Bevezetés	7
1. Gráfimeszek	9
1.1. Motiváció	9
1.1.1. Gráftulajdonságok	9
1.1.2. Információkinyerés óriási hálózatokból	11
1.2. Gráfhomomorfizmusok	13
1.3. Magfüggvények és grafonok	15
1.4. Távolságfogalmak	21
1.4.1. Gráfok vágástávolsága	22
1.4.2. Magfüggvények távolságai	25
1.5. A grafontér kompaktsága	28
1.6. Gráfsorozatok konvergenciája	35
1.6.1. Grafonok mint gráfsorozatok határértéke	39
2. Véletlen gráfok. Klasszikus modellek	43
2.1. Az Erdős-Rényi-féle véletlen gráf	44
2.2. Grafonból sorsolt véletlen gráf	50
2.2.1. Véletlen gráfmodellek mint gráfsorozatok határértéke	52
2.3. Preferential attachement-gráfok	55
2.4. Sztochasztikus blokkmodell	61
2.4.1. Visszanyerési követelmények	65

2.4.2. Néhány topologikus tulajdonság	68
3. Sztochasztikus blokkmodell bővítése a preferential attachment-szabály szerint	70
3.1. A modell leírása	70
3.2. A struktúra vizsgálata nagy n érték esetén	71
3.2.1. Élsűrűség	71
3.2.2. Háromszögek	73
Függelék	78

Jelölések

\mathbb{N}	a természetes számok halmaza
\mathbb{N}^+	a pozitív egész számok halmaza
\mathbb{Z}	az egész számok halmaza
\mathbb{R}	a valós számok halmaza
$[n]$	az $\{1, 2, \dots, n\}$ halmaz, ha $n \in \mathbb{N}^+$
2^H	a H halmaz részhalmazainak halmaza
\sqcup	diszjunkt unió
S_n	az n -edrendű permutációcsoport ($n \in \mathbb{N}^+$)
$a \gg b$	a sokkal nagyobb, mint b (pl. $999999 \gg 1$)
$\lambda_n(\cdot)$	az \mathbb{R}^n feletti Lebesgue-mérték ($n \in \mathbb{N}^+$). Ha a kontextusból kiderül a dimenzió, akkor néha $\lambda(\cdot)$
$\mathcal{L}(A)$	az A -beli Lebesgue-mérhető halmazok halmaza, ahol $A \subset \mathbb{R}^n$ valamely alkalmas $n \in \mathbb{N}^+$ -ra
$\mathbb{P}(E)$	az E esemény valószínűsége (itt \mathbb{P} egy <i>valószínűségi mérték</i>)
$\mathbb{E}(X)$	az X valószínűségi változó várható értéke
$D^2(X)$	az X valószínűségi változó szórásnégyzete
$\stackrel{d}{=}$	eloszlásbeli egyenlőség
\bar{E}	az E esemény komplementere
\mathcal{U}	az egyenletes eloszlás
$\text{Bin}(n, p)$	az $(n, p) \in \mathbb{N} \times [0, 1]$ paraméterű binomiális eloszlás
\ln	az e alapú logaritmusfüggvény
$\text{ran } f$	az $f: A \rightarrow B$ függvény képhalmaza: $\text{ran } f = \{f(x) \mid x \in A\} \subset B$
f^+	az f valós függvény pozitív része, azaz $\max(f, 0)$
f^-	az f valós függvény negatív része, azaz $\max(-f, 0)$
$\mathcal{M}_n(\mathbb{K})$	a \mathbb{K} test feletti $n \times n$ -es mátrixok halmaza ($n \in \mathbb{N}^+$)
$\mathbb{K}^{n \times m}$	a \mathbb{K} test feletti $n \times m$ -es mátrixok halmaza ($n, m \in \mathbb{N}^+$, $n \neq m$)
$V(G)$	a G gráf csúcshalmaza
$E(G)$	a G gráf élhalmaza
$v(G)$	a G gráf csúcsainak száma
$e(G)$	a G gráf éleinek száma
χ_A	az A halmaz indikátorfüggvénye
$\text{Aut}(G)$	a G gráf automorfizmusainak halmaza

Bevezetés

A bonyolult hálózatok tanulmányozása egyre fontosabb szerepet játszik a tudományban. Ilyen hálózatokra jó példa lehet az elektromos hálózatok, a telekommunikációs hálózatok, a világháló és az internet, a szociális kapcsolatok, az agyi neuronok hálózata, de akár a tudósok cikkeinek hivatkozási hálója is. Szükség van arra, hogy megértsük ezeket a struktúrákat és a rajtuk zajló folyamatokat. Ezek tanulmányozásával választ kaphatunk olyan érdekes kérdésekre, mint hogy milyen gyorsan tud egy információ vagy vírus terjedni egy (kisebb vagy nagyobb) társaságban. Mi a legrosszabb társadalmi struktúra egy járványkitörés esetén? Kit oltunk be először? Mikor tetőzik a járvány? Mennyi ideig tart? Mennyire sebezhető a hálózat (pl. telekommunikáció) egy támadás esetén?

A modern számítógépek fejlődése pedig egyre nagyobb érdeklődést váltott ki az ilyen óriási hálózatok kutatása iránt. A megnövekedett számítási kapacitásnak köszönhetően ugyanis sokkal könnyebben tárolhatók és dolgozhatók fel nagy adathalmazok, ami jelentős hatással volt a kapcsolódó kutatásokra. Egy közös tulajdonsága ezeknek a hálózatoknak, hogy *hatalmasak*. A méretük és a bonyolultságuk miatt pedig nagyon nehéz őket globálisan leírni, megérteni. Egy módszer ezen nehézségek leküzdésére, hogy a globális összkép helyett csak lokális tulajdonságokat vizsgálunk. Ehhez pedig a *véletlenszerűség*, a valószínűségszámítás szolgáltat igencsak hatásos eszközöket. Például a hálózat teljes struktúrájának ismerete helyett vizsgálhatjuk, hogy egy egyenletesen választott csúcsnak hány szomszédja van, mennyi a várható élsűrűség, stb.

Ezek a módszerek vezetnek a *véletlen gráfok* fogalmához. Egy n pontú véletlen gráf alatt egy, az n pontú gráfok családján vett valószínűségeloszlást értünk, azaz egy ilyen gráf szerkezete a véletlentől függ. Ilyen gráfokról először Erdős Pál és Rényi Alfréd magyar matematikusok publikáltak 1959-ben, ám a 2000-es évek óta a témának igen komoly szakirodalma lett.

E dolgozat fő témái a gráfok *konvergenciafogalma*, illetve a véletlen gráfok, és ezek konvergenciája. Az első fejezetben a releváns szakirodalom alapján felépítjük a gráfok konvergenciaelméletének alapjait: definiáljuk a homomorfizmus-sűrűség különböző faj-

táit, illetve az ún. magfüggvények vagy kernelek elméletének alapjaiba vezetjük be az olvasót. Az előbbi maga a konvergenciafogalom értelmezésénél, utóbbi a határérték-objektum legkézenfekvőbb változatához lesz segítségünkre.

A második fejezet a véletlen gráfokkal foglalkozik részletesebben: a véletlen gráfok precíz definíciója után bemutatásra kerül néhány klasszikus véletlen gráfmodell, megvizsgálva ezek konvergenciáját is.

Végül a harmadik fejezetben egy két előzőleg vizsgált gráfmodell, a *sztochasztikus blokkmodell* és a *preferential attachment-modell* kombinációjából adódó véletlenül növő modell tulajdonságait kezdjük el vizsgálni elméleti számítások valamint számítógépes szimulációk segítségével. A szimulációkhoz használt számítógépes programok forráskódjait listázza a dolgozat végén található függelék.

1. fejezet

Gráfimeszek

1.1. Motiváció

Alább látni fogjuk, hogy sok valódi hálózatnak speciális sajátosságai vannak. A hálózat-kutatás célja, hogy megértsük, számos hálózat miért osztozik ezeken a tulajdonságokon, másrészt hogy ezek milyen hatással vannak például a különböző betegségek terjedésére, útvonaltervezésre, stb.

1.1.1. Gráftulajdonságok

Ebben a részben [21] alapján nagyrészt informálisan bemutatunk néhány valódi hálózaton megfigyelt tulajdonságot, hogy ezzel is illusztráljuk ezen kutatások fontosságát.

Fokszámeloszlások

A véletlen gráfok tanulmányozásában rendkívül fontos szerepe van a *fokszámeloszlásoknak*, mivel ezek több más tulajdonságot is alapvetően befolyásolnak. Ez nem más, mint a fokszámok valószínűség-eloszlása az egész gráf felett, azaz egy n csúcsú gráf esetén

$$P_k^{(n)} = \frac{1}{n} \sum_{i=1}^n \chi_{\{\deg(v_i)=k\}},$$

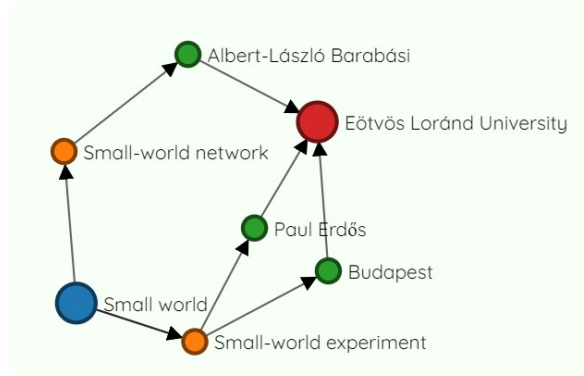
ahol $\{v_1, \dots, v_n\}$ a gráf csúcshalmaza.

A tapasztalati esetben gyakran gráfoknak egy sorozatát vizsgáljuk, és sokszor felteesszük, hogy ez *ritka*. Ha $(G_n)_{n \in \mathbb{N}}$ egy olyan gráfsorozat, hogy minden $n \in \mathbb{N}$ -re $v(G_n) = n$, akkor ez annyit jelent, hogy $P_k^{(n)} \rightarrow p_k$ (ha $n \rightarrow \infty$), ahol $(p_k)_{k \geq 0}$ egy valószínűség-eloszlás. Megjegyzendő, hogy ezt a definíciót főleg ún. *véletlen* gráfsorozatoknál használják, így ezek esetén $P_k^{(n)}$ egy valószínűségi változó, míg p_k egy determinisztikus mennyiség, tehát a konvergenciát sztochasztikusan szokás érteni. Fontos még azt is észrevenni, hogy mivel $(p_k)_{k \geq 0}$ valószínűség-eloszlás, ezért $\sum_{k=0}^{\infty} p_k = 1$, tehát nagyon nagy n -re a csúcsok többsége korlátos fokszámú lesz. Innen ered a *ritka* elnevezés.

Kisvilág-tulajdonság

Egy másik érdekes gyakran megfigyelt sajátossága a tapasztalati hálózatoknak a *kisvilág-tulajdonság*, ami nagy vonalakban azt jelenti, hogy a csomópontok "közel" vannak egymáshoz. A legjobban a szociális, ismeretségeket leíró hálózatokon képzelhető el ez a jelenség: például a Stanley Milgram pszichológustól származó híres *hat lépés távolság elmélet* azt állítja, hogy a világon bárki kapcsolatba hozható bárkivel egy ismeretségi láncon keresztül, amiben a két végpont között legtöbb 5 elem van. Egyébként tőle származik a jelenség elnevezése is. 1967-ben Milgram egy érdekes kísérletet hajtott végre: 60 levelet küldött különböző önként jelentkezőknek a kansasi Wichitába. A címzeteknek az volt a feladatuk, hogy eljuttassák a küldeményt egy hittantudomány-hallgató feleségének úgy, csak személyes ismerőseiknek adhatták át a levelet, más szóval csak a saját ismeretségi hálójukon keresztül juttathatták célba a küldeményt. Bár először csak 3 levél ért célba, ami kb. 5%-os sikerességi arány, későbbi kísérletek során ezt az arányt 95%-osra is fel tudta tornászni különböző trükkök bevetésével (pl. elhitette, hogy a küldemény nagy értékű, több információt adott meg a címzetről, stb.). A kísérletet Milgram *kisvilág-kísérletnek* nevezte (small-world experience). A további részletekért lásd [19]-t. Az ötletet Karinthy Frigyes *Láncszemek* című novellájában egyébként már 1929-ben felvetette.

Maga a kisvilág-tulajdonság a következőképpen tehető precízebbé: tekintsük a $(G_n)_{n \in \mathbb{N}}$



1.1. ábra. A *Kisvilág* és az *Eötvös Loránd Tudományegyetem* Wikipédia-oldalak közötti legrövidebb utak a hiperhivatkozásokon keresztül. Az ábra [23] segítségével készült.

gráfsorozatot, és minden $n \in \mathbb{N}$ -re legyen $\text{dist}_n(u, v)$ az $u, v \in V(G_n)$ csúcsok közötti legrövidebb út hossza a G_n gráfban (lehet ∞ is, ha az u és v között nincs út). Ekkor ha U_1, U_2 két egyenletesen választott csúcs, akkor legyen $H_n = \text{dist}_n(U_1, U_2)$ a G_n gráf *tipikus távolsága*. Ekkor a $(G_n)_{n \in \mathbb{N}}$ gráfsorozatot *kisvilágnak* nevezzük, ha létezik olyan $0 < C < \infty$ konstans, hogy $\lim_{n \rightarrow \infty} \mathbb{P}(H_n \leq C \ln n) = 1$. Ez azt jelenti, hogy a tipikus távolság logaritmikus n -ben, azaz minden egyes lépésben megtöbbszöröződik az elérhető csúcsok száma. Egyes modellekben a tipikus távolság még $\ln n$ -nél is kisebb, ami indokolja az ún. *ultra-kisvilág tulajdonság* létjogosultságát: egy gráfsorozat, ha bármely $\varepsilon > 0$ -ra $\lim_{n \rightarrow \infty} \mathbb{P}(H_n \leq \varepsilon \ln n) = 1$. Sőt, egyes esetekben még olyan $0 < C < \infty$ konstans is van, amelyre $\lim_{n \rightarrow \infty} \mathbb{P}(H_n \leq C \ln \ln n) = 1$ teljesül ([14] 1.4.3.).

1.1.2. Információkinyerés óriási hálózatokból

A "tantermi" problémákkal ellentétben a tapasztalati hálózatok gyakran nem teljesen ismertek, sőt néha még csak nem is jól definiáltak. Gyakran még a csúcsok számát sem tudjuk pontosan. Alább [16] 1.3. alfejezete alapján leírunk néhány módszert arra, hogy hogyan lehet ilyen óriási hálózatokból információt kinyerni.

Mintavételezés

Egy lehetséges megoldás, hogy egymástól függetlenül, visszatevés nélkül, egyenletesen kiválasztunk k db. csúcsot a gráfból ($k \in \mathbb{N}$), és meghatározzuk a köztük futó éleket. Így egy úgynevezett *véletlen indukált részgráfot* kapunk. (Persze mindez feltételezi, hogy van eszközünk arra, hogy az előbbi véletlen kiválasztást megvalósítsuk.) Adott $k \in \mathbb{N}$ -re minden k csúcsú gráfra megvan annak a valószínűsége, hogy a k csúcs mintavételezése során előfordul mint véletlen indukált részgráf, ez meghatároz egy valószínűség-eloszlást. Ez a minta már elég információt tartalmaz sok tulajdonság eldöntésére, persze bizonyos fokú hibával. A hiba csökkenthető k növelésével.

Ha G és G' két gráf, akkor minden rögzített k -ra összehasonlíthatjuk a k csúcsú véletlen indukált részgráfjaik eloszlását az úgynevezett *szórástávolság* segítségével ([16] 1.5). Két azonos halmazon értelmezett $\alpha, \beta: \mathcal{A} \rightarrow \mathbb{R}$ eloszlás esetén:

$$d_{\text{var}}(\alpha, \beta) = \sup_{X \in \mathcal{A}} |\alpha(X) - \beta(X)| \quad (1.1)$$

Ekkor a két gráf *mintavételi távolsága* ([16] (1.2)):

$$\delta_{\text{samp}}(G, G') = \sum_{k=1}^{\infty} \frac{1}{2^k} d_{\text{var}}(\sigma_{G,k}, \sigma_{G',k}), \quad (1.2)$$

ahol $\sigma_{G,k}$ és $\sigma_{G',k}$ a G illetve G' gráfok k csúcsú véletlen indukált részgráfjainak az eloszlásai.

Globális megfigyelések

Véletlenszerűen osszuk a csúcsokat osztályokra, azaz vegyük a csúcsok egy véletlen partícióját. A kapott osztályokat "olvasszuk egybe", tekintsünk rájuk úgy, mintha egyetlen csúcsot alkotnának. Így olyan globális tulajdonságokról nyerhetünk információkat, mint például a maximális vágás, a független csúcshalmazok száma, stb. Megjegyzendő, hogy ezek a globális megfigyelések ugyanannyi információt hordoznak, mint a lokális mintavételezés ([16] 12.3., 20.2.).

Homomorfizmusok

Elméleti kutatásokban gyakran használják a *gráfhomomorfizmusokat* (illeszkedéstartó leképezések) a véletlen mintavételezések helyett. Ennek számos előnye van, például:

- Az ún. homomorfizmusszámok (lásd később) algebrailag sokkal jobban kezelhetők.
- Egy óriási G gráf tanulmányozásakor vizsgálhatjuk a lokális tulajdonságokat úgy, hogy több kisebb L gráfot próbálunk meg beleképezni a G -be. Ugyanígy, globális tulajdonságokat vizsgálhatunk azzal, hogy a G -t képezzük több kisebb J gráfba.

$$B \rightarrow G \rightarrow J$$

1.2. Gráfhomomorfizmusok

Amint azt már korábban említettük, az ilyen hálózatok tanulmányozásakor sokszor célszerű a globális struktúra vizsgálata mellett lokális tulajdonságokat keresni. Ehhez egy hasznos eszköz a *gráfhomomorfizmusok* fogalma, amellyel ez az alfejezet foglalkozik ([16] 5. fejezete alapján).

Legyen $G = (V(G), E(G))$ és $H = (V(H), E(H))$ két egyszerű gráf. Ekkor $G \rightarrow H$ azt jelöli, hogy létezik egy $f: V(G) \rightarrow V(H)$ homomorfizmus. A homomorfizmus létezése korántsem egy triviális probléma, de ebben a dolgozatban ezzel nem foglalkozunk (részletekért lásd: [16] 5.1.).

Legyen a G és H közötti gráfhomomorfizmusok halmaza

$$\text{Hom}(G, H) = \{f: V(G) \rightarrow V(H) \mid \{f(x), f(y)\} \in E(H) \quad \forall e = \{x, y\} \in E(G)\}$$

és legyen $\text{hom}(G, H) = |\text{Hom}(G, H)|$ homomorfizmusok száma.

Ezenkívül jelölje $\text{inj}(G, H) = |\text{Inj}(G, H)|$, ahol

$$\text{Inj}(G, H) = \{f \in \text{Hom}(G, H) \mid f \text{ injektív}\}$$

az injektív homomorfizmusok számát, $\text{ind}(G, H) = |\text{Ind}(G, H)|$ pedig a G gráf *beágyazásainak* a számát H -ba, ahol

$$\text{Ind}(G, H) = \{f \in \text{Hom}(G, H) \mid f \text{ injektív és } \{f(x), f(y)\} \notin E(H) \ \forall e = \{x, y\} \notin E(G)\}$$

Néha a $\text{surj}(G, H) = |\text{Surj}(G, H)|$ (ahol $\text{Surj}(G, H) = \{f \in \text{Hom}(G, H) \mid f \text{ szürjektív}\}$) jelölés is hasznos lesz. Ez úgy értendő, hogy a csúcsokra és az élekre nézve is ki kell tenni a szürjektivitást.

Ezeket a homomorfizmusszámokat gyakran normáljuk, így eljutva a *homomorfizmus-sűrűség* fogalmához.

1.2.1. Definíció (homomorfizmus-sűrűségek). A G és H egyszerű gráfok közötti homomorfizmus-sűrűség:

$$t(G, H) = \frac{\text{hom}(G, H)}{v(H)^{v(G)}} \quad (1.3)$$

Tehát ha $f: V(G) \rightarrow V(H)$ egy egyenletesen véletlenszerűen kiválasztott függvény, akkor $t(G, H) = \mathbb{P}(f \in \text{Hom}(G, H))$. Hasonlóan,

$$t_{\text{inj}}(G, H) = \frac{\text{inj}(G, H)}{v(H)!/(v(H) - v(G))!} \quad (1.4)$$

azaz $t_{\text{inj}}(G, H) = \mathbb{P}(f \in \text{Inj}(G, H))$, illetve

$$t_{\text{ind}}(G, H) = \frac{\text{ind}(G, H)}{v(H)!/(v(H) - v(G))!} \quad (1.5)$$

vagyis $t_{\text{ind}}(G, H) = \mathbb{P}(f \in \text{Ind}(G, H))$.

Ritka gráfok

Az 1.2.1 definíció jól működik, ha *sűrű* gráfokról van szó, azaz ha a foksámokra nincsen felső korlát. Azonban, ha a foksámok korlátosak, azaz a gráf *ritka*, akkor az így értelmezett homomorfizmus-sűrűségek lényegében semmitmondók. (Itt a *ritka* fogalom – a korábbiakkal ellentétben – azt jelenti, hogy minden csúcs korlátos foksámú.) Ezért a ritka esetben másképp értelmezzük ezeket, ami az eredetitől lényegében annyiban különbözik, hogy másképp van normálva. Ezeket *homomorfizmus-gyakoriságnak* hívjuk:

1.2.2. Definíció (homomorfizmus-gyakoriságok). Legyenek G és H egyszerű gráfok, G összefüggő. Ekkor a homomorfizmus-gyakoriságok:

$$t^*(G, H) = \frac{\text{hom}(G, H)}{v(H)}; \quad t_{\text{inj}}^*(G, H) = \frac{\text{inj}(G, H)}{v(H)}; \quad t_{\text{ind}}^*(G, H) = \frac{\text{ind}(G, H)}{v(H)}$$

Ezt a következőképpen lehet értelmezni: címkézzük fel a G gráf egy tetszőleges csúcsát 1-el. Így megkapjuk a G_1 gráfot. Tetszőleges $v \in V(H)$ -ra $\text{hom}_v(G_1, H)$ jelölje az olyan $f: V(G_1) \rightarrow V(H)$ gráfhomomorfizmusok számát, melyre $f(1) = v$ teljesül. Ekkor, ha v egy egyenletesen véletlenszerűen kiválasztott csúcsa H -nak, akkor

$$t^*(G, H) = \mathbb{E}[\text{hom}_v(G_1, H)]$$

Homomorfizmusszámok és a mintavételezés

Alapvetően a mintavételezést szokás használni arra, hogy információt nyerjünk ki egy óriási gráfból. A sűrű esetben a homomorfizmus-sűrűségek ugyanazt az információt hordozzák magukban, mint a részgráf-mintavételezés által meghatározott valószínűség-eloszlás. Ezt az alábbi, könnyen látható állítás mutatja:

1.2.1. Állítás. *Legyen G és H két egyszerű gráf. Ekkor $t_{\text{ind}}(G, H)$ annak a valószínűsége, hogy ha a H gráfból $v(G)$ csúcsot mintavételezünk, akkor a véletlen indukált részgráf izomorf lesz G -vel.*

A ritka esetben is felírható hasonló állítás, ennek bizonyítása viszont bonyolultabb, ebben a dolgozatban vele nem foglalkozunk.

1.3. Magfüggvények és grafonok

Ebben az alfejezetben [16] 7. fejezete alapján bevezetjük magfüggvények és a grafonok fogalmát. Erre azért van szükség, mert a későbbiekben ezek határérték-objektumokként fognak szolgálni sűrű gráfsorozatok limeszének vizsgálatakor. Ezek az analitikus objektumok a gráf fogalom egyfajta kiterjesztésül szolgálnak, és több eredmény bizonyítását egyszerűbb az általános esetben, ezekkel elvégezni.

1.3.1. Definíció. Értelmezzük a következő halmazt:

$$\mathcal{W} = \{W: [0, 1]^2 \rightarrow \mathbb{R} \mid W \text{ korlátos, Lebesgue-mérhető és } W(x, y) = W(y, x) \forall x, y \in [0, 1]\}$$

Ekkor \mathcal{W} elemeit *magfüggvényeknek* vagy más szóval *kerneleknek* nevezzük. Amennyiben egy $W \in \mathcal{W}$ -re $\text{ran } W \subset [0, 1]$, akkor W -t *grafonnak* nevezzük. A grafonok halmazát \mathcal{W}_0 jelöli. Használatos még a $\mathcal{W}_1 = \{W \in \mathcal{W}: \text{ran } W \in [-1, 1]\}$ jelölés.

Az előbbi definícióban a mérhetőség a Lebesgue-mértékre nézve értendő. Mivel jellemzően nem teszünk különbséget a Lebesgue-mértékre nézve majdnem mindenütt egyenlő függvények között, ezért \mathcal{W} nem más, mint az $L^\infty([0, 1]^2)$ -beli szimmetrikus függvények tere.

Megjegyezzük, hogy a $[0, 1]$ intervallum helyett vehettünk volna bármilyen $(\Omega, \mathcal{A}, \mathbb{P})$ Kolmogorov-féle valószínűségi mezőt is, és tekinthettük volna ezen a $W: \Omega \times \Omega \rightarrow [0, 1]$ szimmetrikus függvényeket is, de ettől nem lenne sokkal nagyobb az általánosság ([16] 116. oldal).

A *grafon* elnevezés a gráf-függvény szavak angol verzióinak összeolvadásából ered (graph-function). A grafonokra tényleg tekinthetünk úgy, mint a gráfok fogalmának egy általánosítására: ha egy $W \in \mathcal{W}_0$ -ra $\text{ran } W = \{0, 1\}$, akkor W -ről beszélhetünk úgy, mint egy $[0, 1]$ csúcshalmazú gráfról.

A *magfüggvények* a súlyozott gráfok egyfajta általánosításai.

1.3.2. Definíció. Egy $W \in \mathcal{W}$ kernelt *lépcsősnek* nevezünk, ha léteznek olyan H_1, \dots, H_n ($n \in \mathbb{N}$) Lebesgue-mérhető halmazok, melyekre $\bigcup_{k=0}^n H_k = [0, 1]$ és W bármely $1 \leq i, j \leq n$ egészre konstans a $H_i \times H_j$ halmazon. A H_k ($1 \leq k \leq n$) halmazokat a W kernel *lépcsőinek* nevezzük.

Legyen G egy súlyozott gráf, melyre $V(G) = [n]$ ($n \in \mathbb{N}$), és egy $v \in V(G)$ csúcs súlya $\alpha(v)$, egy $e \in E(G)$ él súlya pedig $\beta(e)$. Ehhez hozzárendelhetünk egy $W_G \in \mathcal{W}$ lépcsős függvényt: bontsuk fel a $[0, 1]$ intervallumot $I_1 \sqcup \dots \sqcup I_n$ intervallumok diszjunkt uniójára, melyeknek Lebesgue-mérték szerinti hossza $\lambda(I_i) = \frac{\alpha(i)}{\sum_{j=1}^n \alpha(j)}$, és $\mathbf{x} \in I_i \times I_j$ -re legyen $W(\mathbf{x}) = \beta(\{i, j\})$ (ez 0, ha i és j között nem fut él a gráfban). Megfordítva,

minden $W \in \mathcal{W}$ lépcsős függvényhez rendelhető egy súlyozott gráf. Ha W lépcsői H_1, \dots, H_n , akkor legyen $V(G) = [n]$, és az $\{i, j\}$ él súlya legyen $\beta(\{i, j\}) = W(\mathbf{x})$, ahol $\mathbf{x} \in H_i \times H_j$ tetszőleges (ha a súly 0 lenne, akkor i és j között ne is fusson él a gráfban).

Ha G egy egyszerű (súlyozatlan) gráf, akkor tekinthetünk rá úgy, mintha minden él és csúcs súlya 1 lenne. Ekkor W_G egy olyan *grafon*, amelyre $\text{ran } W_G \subset \{0; 1\}$. W_G valamilyen értelemben a szomszédsági mátrix kiterjesztése függvényé.

Ezek a megfeleltetések lehetővé teszik, hogy a gráfokkal kapcsolatos mennyiségeket grafonokra (magfüggvényekre) is kiterjesszük. Például a gráfokra vonatkozó $\frac{\text{deg}_G(x)}{v(G)}$ mennyiségnek a grafonokra vonatkozó *fokszámfüggvény* felel meg:

$$d_W(x) = \int_0^1 W(x, y) dy$$

Ezután a homomorfizmus-sűrűség fogalmát terjesztjük ki magfüggvényekre. Ez később nagyon fontos lesz a konvergenciafogalom szempontjából.

1.3.3. Definíció. Legyen G egy véges egyszerű gráf, és legyen $W \in \mathcal{W}$. Ekkor

$$t(G, W) = \int_{[0,1]^{v(G)}} \prod_{\{u,v\} \in E(G)} W(x_u, x_v) \prod_{v \in V(G)} dx_v$$

$$t_{\text{ind}}(G, W) = \int_{[0,1]^{v(G)}} \prod_{\{u,v\} \in E(G)} W(x_u, x_v) \prod_{\{u,v\} \notin E(G)} (1 - W(x_u, x_v)) \prod_{v \in V(G)} dx_v$$

Itt érdemes a magfüggvényre úgy gondolni, mint egy $[0, 1]$ csúcshalmazú súlyozott gráfra, ahol az $\{x, y\}$ él súlya $W(x, y)$. Vegyük észre, hogy az injektív homomorfizmus-sűrűséget nincs értelme definiálni, mivel egy $i \mapsto x_i \in [0, 1]$ véletlen hozzárendelés 1 valószínűséggel injektív.

Az alábbi állítás ([16] 7.2.) mutatja, hogy ez valóban a gráfok homomorfizmus-sűrűségeinek egy általánosítása.

1.3.1. Állítás. Legyen G és H két véges egyszerű gráf. Ekkor fennáll, hogy

$$t(G, H) = t(G, W_H).$$

Bizonyítás. Legyen $V(G) = [m - 1] \cup \{0\}$ és $V(H) = [n - 1] \cup \{0\}$ (ahol m és n pozitív egész számok). Egy $\mathbf{x} = (x_1; \dots; x_m) \in [0, 1]^m$ pontnak feleltessük meg azt az $f_{\mathbf{x}}: V(G) \rightarrow V(H)$ leképezést, amelyre $f_{\mathbf{x}}(v) = \lfloor nx_v \rfloor$ bármely $v \in V(G)$ esetén teljesül. Így egy $\mathcal{U}_{[0,1]^m}$ egyenletes eloszlás megfelel a $V(G) \rightarrow V(H)$ függvények közötti egyenletes kiválasztásnak.

Másrészt, $W_H(x, y) = 1$ akkor és csakis akkor teljesül, ha $\{\lfloor nx \rfloor, \lfloor ny \rfloor\} \in E(H)$ is fennáll. Emiatt egy $\mathbf{x} = (x_1; \dots; x_m) \in [0, 1]^m$ -re $\prod_{\{u,v\} \in E(G)} W_H(x_u, x_v) = 1$, akkor és csakis akkor áll fenn, ha $f_{\mathbf{x}}$ gráfhomomorfizmus. Tehát:

$$\begin{aligned} t(G, W_H) &= \int \cdots \int_{[0,1]^m} \prod_{\{u,v\} \in E(G)} W_H(x_u, x_v) \prod_{v \in V(G)} dx_v = \\ &= \int_{[0,1]^m} \chi_{\text{Hom}(G,H)}(f_{\mathbf{x}}) d\mathbf{x} = \\ &= \int_{\Omega} \chi_{\{f_{\mathbf{x}} \in \text{Hom}(G,H)\}} d\mathbb{P} = \\ &= \mathbb{P}(f_{\mathbf{x}} \in \text{Hom}(G, H)) = \\ &= t(G, H) \end{aligned}$$

ahol $\mathbf{X} \sim \mathcal{U}_{[0,1]^m}$ és $(\Omega = [0, 1]^m; \mathcal{A} = \mathcal{L}([0, 1]^m), \mathbb{P} = \lambda_m|_{[0,1]^m})$ a valószínűségi mező. \square

Később hasznos lesz számunkra a $t(\cdot, \cdot)$ egy további általánosítása, amelynek [16] 7.2. alfejezetéből származó definícióját alább ismertetjük:

1.3.4. Definíció. Legyen $\mathcal{A} \subset \mathcal{W}$. Egy \mathcal{A} -címkézett gráf alatt olyan véges egyszerű G gráfot értünk, amelyben minden $e \in E(G)$ él fel van címkézve egy $W_e \in \mathcal{A}$ magfüggvénnyel. A felhasznált címkéket $w = (W_e: e \in E(G))$ jelöli. Ekkor minden \mathcal{A} -címkézett (G, w) gráf esetén

$$t(G, w) = \int \cdots \int_{[0,1]^{v(G)}} \prod_{\{u,v\} \in E(G)} W_{\{u,v\}}(x_u, x_v) \prod_{v \in V(G)} dx_v$$

Ez a definíció jól használható arra, hogy magfüggvények összegére vonatkozó homomorfizmus-sűrűségeket kifejezzünk. Például $W_1, \dots, W_n \in \mathcal{W}$ ($n \in \mathbb{N}$) esetén fennáll, hogy

$$t\left(G, \sum_{i=1}^n W_i\right) = \sum_w t(G, w),$$

ahol a jobb oldali összegzés során w befutja a G összes lehetséges $\{W_1, \dots, W_n\}$ -címkézését.

Kernel-operátorok

A magfüggvények elnevezése onnan származik, hogy úgynevezett *kernel-operátorokat* lehet velük meghatározni:

1.3.5. Definíció. Egy $W \in \mathcal{W}$ kernel meghatározza a $T_W: L^2([0, 1]) \rightarrow L^2([0, 1])$ *kernel-operátort*, amire

$$(T_W f)(x) = \int_0^1 W(x, y) f(y) dy$$

bármely $f \in L^2([0, 1])$ esetén.

1.3.1. Megjegyzés. T_W -t lehetséges $L^1([0, 1]) \rightarrow L^\infty([0, 1])$ vagy $L^\infty([0, 1]) \rightarrow L^1([0, 1])$ operátorként is értelmezni. Ha úgy definiáljuk, mint 1.3.5-ban, akkor T_W egy Hilbert-Schmidt operátor lesz bármely $W \in \mathcal{W}$ esetén.

1.3.2. Állítás. Tetszőleges $W \in \mathcal{W}$ esetén $T_W: L^2([0, 1]) \rightarrow L^2([0, 1])$ Hilbert-Schmidt operátor.

Bizonyítás. Csak annyit bizonyítunk, hogy T_W korlátos operátor.¹ A Cauchy-Bunyakovszkij-Schwarz egyenlőtlenség alapján:

$$|(T_W f)(y)|^2 \leq \int_0^1 |W(x, y)|^2 dx \int_0^1 |f(x)|^2 dx$$

¹Bővebb tárgyalás olvasható [2] 2.8.5. példájában.

Integrálással:

$$\begin{aligned} \int_0^1 |(T_W f)(y)|^2 dy &\leq \int_0^1 \left(\int_0^1 |W(x, y)|^2 dx \int_0^1 |f(x)|^2 dx \right) dy \\ &= \int_0^1 \int_0^1 |W(x, y)|^2 dx dy \int_0^1 |f(x)|^2 dx, \end{aligned}$$

azaz

$$\|T_W f\|_2 \leq \left(\int_0^1 \int_0^1 |W(x, y)|^2 dx dy \right)^{\frac{1}{2}} \|f\|_2,$$

$$\text{és } \left(\int_0^1 \int_0^1 |W(x, y)|^2 dx dy \right)^{\frac{1}{2}} = \|W\|_2 \leq \|W\|_\infty < \infty \quad \square$$

Ez nagyon jó, hiszen így használható számos eredmény a Hilbert-Schmidt operátorok elméletéből. Például, mint minden Hilbert-Schmidt operátor, T_W kompakt, sajátértékeinek multihalmaza megszámlálható, sajátértékei pedig nem nulla valósakból álló zérussorozatot alkotnak. Sőt, érvényesek a

$$\begin{aligned} (T_W f)(x) &= \sum_n \lambda_n \left(\int_0^1 f(y) \phi_n(y) dy \right) \phi_n(x) \\ W(x, y) &= \sum_n \lambda_n \phi_n(x) \phi_n(y) \end{aligned}$$

kifejtések, ahol $\{\phi_n\}$ a T_W operátor $\lambda_n \neq 0$ sajátértékeihez tartozó sajátfüggvények alkotta ortonormált rendszer. Emellett $U, W \in \mathcal{W}$ esetén

$$\begin{aligned} \langle U, W \rangle &= \iint_{[0,1]^2} U(x, y) W(x, y) dx dy = \\ &= \iint_{[0,1]^2} U(x, y) \left(\sum_n \lambda_n \phi_n(x) \phi_n(y) \right) dx dy = \\ &= \sum_n \lambda_n \iint_{[0,1]^2} U(x, y) \phi_n(x) \phi_n(y) dx dy = \sum_n \lambda_n \langle \phi_n, U \phi_n \rangle \end{aligned}$$

is fennáll.

Gyenge izomorfia

Egyik hátránya a gráfokról grafonokra való áttérésnek, hogy az izomorfianak nincsen olyan nyilvánvaló definíciója grafonok között. Egy lehetőség azokat a grafonokat izomorfoknak tekinteni, amelyek majdnem mindenütt egyenlők, de ez túl erős feltétel arra, hogy két grafon "lényegében ugyanaz". Ehelyett a *gyenge izomorfia* fogalmát vezetjük be általánosan, kernelekre:

1.3.6. Definíció. Azt mondjuk, hogy a W és W' kernelek egymással *gyengén izomorfak*, ha bármely véges egyszerű F gráfra $t(F, W) = t(F, W')$ teljesül.

A gyenge izomorfia ezzel a definícióval nyilván ekvivalenciareláció. Az ekvivalenciaosztályok elemeinek egyesítése később fontos szerepet fog játszani.

1.4. Távolságfogalmak

²Amikor a gyakorlatban nagyon nagy hálózatokat vizsgálunk, gyakran célszerű azoknak egy lekicsinyített változatát tanulmányozni, amelynek a struktúrája *hasonlít* az eredetihez. Ahhoz, hogy ezt precízzé tegyük, szükséges definiálni, hogy mit értünk az alatt, hogy két gráf *közel* van egymáshoz. Ez a *konvergenciafogalom* bevezetéséhez is fontos.

Azonos csúcshalmazú gráfoknál egy lehetőség a gráfok úgynevezett *szerkesztési távolságát* tekinteni, ami gyakorlatilag azt méri, hogy hány élt kell kicserélni, hogy az egyik gráfból megkapjuk a másikat. Ezzel több gond is van: egyrészt csak a megegyező csúcshalmazú gráfokon értelmezett, másrészt nem is igazán tükrözi a strukturális hasonlóságokat. Egy lehetséges megoldási ötlet ezekre a problémákra, hogy a gráfokból való mintavételezésből (lásd 1.1.2) származó eloszlások távolságát mérjük valamilyen módon. Ennek a távolságfogalomnak a formalizálásába most nem megyünk bele, mert bár a csúcshalmazokkal kapcsolatos gondot kiküszöböli, sajnos a strukturális jellemzőket ez sem tudja jól elkapni.

²Rövid motiváció [16] 1.5. alapján

Ami igazán hatékony lesz számunkra, az az úgynevezett *vágástávolság*. Ezt a fogalmat több lépésben fogjuk bevezetni, [16] 8. fejezete alapján.

1.4.1. Gráfok vágástávolsága

Mátrixnormák

A már jól ismert mátrixnormák mellett (ℓ_1 -, ℓ_2 -, ℓ_∞ -norma), nekünk egy kevésbé közismert, úgynevezett *vágásnormára*³ lesz szükségünk.

1.4.1. Definíció. Legyen $n \in \mathbb{N}^+$ és $A = (a_{ij})_{1 \leq i, j \leq n} \in \mathcal{M}_n(\mathbb{K})$ (ahol $\mathbb{K} = \mathbb{C}$ vagy \mathbb{R}). Az A mátrix *vágásnormája*:

$$\|A\|_{\square} = \frac{1}{n^2} \max_{S, T \subset [n]} \left| \sum_{i \in S, j \in T} a_{ij} \right| \quad (1.6)$$

1.4.1. Állítás. A fent definiált $\|\cdot\|_{\square}$ vágásnorma valóban norma.

Bizonyítás. Tetszőleges $n \in \mathbb{N}^+$ és $A = (a_{ij})_{1 \leq i, j \leq n}, B = (b_{ij})_{1 \leq i, j \leq n} \in \mathcal{M}_n(\mathbb{K})$ (ahol $\mathbb{K} = \mathbb{C}$ vagy \mathbb{R}) esetén:

(1) Nyilvánvaló, hogy $\|A\|_{\square} \geq 0$.

(2) $\|A\|_{\square} = 0 \iff \left| \sum_{i \in S, j \in T} a_{ij} \right| = 0 \quad \forall S, T \subset [n] \iff A = 0_n$

(3) Írhatjuk, hogy

$$\begin{aligned} \|A + B\|_{\square} &= \frac{1}{n^2} \max_{S, T \subset [n]} \left| \sum_{i \in S, j \in T} (a_{ij} + b_{ij}) \right| = \\ &= \frac{1}{n^2} \max_{S, T \subset [n]} \left| \sum_{i \in S, j \in T} a_{ij} + \sum_{i \in S, j \in T} b_{ij} \right| \leq \\ &\leq \frac{1}{n^2} \max_{S, T \subset [n]} \left(\left| \sum_{i \in S, j \in T} a_{ij} \right| + \left| \sum_{i \in S, j \in T} b_{ij} \right| \right) = \|A\|_{\square} + \|B\|_{\square} \end{aligned}$$

(4) $\|\lambda A\|_{\square} = |\lambda| \|A\|_{\square}$ is teljesül bármely $\lambda \in \mathbb{K}$ esetén.

³Először Frieze és Kannan vezette be 1999-ben ([11])

Tehát $\|\cdot\|_{\square} : \mathcal{M}_n(\mathbb{K}) \rightarrow \mathbb{R}$ valóban norma. □

1.4.1. Megjegyzés. Nyilvánvaló, hogy $\|A\|_{\square} \leq \|A\|_1$, tehát tudjuk, hogy

$$\|A\|_{\square} \leq \|A\|_1 \leq \|A\|_2 \leq \|A\|_{\infty}.$$

Azonos csúcshalmazú gráfok

Ahogy azt fent már leírtuk, vizsgálhatjuk a megegyező csúcshalmazú gráfok *szerkesztési távolságát*. Ez az ℓ_1 -normára vezetődik vissza.

1.4.2. Definíció. Legyen F illetve G két egyszerű gráf, melyekre $V(F) = V(G) = [n]$ ($n \in \mathbb{N}^+$). E két gráf *szerkesztési távolsága*

$$d_1(F, G) = \frac{|E(F) \Delta E(G)|}{n^2} = \|A_F - A_G\|_1,$$

ahol Δ két halmaz szimmetrikus differenciáját, A_G pedig a G gráf szomszédsági mátrixát jelöli.

Ennél fontosabb lesz számunkra a *vágástávolság*.

1.4.3. Definíció. Legyen F illetve G két egyszerű gráf, melyekre $V(F) = V(G) = [n]$ ($n \in \mathbb{N}^+$). E két gráf *vágástávolsága*

$$d_{\square}(F, G) = \max_{S, T \subseteq [n]} \frac{|e_F(S, T) - e_G(S, T)|}{n^2} = \|A_F - A_G\|_{\square},$$

ahol $e_G(S, T) = |\{\{u, v\} \in E(G) : u \in S, v \in T\}|$.

Azonnal adódik, hogy $d_{\square}(F, G) \leq d_1(F, G)$. Sőt, általában is különbözik a két távolság. Például ha F és G is két *véletlen* gráf az $[n]$ csúcshalmazon, melyekben bármely $u, v \in [n]$ különböző csúcsra $\mathbb{P}(\{u, v\} \in E) = \frac{1}{2}$, akkor nagy valószínűséggel $d_1(F, G) \approx \frac{1}{2}$, míg viszont $d_{\square}(F, G) = O\left(\frac{1}{\sqrt{n}}\right)$.

Ugyanannyi csúcsú gráfok

Abban az esetben, ha ugyanannyi csúcsa van a gráfoknak, annyit teszünk, hogy megszámozzuk a csúcsokat az $[n]$ halmaz elemeivel, és a *legjobb* számozással vesszük az előző értelemben vett távolságukat.

1.4.4. Definíció. Legyen F és G két egyszerű gráf, melyekre $v(F) = v(G) = n \in \mathbb{N}^+$. Ekkor a két gráf távolsága

$$\widehat{\delta}_{\square}(F, G) = \min_{\widehat{F}, \widehat{G}} d_{\square}(\widehat{F}, \widehat{G}),$$

ahol a minimalizálás során \widehat{F} és \widehat{G} rendre végigfut az F és G gráfok összes lehetséges $[n]$ -címkézésén. Ekkor úgy tekintjük, hogy $V(\widehat{F}) = V(\widehat{G}) = [n]$.

Végül rátérünk az általános esetre.

Tetszőleges gráfok

Szükségünk van arra, hogy "azonos csúcsszámra hozzuk" a két gráfot. Ehhez bevezetjük egy gráf egy *felrobbantottjának* fogalmát:

1.4.5. Definíció. Legyen G egy gráf, és $k \in \mathbb{N}^+$. Ekkor a G gráf egy $G(k)$ *felrobbantottján* azt a gráfot értjük, amelyet úgy kapunk, hogy a G gráf minden csúcsát k db. csúcsra cseréljük, és két új csúcs akkor és csakis akkor szomszédos, ha az őseik is szomszédosak voltak.

Ezzel az operációval már tudunk azonos csúcsszámú gráfokat gyártani az eredetiekéből: ha $v(F) = n$ és $v(G) = m$ ($n, m \in \mathbb{N}^+$), akkor bármely $k \in \mathbb{N}^+$ -ra $v(F(km)) = v(G(kn))$ teljesül. Így már definiálhatjuk a *vágástávolságot* két tetszőleges gráf esetén:

1.4.6. Definíció. Legyen F és G két véges egyszerű gráf. Ekkor a két gráf *vágástávolsága*

$$\delta_{\square}(F, G) = \lim_{n \rightarrow \infty} \widehat{\delta}_{\square}(F(nv(G)), G(nv(F)))$$

Megadható ugyanerre egy határérték nélküli definíció is.⁴ Az F és G gráfok egy *törtrészes átfedése* legyen az a nemnegatív $X = (x_{ij}) \in \mathbb{R}^{v(F) \times v(G)}$ mátrix, amelyre $\sum_{j=1}^{v(G)} x_{ij} = \frac{1}{v(F)}$ és $\sum_{i=1}^{v(F)} x_{ij} = \frac{1}{v(G)}$. Jelölje $\chi(F, G)$ az összes ilyen mátrix halmazát. Ekkor

$$\delta_{\square}(F, G) = \min_{X \in \chi(F, G)} \max_{Y, Z \subseteq E} \left| \sum_{\substack{\{i, u\} \in Y \\ \{j, v\} \in Z \\ \{i, j\} \in E(F)}} x_{iu} x_{jv} - \sum_{\substack{\{i, u\} \in Y \\ \{j, v\} \in Z \\ \{u, v\} \in E(G)}} x_{iu} x_{jv} \right|$$

ahol $E = \{\{i, v\} \mid i \in V(F), v \in V(G)\}$.

1.4.2. Magfüggvények távolságai

Gráftávolságok után most áttérünk az általánosabb fogalomra, azaz a magfüggvények között szeretnénk távolságfogalmat definiálni. Ez az általánosság ellenére még egyszerűbb is lesz, mint a gráfok esetében.

Vágásnorma

1.4.7. Definíció. A $W \in \mathcal{W}$ magfüggvény *vágásnormája*

$$\|W\|_{\square} = \sup_{S, T \in \mathcal{L}([0, 1])} \left| \int_{S \times T} W(\mathbf{x}) \, d\mathbf{x} \right|,$$

ahol $\mathcal{L}([0; 1])$ a $[0; 1]$ intervallum Lebesgue-mérhető részhalmazainak halmaza. A vágásnorma által indukált metrika:

$$d_{\square}(V, W) = \|V - W\|_{\square}$$

1.4.2. Állítás. A $\|\cdot\|_{\square}$ vágásnorma valóban norma (és így $d_{\square}(\cdot, \cdot)$ valóban metrika)

Bizonyítás. Tetszőleges $V, W \in \mathcal{W}$ magfüggvények és $\lambda \in \mathbb{R}$ esetén:

(1) Nyilvánvaló, hogy $\|W\|_{\square} \geq 0$.

⁴Az alábbi rész [15] 4.3. részén alapszik.

$$(2) \|W\|_{\square} = 0 \iff \int_{S \times T} W(\mathbf{x}) \, d\mathbf{x} = 0 \quad \forall S, T \in \mathcal{L}([0, 1]) \iff W(\mathbf{x}) = 0 \quad \forall \mathbf{x} \in [0, 1]^2$$

(3) Írhatjuk, hogy

$$\begin{aligned} \|V + W\|_{\square} &= \sup_{S, T \in \mathcal{L}([0, 1])} \left| \int_{S \times T} (V(\mathbf{x}) + W(\mathbf{x})) \, d\mathbf{x} \right| = \\ &= \sup_{S, T \in \mathcal{L}([0, 1])} \left| \int_{S \times T} V(\mathbf{x}) \, d\mathbf{x} + \int_{S \times T} W(\mathbf{x}) \, d\mathbf{x} \right| \leq \\ &\leq \sup_{S, T \in \mathcal{L}([0, 1])} \left(\left| \int_{S \times T} V(\mathbf{x}) \, d\mathbf{x} \right| + \left| \int_{S \times T} W(\mathbf{x}) \, d\mathbf{x} \right| \right) = \|V\|_{\square} + \|W\|_{\square} \end{aligned}$$

$$(4) \|\lambda W\|_{\square} = |\lambda| \|W\|_{\square} \text{ is mindig teljesül.}$$

Tehát $\|\cdot\|_{\square}: \mathcal{W} \rightarrow \mathbb{R}$ valóban norma. □

1.4.2. Megjegyzés. *Tetszőleges $W \in \mathcal{W}$ esetén fennáll, hogy*

$$\|W\|_{\square} \leq \|W\|_1 \leq \|W\|_2 \leq \|W\|_{\infty}.$$

Egy hátránya annak, hogy grafonokkal dolgozunk gráfok helyett, hogy amikor egy függvény maximumát szeretnénk egy bizonyos halmaz felett megkapni, ez nem feltétlenül lesz *maximum*, hanem sokszor csak *szuprémum*. [16] 8.9. lemmája, amelyet ebben a dolgozatban nem bizonyítottunk, azt állítja, hogy a 1.4.7 definícióban szereplő szuprémum valójában egy maximum:

1.4.1. Lemma. *Bármely $W \in \mathcal{W}$ kernelre a*

$$\sup_{S, T \in \mathcal{L}([0, 1])} \left| \int_{S \times T} W(\mathbf{x}) \, d\mathbf{x} \right|$$

és

$$\sup_{f, g: [0, 1] \rightarrow [0, 1]} \left| \iint_{[0, 1]^2} f(x)g(y)W(x, y) \, dx dy \right|$$

szuprémumok felvételnek, és mindkettő egyenlő a $\|W\|_{\square}$ értékkel.

1.4.3. Megjegyzés. 1.4.1-ben a második szuprémum úgy értendő, hogy f és g a Lebesgue-mérhető $[0, 1] \rightarrow [0, 1]$ függvényeket futják be.

Vágástávolság

A $[0, 1]^2$ -en értelmezett kernelek címkézett (él- és csúcssúlyozott) gráfoknak feleltethetők meg (lásd: 1.3 alfejezet). Bevezethetjük ezen fogalom *címkézetlen* változatát az alábbiak szerint. Jelölje $\bar{S}_{[0,1]}$ a (Lebesgue-)mértéktartó $[0, 1] \rightarrow [0, 1]$ leképezések halmazát, illetve $S_{[0,1]}$ tartalmazza $\bar{S}_{[0,1]}$ elemei közül az invertálhatókat. (Nyilván, ekkor $S_{[0,1]}$ csoport a függvények kompozíciójára mint műveletre nézve.)

1.4.8. Definíció. Két $V, W \in \mathcal{W}$ kernel *vágástávolsága* alatt a

$$\delta_{\square}(V, W) = \inf_{\varphi \in S_{[0,1]}} d_{\square}(V, W^{\varphi})$$

mennyiséget értjük, ahol $W^{\varphi}(x, y) = W(\varphi(x), \varphi(y))$.

Az persze könnyen látható, hogy a

$$\delta_{\square}(V, W) = \inf_{\varphi \in \bar{S}_{[0,1]}} d_{\square}(V^{\varphi}, W) = \inf_{\varphi \in \bar{S}_{[0,1]}} d_{\square}(V, W^{\varphi}) = \inf_{\varphi, \psi \in \bar{S}_{[0,1]}} d_{\square}(V^{\varphi}, W^{\psi})$$

egyenlőségek mindenike teljesül.

Fontos megjegyezni, hogy $\delta_{\square}: \mathcal{W} \times \mathcal{W} \rightarrow \mathbb{R}$ csak egy félmetrika, hiszen különböző magfüggvényeknek lehet 0 a távolsága. Például ha $\varphi: [0, 1] \rightarrow [0, 1]$ olyan függvény, melyre $\varphi(x) = 2x \bmod 1$ bármely $x \in [0, 1]$ esetén, akkor $W_1(x, y) = xy$ mellett legyen $W_2 = W_1^{\varphi}$ (mindkét függvény $[0, 1]^2 \rightarrow \mathbb{R}$). Ekkor $W_1, W_2 \in \mathcal{W}_0$, különbözőek, de $\delta_{\square}(W_1, W_2) = 0$.

Hogyha az egymástól vágástávolságban 0 távol lévő elemeket azonosítjuk, akkor a *címkézetlen magfüggvények* $\widetilde{\mathcal{W}}$ halmazát kapjuk. Hasonlóan definiáljuk a $\widetilde{\mathcal{W}}_0$ és $\widetilde{\mathcal{W}}_1$ halmazokat is. Később meg fogjuk látni, hogy az ezen terek elemei pontosan a gyenge izomorfia mint ekvivalenciareláció szerinti ekvivalenciaosztályok lesznek.

Amikor gráfokról a megfelelő grafonokra térünk át, felmerül a kérdés, hogy mennyi információt veszítünk el az eredeti gráf struktúrájáról, azaz mekkora a *hiba*. Néhány

esetben ez a hiba 0, például az 1.3.1 állítás arra vonatkozott, hogy a $G \mapsto W_G$ megfeleltetés által nem vétünk hibát, ha a homomorfizmus-sűrűség érdekel bennünket, az az áttérés során változatlan marad. Ki lehet mutatni⁵, hogy a helyzet a vágástávolsággal is ugyanez:

1.4.2. Lemma. *Bármely két véges egyszerű G és H gráfra*

$$\delta_{\square}(G, H) = \delta_{\square}(W_G, W_H)$$

Ez a komplikált áttérés gráfokról grafonokra, a vágástávolság bonyolult definiálása a vágásnorma által indukált metrika mérhető függvényeken való minimalizálásával azért szükséges, mert az következő alfejezetben tárgyalt igen fontos tétel szerint a vágástávolság kompakt metrikus teret definiál a címkézetlen grafonokon.

1.5. A grafontér kompaktsága

Ennek a résznek a fő célja az lesz, hogy bebizonyítsuk a $(\widetilde{\mathcal{W}}_0, \delta_{\square})$ metrikus tér kompaktságát. [16] 9. fejezete alapján haladva, előbb néhány regularitási lemmát írunk le. Először azonban bevezetjük a *léptető operátor* fogalmát.

1.5.1. Definíció. Legyen $W \in \mathcal{W}$ és $\mathcal{P} = \{H_1, \dots, H_q\}$ a $[0, 1]$ intervallum egy partíciója (ahol $q \in \mathbb{N}^+$, és H_1, \dots, H_q Lebesgue-mérhetők – amikor partícióról beszélünk, mindig mérhető particionáló halmazokra fogunk gondolni). Ekkor definiáljuk a $W_{\mathcal{P}}: [0, 1]^2 \rightarrow \mathbb{R}$ függvényt, amelyre

$$W_{\mathcal{P}}(\mathbf{x}) = \sum_{i,j=1}^q \frac{\chi_{H_i \times H_j}(\mathbf{x})}{\lambda(H_i)\lambda(H_j)} \int_{H_i \times H_j} W(\mathbf{v}) \, d\mathbf{v}$$

Ekkor a $W_{\mathcal{P}}$ függvényt W *léptetésének*, valamint a $\varphi_{\mathcal{P}}: \mathcal{W} \rightarrow \mathcal{W}$ operátort, amelyre $\varphi_{\mathcal{P}}(W) = W_{\mathcal{P}}$, *léptető operátornak* nevezzük.

Ekkor tehát egy \mathcal{P} partícióra $W_{\mathcal{P}}$ egy lépcsős függvény lesz, amit a W magfüggvény $H_i \times H_j$ halmazokon ($i, j \in [|\mathcal{P}|]$) való kiátlagolásából kapunk. Ezért úgy is lehet erre gondolni, mint a \mathcal{P} által generált σ -algebra szerint vett feltételes várható érték.

⁵A lemmát itt nem bizonyítjuk, egy bizonyítás olvasható [16] 8.9. lemmájánál.

1.5.1. Állítás ([16] 9.17. feladat). A $[0, 1]$ tetszőleges $\mathcal{P} = \{H_1, \dots, H_q\}$ ($q \in \mathbb{N}^+$) partíciója esetén a $\varphi_{\mathcal{P}}$ léptető operátor kontrakció a $\|\cdot\|_{\square}$ vágásnormára nézve.

Bizonyítás. Csupán nemnegatív értékű kernelek esetén bizonyítunk. Tetszőleges $W \in \mathcal{W}$ -re valamint $f, g: [0, 1] \rightarrow [0, 1]$ mérhető függvényekre írhatjuk, hogy

$$\begin{aligned} \iint_{[0,1]^2} f(x)g(y)W_{\mathcal{P}}(x,y) \, dx dy &= \iint_{[0,1]^2} f(x)g(y) \sum_{i,j=1}^q \left(\frac{\chi_{H_i \times H_j}(x,y)}{\lambda(H_i)\lambda(H_j)} \int_{H_i \times H_j} W(\mathbf{v}) \, d\mathbf{v} \right) \, dx dy = \\ &= \sum_{i,j=1}^q \frac{1}{\lambda(H_i)\lambda(H_j)} \iint_{H_i \times H_j} f(x)g(y) \int_{H_i \times H_j} W(\mathbf{v}) \, d\mathbf{v} \, dx dy = \\ &= \sum_{i,j=1}^q \frac{1}{\lambda(H_i)\lambda(H_j)} \left(\int_{H_i \times H_j} W(\mathbf{v}) \, d\mathbf{v} \right) \left(\iint_{H_i \times H_j} f(x)g(y) \, dx dy \right) \end{aligned}$$

Tehát 1.4.1 miatt, és mivel $f, g \geq 0$, fennáll, hogy

$$\begin{aligned} \|\varphi_{\mathcal{P}}W\|_{\square} &= \|W_{\mathcal{P}}\|_{\square} = \sup_{f,g} \left| \iint_{[0,1]^2} f(x)g(y)W_{\mathcal{P}}(x,y) \, dx dy \right| = \\ &= \sup_{f,g} \left| \sum_{i,j=1}^q \frac{1}{\lambda(H_i)\lambda(H_j)} \left(\int_{H_i \times H_j} W(\mathbf{v}) \, d\mathbf{v} \right) \left(\iint_{H_i \times H_j} f(x)g(y) \, dx dy \right) \right| \leq \\ &\leq \sup_{f,g} \sum_{i,j=1}^q \frac{1}{\lambda(H_i)\lambda(H_j)} \left| \int_{H_i \times H_j} W(\mathbf{v}) \, d\mathbf{v} \right| \left| \iint_{H_i \times H_j} f(x)g(y) \, dx dy \right| = \\ &= \sum_{i,j=1}^q \left| \int_{H_i \times H_j} W(\mathbf{v}) \, d\mathbf{v} \right|. \end{aligned}$$

Azaz $W \geq 0$ esetén

$$\|\varphi_{\mathcal{P}}W\|_{\square} \leq \iint_{[0,1]^2} W(x,y) \, dx dy \leq \sup_{S,T \in \mathcal{L}([0;1])} \iint_{S \times T} W(x,y) \, dx dy = \|W\|_{\square},$$

vagyis az operátornorma

$$\|\varphi_{\mathcal{P}}\| = \sup_W \frac{\|\varphi_{\mathcal{P}}W\|_{\square}}{\|W\|_{\square}} \leq 1,$$

tehát $\varphi_{\mathcal{P}}$ kontrakció. □

Most térjünk rá a gyenge regularitási lemmára, amit először Frieze és Kannan bizonyított⁶. A lemma arról szól, hogy minden kernel vágásnormában tetszőlegesen jól közelíthető elég sok lépcsőjű lépcsős függvénnyel:

1.5.1. Lemma (Gyenge regularitási lemma). *Tetszőleges $W \in \mathcal{W}$ kernelre és $k \in \mathbb{N}^+$ -ra létezik olyan $U \in \mathcal{W}$ k lépcsőjű lépcsős függvény, melyre*

$$\|W - U\|_{\square} < \frac{2}{\sqrt{\log_2 k}} \|W\|_2$$

Ha egy előre rögzített $\varepsilon > 0$ hiba a cél, akkor a szükséges partícióosztályok száma $2^{\frac{4}{\varepsilon^2}}$, azaz $\frac{1}{\varepsilon^2} \gg 0$ -ban még mindig exponenciális. Az közelítés egy alábbi finomított verziója, amivel a közelítő lépcsős függvény mérete már polinomiálisra csökken, szintén Frieze és Kannantól származik.

1.5.2. Lemma. *Tetszőleges $W \in \mathcal{W}$ kernelre és $k \in \mathbb{N}^+$ -ra léteznek olyan $S_i, T_i \subset [0, 1]$ mérhető halmazok és $a_i \in \mathbb{R}$ számok ($i \in [k]$), melyekre*

$$\left\| W - \sum_{i=1}^k a_i \chi_{S_i \times T_i} \right\|_{\square} < \frac{1}{\sqrt{k}}$$

Mivel $\sum_{i=1}^k a_i \chi_{S_i \times T_i}$ egy lépcsős függvény, ez szimmetrikussá tehető, ha a $\sum_{i=1}^k a_i \chi_{T_i \times S_i}$ -vel vett átlagát tekintjük (ez összesen $2k$ tagú). Az így kapott szimmetrikus lépcsős függvénynek legtöbb 2^{2k} lépcsője van, tehát az 1.5.1 lemma következik az 1.5.2 lemmából (ha k helyett 2^{2k} szerepel). Ahhoz, hogy a 1.5.2 lemmát belássuk, szükségünk van a következő lemmára ([16] 9.11a)

1.5.3. Lemma. *Tetszőleges $W \in \mathcal{W}$ kernelre vannak olyan $S, T \subset [0, 1]$ mérhető halmazok és $a \in [0, \|W\|_{\infty}]$ szám, melyekre*

$$\|W - a \chi_{S \times T}\|_2^2 \leq \|W\|_2^2 - \|W\|_{\square}^2$$

⁶lásd [16] 9.2.2-t, és az ottani hivatkozásokat

Bizonyítás. A 1.4.1 lemma szerint létezik olyan $S, T \in \mathcal{L}([0, 1])$ halmazok, melyekre $\|W\|_{\square} = \left| \int_{S \times T} W(\mathbf{x}) \, d\mathbf{x} \right|$ teljesül, és feltehetjük, hogy $\int_{S \times T} W(\mathbf{x}) \, d\mathbf{x} \geq 0$, hiszen ha W -t pozitív illetve negatív részre bontjuk, akkor ennek a kettőnek a szorzata 0, ezért a négyzetösszegben $\|W^+ + W^-\|_2^2 = \|W^+\|_2^2 + \|W^-\|_2^2$.

Legyen $a = \frac{\|W\|_{\square}}{\lambda(S)\lambda(T)}$. Ekkor

$$\begin{aligned} \|W - a\chi_{S \times T}\|_2^2 &= \int_{[0,1]^2} |W(\mathbf{x}) - a\chi_{S \times T}(\mathbf{x})|^2 \, d\mathbf{x} = \\ &= \int_{[0,1]^2} |W(\mathbf{x})|^2 \, d\mathbf{x} - 2a \int_{S \times T} W(\mathbf{x}) \, d\mathbf{x} + a^2 \lambda(S)\lambda(T) = \\ &= \|W\|_2^2 - \frac{\|W\|_{\square}^2}{\lambda(S)\lambda(T)} \leq \|W\|_2^2 - \|W\|_{\square}^2 \end{aligned}$$

□

Most már rátérhetünk 1.5.2 bizonyítására.

Bizonyítás. W -re alkalmazva 1.5.3-öt, kapjuk az $S_1, T_1 \in \mathcal{L}([0, 1])$ halmazokat és az a_1 valós számot. Legyen $W_1 = W - a_1\chi_{S_1 \times T_1}$. Ekkor fennáll, hogy

$$\|W_1\|_2^2 \leq \|W\|_2^2 - \|W\|_{\square}^2$$

Most W_1 -re alkalmazva 1.5.3-öt, adódnak az $S_2, T_2 \in \mathcal{L}([0, 1])$ halmazok, és az a_2 valós szám, melyekre a $W_2 = W_1 - a_2\chi_{S_2 \times T_2}$ jelöléssel fennáll, hogy

$$\|W_2\|_2^2 \leq \|W_1\|_2^2 - \|W_1\|_{\square}^2 \leq \|W\|_2^2 - \|W\|_{\square}^2 - \|W_1\|_{\square}^2$$

Ezt folytatva matematikai indukcióval igazolható, hogy ha $W_j = W - \sum_{i=1}^j a_i\chi_{S_i \times T_i}$ bármely $j \in \mathbb{N}^+$ esetén, illetve $W_0 = W$, akkor bármely $n \in \mathbb{N}^+$ -ra teljesül, hogy

$$\|W_n\|_2^2 \leq \|W\|_2^2 - \sum_{i=1}^{n-1} \|W_i\|_{\square}^2$$

Mivel a baloldal nemnegatív, így a jobboldal is nemnegatív kell maradjon, ezért bármely $k \in \mathbb{N}^+$ -ra létezik olyan $0 \leq i_0 < k$ egész, amelyre igaz, hogy $\|W_{i_0}\|_{\square}^2 < \frac{1}{k}$. Tehát teljesül, hogy

$$\left\| W - \sum_{i=1}^k \alpha_i \chi_{S_i \times T_i} \right\|_{\square} < \frac{1}{\sqrt{k}},$$

ahol $\alpha_i = a_i$, ha $i \in [i_0]$, és $\alpha_i = 0$, ha $i > i_0$. Ezzel a lemmát bebizonyítottuk. \square

Felvetődik a kérdés, hogy egy $W \in \mathcal{W}$ kernel 1.5.1 általi lépcsős függvénnyel való közelítése a W léptetése-e. Bár ez nincs mindig így, Frieze és Kannan megmutatták, hogy a léptetések elég közel vannak a jó becsléshez:

1.5.4. Lemma. *Tetszőleges $W \in \mathcal{W}$ kernelre, $U \in \mathcal{W}$ lépcsős függvényre, ha \mathcal{P} a $[0, 1]$ -nek U lépcsői szerinti partíciója, akkor fennáll, hogy*

$$\|W - W_{\mathcal{P}}\|_{\square} \leq 2\|W - U\|_{\square}$$

Bizonyítás. Mivel $U = U_{\mathcal{P}}$ és a léptető operátor kontrakció a vágásnormára nézve (1.5.1), ezért írhatjuk, hogy

$$\begin{aligned} \|W - W_{\mathcal{P}}\|_{\square} &= \|W - U + U - W_{\mathcal{P}}\|_{\square} \leq \\ &\leq \|W - U\|_{\square} + \|U - W_{\mathcal{P}}\|_{\square} = \\ &= \|W - U\|_{\square} + \|U_{\mathcal{P}} - W_{\mathcal{P}}\|_{\square} = \\ &= \|W - U\|_{\square} + \|(U - W)_{\mathcal{P}}\|_{\square} \leq \\ &\leq \|W - U\|_{\square} + \|U - W\|_{\square} = \\ &= 2\|W - U\|_{\square}, \end{aligned}$$

ami bizonyítja a lemmát. \square

1.5.5. Lemma. *Tetszőleges $W \in \mathcal{W}$ kernelre, és $1 \leq m < k$ egészekre*

(a) *bármely m -osztályú \mathcal{Q} partíciójára a $[0, 1]$ -nek van olyan \mathcal{Q} -t finomító k -osztályú \mathcal{P} partíciója $[0, 1]$ -nek, melyekre fennáll, hogy*

$$\|W - W_{\mathcal{P}}\|_{\square} \leq \frac{2}{\sqrt{\log_2 \frac{k}{m}}}$$

(b) *bármely m -osztályú \mathcal{Q} partíciójára a $[0, 1]$ -nek van olyan egyenlő mértékű osztályokból álló k -osztályú \mathcal{P} partíciója (ekvipartíciója) a $[0, 1]$ -nek, hogy*

$$\|W - W_{\mathcal{P}}\|_{\square} \leq 2\|W - W_{\mathcal{Q}}\|_{\square} + \frac{2m}{k}$$

Bizonyítás. Az (a) rész hasonló 1.5.1-hoz, most nem bizonyítjuk. A (b) részben osszuk fel \mathcal{Q} osztályait $\frac{1}{k}$ mértékű részekre úgy, hogy legfeljebb egy rész lehet kisebb. Tekintsük ezeknek a kisebb osztályoknak az unióját, és particionáljuk ezt úgy, hogy szintén $\frac{1}{k}$ mértékű osztályok legyenek, így megkapjuk a \mathcal{P} partíciót. Tekintsük a két partíció közös finomítását, azaz $\mathcal{R} = \mathcal{P} \wedge \mathcal{Q}$ -t. Ekkor $W_{\mathcal{R}}$ és $W_{\mathcal{P}}$ egy legfeljebb $2 \cdot \frac{m}{k}$ mértékű halmazban térnek el egymástól, tehát

$$\|W - W_{\mathcal{P}}\|_{\square} \leq \|W - W_{\mathcal{R}}\|_{\square} + \frac{2m}{k}$$

Másrészt 1.5.4 szerint $\|W - W_{\mathcal{R}}\|_{\square} \leq \|W - W_{\mathcal{Q}}\|_{\square}$, ami bizonyítja a lemmát. \square

Most már megvannak az eszközeink ahhoz, hogy bebizonyítsuk a grafontér kompaktságát. A tétel Lovásztól és Szegedytől származik [18], és gyakorlatilag a gyenge regularitási lemma következménye.

1.5.1. Tétel. *A $(\widetilde{\mathcal{W}}_0, \delta_{\square})$ metrikus tér kompakt.*

Bizonyítás. Mivel metrikus térben vagyunk, elég annyit bizonyítani, hogy minden $\widetilde{\mathcal{W}}_0$ -beli sorozatnak létezik konvergens részsorozata.

Legyen $(W_n)_{n \in \mathbb{N}} \subset \widetilde{\mathcal{W}}_0$ egy sorozat. 1.5.5 szerint bármely $n \in \mathbb{N}$ -re és $k \in \mathbb{N}^+$ -ra léteznek a $[0, 1]$ intervallumnak olyan $\mathcal{P}_{n,k}$ partíciói, amelyekre teljesül, hogy ha a megfelelő léptetések $W_{n,k} = (W_n)_{\mathcal{P}_{n,k}} \in \mathcal{W}_0$ ($\forall n \in \mathbb{N}, k \in \mathbb{N}^+$), akkor

(a) $\|W_n - W_{n,k}\|_{\square} \leq \frac{1}{k} \quad \forall n \in \mathbb{N}, k \in \mathbb{N}^+$

(b) $\mathcal{P}_{n,k+1}$ a $\mathcal{P}_{n,k}$ finomítása ($\forall n \in \mathbb{N}, k \in \mathbb{N}^+$)

(c) $|\mathcal{P}_{n,k}| = m_k$ csak k -től függ.

Ezek után egy mértéktartó bijekcióval átrendezhetők $[0, 1]$ pontjai oly módon, hogy minden partíció osztályai egy-egy intervallumot alkossanak.

1.5.2. Állítás. *$(W_n)_{n \in \mathbb{N}}$ -nek létezik olyan $(W_{n_m})_{m \in \mathbb{N}}$ részsorozata, melyre alkalmas $U_k \in \mathcal{W}_0$ ($k \in \mathbb{N}^+$) lépcsős függvényekre*

$$\lim_{m \rightarrow \infty} W_{n_m, k} = U_k$$

minden $k \in \mathbb{N}^+$ -re Lebesgue-majdnem mindenütt teljesül.

Bizonyítás. Kiválasztható olyan $(W_{n_m})_{m \in \mathbb{N}}$ részsorozata $(W_n)_{n \in \mathbb{N}}$ -nek, amire a $\mathcal{P}_{n_m,1}$ partíció i . intervallumának hossza minden $i \in [m_1]$ -re konvergens, és emellett $W_{n_m,1}$ értéke az i . és a j . intervallum Descartes-szorzatán is konvergál, minden $i, j \in [m_1]$ -re (midőn $m \rightarrow \infty$) Tehát ez a $W_{n_m,1}$ részsorozat Lebesgue-majdnem mindenütt tart egy U_1 lépcsős függvénybe, aminek m_1 lépcsője van.

Ugyanezt az eljárást megismételve $k = 2, 3, \dots$ -ra is, olyan részsorozatokat kapunk, aminek a léptetései majdnem mindenütt U_k -ba konvergálnak, ahol U_k lépcsős függvény, aminek m_k lépcsője van. Vegyük azt a sorozatot, amelynek i . eleme az i . részsorozat i . eleme ($i \in \mathbb{N}$). Ez megfelel az állításnak. \square

Tehát $(W_n)_{n \in \mathbb{N}}$ -t kicserélhetjük erre a részsorozatra, azaz feltehetjük, hogy teljesíti az előbbi állítás részsorozatra vonatkozó feltételeit. Legyen \mathcal{P}_k az így kapott U_k limesz-függvények lépcsői szerinti partíciója a $[0, 1]$ -nek. Ha $k < l$ pozitív egészek, akkor $\mathcal{P}_{n,l}$ a $\mathcal{P}_{n,k}$ finomítása, tehát $W_{n,k} = (W_{n,l})_{\mathcal{P}_{n,k}}$. Nem nehéz meggondolni, hogy ez a reláció határérték-számítás után is megmarad, azaz írhatjuk, hogy ilyenkor $U_l = (U_k)_{\mathcal{P}_k}$.

Legyen $X, Y \sim \mathcal{U}_{[0,1]}$ két független valószínűségi változó. Ekkor az előző miatt

$$(U_1(X, Y), U_2(X, Y), \dots)$$

martingál lesz. Mivel tetszőleges $k \in \mathbb{N}^+$ -ra az $U_k(X, Y)$ valószínűségi változó korlátos marad, ezért a martingál-konvergencia tétel miatt ez a sorozat majdnem biztosan konvergens. Más szóval az $(U_k)_{k \in \mathbb{N}^+} \subset \mathcal{W}_0$ függvénysorozat majdnem mindenütt konvergens. Legyen $U \in \mathcal{W}_0$ a határértéke, és legyen $\varepsilon > 0$ tetszőleges pozitív valós szám. Ekkor elég nagy $k \in \mathbb{N}^+$ -ra $\|U - U_k\|_1 < \frac{\varepsilon}{3}$ és $\delta_{\square}(W_{n,k} - W_n) < \frac{\varepsilon}{3}$ is teljesül. Erre a k -ra van olyan $N(\varepsilon) \in \mathbb{N}$, hogy $\|W_{n,k} - U_k\|_1 < \frac{\varepsilon}{3}$ valamennyi $n \geq N(\varepsilon)$ -ra teljesül. Ennélfogva írhatjuk, hogy valamennyi $n \geq N(\varepsilon)$ -ra

$$\begin{aligned} \delta_{\square}(U, W_n) &\leq \delta_{\square}(U, U_k) + \delta_{\square}(U_k, W_{n,k}) + \delta_{\square}(W_{n,k}, W_n) \leq \\ &\leq \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \delta_{\square}(W_{n,k} - W_n) < \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

Tehát $W_n \rightarrow U$, ami bizonyítja a tételt. \square

A tétel igaz marad akkor is, ha $\widetilde{\mathcal{W}}_0$ helyett $\widetilde{\mathcal{W}}$ más egyenletesen korlátos részhalmazára mondjuk ki. Például a $(\widetilde{\mathcal{W}}_1, \delta_{\square})$ metrikus tér is kompakt. Ezt abból is lehet látni, hogy a $W \rightarrow 2W - 1$ leképezés folytonos és szürjektív $\widetilde{\mathcal{W}}_0$ és $\widetilde{\mathcal{W}}_1$ között, az ilyen leképezések pedig megőrzik a kompaktságot.

1.6. Gráfsorozatok konvergenciája

A dolgozat egyik fő témája a gráfsorozatok konvergenciájának tárgyalása.⁷ Egy gráfokból álló sorozatot akkor mondunk konvergensnek, ha a későbbi és későbbi tagok valamilyen értelemben egyre jobban kezdenek szerkezetileg hasonlítani egymásra. Ezt formalizálja az alábbi definíció.

1.6.1. Definíció. Azt mondjuk, hogy egy $(G_n)_{n \in \mathbb{N}}$ gráfsorozat, melyre $\lim_{n \rightarrow \infty} v(G_n) = \infty$ *konvergens*, ha bármely F véges gráfra a $(t_{\text{ind}}(F, G_n))_{n \in \mathbb{N}} \subset \mathbb{R}$ számsorozat konvergens.

Két tetszőleges F, G gráfra teljesül, hogy $\text{inj}(F, G) = \sum_{F' \supset F} \text{ind}(F', G)$ (ahol F' az összes F -ből él hozzáadása által megkapható gráfokat futja be). Másrészt, a szitaformula szerint fennáll, hogy $\text{ind}(F, G) = \sum_{\substack{F' \supset F \\ V(F')=V(F)}} (-1)^{e(F')-e(F)} \text{inj}(F', G)$, ezért a megfelelő homomorfizmus-sűrűségekre is igaz, hogy

$$t_{\text{inj}}(F, G) = \sum_{F' \supset F} t_{\text{ind}}(F', G) \quad \text{illetve} \quad t_{\text{ind}}(F, G) = \sum_{\substack{F' \supset F \\ V(F')=V(F)}} (-1)^{e(F')-e(F)} t_{\text{inj}}(F', G).$$

Ezekből következik, hogy tetszőleges F véges gráf esetén $t_{\text{ind}}(F, G_n)$ akkor és csak akkor konvergál, ha $t_{\text{inj}}(F, G_n)$ is (midőn $n \rightarrow \infty$), tehát a 1.6.1 definícióban t_{ind} helyett írhatnánk t_{inj} -et is, a konvergenciafogalom nem változna. A következő lemma (bizonyítás [17] szerint) gyakorlatilag azt mondja, hogy t is ugyanúgy használható a konvergencia értelmezésénél.

⁷Az alfejezet javarészt [16] 11. fejezetén alapszik.

1.6.1. Lemma. *Bármely F és G véges egyszerű gráfra fennáll, hogy*

$$|t_{\text{inj}}(F, G) - t(F, G)| \leq \frac{1}{v(G)} \binom{v(F)}{2}$$

Bizonyítás. Triviális, hogy $\text{hom}(F, G) \geq \text{inj}(F, G)$, és ezért

$$t(F, G) = \frac{\text{hom}(F, G)}{v(G)^{v(F)}} \geq \frac{\text{inj}(F, G)}{v(G)^{v(F)}} = t_{\text{inj}}(F, G) \cdot \frac{v(G)!/(v(G) - v(F))!}{v(G)^{v(F)}}$$

Mivel

$$\frac{v(G)!/(v(G) - v(F))!}{v(G)^{v(F)}} = \prod_{i=1}^{v(F)-1} \left(1 - \frac{i}{v(G)}\right) \geq 1 - \binom{v(F)}{2} \frac{1}{v(G)},$$

ezért

$$t(F, G) \geq t_{\text{inj}}(F, G) \left(1 - \binom{v(F)}{2} \frac{1}{v(G)}\right).$$

Másrészt, a szitaformula elejéből adódik, hogy

$$\text{inj}(F, G) = \text{hom}(F, G) - \sum_{F'} \text{hom}(F', G),$$

ahol F' az olyan gráfokat futja be, amik F -ből két csúcs egybeolvasztásával megkaphatók (ilyen gráfból összesen $\binom{v(F)}{2}$ van). Tehát mindezt összegezve írhatjuk, hogy

$$\begin{aligned} t_{\text{inj}}(F, G) &= \frac{\text{inj}(F, G)}{v(G)!/(v(G) - v(F))!} \geq \frac{\text{inj}(F, G)}{v(G)^{v(F)}} \geq \frac{\text{hom}(F, G)}{v(G)^{v(F)}} - \sum_{F'} \frac{\text{hom}(F', G)}{v(G)^{v(F)}} = \\ &= t(F, G) - \frac{1}{v(G)} \sum_{F'} t(F', G) = \\ &\geq t(F, G) - \frac{1}{v(G)} \binom{v(F)}{2}, \end{aligned}$$

ami bizonyítja a lemmát. □

Szeretnénk ezt a definíciót összekapcsolni valamilyen metrikus térrel, valamilyen metrikával, normával. Természetesen erre a δ_{\square} vágástávolság lesz a legalkalmasabb. A Lovásztól és Szegedytől származó [17] *számlálási lemmák* megállapítják, hogy a homomorfizmus-sűrűségek és a vágástávolságok milyen szoros kapcsolatban állnak egymással.

1.6.2. Lemma (Számlálási lemma gráfokra). *Bármely F, G_1 és G_2 véges egyszerű gráf esetén fennáll, hogy*

$$|t(F, G_1) - t(F, G_2)| \leq e(F)\delta_{\square}(G_1, G_2)$$

A lemma grafonokra is kiterjeszthető.

1.6.3. Lemma (Számlálási lemma grafonokra). *Bármely F véges egyszerű gráfra és $W_1, W_2 \in \mathcal{W}_0$ grafonra fennáll, hogy*

$$|t(F, W_1) - t(F, W_2)| \leq e(F)\delta_{\square}(W_1, W_2)$$

Igazából az az állítás, hogy ha adott egy véges egyszerű F gráf, akkor a $W \mapsto t(F, W)$ függvény Lipschitz-tulajdonságú a \mathcal{W}_0 térben.

A számlálási lemmát a \mathcal{W}_0 -címkézett gráfok sokkal általánosabb esetében bizonyítjuk be (emlékeztetőül lásd a 1.3.4 definíciót), de a bizonyítás így jóval egyszerűbb lesz (már többször említettük, hogy ezeket a látszólag bonyolult általánosításokat pontosan a bizonyítások leegyszerűsítése érdekében vezettük be).

1.6.4. Lemma (Számlálási lemma \mathcal{W}_0 -címkézett gráfokra). *Legyen (G, w) és (G, w') két \mathcal{W}_0 -címkézett gráf (ugyanazzal az alapgráffal), ahol $w = (W_e : e \in E(G))$ valamint $w' = (W'_e : e \in E(G))$. Ekkor teljesül, hogy*

$$|t(G, w) - t(G, w')| \leq \sum_{e \in E(G)} \|W_e - W'_e\|_{\square}$$

Bizonyítás. Ha $E(G) = \emptyset$, akkor az egyenlőtlenség triviálisan teljesül. Feltehető tehát, hogy $E(G) \neq \emptyset$.

Elegendő arra az esetre bizonyítani az állítást, amikor a címkéző grafonok egy él kivételével minden élen megegyeznek. Ugyanis ha képezünk w -tól w' -ig egy w_1, w_2, \dots, w_k sorozatot ($k \in \mathbb{N}^+$) úgy, hogy minden $i \in [k] \cup \{0\}$ -ra w_i csak egyetlen élen tér el w_{i+1} -től, ahol $w_0 = w$ illetve $w_{k+1} = w'$ (azaz w -t addig módosítjuk élenként, amíg w' -et nem kapunk), akkor fennáll, hogy

$$|t(G, w) - t(G, w')| \leq \sum_{i=0}^k |t(G, w_i) - t(G, w_{i+1})|.$$

Legyen $e = \{u, v\} \in E(G)$ olyan, hogy $W_{\{u,v\}} \neq W'_{\{u,v\}}$, illetve legyen $E' = E(G) \setminus \{e\}$. Ekkor írhatjuk, hogy

$$\begin{aligned} t(G, w) - t(G, w') &= \int_{[0,1]^{v(G)}} \cdots \int_{[0,1]^{v(G)}} \prod_{\{s,t\} \in E'} W_{\{s,t\}}(x_s, x_t) (W_{\{u,v\}}(x_u, x_v) - W'_{\{u,v\}}(x_u, x_v)) \prod_{i \in V(G)} dx_i = \\ &= \int_{[0,1]^{v(G)}} \cdots \int_{[0,1]^{v(G)}} \varphi(\mathbf{x}) \psi(\mathbf{x}) (W_{\{u,v\}}(x_u, x_v) - W'_{\{u,v\}}(x_u, x_v)) \prod_{i \in V(G)} dx_i, \end{aligned}$$

ahol $\mathbf{x} = (x_i : i \in V(G))$ a csúcsok valamilyen sorba rendezése szerint, továbbá

$$\varphi(\mathbf{x}) = \prod_{\{s,t\} \in \nabla(u) \setminus \{e\}} W_{\{s,t\}}(x_s, x_t) \quad \text{illetve} \quad \psi(\mathbf{x}) = \prod_{\{s,t\} \in E(G) \setminus \nabla(u)} W_{\{s,t\}}(x_s, x_t).$$

Ezekre igaz, hogy $\varphi: [0,1]^{v(G)} \rightarrow \mathbb{R}$ illetve $\psi: [0,1]^{v(G)} \rightarrow \mathbb{R}$ nem függ x_v -től, illetve x_u -től, továbbá $\text{ran } \varphi \subset [0,1] \supset \text{ran } \psi$. Ekkor 1.4.1-t alkalmazva adódik, hogy

$$\left| \iint_{[0,1]^2} \varphi(\mathbf{x}) \psi(\mathbf{x}) (W_{\{u,v\}}(x_u, x_v) - W'_{\{u,v\}}(x_u, x_v)) dx_u dx_v \right| \leq \|W_{\{u,v\}} - W'_{\{u,v\}}\|_{\square}$$

Ezt integrálva a többi változó szerint a bizonyítandó egyenlőtlenséget kapjuk. \square

Ebből már nyilván következik 1.6.3 és 1.6.2 is.

A számlálási lemma fordított változata [7] (amit most nem bizonyítunk) arról szól, hogy ha két nagy méretű gráf *lokálisan* (pl. homomorfizmus-sűrűségben) közel van, akkor *globálisan* (vágástávolságban) is közel van egymáshoz.

1.6.5. Lemma (Fordított számlálási lemma). *Legyen $k \in \mathbb{N}^+$ és legyen $U, W \in \mathcal{W}_0$ két grafon. Tegyük fel, hogy tetszőleges k csúcsú G egyszerű gráf esetén igaz, hogy*

$$|t(G, U) - t(G, W)| \leq \frac{1}{2^{k^2}}$$

Ekkor teljesül, hogy

$$\delta_{\square}(U, W) \leq \frac{50}{\sqrt{\log_2 k}}$$

Az eddigiek következménye az alábbi tétel⁸:

⁸[16], [7]

1.6.1. Tétel. *Az egyszerű gráfokból álló $(G_n)_{n \in \mathbb{N}}$ gráfsorozat, amelyre $\lim_{n \rightarrow \infty} v(G_n) = \infty$, akkor és csak akkor konvergens, ha δ_{\square} szerint Cauchy-sorozat.*

Bizonyítás. Ha $(G_n)_{n \in \mathbb{N}}$ Cauchy-sorozat, akkor az 1.6.2 számlálási lemma szerint konvergens is.

Megfordítva, ha $(G_n)_{n \in \mathbb{N}}$ konvergens, akkor az 1.6.5 fordított számlálási lemma szerint elég nagy $n, m \in \mathbb{N}$ -re W_{G_n} és W_{G_m} vágástávolságban tetszőlegesen közel vannak egymáshoz, amiből 1.4.2 szerint következik, hogy $(G_n)_{n \in \mathbb{N}}$ Cauchy-sorozat. \square

1.6.1 általánosítható grafonokra. A bizonyítás hasonlóan történik, csak a grafonokra kimondott számlálási lemmát kell alkalmazni.

1.6.2. Tétel. *Legyen $(W_n)_{n \in \mathbb{N}} \subset \mathcal{W}_0$ egy grafonokból álló sorozat, és legyen $W \in \mathcal{W}_0$. Ekkor a $(t(G, W_n))_{n \in \mathbb{N}} \subset \mathbb{R}$ számsorozat akkor és csak akkor lesz bármely véges egyszerű G gráf esetén konvergens, ha $(W_n)_{n \in \mathbb{N}}$ Cauchy-sorozat \mathcal{W}_0 -ban. Sőt, $t(G, W_n) \rightarrow t(G, W)$ (midőn $n \rightarrow \infty$) akkor és csak akkor teljesül bármely véges egyszerű G gráf esetén, ha $\delta_{\square}(W_n, W) \rightarrow 0$ (midőn $n \rightarrow \infty$).*

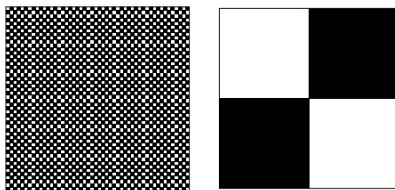
1.6.1. Megjegyzés. *Az 1.6.3 számlálási lemmából, valamint az 1.6.5 fordított számlálási lemmából következik, hogy két W, W' grafon akkor és csak akkor gyengén izomorf egymással, ha $\delta_{\square}(W, W') = 0$. Valamivel hosszabban [16] az is belátható, hogy két kernel pontosan akkor lesz gyengén izomorf egymással, ha a vágástávolságuk 0.*

1.6.1. Grafonok mint gráfsorozatok határértéke

Felvetődik a kérdés, hogy jó, hogy bizonyos gráfsorozatok konvergálnak, de *hova*? Bár ezt a problémát többféleképpen is meg lehet közelíteni, az derül ki, hogy célszerű a grafonokat használni limeszobjektumnak (az egyéb megközelítésekről a következő fejezetben majd még szót ejtünk).

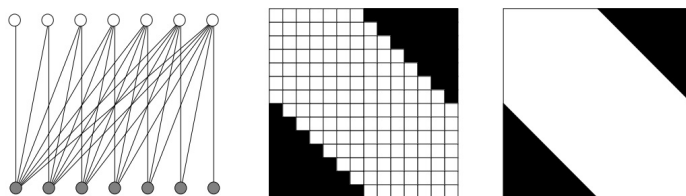
Intuitíve, képzeljük el egy $(G_n)_{n \in \mathbb{N}}$ gráfsorozat elemeinek szomszédsági mátrixainak pixelképét. Amennyiben a sorozat konvergál, azaz valamilyen értelemben a nagyobb és nagyobb indexű elemek egyre jobban kezdenek hasonlítani egymásra, akkor a pixelké-

pekben is egy bizonyos mintázat lesz egyre jobban kivehető. Mindez persze nagyban függ a csúcsok számozásától. Erre az 1.2. ábra⁹ egy jó példa.



1.2. ábra. Egy sakktáblaminta, és a sorok és oszlopok átrendezésével kapott pixelkép.

Ha $n \rightarrow \infty$, akkor a pixelek "végtelenül besűrűsödnek", és gyakorlatilag pontosan egy grafon szürkeárnyalatos képét fogják adni. Például az 1.3. ábra¹⁰ illusztrálja azt,



1.3. ábra. Egy félgráf, a pixelképe és a limeszgrafon

hogy egy félgráfokból álló (növekvő csúcsszámú) sorozat konvergens. A határértéke pedig nem lesz más, mint a $W: [0, 1]^2 \rightarrow [0, 1]$, $W(x, y) = \chi_{\{y \geq x + \frac{1}{2}\}} \cup \{x \geq y + \frac{1}{2}\}}(x, y)$ függvény.

Azért is célszerű ezt a határérték-objektumot használni, mert a grafonok jól megöröklék a gráfsorozat fontos tulajdonságait, másrészt analitikusan könnyebb velük dolgozni, mint magukkal a gráfokkal.

Az intuíció után jöjjön az egzakt tétel [17].

1.6.3. Tétel. Minden $(G_n)_{n \in \mathbb{N}}$ konvergens gráfsorozat esetén létezik egy W grafon, amelyre bármely véges egyszerű F gráf esetén teljesül, hogy

$$\lim_{n \rightarrow \infty} t(F, G_n) = t(F, W)$$

⁹[16] 1.5. ábra

¹⁰[16] 1.7. ábra

Ilyenkor azt mondjuk, hogy ez a grafon a gráfsorozat *határértéke*, és ezt $G_n \rightarrow W$ -vel jelöljük. Felvetődik a kérdés, hogy miért van szükség az összes $[0, 1]^2 \rightarrow [0, 1]$ mérhető függvényre, miért nem elég például a monoton vagy a folytonos függvényeket tekinteni. Egy későbbi állításból majd következni fog, hogy minden grafon előáll mint egy bizonyos gráfsorozat határértéke.

A tétel bizonyítása következik, [16] alapján.

Bizonyítás. 1.5.1 szerint $(\widetilde{\mathcal{W}}_0, \delta_\square)$ kompakt metrikus tér, tehát sorozatkompakt, ezért a $(W_{G_n})_{n \in \mathbb{N}} \subset \widetilde{\mathcal{W}}_0$ sorozatnak létezik egy $(W_{G_{n_k}})_{k \in \mathbb{N}}$ konvergens részsorozata. Azaz létezik egy $W \in \widetilde{\mathcal{W}}_0$ grafon, amelyre $\lim_{k \rightarrow \infty} \delta_\square(W_{G_{n_k}}, W) = 0$.

1.6.3 szerint minden véges egyszerű F gráfra és minden $k \in \mathbb{N}$ -re fennáll, hogy

$$|t(F, W_{G_{n_k}}) - t(F, W)| \leq e(G)\delta_\square(W_{G_{n_k}}, W) \longrightarrow 0 \quad (k \rightarrow \infty)$$

tehát $\lim_{k \rightarrow \infty} t(F, G_{n_k}) = \lim_{k \rightarrow \infty} t(F, W_{G_{n_k}}) = t(F, W)$ bármely véges egyszerű F gráfra. De mivel $(G_n)_{n \in \mathbb{N}}$ konvergens, ezért az 1.6.1 definíció szerint a $(t(F, G_n))_{n \in \mathbb{N}} \subset \mathbb{R}$ sorozat minden véges egyszerű F gráf mellett konvergens, így minden részsorozata is konvergens ugyanazzal a határértékkal, amiből következik, hogy bármely véges egyszerű F gráfra

$$\lim_{n \rightarrow \infty} t(F, G_n) = t(F, W)$$

□

Gond lehetne, hogy sehol sincs kizárva, hogy egy konvergens sorozatnak két különböző határértéke legyen. Például a limeszfüggvényt nullmértékű helyen megváltoztatva, még mindig teljesül a konvergencia. Viszont belátható, hogy ezek a különböző függvények nem *lényegesen* különböznek:

1.6.1. Állítás. *Legyen $(G_n)_{n \rightarrow \infty}$ egy konvergens gráfsorozat, és $W, W' \in \mathcal{W}_0$ két grafon. Ha $G_n \rightarrow W$, akkor $G_n \rightarrow W'$ akkor és csak akkor teljesül, ha W és W' egymással gyengén izomorfak.*

Bizonyítás. \Leftarrow : Ha W és W' egymással gyengén izomorfak, akkor a gyenge izomorfia definíciója szerint bármely véges egyszerű F gráfra $t(F, W) = t(F, W')$, azaz $G_n \rightarrow W'$.

\implies : Ha $G_n \rightarrow W$ és $G_n \rightarrow W'$, akkor bármely véges egyszerű F gráfra

$$t(F, W) = \lim_{n \rightarrow \infty} t(F, W_{G_n}) = t(F, W'),$$

így W és W' gyengén izomorfak. □

Ez azt jelenti, hogy $\widetilde{\mathcal{W}}_0$ felett egyértelmű a gráflimesz fogalma.

Egy gráfsorozat valamely grafonhoz való konvergenciája a vágástávolsággal is jellemezhető [16]:

1.6.4. Tétel. *Legyen $(G_n)_{n \in \mathbb{N}}$ véges egyszerű gráfoknak egy olyan sorozata, amelyre $\lim_{n \rightarrow \infty} v(G_n) = \infty$, és legyen W egy grafon. Ekkor $G_n \rightarrow W$ akkor és csak akkor, ha $\delta_{\square}(W_{G_n}, W) \rightarrow 0$.*

Bizonyítás. \Leftarrow : Ha $\delta_{\square}(W_{G_n}, W) \rightarrow 0$, akkor az 1.6.3 számlálási lemmából következik, hogy $G_n \rightarrow W$, mint az 1.6.3 tétel bizonyításában.

\implies : Ha $G_n \rightarrow W$, akkor az 1.6.5 fordított számlálási lemma alapján bármely $k \in \mathbb{N}^+$ -ra elég nagy (k -tól függő) $n \in \mathbb{N}$ esetén írhatjuk, hogy

$$\delta_{\square}(W_{G_n}, W) \leq \frac{50}{\sqrt{\log_2 k}},$$

ami $k \rightarrow \infty$ mellett igazolja a tételt. □

1.6.2. Megjegyzés. *A tételt az alábbi érveléssel is be lehetni bizonyítani:*

Bizonyítás. Mivel a $(\widetilde{\mathcal{W}}_0, \delta_{\square})$ tér kompakt, és a $W \mapsto (t(F, W): F \in \mathcal{F})$ leképezés folytonos és injektív, ezért az inverze is folytonos, ami bizonyítja a tételt. □

Ez az érvelés azonban nem mutatja az 1.6.1-ban definiált konvergencia és a vágásmetrikában való konvergencia közötti kapcsolat mélységét.

2. fejezet

Véletlen gráfok. Klasszikus modellek

A véletlen gráfokkal először Erdős és Rényi kezdett el önállóan foglalkozni az 1900-as évek közepén, miután Erdős rájött, hogy a valószínűségi módszerek gyakran nagyon hasznosak bizonyos extremális gráfelméleti problémák megoldásában. Erdős ilyen eszközökkel megmutatta például, hogy bármely $m \geq 3$ és $n \geq 3$ egész szám esetén létezik olyan gráf, melynek kromatikus száma m , a benne található legrövidebb kör pedig n hosszú. [4]

Erdős és Rényi publikációja [9] volt az első olyan tanulmány, amely bemutatta a véletlen gráfok témájára jellemző alapvető jelenségeket.

Az, hogy *mit értünk véletlen gráf alatt*, nem teljesen egyértelmű, mondhatni szerzője válogatja. Egy viszonylag általános definíció az alábbi:

2.0.1. Definíció. Ha $n \in \mathbb{N}^+$, akkor egy n csúcsú véletlen gráf alatt egy σ valószínűség-eloszlást értünk az n csúcsú gráfokon, ami invariáns a csúcsok átrendezésére. Más szóval, ez egy \mathbf{G} valószínűségi változó, aminek az értékei n csúcsú gráfok, és izomorf gráfokat egyenlő valószínűséggel vesz fel.

Az elkövetkezőkben bemutatunk néhányat a legalapvetőbb véletlen gráfok közül. Látni fogjuk, hogy ezek a gráfimeszek témájához is szorosan kapcsolódnak.

2.1. Az Erdős-Rényi-féle véletlen gráf

A modellt Erdős és Rényi [9] alkotta meg 1960-ban. Ez az egyik legegyszerűbb véletlen gráf, melyet kétféleképpen is szokás definiálni:

2.1.1. Definíció ($G(n, M)$ -modell). Legyen $n, M \in \mathbb{N}$. A $G(n, M)$ -modell szerinti Erdős-Rényi-gráf egy $[n]$ csúcshalmazú, M élű véletlen gráf, amely a lehetséges $\binom{\binom{n}{2}}{M}$ gráf közül egyenletes eloszlás szerint van kiválasztva. Jelölése: $\text{ER}_n^{(e)}(M)$

2.1.2. Definíció ($G(n, p)$ -modell). Legyen $n \in \mathbb{N}$ és $p \in [0, 1]$. A $G(n, p)$ -modell szerinti Erdős-Rényi-gráf egy $[n]$ csúcshalmazú véletlen gráf, melyben minden lehetséges élt egymástól függetlenül p valószínűséggel húzunk be. Jelölése: $\text{ER}_n(p)$ ¹

A fenti két változat szoros összefüggésben áll egymással. Ha $n \in \mathbb{N}$ és $p \in [0, 1]$ esetén $G \sim \text{ER}_n(p)$, akkor $e(G) \sim \text{Bin}\left(\binom{n}{2}, p\right)$, így $\text{ER}_n^{(e)}(M)$ esetén M nagyjából $\binom{n}{2}p$ -nek felel meg. Továbbá, ha $\text{ER}_n(p)$ és $\text{ER}_n^{(e)}(M)$ eloszlásainak valószínűség-függvényei rendre \mathbb{P}_p és \mathbb{P}_M , akkor ezek között tetszőleges E eseményre fennáll a

$$\mathbb{P}_p(E) = \sum_{m=1}^{\binom{n}{2}} \mathbb{P}_M(E) \mathbb{P}\left(\text{Bin}\left(\binom{n}{2}, p\right) = m\right)$$

összefüggés.² Az alábbi állítás is egy egyszerű észrevétel a két változat közötti kapcsolatról:

2.1.1. Állítás ([12] 1.1. lemma). *Ha $n \in \mathbb{N}^+$ és $p \in [0, 1]$, akkor egy $\text{ER}_n(p)$ véletlen gráf, feltéve, hogy éleinek száma $M \in \mathbb{N}$, egyenlő valószínűséggel egyike a $\binom{\binom{n}{2}}{M}$ db. M élű gráfnak.*

¹Ez a definíció eredetileg Edgar Nelson Gilberttől [13] származik, Erdős és Rényi [9] az 2.1.1 definíciót adták meg. A véletlen gráfok elméletében elért áttöréseikért, tiszteletből nevezték el mindkét – szorosan kapcsolódó – modellt a magyar matematikuspárosról. [14]

²[14] (4.6.1)

Bizonyítás. Legyen $\mathbf{G} \sim \text{ER}_n(p)$ egy Erdős-Rényi véletlen gráf, és $G_0 = ([n], E)$ egy egyszerű gráf, amelyre $e(G_0) = |E| = M$. Ekkor, mivel $\{\mathbf{G} = G_0\} \subset \{e(\mathbf{G}) = M\}$, így

$$\begin{aligned} \mathbb{P}_p(\mathbf{G} = G_0 \mid e(\mathbf{G}) = M) &= \frac{\mathbb{P}_p(\mathbf{G} = G_0, e(\mathbf{G}) = M)}{\mathbb{P}_p(e(\mathbf{G}) = M)} = \\ &= \frac{\mathbb{P}_p(\mathbf{G} = G_0)}{\mathbb{P}_p(e(\mathbf{G}) = M)} = \frac{p^M (1-p)^{\binom{n}{2}-M}}{\binom{\binom{n}{2}}{M} p^M (1-p)^{\binom{n}{2}-M}} = \binom{\binom{n}{2}}{M}^{-1} \end{aligned}$$

□

Tehát

$$\text{ER}_n(p)|_{e=M} \stackrel{d}{=} \text{ER}_n^{(e)}(M)$$

A következő megfigyelés a $\text{ER}_n(p)$ és $\text{ER}_n^{(e)}(M)$ változatok közötti általános kapcsolatot mutatja annak a valószínűségére nézve, hogy egy gráfnak megvan-e egy bizonyos \mathcal{P} tulajdonsága.

2.1.2. Állítás ([12] 1.2. lemma). *Legyen \mathcal{P} egy gráftulajdonság (azaz az $[n]$ csúcshalmazú gráfok halmazának egy részhalmaza), és $p = \frac{M}{\binom{n}{2}}$, ahol $M = M(n) \rightarrow \infty$ és $\binom{n}{2} - M \rightarrow \infty$, amint $n \rightarrow \infty$. Ekkor elég nagy $n \in \mathbb{N}^+$ -ra*

$$\mathbb{P}_n(\text{ER}_n^{(e)}(M) \in \mathcal{P}) \leq 10\sqrt{M} \mathbb{P}_n(\text{ER}_n(p) \in \mathcal{P})$$

Bizonyítás. Legyen $\mathbf{G}_n \sim \text{ER}_n(p)$ és $\mathbf{H}_n \sim \text{ER}_n^{(e)}(M)$ két véletlen gráf ($n \in \mathbb{N}$). A

teljes valószínűség tétele alapján írhatjuk, hogy

$$\begin{aligned}
\mathbb{P}_n(\mathbf{G}_n \in \mathcal{P}) &= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}_n(\mathbf{G}_n \in \mathcal{P} \mid e(\mathbf{G}_n) = k) \mathbb{P}_n(e(\mathbf{G}_n) = k) = \\
&= \sum_{k=0}^{\binom{n}{2}} \frac{\mathbb{P}_n(\mathbf{G}_n \in \mathcal{P}, e(\mathbf{G}_n) = k)}{\mathbb{P}_n(e(\mathbf{G}_n) = k)} \cdot \mathbb{P}_n(e(\mathbf{G}_n) = k) = \\
&= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}_n(e(\mathbf{G}_n) = k) \sum_{\substack{G \in \mathcal{P} \\ e(G)=k \\ v(G)=n}} \frac{p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \\
&= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}_n(e(\mathbf{G}_n) = k) \sum_{\substack{G \in \mathcal{P} \\ e(G)=k \\ v(G)=n}} \binom{n}{k}^{-1} = \\
&= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}_n(e(\mathbf{G}_n) = k) \mathbb{P}_n(\text{ER}_n^{(e)}(k) \in \mathcal{P}) \geq \\
&\geq \mathbb{P}_n(e(\mathbf{G}_n) = M) \mathbb{P}_n(\mathbf{H}_n \in \mathcal{P})
\end{aligned}$$

Továbbá $e(\mathbf{G}_n) \sim \text{Bin}(\binom{n}{2}, p)$, így a Stirling-formula szerint

$$k! = (1 + o(1)) \left(\frac{k}{e}\right)^k \sqrt{2\pi k},$$

ezért az $N := \binom{n}{2}$ jelöléssel élve

$$\begin{aligned}
\mathbb{P}_n(e(\mathbf{G}_n) = M) &= \binom{N}{M} p^M (1-p)^{N-M} = \\
&= (1 + o(1)) \frac{N^N \sqrt{2\pi N} p^M (1-p)^{N-M}}{M^M (N-M)^{N-M} 2\pi \sqrt{M(N-M)}} = \\
&= (1 + o(1)) \sqrt{\frac{N}{2\pi M(N-M)}},
\end{aligned}$$

amiből következik, hogy elég nagy n -re $\mathbb{P}_n(e(\mathbf{G}_n) = M) \geq \frac{1}{10\sqrt{M}}$, ami már bizonyítja az állítást. \square

Mivel független élekkel gyakran könnyebb dolgozni, főleg az $\text{ER}_n(p)$ változatot fogjuk használni. Ez az egyik legegyszerűbb véletlen gráf, mégis érdekes jelenségek figyelhetők meg rajta, ha n elég nagy [21]. Ha $p_n = \frac{\lambda}{n}$ ($\lambda > 0$), akkor $\text{ER}_n(p_n)$ egy csúcs, pl.

az 1 fokszáma $\text{Bin}(n-1, p)$ binomiális eloszlású. Ismert, hogy tetszőleges $k \in \mathbb{N}$ esetén

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\text{Bin} \left(n, \frac{\lambda}{n} \right) = k \right) = e^{-\lambda} \frac{\lambda^k}{k!}$$

és itt $e^{-\lambda} \frac{\lambda^k}{k!}$ pontosan a λ paraméterű Poisson-eloszlás valószínűség-függvénye a k helyen. Bizonyítás nélkül kimondjuk az alábbi tételt, amely szerint ennél több is igaz:

2.1.1. Tétel ([14] 5.12 tétel). *Legyen $\lambda > 0$, $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ bármely $k \in \mathbb{N}$ -re. Ekkor ha $(\varepsilon_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ olyan sorozat, amelyre $\lim_{n \rightarrow \infty} n\varepsilon_n^2 = \infty$, akkor*

$$\lim_{n \rightarrow \infty} \mathbb{P}_\lambda \left(\max_{k \in \mathbb{N}} \left| P_k^{(n)} - p_k \right| \geq \varepsilon_n \right) = 0,$$

ahol $P_k^{(n)}$ egy $\text{ER}_n(\frac{\lambda}{n})$ Erdős-Rényi véletlen gráf fokszámeloszlása (lásd: 1.1.1 alfejezet).

A modell leghíresebb tulajdonsága azonban az, hogy fázisátmenetet mutat – van egy éles paraméterküszöb, ahol a gráf tulajdonságai drasztikusan megváltoznak: a legnagyobb összefüggő komponensben lévő csúcsok számára, amelyet $|\mathcal{C}_{\max}(n, p)|$ -vel jelölünk, a következő klasszikus eredményt kapjuk:

2.1.2. Tétel ([21] 1. tétel). *Legyen $\lambda > 0$ és $p_n = \frac{\lambda}{n}$ bármely $n \in \mathbb{N}^+$ esetén.*

(a) *Ha $\lambda < 1$, akkor $\frac{|\mathcal{C}_{\max}(n, p_n)|}{\ln n} \xrightarrow{p} \frac{1}{\lambda - 1 - \ln \lambda}$ sztochasztikusan, amint $n \rightarrow \infty$.*

(b) *Ha $\lambda > 1$, akkor $\frac{|\mathcal{C}_{\max}(n, p_n)|}{n} \xrightarrow{p} \zeta$ sztochasztikusan, amint $n \rightarrow \infty$, ahol ζ a λ paraméterű Poisson-folyamat túlélési valószínűsége.*

A kritikus eset, azaz ha $\lambda = 1$, is érdekes. Ekkor belátható ([14] 5.1.), hogy közel 1 valószínűséggel $|\mathcal{C}_{\max}(n, p_n)| = O\left(n^{\frac{2}{3}}\right)$.

A modell definícióját vizsgálva sejthetjük, hogy $n \rightarrow \infty$ esetén az $\text{ER}_n(p)$ Erdős-Rényi-féle véletlen gráf *konvergálni fog*.³

2.1.3. Állítás. *Legyen $p \in [0, 1]$, és $W: [0, 1]^2 \rightarrow [0, 1]$, $W \equiv p$ az azonosan p grafon. Ekkor 1 valószínűséggel $\text{ER}_n(p) \rightarrow W$, amint $n \rightarrow \infty$.*

³[6] 5. példa

Bizonyítás. Legyen F egy rögzített véges egyszerű gráf (melyre $v(F) \geq 2$), illetve minden $n \in \mathbb{N}^+$ -ra $\mathbf{G}_n \sim \text{ER}_n(p)$. Ekkor

$$\text{ind}(F, \mathbf{G}_n) = \sum_H \mathbf{1}_H,$$

ahol H a K_n teljes gráf összes F -részgráfján fut végig, míg $\mathbf{1}_H$ annak az indikátora, hogy H megtalálható-e \mathbf{G}_n -ben beágyazott részgráfként. Ekkor minden ilyen H -ra teljesül, hogy $\mathbb{E}[\mathbf{1}_H] = p^{v(F)}(1-p)^{\binom{v(F)}{2}-e(F)}$, így egyszerű leszámolások miatt

$$\mathbb{E}[\text{ind}(F, \mathbf{G}_n)] = \binom{n}{v(F)} |\text{Aut}(F)| p^{v(F)} (1-p)^{\binom{v(F)}{2}-e(F)} = t_{\text{ind}}(F, W),$$

vagyis

$$\mathbb{E}[\text{ind}(F, \mathbf{G}_n)] = \Theta(n^{v(F)}). \quad (2.1)$$

Másrészt,

$$\begin{aligned} D^2(\text{ind}(F, \mathbf{G}_n)) &= \mathbb{E}[\text{ind}(F, \mathbf{G}_n)^2] - \mathbb{E}[\text{ind}(F, \mathbf{G}_n)]^2 = \\ &= \sum_{H, H'} (\mathbb{E}[\mathbf{1}_H \mathbf{1}_{H'}] - \mathbb{E}[\mathbf{1}_H] \mathbb{E}[\mathbf{1}_{H'}]) = \\ &= \sum_{H=H'} (\mathbb{E}[\mathbf{1}_H \mathbf{1}_{H'}] - \mathbb{E}[\mathbf{1}_H] \mathbb{E}[\mathbf{1}_{H'}]) + \sum_{H \neq H'} (\mathbb{E}[\mathbf{1}_H \mathbf{1}_{H'}] - \mathbb{E}[\mathbf{1}_H] \mathbb{E}[\mathbf{1}_{H'}]) \leq \\ &\leq \sum_H \mathbb{E}[\mathbf{1}_H] + \sum_{H \neq H'} (\mathbb{E}[\mathbf{1}_H \mathbf{1}_{H'}] - \mathbb{E}[\mathbf{1}_H] \mathbb{E}[\mathbf{1}_{H'}]) = \\ &= \mathbb{E}[\text{ind}(F, \mathbf{G}_n)] + \sum_{H \neq H'} (\mathbb{E}[\mathbf{1}_H \mathbf{1}_{H'}] - \mathbb{E}[\mathbf{1}_H] \mathbb{E}[\mathbf{1}_{H'}]) \end{aligned}$$

Ha H és H' nem egyeznek meg legalább két csúcsban, akkor az utolsó összeg 0, hiszen ekkor $\mathbf{1}_H$ és $\mathbf{1}_{H'}$ függetlenek, így $\mathbb{E}[\mathbf{1}_H \mathbf{1}_{H'}] = \mathbb{E}[\mathbf{1}_H] \mathbb{E}[\mathbf{1}_{H'}]$. Ezért

$$\begin{aligned} D^2(\text{ind}(F, \mathbf{G}_n)) &\leq \mathbb{E}[\text{ind}(F, \mathbf{G}_n)] + \sum_{\substack{H \neq H' \\ |V(H) \cap V(H')| \geq 2}} (\mathbb{E}[\mathbf{1}_H \mathbf{1}_{H'}] - \mathbb{E}[\mathbf{1}_H] \mathbb{E}[\mathbf{1}_{H'}]) \leq \\ &\leq \mathbb{E}[\text{ind}(F, \mathbf{G}_n)] + \sum_{\substack{H \neq H' \\ |V(H) \cap V(H')| \geq 2}} \mathbb{E}[\mathbf{1}_H \mathbf{1}_{H'}] = \\ &= O(n^{v(F)}) + O(n^{v(F)+v(F)-2}) = O(n^{2v(F)-2}) \end{aligned}$$

Ha $X_n = \text{ind}(F, \mathbf{G}_n)$, akkor mivel $t_{\text{ind}}(F, \mathbf{G}_n) = \frac{X_n}{n!/(n-v(F))!}$, így alkalmazva a Csebisev-egyenlőtlenséget, kapjuk, hogy tetszőleges pozitív ε_n -ek ($n \in \mathbb{N}^+$) esetén

$$\begin{aligned} \mathbb{P}(|t_{\text{ind}}(F, \mathbf{G}_n) - \mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_n)]| > \varepsilon_n \mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_n)]) &= \mathbb{P}(|X_n - \mathbb{E}[X_n]| > \varepsilon_n \mathbb{E}[X_n]) \leq \\ &\leq \frac{D^2(X_n)}{\varepsilon_n^2 \mathbb{E}[X_n]^2} = O\left(\frac{1}{\varepsilon_n^2 n^2}\right), \end{aligned}$$

ε_n -t válasszuk $\frac{1}{\sqrt[4]{n}}$ -nek ($n \in \mathbb{N}^+$). Ekkor $\frac{1}{\varepsilon_n^2 n^2} = \frac{1}{n\sqrt{n}}$, így (például az integrálkritérium miatt) a $\sum_{n \in \mathbb{N}^+} \frac{1}{\varepsilon_n^2 n^2}$ sor konvergens, ezért a $\sum_{n \in \mathbb{N}^+} \frac{D^2(X_n)}{\varepsilon_n^2 \mathbb{E}[X_n]^2}$ sor is konvergens lesz. Így a Borel-Cantelli lemma szerint

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|t_{\text{ind}}(F, \mathbf{G}_k) - \mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_k)]| > \varepsilon_k \mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_k)]\}\right) = 0,$$

vagyis 1 valószínűséggel $|t_{\text{ind}}(F, \mathbf{G}_n) - \mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_n)]| > \varepsilon_n \mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_n)]$ csak véges sok $n \in \mathbb{N}^+$ -re teljesül. Mivel (2.1) miatt $\varepsilon_n \mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_n)] \rightarrow 0$, amint $n \rightarrow \infty$, ez azt jelenti, hogy 1 valószínűséggel $t_{\text{ind}}(F, \mathbf{G}_n) \rightarrow \mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_n)] = t_{\text{ind}}(F, W)$, amint $n \rightarrow \infty$, azaz 1 valószínűséggel

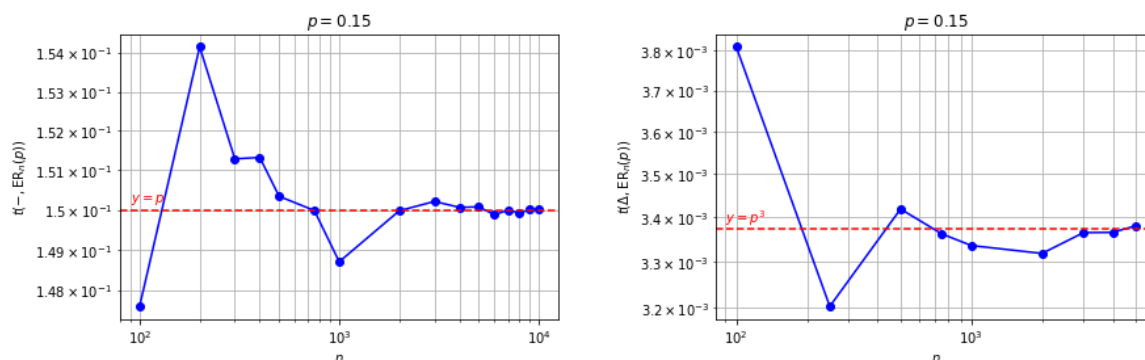
$$\mathbf{G}_n \rightarrow W \equiv p,$$

amint $n \rightarrow \infty$. □

2.1.1. Megjegyzés. Ha $p_n = \frac{\lambda}{n}$ valamely $\lambda > 0$ mellett, akkor 1 valószínűséggel teljesül, hogy $\text{ER}_n(p_n) \rightarrow 0$, amint $n \rightarrow \infty$ (azaz az így képzett Erdős-Rényi gráfsorozat 1 valószínűséggel az azonosan 0 grafonhoz tart).

2.1.2. Megjegyzés. Később látni fogjuk, hogy $\text{ER}_n(p)$ sajátos esete az ún. sztochasztikus blokkmodellnek, ami pedig a következő szakaszban tárgyalt grafonból sorsolt véletlen gráfoknak egy sajátos esete. Emiatt az utóbbiak általános konvergenciatulajdonsága miatt is azonnal adódik az Erdős-Rényi gráfok konvergenciája.

A 2.1 ábrán jól látható, $F = K_2$ illetve $F = K_3$ teljes gráfok esetén, hogy amint $n \rightarrow \infty$, $t(F, \text{ER}_n(p)) \rightarrow p^{e(F)}$



2.1. ábra. Erdős-Rényi-modell szerint sorsolt gráfok él- illetve háromszögsűrűségének alakulása, amint n növekszik (3.1, 3.2)

2.2. Grafonból sorsolt véletlen gráf

Grafonokkal az Erdős-Rényi gráfoknál jóval általánosabb véletlen gráfokat is gyárthatunk. A konstrukciót Lovász és Szegedy 2006-ban vezette be [17], de tőlük függetlenül mások is megfogalmazták. ([16] 10.1).

2.2.1. Definíció. Legyen W egy grafon, $n \in \mathbb{N}^+$, $S = (x_1, \dots, x_n) \in [0, 1]^n$. Ekkor:

- a $\mathbb{G}(S, W)$ egyszerű véletlen gráfot, melyre $V(\mathbb{G}(S, W)) = [n]$, és minden $u, v \in [n], u \neq v$ -re, egymástól függetlenül, $\{u, v\} \in E(\mathbb{G}(S, W))$ -nek $W(x_u, x_v)$ a valószínűsége, a W -ből sorsolt véletlen gráfnak, vagy röviden W -random gráfnak nevezzük.
- Röviden $\mathbb{G}(n, W)$ -vel azt a $\mathbb{G}(S, W)$ W -random gráfot jelöljük, ahol $S = (X_1, \dots, X_n)$, és X_1, \dots, X_n a $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változók.
- $\mathbb{G}(W)$ -vel azt a megszámlálható egyszerű véletlen gráfot jelöljük, melynek csúcshalmaza \mathbb{N} , élhalmaza E , és két különböző $u, v \in \mathbb{N}$ esetén $\{u, v\} \in E$ valószínűsége $W(X_u, X_v)$, ahol $(X_i)_{i \in \mathbb{N}}$ egy független, $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változókból álló sorozat.

Például egy véges egyszerű G gráf esetén, ha feltesszük, hogy az x_1, \dots, x_k ($k \leq v(G)$ pozitív egész) számok a W_G különböző lépcsőiből valók, akkor $(G)(k, W_G)$ a G gráfnak

egy k csúcsú véletlen indukált részgráfja (lásd: 1.1.2), amit $\mathbb{G}(k, G)$ -vel jelölünk. Ez a feltétel pontosan azokat az x_1, \dots, x_k szám k -asokat zárják ki, amelyek tartalmaznak ismétlődést, és ezeknek a (valószínűségi) mértéke $\frac{\binom{k}{2}}{v(G)}$, ezért a két véletlen gráf eloszlásainak (1.1) szórástávolságára fennáll, hogy

$$d_{\text{var}}(\mathbb{G}(k, G), \mathbb{G}(k, W_G)) \leq \frac{\binom{k}{2}}{v(G)}$$

2.2.1. Megjegyzés. *Itt – talán helyesebben – írhattunk volna $d_{\text{var}}(\sigma_{G,k}, \sigma_{W_G,k})$ -t is, de a 2.0.1 definíció szerint magát a véletlen gráf fogalmát eloszlásként is lehet használni. Az utóbbi jelölést inkább akkor fogjuk használni, amikor az eloszlás (mint valószínűségfüggvény) értékére hivatkozunk egy bizonyos helyen.*

Vizsgáljuk meg a modell *konvergenciáját*. Legyen W egy grafon, és tetszőleges F véges egyszerű gráf esetén tekintsük a $(t_{\text{ind}}(F, \mathbb{G}(n, W)))_{n \in \mathbb{N}^+}$ sorozatot. $n \geq v(F)$ mellett 1.3.1-hez hasonlóan belátható, hogy

$$t_{\text{ind}}(F, \mathbb{G}(n, W)) = t_{\text{ind}}(F, W_{\mathbb{G}(n, W)}),$$

így várható értékre térve könnyen adódik, hogy

$$\mathbb{E}[t_{\text{ind}}(F, \mathbb{G}(n, W))] = t_{\text{ind}}(F, W)$$

Másrészt, 2.2.1 bizonyításából látni fogjuk, hogy $t_{\text{ind}}(F, \mathbb{G}(n, W))$ erősen koncentrált a várható értéke körül, azaz tetszőleges $\varepsilon > 0$ esetén 1 valószínűséggel

$$|t_{\text{ind}}(F, \mathbb{G}(n, W)) - \mathbb{E}[t_{\text{ind}}(F, \mathbb{G}(n, W))]| > \varepsilon$$

csak véges sok n -re teljesül. Ennélfogva 1 valószínűséggel $t_{\text{ind}}(F, \mathbb{G}(n, W)) \rightarrow t_{\text{ind}}(F, W)$ (amint $n \rightarrow \infty$), vagyis igaz az alábbi állítás:

2.2.1. Állítás. *Legyen W egy tetszőleges grafon. Ekkor $\mathbb{G}(n, W) \rightarrow W$ (amint $n \rightarrow \infty$) 1 valószínűséggel teljesül.*

2.2.1. Véletlen gráfmodellek mint gráfsorozatok határértéke

Az előző fejezet 1.6.1 részében szó esett arról, hogy a gráfsorozatok konvergenciájának tanulmányozása során nem csak a grafonok használhatók határérték-objektumként. Most bővebben kifejtjük azt a lehetőséget, amikor a gráfsorozatok limeszének véletlen gráfmodelleket használunk. Ezt a konstrukciót fogjuk *gyenge határértéknek* nevezni. Egy egyszerű G gráf és egy $k < v(G)$ pozitív egész szám esetén jelölje a $\mathbb{G}(k, G)$ véletlen indukált részgráf eloszlását $\sigma_{G,k}$. Ekkor bármely véges egyszerű F gráf esetén fennáll, hogy $\sigma_{G,k}(F) = t_{\text{ind}}(F, G)$. Amennyiben egy konvergens $(G_n)_{n \in \mathbb{N}}$ sorozatot tekintünk, akkor látni fogjuk, hogy a $\sigma_{G_n,k}$ eloszlások minden k -ra tartani fognak egy σ_k eloszláshoz, és fordítva, ha a véletlen indukált részgráfok eloszlása konvergál, akkor a gráfsorozat is konvergálni fog.

2.2.2. Definíció. *Véletlen gráfmodellnek* egy olyan σ_k valószínűségeloszlást nevezünk a $[k]$ csúcshalmazú gráfokon, minden $k \in \mathbb{N}^+$ -ra, ami invariáns a csúcsok átrendezésére nézve. Más szóval, ez valószínűségi változóknak egy olyan $(\mathbf{G}_k)_{k \in \mathbb{N}^+}$ sorozata, amely minden \mathbf{G}_k tagjának a lehetséges értékei minden $k \in \mathbb{N}^+$ -ra $[k]$ csúcshalmazú gráfok, és izomorf gráfokat azonos valószínűséggel vesz fel.

2.2.3. Definíció. Azt mondjuk, hogy egy véletlen gráfmodell *konzisztens*, ha bármely $k - 1$ csúcsú H gráfra

$$\sigma_{k-1}(H) = \sum_{F: F'=H} \sigma_k(H),$$

ahol F' az F gráfból a k csúcs kitörlésével keletkező gráf.

Más szóval, a véletlen gráfmodell konzisztens, ha \mathbf{G}_k -ből kitörölve a k csúcsot, a keletkező gráfnak ugyanaz az eloszlása, mint \mathbf{G}_{k-1} -nek.

2.2.4. Definíció. Azt mondjuk, hogy egy véletlen gráfmodell *lokális*, ha bármely két $S, T \subset [k]$ diszjunkt halmazra \mathbf{G}_k -nak az S és a T által kifizített részgráfjai mint valószínűségi változók függetlenek.

Triviális példa a fenti tulajdonságokra a W -random gráf. Tetszőleges W grafon esetén a $\mathbf{G}_k = \mathbb{G}(k, W)$ ($k \in \mathbb{N}^+$) véletlen gráfmodell *konzisztens* és *lokális* is egyben

2.2.2. Megjegyzés. A konzisztenciából és a csúcsok átrendezésére való invarianciából együtt következik, hogy ha $n \geq k$, akkor tetszőleges véges egyszerű k csúcsú F gráfra az

$$\mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_n)] = \sigma_k(F)$$

várható érték független n -től.

A fontosabb tétel előtt szükséges az alábbi lemma:

2.2.1. Lemma ([16] 11.8. lemma). *Legyen $(\sigma_n)_{n \in \mathbb{N}^+}$ egy lokális és konzisztens véletlen gráfmodell, és legyen \mathbf{G}_n a σ_n eloszlás szerint különböző n -ekre függetlenül sorsolt véletlen gráf. Ekkor $(\mathbf{G}_n)_{n \in \mathbb{N}^+}$ 1 valószínűséggel konvergens.*

Bizonyítás. Először megmutatjuk, hogy ha $k \in \mathbb{N}$, akkor bármely $[k]$ csúcshalmazú F egyszerű gráfra és $n \geq k$ egészre

$$\mathbb{E}[t_{\text{ind}}(F, \mathbf{G}_n)] = \sigma_k(F) \tag{2.2}$$

Valóban, tekintsünk egy $f: [k] \rightarrow v(\mathbf{G}_n)$ leképezést. σ_n izomorfia-invarianciája miatt annak valószínűsége, hogy $f \in \text{Ind}(F, \mathbf{G}_n)$ minden f -re ugyanaz, tehát elegendő ezt a valószínűséget arra az esetre kiszámolni, amikor f az identikus függvény. A véletlen gráfmodell konzisztenciája miatt ez a valószínűség nem más, mint $\mathbb{P}(\mathbf{G}_k = F) = \sigma_k(F)$.

A továbbiakban legyen $k < \sqrt{n}$ és $S_1, S_2, S_3, S_4 \subset [n]$ egymástól független véletlen rendezett k -asok. Legyen emellett minden $1 \leq i \leq 4$ indexre $X_i = \chi_{\{\mathbf{G}_n[S_i]=F\}} - \sigma_k(F)$. Ekkor 2.2 miatt $\mathbb{E}[X_i] = 0$ minden $1 \leq i \leq 4$ -re. Sőt, még $\mathbb{E}[X_i | S_i] = 0$ is teljesül, mivel $\mathbf{G}_n[S]$ eloszlása bármely $S \subset [n]$ rendezett k -as esetén ugyanaz.

Mivel ha \mathbf{G}_n adott, akkor az X_i -k egymástól függetlenek, így

$$\begin{aligned} \mathbb{E}[X_1 X_2 X_3 X_4] &= \mathbb{E}[\mathbb{E}[X_1 X_2 X_3 X_4 | \mathbf{G}_n]] = \\ &= \mathbb{E} \left[\prod_{i=1}^4 \mathbb{E}[X_i | \mathbf{G}_n] \right] = \mathbb{E} [(t_{\text{ind}}(F, \mathbf{G}_n) - \sigma_k(F))^4] \end{aligned}$$

Legyen $E = \{S_i \cap S_j \neq \emptyset \quad \forall 1 \leq i < j \leq 4\}$ egy esemény. Ha az S_i -k olyanok, hogy mondjuk S_1 diszjunkt a többitől, akkor X_1 független $\{X_2, X_3, X_4\}$ -től, és – mint ahogy

azt már megjegyeztük – a várható értéke 0 (itt használjuk, hogy a modell lokális!). Ennélfogva $\mathbb{E}[X_1 X_2 X_3 X_4 | \bar{E}] = 0$, tehát a teljes várható érték tétele szerint

$$\begin{aligned} \mathbb{E}[X_1 X_2 X_3 X_4] &= \mathbb{E}[X_1 X_2 X_3 X_4 | \bar{E}] \mathbb{P}(\bar{E}) + \mathbb{E}[X_1 X_2 X_3 X_4 | E] \mathbb{P}(E) \leq \\ &\leq \mathbb{P}(E) \leq \frac{7k^4}{n^2} \end{aligned}$$

ahol az utolsó egyenlőtlenség egyszerű kombinatorikai okokból igaz. Így a Markov-egyenlőtlenséget használva írhatjuk, hogy tetszőleges $\varepsilon > 0$ mellett

$$\begin{aligned} \mathbb{P}(|t_{\text{ind}}(F, \mathbf{G}_n) - \sigma_k(F)| > \varepsilon) &= \mathbb{P}((t_{\text{ind}}(F, \mathbf{G}_n) - \sigma_k(F))^4 > \varepsilon^4) \leq \\ &\leq \frac{\mathbb{E}[(t_{\text{ind}}(F, \mathbf{G}_n) - \sigma_k(F))^4]}{\varepsilon^4} \leq \frac{7k^4}{\varepsilon^4 n^2} \end{aligned}$$

Az előzőt minden $n \in \mathbb{N}^+$ -ra összeadva kapjuk, hogy a $\sum_{n \in \mathbb{N}^+} \mathbb{P}(|t_{\text{ind}}(F, \mathbf{G}_n) - \sigma_k(F)| > \varepsilon)$ sor konvergens, így a Borel-Cantelli lemma alapján

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} \{|t_{\text{ind}}(F, \mathbf{G}_j) - \sigma_k(F)| > \varepsilon\}\right) = 0,$$

vagyis 1 valószínűséggel $|t_{\text{ind}}(F, \mathbf{G}_n) - \sigma_k(F)| > \varepsilon$ csak véges sok n -re teljesül, vagyis 1 valószínűséggel $t_{\text{ind}}(F, \mathbf{G}_n) \rightarrow \sigma_k(F)$, amint $n \rightarrow \infty$. Tehát $(\mathbf{G}_n)_{n \in \mathbb{N}^+}$ 1 valószínűséggel konvergens. \square

Ezek után, egy kis lokalitásra és konzisztenciára vonatkozó ügyeskedéssel már bebizonyítható az alábbi tétel:

2.2.1. Tétel ([16] 11.7. tétel). *Ha egy $(G_n)_{n \in \mathbb{N}^+}$ gráfsorozat konvergens, akkor a $\sigma_k = \lim_{n \rightarrow \infty} \sigma_{k, G_n}$ eloszlások lokális és konzisztens véletlen gráfmodellt határoznak meg. Sőt, minden lokális és konzisztens véletlen gráfmodell előáll egy konvergens gráfsorozat véletlen indukált részgráfjainak eloszlásainak határértékeként.*

Bizonyítás. Adott G véges egyszerű gráfra legyen $\sigma'_{k, G}(F) = t_{\text{ind}}(F, W_G)$ bármely véges egyszerű F gráfra. Ekkor a $(\sigma'_{k, G})_{k \in \mathbb{N}^+}$ modell lokális és konzisztens is, az ezekből az eloszlásokból rendre sorsolt véletlen gráfok a $(\mathbb{G}(k, W_G))_{k \in \mathbb{N}^+}$ sorozatot alkotják.

Másrészt, ha $\mathbb{G}(k, W_G)$ -ből úgy veszünk mintát (lásd: 1.1.2), hogy csak akkor tartjuk meg az eredményt, ha a kapott pontok megfelelnek a G gráf *különböző* csúcsainak (ellenben újból mintát veszünk), akkor a $\mathbb{G}(k, G)$ eloszlású mintát kapunk, és

$$d_{\text{var}}(\sigma'_{k,G}, \sigma_{k,G}) \leq 1 - \frac{v(G)!}{(v(G) - k)! v(G)^k} < \frac{1}{v(G)} \binom{k}{2} \rightarrow 0 \quad (v(G) \rightarrow \infty)$$

Ebből a σ_{k,G_n} -ek $n \rightarrow \infty$ melletti határértékeinek lokalitása és konzisztenciája következik.

Fordítva, tekintsük a $(\sigma_n)_{n \in \mathbb{N}^+}$ lokális és konzisztens véletlen gráfmodellt, és az ennek elemei szerint különböző n -ekre függetlenül sorsolt \mathbf{G}_n ($n \in \mathbb{N}^+$) véletlen gráfokból álló sorozatot. 2.2.1 szerint ez a gráfsorozat 1 valószínűséggel konvergens, (2.2)-ből pedig következik, hogy ez a megfelelő gráfmodellt reprodukálja. \square

2.3. Preferential attachment-gráfok

Az előző modellek *statikusak* voltak, azaz a gráf mérete mindkét esetben rögzítve volt. A valódi hálózatoknak jó része viszont *dinamikus*, ami azt jelenti, hogy a gráf (mérete) az időben változik. Jó példa lehet erre a világháló (WWW), biológiai hálózatok vagy a tudományos cikkek hivatkozásainak hálózata.

Egy másik példa a szociális hálózatok modellezése. Ezt akár az Erdős-Rényi-gráf dinamikus megfogalmazásával (minden pillanatban az új csúcsnak a már meglévők közül azonos valószínűséggel választunk szomszédot [9]) is megtehetnénk, de ez nem biztos, hogy jól tükrözi a valóságot: abban az esetben, ha egy adott társaságba egy új ember kerül be, inkább azt várjuk, hogy az új személy egy olyan emberrel köt ismeretséget, akinek már több ismerőse van. Ezért az alábbiakban definiálásra kerülő⁴ ún. *preferential attachment-gráfok* valamivel jobban leírják egy bővülő szociális hálózat tulajdonságait. Ezt a modellt először Barabási és Albert vezették be.

⁴[14] 8. fejezete alapján

A modell egy gráfsorozatot fog képezni, amelyet $\left(\text{PA}_t^{(m,\delta)}\right)_{t \in \mathbb{N}^+}$ -vel fogunk jelölni. E sorozat minden t . tagjának ($t \in \mathbb{N}^+$) t db. csúcsa és mt db. éle lesz valamilyen $m \in \mathbb{N}^+$ esetén.

Először az $m = 1$ esetet írjuk le, amikor a gráfsorozatunk fákból fog állni. $\text{PA}_1^{(1,\delta)}$ álljon egyetlen csúcsból és egy hurokélból. $\text{PA}_t^{(1,\delta)}$ csúcsait jelöljük $v_1^{(1)}, \dots, v_t^{(1)}$ -el, a $v_i^{(1)}$ csúcs $\text{PA}_t^{(1,\delta)}$ -beli fokszámát jelölje $D_i(t)$ (az a konvenció, hogy egy hurokél kettővel növeli az adott csúcs fokszámát).

Ezután minden $t \in \mathbb{N}^+$ -ra, feltéve, hogy a $\text{PA}_t^{(1,\delta)}$ gráf ismert, a $\text{PA}_{t+1}^{(1,\delta)}$ gráf a következőképpen képződik: $\text{PA}_t^{(1,\delta)}$ -hoz hozzávesszük a $v_{t+1}^{(1)}$ csúcsot, és egy élet, ami illeszkedik erre az új csúcsra. A másik végpontot úgy választjuk, hogy annak valószínűsége, hogy ez maga $v_{t+1}^{(1)}$ (vagyis az új él hurokél) $\frac{1+\delta}{t(2+\delta)+1+\delta}$, annak valószínűsége, hogy ez (valamely $i \in [t]$ -re) $v_i^{(1)}$ pedig $\frac{D_i(t)+\delta}{t(2+\delta)+1+\delta}$, ahol $\delta \in [-1, \infty)$ a modell egy paramétere. Ha (valamely $i \in [t+1]$ esetén) a $v_i^{(1)}$ csúcsnak az új él második végpontjának való választását mint eseményt $v_{t+1}^{(1)} \rightarrow v_i^{(1)}$ jelöli, akkor

$$\mathbb{P}\left(v_{t+1}^{(1)} \rightarrow v_i^{(1)} \mid \text{PA}_t^{(1,\delta)}\right) = \begin{cases} \frac{1+\delta}{t(2+\delta)+1+\delta} & \text{ha } i = t+1 \\ \frac{D_i(t)+\delta}{t(2+\delta)+1+\delta} & \text{ha } i \in [t] \end{cases}$$

Az $m > 1$ esetet az $m = 1$ eset alapján értelmezzük, a következőképpen. Tekintsük az előzőek szerint generált $\text{PA}_{mt}^{(1,\frac{\delta}{m})}$ gráfot, melynek csúcsait jelölje rendre $v_1^{(1)}, \dots, v_{mt}^{(1)}$. Vegyük ebben a gráfban a $v_1^{(1)}, \dots, v_m^{(1)}$ csúcsokat, és olvasszuk őket egybe a $\text{PA}_t^{(m,\delta)}$ gráf $v_1^{(m)}$ csúcsává úgy, hogy minden olyan $\text{PA}_{mt}^{(1,\frac{\delta}{m})}$ -beli élre, amelyre a $v_1^{(1)}, \dots, v_m^{(1)}$ csúcsok valamelyike illeszkedett, ezek helyett illeszkedjen $\text{PA}_t^{(m,\delta)}$ -ben $v_1^{(m)}$. Ugyanígy olvasszuk össze a $v_{m+1}^{(1)}, \dots, v_{2m}^{(1)}$ csúcsokat a $\text{PA}_t^{(m,\delta)}$ gráf $v_2^{(m)}$ csúcsává, stb., általában a $v_{(j-1)m+1}^{(1)}, \dots, v_{jm}^{(1)}$ csúcsokat a $\text{PA}_t^{(m,\delta)}$ gráf $v_j^{(m)}$ csúcsává minden $j \in [t]$ esetén. Eredményül egy olyan $\text{PA}_t^{(m,\delta)}$ multigráfot kapunk, aminek pontosan t csúcsa és mt éle van.

Meggondolható ([14] 8.2.), hogy annak valószínűsége, hogy $\text{PA}_t^{(m,\delta)}$ -ben valamely $v_j^{(m)}$ csúcsra élt illesztünk, arányos a $v_j^{(m)}$ csúcs $\text{PA}_t^{(m,\delta)}$ -beli fokszámánál δ -val nagyobb mennyiséggel. Azt is érdemes megjegyezni, hogy a fenti konstrukcióban a fokszámokat

minden él bekötése után frissítjük. Ezt *a foksámok időközi frissítésének* nevezzük.

Kulcstulajdonsága ennek a modellnek az, hogy az újonnan megjelenő élek nagy valószínűséggel a már nagy foksámú csúcsokhoz fognak csatlakozni, ezek foksámát ezzel még nagyobbá téve. Ezt a jelenséget nevezik *preferential attachment*-nek. A modellt néha még *Rich-get-Richer* modellnek (*a gazdagok gazdagabbak lesznek*) modellnek is nevezik.

Az $m \geq 2$ eseteket direkt módon is definiálhatjuk a következő módon: legyen $\text{PA}_1^{(m,\delta)}$ egy egycsúcsú gráf, m darab hurokéllel. Ezután minden $t \in \mathbb{N}^+$ -ra konstruáljuk $\text{PA}_t^{(m,\delta)}$ -ből $\text{PA}_{t+1}^{(m,\delta)}$ -t úgy, hogy az előbbihez hozzáveszünk egy új csúcsot, m darab ehhez csatlakozó éllel. Az éleket sorban, a foksámok időközi frissítésével húzzuk be úgy, hogy az e . él a $v_i^{(m)}$ csúcshoz $i \in [t]$ esetén a $D_i(e-1, t) + \delta$ mennyiséggel arányos valószínűséggel csatlakozzon (itt $D_i(e, t)$ a $v_i^{(m)}$ csúcs foksáma az e . $v_{t+1}^{(m)}$ csúcshoz kapcsolódó él behúzása után), míg $i = t+1$ esetén ez a valószínűség legyen a $D_{t+1}(e-1, t) + 1 + \frac{e\delta}{m}$ mennyiséggel arányos.

2.3.1. Megjegyzés. *Ha $\delta = 0$, akkor az alábbi modellt kapjuk: a kiindulási gráf egy összefüggő $m_0 \in \mathbb{N}^+$ csúcsú gráf. Minden lépésben egy új csúcsot adunk a már meglévő gráfhoz, minden új csúcs a már meglévők közül $m \leq m_0$ -hoz csatlakozik éllel úgy, hogy annak valószínűsége, hogy az új csúcs az i . csúcshoz csatlakozik $p_i = \frac{d_i}{\sum_j d_j}$, ahol d_i az i . csúcs foksáma, a nevezőbeli összegzés pedig az összes már eddig is létező csúcs szerint értendő. Ezt a modellt nevezzük **Barabási-Albert-modellnek**.*

A preferential attachment-modellben egy rögzített csúcs foksámáról az alábbi tétel mondható ki:

2.3.1. Tétel ([14] 8.2. tétel). *Legyen $m = 1$ és $\delta \in (-1, \infty)$. Ekkor a $D_i(t)t^{-\frac{1}{2+\delta}}$ valószínűségi változó 1 valószínűséggel konvergál egy ξ_i valószínűségi változóhoz (amint $t \rightarrow \infty$), továbbá*

$$\mathbb{E}[D_i(t) + \delta] = (1 + \delta) \frac{\Gamma(t+1)\Gamma\left(i - \frac{1}{2+\delta}\right)}{\Gamma\left(t + \frac{1+\delta}{2+\delta}\right)\Gamma(i)} \quad (2.3)$$

Bizonyítás. Legyen $t \geq i$. Ekkor írhatjuk, hogy

$$\mathbb{E}[D_i(t+1) + \delta \mid D_i(t)] = D_i(t) + \delta + \mathbb{E}[D_i(t+1) - D_i(t) \mid D_i(t)] = \quad (2.4)$$

$$= D_i(t) + \delta + \mathbb{E} \left[\chi_{v_{t+1}^{(1)} \rightarrow v_i^{(1)}} \mid D_i(t) \right] = \quad (2.5)$$

$$= D_i(t) + \delta + \frac{D_i(t) + \delta}{t(2+\delta) + 1 + \delta} = \quad (2.6)$$

$$= (D_i(t) + \delta) \frac{(2+\delta)(1+t)}{t(2+\delta) + 1 + \delta} \quad (2.7)$$

Másrészt,

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

Ebből a kettőből már matematikai indukcióval következik (2.3).

(2.7)-et újra felhasználva kapjuk, hogy az az $(M_i(t))_{t \geq i}$ sorozat, amelyre

$$M_i(t) = \frac{D_i(t) + \delta}{1 + \delta} \prod_{j=i-1}^{t-1} \frac{j(2+\delta) + 1 + \delta}{(2+\delta)(1+j)} \quad \forall t \geq i \text{ egészre}$$

egy 1 várható értékű nemnegatív martingál, tehát a martingál konvergencia-tétel szerint $M_i(t)$ 1 valószínűséggel konvergál egy M_i valószínűségi változóhoz, amint $t \rightarrow \infty$. Mivel

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

ezért

$$M_i(t) = \frac{D_i(t) + \delta}{1 + \delta} \cdot \frac{\Gamma(t + \frac{1+\delta}{2+\delta}) \Gamma(i)}{\Gamma(t+1) \Gamma(i - \frac{1}{2+\delta})}$$

A Stirling-formulával könnyen megmutatható, hogy $\frac{\Gamma(t+a)}{\Gamma(t)} = t^a (1 + O(\frac{1}{t}))$, amint $t \rightarrow \infty$. Ezek szerint

$$(D_i(t) + \delta) t^{-\frac{1}{2+\delta}} = M_i(t) \frac{(2+\delta) \Gamma(i - \frac{1}{2+\delta})}{\Gamma(i)} \left(1 + O\left(\frac{1}{t}\right) \right) \rightarrow M_i \frac{(2+\delta) \Gamma(i - \frac{1}{2+\delta})}{\Gamma(i)} =: \xi_i,$$

1 valószínűséggel, amint $t \rightarrow \infty$. Mivel $\frac{1}{2+\delta} > 0$, ezért $\delta t^{-\frac{1}{2+\delta}} \rightarrow 0$, amint $t \rightarrow \infty$, 1 valószínűséggel $D_i(t) t^{-\frac{1}{2+\delta}} \rightarrow \xi_i$ (amint $t \rightarrow \infty$) is teljesül. \square

2.3.2. Megjegyzés. Az előzőekben felhasználtuk a $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ gammafüggvény $\Gamma(x) = (x-1)\Gamma(x-1) \forall x \in (1, \infty)$ tulajdonságát.

2.3.3. Megjegyzés. $PA_t^{(m,\delta)}$ és $PA_t^{(1, \frac{\delta}{m})}$ kapcsolata alapján (lásd fennebb) az $m \geq 1$ esetre az

$$\mathbb{E}_m^\delta[D_i(t)] = \sum_{j=1}^m \mathbb{E}_1^{\frac{\delta}{m}}[D_{m(i-1)+j}(mt)]$$

összefüggés írható föl, ahol m alsó illetve δ felső index a modell paramétereit jelölik.

Egy másik fő eredmény [14] a preferential attachement-gráfok $P_k^{(t)}$ fokszámeloszlására vonatkozik (lásd: 1.1.1). Ehhez vezessünk be előbb egy jelölést: legyen $m \in \mathbb{N}^+$, $\delta \in (-m, \infty)$ és definiáljuk a $(p_k)_{k \in \mathbb{N}}$ sorozatot úgy, hogy $p_k = 0$ bármely $k \in [m-1] \cup \{0\}$ esetén, illetve ha $k \geq m$ egész, akkor

$$p_k = \left(2 + \frac{\delta}{m}\right) \frac{\Gamma(k+\delta)\Gamma\left(m+2+\delta+\frac{\delta}{m}\right)}{\Gamma(m+\delta)\Gamma\left(k+3+\delta+\frac{\delta}{m}\right)}. \quad (2.8)$$

2.3.1. Állítás ([14]). *Az így definiált $(p_k)_{k \in \mathbb{N}}$ egy valószínűségeloszlás.*

Bizonyítás. A 2.3.2 megjegyzés miatt fennáll, hogy

$$\frac{\Gamma(k+a)}{\Gamma(k+b)} = \frac{1}{b-a-1} \left(\frac{\Gamma(k+a)}{\Gamma(k-1+b)} + \frac{\Gamma(k+1+a)}{\Gamma(k+b)} \right) \quad (2.9)$$

(2.9)-ba $a = \delta$ -t és $b = 3 + \delta + \frac{\delta}{m}$ -et helyettesítve kapjuk, hogy $k \geq m$ esetén

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

Eszerint a lenti összeg teleszkopikus lesz, mégpedig

$$\sum_{k=0}^{\infty} p_k = \sum_{k=m}^{\infty} p_k = \frac{\Gamma\left(m+2+\delta+\frac{\delta}{m}\right)}{\Gamma(m+\delta)} \cdot \frac{\Gamma(m+\delta)}{\Gamma\left(m+2+\delta+\frac{\delta}{m}\right)} = 1,$$

és mivel $p_k \geq 0$ minden $k \in \mathbb{N}$ -re teljesül, ezért $(p_k)_{k \in \mathbb{N}}$ egy valószínűség-eloszlás. \square

Hogy talán kevésbé légből kapottnak tűnjön p_k definíciója, gondoljuk meg az alábbiakat. Legyen $X \sim \text{NB}(r, p)$ egy negatív binomiális eloszlású valószínűségi változó,

$r \in (0, \infty)$ és $p \in [0, 1]$ paraméterekkel. Ekkor, ha $r \in \mathbb{Z}$, akkor valamely $k \in \mathbb{N}$ -re

$$\mathbb{P}(X = k) = \binom{k}{r} p^r (1-p)^{k-r} \quad (2.10)$$

annak valószínűsége, hogy egy olyan kísérletet, ami p valószínűséggel sikeres, egymás után többször elvégezve, a k . kísérlet lesz az r . sikeres. A 2.3.2 megjegyzés alapján (2.10) így is írható:

$$\mathbb{P}(X = k) = \frac{\Gamma(r+k)}{k! \Gamma(r)} p^r (1-p)^r$$

Ha $r = m + \delta$, és a p paramétert úgy véletlenítjük, hogy $p = Z^{\frac{1}{2+\frac{\delta}{m}}}$, ahol $Z \sim \mathcal{U}_{[0,1]}$, akkor azt kaphatjuk⁵, hogy

$$p_k = \mathbb{E}[\mathbb{P}(X = k)],$$

ahol a várható érték Z szerint értendő.

Az alábbi tétel mutatja, hogy adott $k \in \mathbb{N}$ mellett a $\text{PA}_t^{(m,\delta)}$ preferential attachment-gráfok $P_k^{(t)}$ fokszámeloszlása sztochasztikusan tart p_k -hoz:

2.3.2. Tétel ([14] 8.3. tétel). *Legyen $m \in \mathbb{N}^+$ és $\delta \in (-m, \infty)$. Ekkor létezik egy m -től és δ -tól függő $C > 0$, melyre $t \rightarrow \infty$ mellett*

$$\mathbb{P} \left(\max_{k \in \mathbb{N}} \left| P_k^{(t)} - p_k \right| \geq C \sqrt{\frac{\ln t}{t}} \right) = o(1)$$

Emiatt $(p_k)_{k \in \mathbb{N}}$ -t a $\text{PA}_t^{(m,\delta)}$ gráf *aszimptotikus fokszámeloszlásának* nevezzük.

Ha jobban megvizsgáljuk p_k aszimptotikáját, akkor (2.8) illetve $\frac{\Gamma(t+a)}{\Gamma(t)} = t^a (1 + O(\frac{1}{t}))$ (amint $t \rightarrow \infty$) miatt $k \rightarrow \infty$ mellett

$$p_k = c_{m,\delta} k^{-\tau} \left(1 + O\left(\frac{1}{k}\right) \right) \quad (2.11)$$

adódik, ahol $\tau = 3 + \frac{\delta}{m} > 2$, és $c_{m,\delta} = \left(2 + \frac{\delta}{m}\right) \frac{\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)}$. Ekkor 2.3.2 és (2.11) miatt $\text{PA}_t^{(m,\delta)}$ fokszámeloszlása közel van egy $\tau = 3 + \frac{\delta}{m}$ kitevőjű *hatványtörvény* típusú valószínűség-eloszláshoz.

A preferential attachment-gráfok *konvergenciája* egy elég bonyolult kérdés, és több cikk is foglalkozik a témával. [22]-ban a szerzők a modell *sűrű* limeszét határozzák

⁵[14] 8.16. feladat

meg olyan eszközökkel, mint a gráflimeszek és a részben felcserélhető véletlen tömbök elmélete közti kapcsolat, illetve a véletlen gráf egy olyan explicit konstrukciója, aminek az urnamodellekhez is köze van. A [3] publikáció a *ritka* preferential attachment gráfok határértékével foglalkozik, ami tartalmazza az összes lokális információt a modellről, és számos olyan számítást tesz lehetővé, amelyek egyébként nehezek.

2.4. Sztochasztikus blokkmodell

Számos hálózatban vannak *közösségek*. Egy adott hálózatnak több közössége is lehet úgy, hogy a közösségen belüli csomópontok sűrűn kapcsolódnak egymáshoz. A több közösségben lévő csomópontok átfedhetik egymást. Gondolhatunk a közösségi oldalakon fenntartott fiókjainkra, és arra, hogy kivel lépünk kapcsolatba naponta. Lehet, hogy erősen interakcióban vagyunk a barátainkkal, kollégáinkkal, családtagjainkkal és néhány más fontos emberrel az életünkben. Ők egy nagyon sűrű közösséget alkotnak a közösségi hálózatunkon belül.

A különböző hálózatok elemzésekor fontos lehet a bennük lévő közösségek felfedezése. A közösségfelismerő technikák hasznosak a közösségi média algoritmusai számára, hogy felfedezzék a közös érdeklődési körrel rendelkező embereket, és szoros kapcsolatot tartsanak fenn közöttük. A közösségfelismerés a gépi tanulásban is használható a hasonló tulajdonságokkal rendelkező csoportok felismerésére. Ez a technika például arra használható, hogy felfedezzenek manipulatív csoportokat egy közösségi hálózaton vagy egy tőzsdén belül.

A *sztochasztikus blokkmodell* (SBM) széles körben használják a közösségek felismerésének kanonikus modelljeként.⁶ Vitathatatlanul ez a közösségeket tartalmazó gráf legegyszerűbb modellje (lásd az alábbi definíciókat). Mint minden modell, ez sem feltétlenül reális, de tanulságos - például a vizsgálatából levezethető hatékony algoritmusok alapján ítélve. Az alapmodell a legfontosabb jelenségek némelyikét ragadja meg, és fejlettebb és realiztikusabb finomításokkal (pl. élcímkékkel vagy átfedő közösségekkel)

⁶[1] 1.1. alfejezet

bővíthető. Itt az alap SBM megértésére fogunk összpontosítani.

Az általános sztochasztikus blokkmodell definíciója az alábbi:

2.4.1. Definíció. Legyen $n \in \mathbb{N}^+$ a csomópontok száma, $k \in \mathbb{N}^+$ a közösségek száma, $\mathbf{p} = (p_1, \dots, p_k) \in [0, 1]^k$ egy valószínűségvektor $[k]$ -n (a k közösség prioritása) és $W = (w_{ij})_{i,j \in [k]} \in \mathcal{M}_k([0, 1])$ egy szimmetrikus mátrix (a kapcsolódási valószínűségek). Az (\mathbf{X}, G) pár SBM(n, \mathbf{p}, W) modell szerint van sorsolva, ha $\mathbf{X} = (X_1, \dots, X_n)$ egy n dimenziós független azonos \mathbf{p} eloszlású valószínűségi változókból álló valószínűségi vektorváltozó, és G egy n csomópontú egyszerű gráf, melyre $V(G) = [n]$ és minden $u, v \in [n]$ -re az u és v csomópontok $w_{X_u X_v}$ valószínűséggel fut él. Továbbá a közösség-halmazok alatt az $\Omega_i = \Omega_i(\mathbf{X}) = \{v \in [n] \mid X_v = i\}$ ($i \in [k]$) halmazokat értjük. A G gráfot sztochasztikus blokkgráfnak is nevezzük.

Tehát az (\mathbf{X}, G) pár eloszlásáról $\mathbf{x} = (x_1, \dots, x_n) \in [k]^n$ illetve $\mathbf{y} = (y_e \mid e \in 2^{[n]}) \in \{0; 1\}^{\binom{n}{2}}$ mellett a következők mondhatók:

$$\begin{aligned} \mathbb{P}(\mathbf{X} = \mathbf{x}) &= \prod_{v=1}^n p_{x_v} = \prod_{i=1}^k p_i^{|\Omega_i(\mathbf{x})|} \\ \mathbb{P}(\mathbf{E}(G) = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) &= \prod_{1 \leq u < v \leq n} w_{x_u x_v}^{y_{\{u,v\}}} = \\ &= \prod_{1 \leq i < j \leq k} w_{ij}^{N_{ij}(\mathbf{x}, \mathbf{y})} \left(1 - w_{ij}^{N_{ij}^c(\mathbf{x}, \mathbf{y})}\right) \end{aligned}$$

ahol $\mathbf{E}(G)$ egy $\binom{n}{2}$ dimenziós 0-1 vektor, amelynek bármely különböző $u, v \in [n]$ -re az $\{u, v\}$ -edik eleme pontosan akkor 1, ha u és v csomópont között fut él G -ben, illetve

$$\begin{aligned} N_{ij}(\mathbf{x}, \mathbf{y}) &= \sum_{\substack{u < v \\ x_u = i \\ x_v = j}} \chi_{\{y_{\{u,v\}} = 1\}} \\ N_{ij}^c(\mathbf{x}, \mathbf{y}) &= \begin{cases} \sum_{\substack{u < v \\ x_u = i \\ x_v = j}} \chi_{\{y_{\{u,v\}} = 0\}} = |\Omega_i(\mathbf{x})| |\Omega_j(\mathbf{x})| - N_{ij}(\mathbf{x}, \mathbf{y}) & \text{ha } i \neq j \\ \sum_{\substack{u < v \\ x_u = x_v = i}} \chi_{\{y_{\{u,v\}} = 0\}} = |\Omega_i(\mathbf{x})| \frac{|\Omega_i(\mathbf{x})| - 1}{2} - N_{ii}(\mathbf{x}, \mathbf{y}) & \text{ha } i = j \end{cases} \end{aligned}$$

$i, j \in [k]$ -ra a közösségek között futó élek és nemélek száma.

Tehát az történik, hogy adott n -re minden csúcsra a prioritási eloszlásból kisorsoljuk, hogy melyik közösséghez (blokkhoz) tartozzon, aztán a kapcsolódási valószínűségek W mátrixa alapján kisorsoljuk a gráfunk éleit. Nyilván tipikusan olyan modelleket fogunk nézni, ahol a közösségen belüli élek valószínűsége magasabb lesz, a különböző közösségek között futó élek valószínűsége pedig kisebb, de ennek nem feltétlenül kell így lennie.

Megjegyezzük, hogy amint az n értékét növeljük, általában úgy tekintjük, hogy \mathbf{p} nem skálázódik, míg W igen. Tehát a közösségek száma nem változik n -nel, és a méretük lineáris. Azt is vegyük észre, hogy $n \rightarrow \infty$ mellett a nagy számok törvénye alapján minden $i \in [k]$ -ra 1 valószínűséggel

$$\frac{1}{n}|\Omega_i| \longrightarrow p_i. \quad (2.12)$$

Észrevehetjük, hogy a fenti példában a mátrix *szimmetrikus*. Szokás az SBM szimmetrikus változatát is definiálni, noha a mátrix szimmetrikusságánál kicsit többet követelünk meg:

2.4.2. Definíció. Legyen $n, k \in \mathbb{N}^+$ és $A, B \in [0, 1]$. Az (\mathbf{X}, G) pár az $\text{SSBM}(n, k, A, B)$ modell szerint van sorsolva, ha az $\text{SBM}(n, \mathbf{p}, W)$ szerint van sorsolva, feltéve azt, hogy $\mathbf{p} = \underbrace{\left(\frac{1}{k}, \dots, \frac{1}{k}\right)}_{k\text{-szor}}$ és $W \in \mathcal{M}_k([0, 1])$ olyan mátrix, melynek főátlójában csupa A , míg főátlóján kívül csupa B helyezkedik el.

Vegyük észre, hogy az SBM-modell a grafonból sorsolt véletlen gráfok egy sajátos esete. Ugyanis ha W olyan lépcsős grafon, aminek a lépcsőméretei a \mathbf{p} prioritásnak megfelelően aránylanak egymáshoz, a lépcsőin felvett értékek pedig a $Q = (q_{ij})_{i,j \in [k]} \in \mathcal{M}_k([0, 1])$ mátrix elemeinek felelnek meg, akkor a $\mathbb{G}(n, W)$ modell pontosan az $\text{SBM}(n, \mathbf{p}, Q)$ modellre redukálódik. Pontosabban, ha $W: [0, 1]^2 \rightarrow [0, 1]$ olyan,

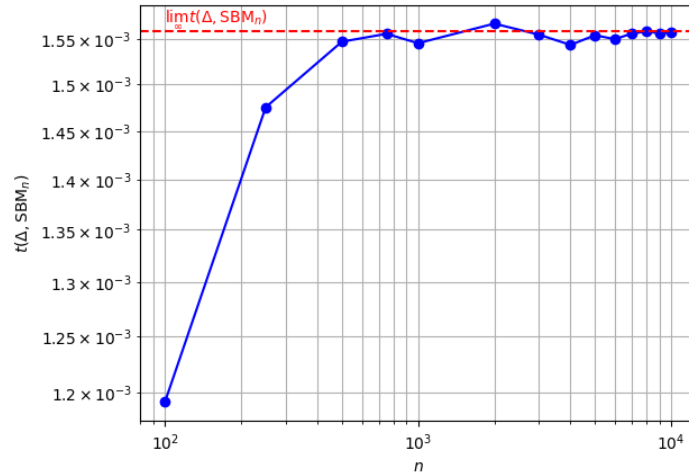
hogy

$$W(x, y) = \begin{cases} q_{ij} & \text{ha } \sum_{r=0}^{i-1} p_r \leq x < \sum_{r=0}^i p_r \text{ és } \sum_{r=0}^{j-1} p_r \leq y < \sum_{r=0}^j p_r \quad (i, j \in [k-1]) \\ q_{ik} & \text{ha } \sum_{r=0}^{i-1} p_r \leq x < \sum_{r=0}^i p_r \text{ és } \sum_{r=0}^{k-1} p_r \leq y \leq 1 \quad (i \in [k-1]) \\ q_{kj} & \text{ha } \sum_{r=0}^{k-1} p_r \leq x \leq 1 \text{ és } \sum_{r=0}^{j-1} p_r \leq y < \sum_{r=0}^j p_r \quad (j \in [k-1]) \\ q_{kk} & \text{ha } \sum_{r=0}^{k-1} p_r \leq x \leq 1 \text{ és } \sum_{r=0}^{k-1} p_r \leq y \leq 1 \end{cases}$$

ahol $p_0 := 0$, akkor $\mathbb{G}(n, W) \stackrel{d}{=} \text{SBM}(n, \mathbf{p}, Q)$. Az így konstruált grafont $W_{\mathbf{p}, Q}$ -val jelöljük.

2.4.1. Következmény. 1 valószínűséggel $\text{SBM}(n, \mathbf{p}, Q) \rightarrow W_{\mathbf{p}, Q}$, amint $n \rightarrow \infty$.

A 2.2 ábrán $\text{SBM}(n, \mathbf{p}, W)$ blokkmodellből sorsolt gráfok háromszögsűrűségének alakulása van ábrázolva, amint n értéke növekszik. Látható, hogy $t(\Delta, \text{SBM}_n)$ konvergál, amint $n \rightarrow \infty$.

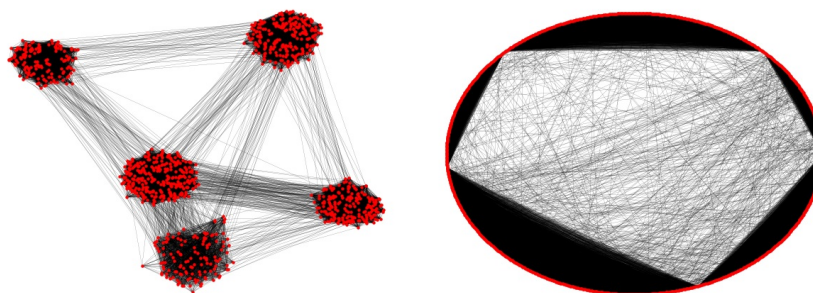


2.2. ábra. A háromszögsűrűség alakulása $n = 10^2, \frac{5}{2} \cdot 10^2, 5 \cdot 10^2, \frac{15}{2} \cdot 10^2, 10^3, 2 \cdot 10^3, 3 \cdot 10^3, 4 \cdot 10^3, 5 \cdot 10^3, 6 \cdot 10^3, 7 \cdot 10^3, 8 \cdot 10^3, 9 \cdot 10^3, 10^4$ esetén (3.3)

Itt

$$\mathbf{p} = \left(\frac{13}{100}, \frac{241}{1000}, \frac{29}{200}, \frac{29}{100}, \frac{97}{500} \right) \quad \text{valamint} \quad W = \begin{pmatrix} \frac{29}{100} & \frac{1}{1000} & \frac{1}{1000} & \frac{1}{200} & \frac{1}{500} \\ \frac{1}{1000} & \frac{3}{10} & \frac{1}{1000} & \frac{1}{1000} & \frac{1}{1000} \\ \frac{1}{1000} & \frac{1}{1000} & \frac{9}{20} & \frac{1}{500} & \frac{1}{1000} \\ \frac{1}{200} & \frac{1}{1000} & \frac{1}{500} & \frac{1}{4} & \frac{3}{1000} \\ \frac{1}{500} & \frac{1}{1000} & \frac{1}{1000} & \frac{3}{1000} & \frac{2}{5} \end{pmatrix}$$

A háromszögsűrűség *várható értékének* határértékét $n \rightarrow \infty$ mellett (piros egyenes) közvetlenül is kiszámoljuk majd a 3.2.2 szakaszban.



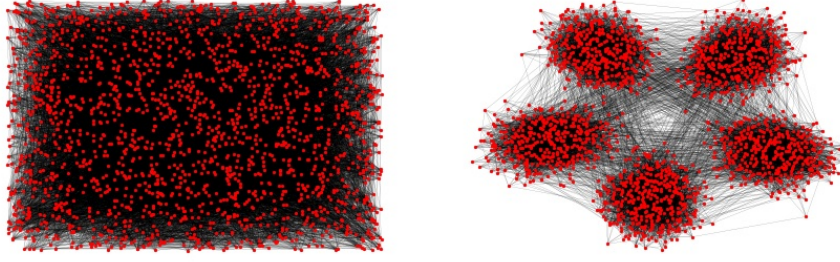
2.3. ábra. Egy, a fenti modellből sorsolt gráf blokkonként illetve körkörösén ábrázolva $n = 1000$ csomópont esetén. (3.4)

2.4.1. Visszanyerési követelmények

Bár ennek a dolgozatnak nem célja a közösségfelismerési algoritmusok bővebb tárgyalása, azért röviden mégis bemutatunk⁷ néhány alapvető fogalmat az alapfeladról. A közösségfelismerés *célja* a csomópontok partícióját, vagyis \mathbf{X} ún. közösségvektort visszanyerni a már kisorsolt G gráf ismeretében.

A 2.4. ábrán látható gráf SSBM $(1000, 5, \frac{1}{25}, \frac{1}{1000})$ modell szerint lett kisorsolva. Láthatjuk, hogy ha nagyon szerencsétlenül helyezük el a gráf csomópontjait a síkban, akkor korántsem könnyű a közösségek beazonosítása, van tehát létjogosultsága a hamarosan definiálásra kerülő közösség-visszanyerési feladat tárgyalásának.

⁷[1] 2.3. szakasz alapján



2.4. ábra. A fenti két ábrán ugyanaz a gráf látható átrendezve. (3.5)

Alapvető kérdés, hogy hogyan mérjük a különböző közösségvektorok megegyezőségét.

2.4.3. Definíció. Két $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n) \in [k]^n$ közösségvektor ún. *agreementjén* az

$$A(\mathbf{x}, \mathbf{y}) = \max_{\sigma \in S_k} \frac{1}{n} \sum_{i=1}^n \chi_{\{x_i = \sigma(y_i)\}}$$

mennyiséget értjük.

$(\mathbf{X}, G) \sim \text{SBM}(n, \mathbf{p}, W)$ esetén mindig megpróbálhatjuk visszanyerni az \mathbf{X} -et a G megfigyelése nélkül is, úgy, hogy egyszerűen a \mathbf{p} -ből mint eloszlásból sorsoljuk ki egymástól függetlenül $\hat{\mathbf{X}}$ elemeit. Ekkor 1 valószínűséggel $A(\mathbf{X}, \hat{\mathbf{X}}) \rightarrow \|\mathbf{p}\|_2^2$, amint $n \rightarrow \infty$.

2.4.4. Definíció. Legyen $(\mathbf{X}, G) \sim \text{SBM}(n, \mathbf{p}, W)$. Az alábbi *visszanyerési követelmények* teljesítetnek, ha létezik olyan algoritmus, amely inputként a G gráfot kapja, outputja pedig olyan $\hat{\mathbf{X}} = \hat{\mathbf{X}}(G)$ közösségvektor, melyre $n \rightarrow \infty$ mellett

- **Pontos visszanyerés:** $\mathbb{P}(A(\mathbf{X}, \hat{\mathbf{X}}) = 1) = 1 - o(1)$
- **Majdnem pontos visszanyerés:** $\mathbb{P}(A(\mathbf{X}, \hat{\mathbf{X}}) = 1 - o(1)) = 1 - o(1)$
- **Részleges visszanyerés:** $\mathbb{P}(A(\mathbf{X}, \hat{\mathbf{X}}) \geq \alpha) = 1 - o(1)$, ahol $\alpha \in (0, 1)$

k közösségű SSBM modell esetén egy olyan algoritmus, amely egyszerűen egymástól függetlenül, egyenletes eloszlás szerint sorsolja $\hat{\mathbf{X}}$ elemeit, $\frac{1}{k}$ pontosságot ér el. A probléma tehát akkor érdekes, amikor $\alpha > \frac{1}{k}$:

2.4.5. Definíció (gyenge visszanyerés). SSBM(n, k, A, B) esetén a gyenge visszanyerési feladat megoldható, ha létezik olyan $\varepsilon > 0$ és olyan algoritmus, amelynek inputja a G gráf, outputja pedig pedig olyan $\hat{\mathbf{X}} = \hat{\mathbf{X}}(G)$ közösségvektor, melyre $n \rightarrow \infty$ mellett

$$\mathbb{P}\left(A(\mathbf{X}, \hat{\mathbf{X}}) \geq \frac{1}{k} + \varepsilon\right) = 1 - o(1)$$

A Kesten-Stigum-küszöb

Az alábbi sejtés a [8] publikációban fogalmazódott meg először, ahol a szerzők a statisztikus fizika heurisztikus eszközeivel motiválták elképzelésüket.

2.4.1. Sejtés. Legyen $a, b > 0, n, k \in \mathbb{N}^+$, illetve $\text{SNR} = \frac{(a-b)^2}{k(a+(k-1)b)}$. Tekintsük az SSBM $(n, k, \frac{a}{n}, \frac{a}{n})$ gráfmodellt. Ekkor

- (a) Ha $k \geq 2$, és $\text{SNR} > 1$ (a Kesten-Stigum(KS)-küszöb), polinomiális időben visszanyerhetők a közösségek.
- (b) Ha $k \geq 4$, akkor $\text{SNR} < 1$ esetén is visszanyerhetők információelméletileg a közösségek (azaz nem feltétlenül polinomiális időben).

Megjegyezzük, hogy a KS-küszöb kifejezés a fákra vonatkozó rekonstrukciós problémából származik [10]. A bináris esetben egy adó egy uniform választott bitet továbbít néhány közvetítő felé, amelyek maguk továbbítják a kapott biteket más közvetítőknek, stb. A cél a gyökérbitelek rekonstrukciója a levélbitekből, ahogy a fa mélysége egyre mélyül. Különösen két közösség esetében redukció végezhető a rekonstrukciós probléma kudarca között fa esetén és a visszanyerésben való sikertelenség között az SBM-ben.⁸ A KS küszöbérték elérése érdekes kihívást jelent a közösségfelismerő algoritmusok szempontjából, mivel a standard klaszterezési módszerek nem érik el a küszöbértéket. Ez

⁸[1] 4.1.

azért probléma, mert a következő tétel szerint ekkor a $k = 2$ esetben a visszanyerés lehetetlen:

2.4.1. Tétel ([1] 8. tétel). *Ha $k = 2$, akkor $\text{SNR} \leq 1$ esetén a gyenge visszanyerési feladat nem megoldható.*

Az előző eredményt is a fákra vonatkozó rekonstrukciós problémára való visszavezetéssel bizonyították [20]-ben.

Abbe és Sandon 2016-ban bizonyították a 2.4.1 sejtést az alábbi tétellel:

2.4.2. Tétel. (a) *Minden $k \geq 2$ -re, ha $\text{SNR} > 1$, akkor a gyenge visszanyerési feladat $O(n \ln n)$ időben megoldható.*

(b) *Minden $k \geq 4$ -re, $\text{SNR} < 1$ esetén a gyenge visszanyerési feladat információelméletileg megoldható.*

A fenti tétel mindkét állítására konkrét algoritmus olvasható [1]-ben.

2.4.2. Néhány topologikus tulajdonság

Fontos megemlíteni az SBM gráfok néhány topologikus tulajdonságát. Ha a W gráf csupa $w \in [0, 1]$ értékből áll, akkor az $\text{SBM}(n, \mathbf{p}, W)$ modell gyakorlatilag nem lesz más, mint egy $\text{ER}_n(p)$ Erdős-Rényi-modell, aminek alapvető tulajdonságait a 2.1.2. tétel írja le.

Az $\text{SSBM}(n, k, A, B)$ modell esetén ezek a jellemzők az alábbi formában állnak fenn ([1] 2.5.):

- Ha $a, b > 0$, az $\text{SSBM}(n, k, \frac{a \ln n}{n}, \frac{b \ln n}{n})$ modellből sorsolt gráf akkor és csakis akkor lesz $1 - o(1)$ valószínűséggel összefüggő, ha $\frac{a+(k-1)b}{k} > 1$
- Az $\text{SSBM}(n, k, \frac{a}{n}, \frac{b}{n})$ modellből sorsolt gráfnak akkor és csakis akkor létezik óriási (n -ben lineáris méretű) komponense, ha $\frac{a+(k-1)b}{k} > 1$
- Ha $\delta < \frac{1}{2}$, akkor egy $\text{SSBM}(n, k, \frac{a}{n}, \frac{b}{n})$ modellből sorsolt gráf egy v csomópontjának $r = \delta \log_{\frac{a+(k-1)b}{k}} n$ mélységű környezete totális variációban tart a $\frac{a+(k-1)b}{k}$ paraméterű Poisson-utódeloszlású Galton-Watson folyamathoz.

Hasonló eredmény igaz az általános SBM-re, bizonyos megkötésekkel: egy SBM $(n, \mathbf{p}, Q^{\frac{\ln n}{n}})$ modellből sorsolt gráf $1 - o(1)$ valószínűséggel összefüggő, ha $\min_{i \in [k]} \|(\text{diag}(p)Q)_i\|_1 > 1$, illetve $1 - o(1)$ valószínűséggel nem összefüggő, ha $\min_{i \in [k]} \|(\text{diag}(p)Q)_i\|_1 < 1$, ahol $(M)_i$ egy M mátrix i . oszlopvektorát jelöli.

3. fejezet

Sztochasztikus blokkmodell bővítése a preferential attachment-szabály szerint

3.1. A modell leírása

Az alapprobléma a következőképpen is megfogalmazható: adott egy n tagú közösség ($n \in \mathbb{N}^+ \setminus \{1\}$), amin belül több alközösség alakult ki, azaz a közösség olyan tagjainak csoportjai, akik egymást jobban ismerik, mint a közösség többi tagját. Gondolhatunk egy osztályközösségre, ahol "klikkek" alakultak ki. A közösségbe új tagok érkeznek, n darab egyesével. Minden új tag $\lfloor \frac{n}{10} \rfloor$ ismeretséget köt. Ez azért fontos, hogy így a régi és új ismeretségek száma azonos nagyságrendű legyen (az $\frac{1}{10}$ -es szorzónak nincs különösebb jelentősége, mást is lehetett volna választani). Feltesszük továbbá, hogy az újonnan érkezők hajlamosabbak előbb az eredeti közösség népszerűbb tagjaival megismerkedni, azaz azokkal, akiknek már addig is több ismerősük volt. Mit mondhatunk az így kialakult társaság ismeretségi struktúráiról nagyon nagy n esetén?

A feladatot véletlen gráfmodellekkel is leírhatjuk: az eredeti közösség ismeretségi gráfja legyen egy $\text{SBM}(n, \mathbf{p}, Q)$ *sztochasztikus blokkmodell* szerint sorsolt G_0 gráf, azaz egy *sztochasztikus blokkgráf*, amelynek csúcsai legyenek v_1, \dots, v_n . Az egyszerűség ked-

véért legyen a közösségek száma egy rögzített $k \in \mathbb{N}^+$ szám, és minden $i \in [k]$ -ra a G_0 gráf csomópontjainak p_i -ed része alkossa az i . közösséget – így \mathbf{X} rögzített, nem függ a véletlentől. Minden $i \in [n]$ esetén a G_i gráfot úgy kapjuk, hogy a G_{i-1} gráfhoz hozzáveszünk egy v_{n+i} csúcsot, és azt összekötjük a $V_{i-1} := V(G_{i-1})$ halmaz $\lfloor \frac{n}{10} \rfloor$ elemével, amelyeket véletlenszerűen, *preferential attachment-szabállyal* választunk, azaz úgy, hogy annak valószínűsége, hogy valamely $j \in [n+i-1]$ -re $\{v_{n+i}, v_j\} \in E_i := E(G_i)$, nem más, mint

$$\mathbb{P}(v_{n+i} \rightarrow v_j \mid G_{i-1}) = \frac{\deg_{i-1}(v_j)}{\sum_{r=1}^{n+i-1} \deg_{i-1}(v_r)} \quad (3.1)$$

ahol $\deg_i(v)$ a $v \in V_i$ csúcs G_i -beli fokszámát jelöli ($i \in [n] \cup \{0\}$). Ez tulajdonképpen a preferential attachment-szabály $\delta = 0$ esete hurokélek nélkül, azaz a *Barabási-Albert-szabály*.

A modell nagy n érték melletti struktúrájáról több kérdés is feltehető, mint például, hogy mennyi az élsűrűség, mennyi a háromszögek számának várható értéke, vagy hogy konvergense-e a modell $n \rightarrow \infty$ mellett. Ezekhez kötődő kérdéseket kezdünk el vizsgálni a következőkben.

3.2. A struktúra vizsgálata nagy n érték esetén

Az alábbiakban a 2.4.1 definíció jelöléseit fogjuk használni.

3.2.1. Élsűrűség

Tekintsük először csak G_0 -t. Legyen $A_0 = \left(a_{ij}^{(0)} \right)_{i,j \in [n]}$ a G_0 gráf szomszédsági mátrixa. Ekkor $e_0 := e(G_0)$ nem lesz más, mint az A_0 mátrix elemei összegének a fele. Mivel G_0 egy sztochasztikus blokkgráf, ezért írhatjuk, hogy ha k a közösségek száma, akkor

$$2 \mathbb{E}[e_0] = \sum_{i,j=1}^n \mathbb{E} \left[a_{ij}^{(0)} \right] = \sum_{i,j=1}^n \mathbb{E} \left[\chi_{\{v_i, v_j\} \in E_0} \right] = \sum_{i,j=1}^n q_{X_i X_j} = \sum_{i,j=1}^k |\Omega_i| |\Omega_j| q_{ij}, \quad (3.2)$$

így $|\Omega_i| = np_i$ (bármely $i \in [k]$ -ra) miatt fennáll, hogy

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[2e_0]}{n^2} = \sum_{i,j=1}^k p_i p_j q_{ij} = \iint_{[0,1]^2} W_{\mathbf{p},Q}(x,y) dx dy = t(-, W_{\mathbf{p},Q}) \quad (3.3)$$

(Itt $W_{\mathbf{p},Q}$ az $\text{SBM}(n, \mathbf{p}, Q)$ -hoz rendelt grafon, lásd: 2.4.1.)

3.2.1. Megjegyzés. (2.12) miatt (3.3) véletlenszerű \mathbf{X} esetén is (1 valószínűséggel) teljesül.

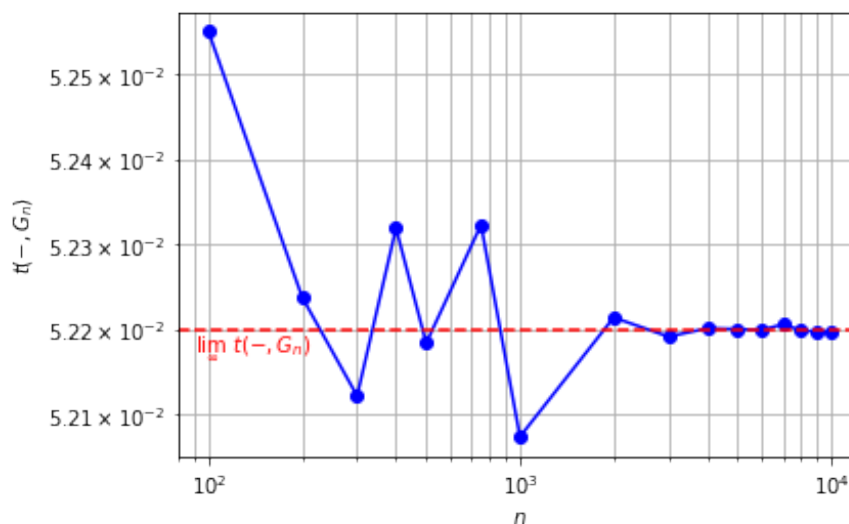
Mivel a modell szerint minden lépésben $\lfloor \frac{n}{10} \rfloor$ új él keletkezik, ezért

$$e_n = e_0 + n \left\lfloor \frac{n}{10} \right\rfloor,$$

így (3.3) alapján

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[2e_n]}{(2n)^2} = \frac{1}{4} \sum_{i,j=1}^k p_i p_j q_{ij} + \frac{1}{20}. \quad (3.4)$$

Maga az élsűrűség is jó eséllyel konvergál, amint $n \rightarrow \infty$, amint azt a 3.1 szimuláció ábrázolja.



3.1. ábra. A G_n gráf élsűrűségének alakulása, amint n értéke növekszik (3.6)

A szimuláció során a kiindulási gráf egy $\text{SSBM}(n, 5, \frac{1}{25}, \frac{1}{1000})$ szimmetrikus sztochasztikus blokkgráf volt, amint n rendre a 100, 200, 300, 400, 500, 750, 1000, 2000, 3000,

4000, 5000, 6000, 7000, 8000, 9000, 10000 értékeket vette fel. A grafikon alapján valószínűsíthető, hogy az élsűrűség konvergál az $5,22 \cdot 10^{-2}$ értékhez, ami pontosan az élsűrűségek *várható értékének* (3.4) határértéke ennek a modellnek az esetén.

3.2.2. Háromszögek

Jelölje G_i -ben a háromszögek számát t_i ($i \in [n] \cup \{0\}$). Kezdjük megint G_0 -al. Aszerint, hogy a háromszögek csúcsai melyik közösségbe tartoznak, írhatjuk, hogy

$$\begin{aligned} \mathbb{E}[t_0] &= \mathbb{E} \left[\frac{1}{6} \sum_{i,j,l=1}^n \chi_{\{\{v_i,v_j\} \in E_0\} \cap \{\{v_j,v_l\} \in E_0\} \cap \{\{v_l,v_i\} \in E_0\}} \right] = \\ &= \sum_{i=1}^k \binom{|\Omega_i|}{3} q_{ii}^3 + \sum_{\substack{i,j=1 \\ i \neq j}}^k \binom{|\Omega_i|}{2} |\Omega_j| q_{ii} q_{ij}^2 + \sum_{\substack{i,j,l=1 \\ i \neq j \neq l \neq i}}^k |\Omega_i| |\Omega_j| |\Omega_l| q_{ij} q_{jl} q_{li}, \end{aligned}$$

így $|\Omega_i| = np_i$ (bármely $i \in [k]$ -ra) miatt fennáll, hogy

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[6t_0 \mid \mathbf{X}]}{n^3} = \sum_{i=1}^k p_i^3 q_{ii}^3 + 3 \sum_{\substack{i,j=1 \\ i \neq j}}^k p_i^2 p_j q_{ii} q_{ij}^2 + 6 \sum_{\substack{i,j,l=1 \\ i \neq j \neq l \neq i}}^k p_i p_j p_l q_{ij} q_{jl} q_{li} = \quad (3.5)$$

$$= \iiint_{[0,1]^3} W_{\mathbf{p},\mathbf{Q}}(x,y) W_{\mathbf{p},\mathbf{Q}}(y,z) W_{\mathbf{p},\mathbf{Q}}(z,x) dx dy dz = t(\Delta, W_{\mathbf{p},\mathbf{Q}}) \quad (3.6)$$

3.2.2. Megjegyzés. (2.12) miatt (3.5) véletlenszerű \mathbf{X} esetén is (1 valószínűséggel) teljesül.

Vizsgáljuk most t_1 -et. v_{n+1} hozzáadása után csak úgy keletkezhet új háromszög a gráfban, ha annak egyik csúcsa v_{n+1} , másik két csúcsa pedig V_0 -beli. Tehát

$$t_1 = t_0 + \frac{1}{2} \sum_{i,j=1}^n \chi_{\{v_{n+1} \rightarrow v_i\} \cap \{v_{n+1} \rightarrow v_j\} \cap \{\{v_i,v_j\} \in E_0\}}$$

Ha adva van G_0 , akkor az új háromszögek $t_1 - t_0$ száma az $(n+1)$. csúcs hozzáadása után:

$$\sum_{\substack{i,j=1 \\ i \neq j}}^n q_{X_i X_j} \cdot \frac{\deg_0(v_i)}{\sum_{s=1}^n \deg_0(v_s)} \cdot \frac{\deg_0(v_j)}{\sum_{s=1}^n \deg_0(v_s)} \cdot \left\lfloor \frac{n}{10} \right\rfloor$$

Ezt úgy közelítjük, hogy minden tényező helyett azok várható értékével számolunk (ez nem az új háromszögek számának várható értéke lesz, csak egy közelítés). G_0 -ban a foksámok várható értéke: bármely $i \in [n]$ -re

$$\mathbb{E}[\deg_0(v_i)] = \mathbb{E}\left[\sum_{j=1}^n a_{ij}^{(0)}\right] = \sum_{j=1}^k |\Omega_j| q_{X_{ij}}$$

Ezért a fent leírt közelítő helyettesítéssel:

$$\sum_{\substack{i,j=1 \\ i \neq j}}^n q_{X_i X_j} \deg_0(v_i) \deg_0(v_j) \approx \sum_{\substack{i,j=1 \\ i \neq j}}^n q_{X_i X_j} \left(\sum_{l=1}^k |\Omega_l| q_{X_{il}}\right) \left(\sum_{s=1}^k |\Omega_s| q_{X_{js}}\right)$$

Legyen minden $i \in [k]$ -ra $s_i = \sum_{j=1}^k |\Omega_j| q_{ij}$. Ekkor

$$\begin{aligned} \sum_{\substack{i,j=1 \\ i \neq j}}^n q_{X_i X_j} \deg_0(v_i) \deg_0(v_j) &\approx \sum_{\substack{i,j=1 \\ i \neq j}}^n q_{X_i X_j} s_{X_i} s_{X_j} = \\ &= \sum_{i,j=1}^n q_{X_i X_j} s_{X_i} s_{X_j} - \sum_{i=1}^n q_{X_i X_i} s_{X_i}^2 = \\ &= \sum_{i,j=1}^k |\Omega_i| |\Omega_j| q_{ij} s_i s_j - \sum_{i=1}^k |\Omega_i| q_{ii} s_i^2 \end{aligned}$$

Így, figyelembe véve (3.2)-t, az új háromszögek száma ezzel az approximációval közelítőleg:

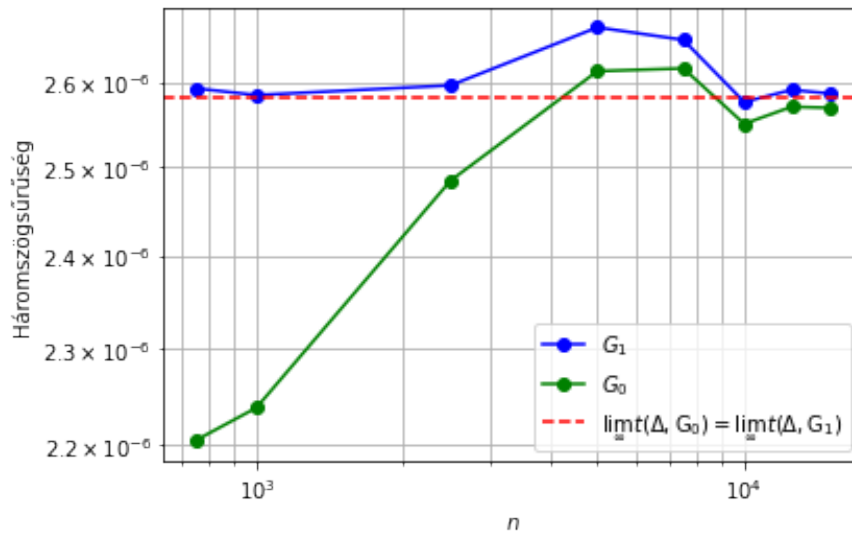
$$\begin{aligned} \frac{n}{10} \cdot \frac{\sum_{i,j=1}^k |\Omega_i| |\Omega_j| q_{ij} s_i s_j - \sum_{i=1}^k |\Omega_i| q_{ii} s_i^2}{\left(\sum_{i,j=1}^k |\Omega_i| |\Omega_j| q_{ij}\right)^2} &= \frac{1}{10} \cdot \frac{n \sum_{i,j=1}^k p_i p_j q_{ij} s_i s_j - \sum_{i=1}^k p_i q_{ii} s_i^2}{n^2 \left(\sum_{i,j=1}^k p_i p_j q_{ij}\right)^2} = \\ &= \frac{1}{10} \left(\frac{\sum_{i,j=1}^k p_i p_j q_{ij} s_i s_j}{n \left(\sum_{i,j=1}^k p_i p_j q_{ij}\right)^2} - \frac{\sum_{i=1}^k p_i q_{ii} s_i^2}{n^2 \left(\sum_{i,j=1}^k p_i p_j q_{ij}\right)^2} \right) = \\ &= \frac{1}{10} \cdot \frac{n \sum_{i,j=1}^k p_i p_j q_{ij} \left(\sum_{r=1}^k p_r q_{ir}\right) \left(\sum_{s=1}^k p_s q_{jr}\right) - \sum_{i=1}^k p_i q_{ii} \left(\sum_{r=1}^k p_r q_{ir}\right)^2}{\left(\sum_{i,j=1}^k p_i p_j q_{ij}\right)^2} \end{aligned}$$

Ez a mennyiség $(n + 1)^3$ -al osztva jelenik meg a G_1 gráf háromszögsűrűségének *közelítésében*, pontosabban, ha az előző mennyiség $x(n)$, akkor

$$\frac{1}{6}t(\Delta, G_1) \approx \frac{t_0}{(n + 1)^3} + \frac{x(n)}{(n + 1)^3} = \frac{n^3}{(n + 1)^3} \cdot \frac{t(\Delta, G_0)}{6} + \frac{x(n)}{(n + 1)^3}$$

és mivel láthatóan $x(n) = O(n)$, ezért $\lim_{n \rightarrow \infty} \frac{x(n)}{(n + 1)^3} = 0$, így

$$\lim_{n \rightarrow \infty} t(\Delta, G_1) \approx \lim_{n \rightarrow \infty} t(\Delta, G_0)$$

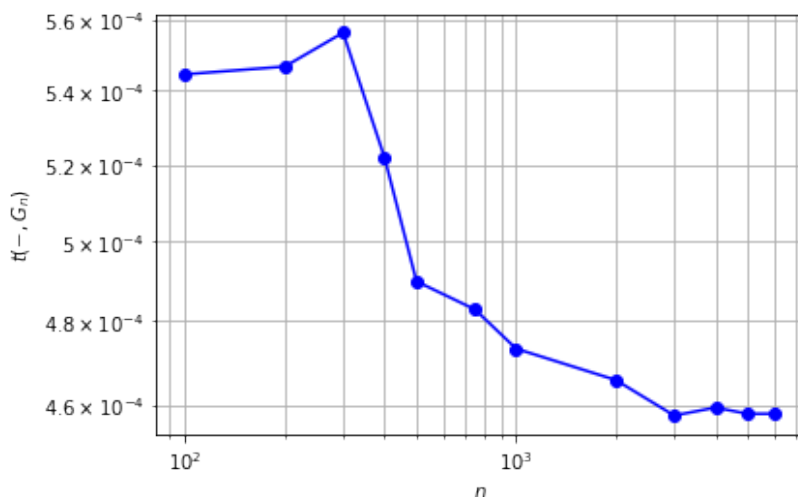


3.2. ábra. A háromszögsűrűség alakulása SBM és Barabási-Albert szabály szerint 1 db. csúccsal bővített SBM esetén, amint n értéke növekszik (3.7)

A szimulációk is alátámasztják ezt a következtetést, amint azt a 3.2 ábrán láthatjuk. Itt SSBM $(1000, 5, \frac{1}{25}, \frac{1}{1000})$ kiindulási gráf esetén az mutatkozik, hogy a háromszögsűrűség 1 darab csúcs Barabási-Albert-módszerrel való hozzáadása után ugyanoda konvergál, mint a kiindulási (SBM) gráf háromszögsűrűsége. Az ábrán a két érték alakulása látható egy sorsolás során, amint n rendre az 500, 750, 1000, 2500, 5000, 7500, 10000, 12500, 15000 értékeket veszi fel.

Ha a G_n gráf háromszögsűrűségét vizsgáljuk számítógéppel, amint n értéke növekszik, azt láthatjuk, hogy elég nagy n -re a ez mintha besűrűsödne egy bizonyos érték környezetébe. A 3.3 ábrán a háromszögsűrűség alakulását láthatjuk, amint n rendre a

100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000 értékeket veszi fel. A kiindulási gráf itt is az SSBM $(1000, 5, \frac{1}{25}, \frac{1}{1000})$ szimmetrikus sztochasztikus blokkgráf volt.



3.3. ábra. A G_n gráf háromszögsűrűségének alakulása, amint n értéke növekszik (3.8)

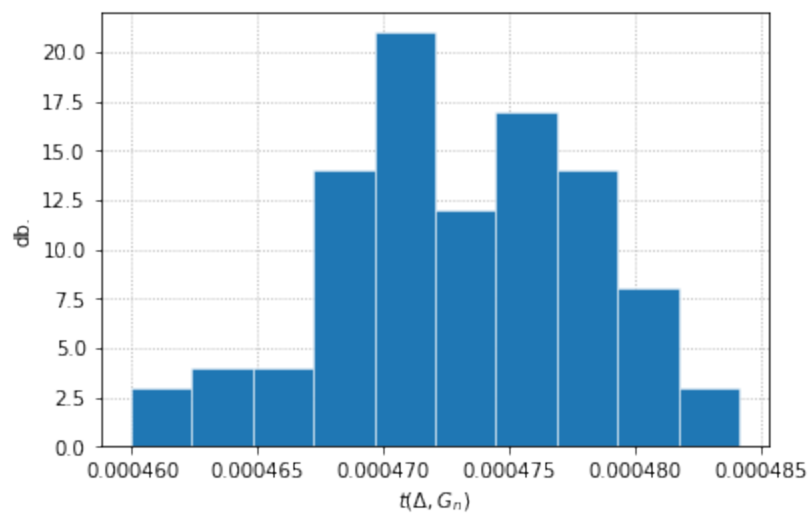
A 3.3 ábrán látható jelenség valamilyen értelemben érthető, hiszen adott kiindulási gráf esetén új háromszög csak úgy tud keletkezni, ha az új csúcs egy olyan él mindkét csúcsához kapcsolódik, ami már az előző gráfban is szerepelt. Azonban, ha az egyik végponthoz hozzá is kötjük az új csúcsot, ettől nem fog nőni annak a valószínűsége, hogy a másikhoz is hozzákötjük, és így háromszög keletkezik. Ahhoz, hogy még jobban megértsük, mi történik, bevezetünk egy új fogalmat. Egy egyszerű G gráf *klaszteresedési együtthatóján* a $C(G) = \frac{\frac{1}{2} \text{hom}(\Delta, G)}{|\{\{e_1, e_2\} \subset E(G) : |e_1 \cap e_2| = 1\}|}$ mennyiséget értjük, vagyis ez nem más, mint a G -ben található háromszögek számának háromszorosa és ún. *cseresznyék* (2 hosszú utak) számának hányadosa. Bollobás és Riordan 2003-ban megmutatták [5], hogy a ritka esetben $m \in \mathbb{N}^+$ esetén

$$\lim_{n \rightarrow \infty} \frac{n}{(\ln n)^2} \mathbb{E} \left[C \left(\text{PA}_n^{(m,0)} \right) \right] = \frac{m-1}{8}$$

ami azt jelenti, hogy ha $m \geq 2$, akkor $\mathbb{E} \left[C \left(\text{PA}_n^{(m,0)} \right) \right] \sim \frac{m-1}{8} \cdot \frac{(\ln n)^2}{n}$, amint $n \rightarrow \infty$. Ez $\lim_{n \rightarrow \infty} \frac{(\ln n)^2}{n} = 0$ miatt pedig megmagyarázhatja, hogy miért csökkenti a preferential attachment dinamika a háromszögsűrűséget, amint n értéke egyre nagyobb lesz. Persze

a mi esetünkben sűrű gráfokról beszélünk, az előbb leírt eredmény pedig ritka gráfokról szól, tehát az nem alkalmazható közvetlenül a mi esetünkre. Ettől függetlenül ez az aszimptotika rámutathat arra, hogy ebben az esetben is valami hasonló jelenség mehet végbe a preferential attachment dinamika által.

Szintén az ugyanilyen szimmetrikus sztochasztikus blokkgráffal indítva a szimulációt – annyi eltéréssel, hogy most $n = 5000$ –, 100-szor függetlenül sorsoltunk G_n gráfokat. A 3.4 ábrán ezek háromszögsűrűségeiből készített hisztogram látható.



3.4. ábra. 100 db. függetlenül sorsolt G_n gráf háromszögsűrűségéből készített hisztogram $n = 5000$ esetén (3.9)

Észrevehetjük, hogy a normális eloszláshoz hasonló alak jelent meg, ezért azt sejtethetjük, hogy a várható értéktől való eltérésnek megfelelő normalizálás után ennek normális eloszlás a határértéke.

Függelék

Ebben a fejezetben a szimulációkat kivitelező számítógépes programok forráskódjai vannak. Minden program a Python nyelvben íródott, és a következő csomagokat használta:

```
1 import networkx as nx
2 import matplotlib.pyplot as plt
3 import numpy as np
```

A 2.1 szimulációkhoz használt programok forráskódjai:

```
1 #Haromszogsuruseg
2 x = np.array([100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000])
3 y = np.array([])
4
5 p=0.15
6 ER_limit = p**3
7 for i in range(len(x)):
8     n = x[i]
9     G = nx.erdos_renyi_graph(n, p)
10    tri_dens = 2*tri(G) / (len(G)**3)
11    y = np.append(y, tri_dens)
12
13 #grafikon
14 fig, ax = plt.subplots()
15
16 ax.plot(x, y, color = 'blue',marker = 'o')
```

```

17 plt.axhline(y = ER_limit, linestyle = '--', color = "red")
18
19 plt.grid(visible = True, which = 'both')
20 ax.loglog()
21
22 ax.set(xlabel = '$n$', ylabel = '$t(\Delta, ER_n(p))$', title = "$p
    =0.15$")
23 plt.annotate("$y=p^3$", (90, ER_limit+0.012*10**(-3)), color = "red")
24
25 plt.show()
26

```

Listing 3.1. A háromszögsűrűség szimulálása Erdős-Rényi-gráf esetén

```

1
2 #elsuruseg
3 x = np.array([100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000,
    5000, 6000, 7000, 8000, 9000, 10000])
4 y = np.array([])
5
6 p=0.15
7 ER_limit = p
8 for i in range(len(x)):
9     n = x[i]
10    G = nx.erdos_renyi_graph(n, p)
11    tri_dens = 2*len(G.edges()) / (len(G)**2)
12    y = np.append(y, tri_dens)
13
14 #grafikon
15 fig, ax = plt.subplots()
16
17 ax.plot(x, y, color = 'blue', marker = 'o')
18
19 plt.axhline(y = ER_limit, linestyle = '--', color = "red")
20
21 plt.grid(visible = True, which = 'both')

```

```

22 ax.loglog()
23
24 ax.set(xlabel = '$n$', ylabel = '$t(-, ER$_n(p))$', title = "$p=0.15$"
        ")
25 plt.annotate("$y=p$", (90, ER_limit+0.02*10**(-2)), color = "red")
26
27 plt.show()

```

Listing 3.2. Az élsűrűség szimulálása Erdős-Rényi-gráf esetén

A 2.2 szimuláció forráskódja:

```

1 def SBM(n, scale, probs):
2     sizes = [int(s*n) for s in scale] #community meretek
3     return nx.stochastic_block_model(sizes, probs)
4
5 def tri(G): #haromszgek szamanak 3-szorosa
6     return sum(list(nx.triangles(G).values()))
7
8 probs = [[0.29, 0.001, 0.001, 0.005, 0.002],
9          [0.001, 0.3, 0.001, 0.001, 0.001],
10         [0.001, 0.001, 0.45, 0.002, 0.0001],
11         [0.005, 0.001, 0.002, 0.25, 0.003],
12         [0.002, 0.001, 0.0001, 0.003, 0.4]]
13
14 scale = [0.13, 0.241, 0.145, 0.29, 0.194]
15
16 #limesz
17 limit = 0
18 for i in range(5):
19     limit+= scale[i]**3 * probs[i][i]**3
20
21 for i in range(5):
22     for j in range(5):
23         limit += (3* scale[i]**2 * scale[j] * probs[i][i] * probs[i][j]
24                 **2) * int(bool(i!=j))

```



```

25 for i in range(5):
26     for j in range(5):
27         for k in range(5):
28             limit += ( 6* scale[i] * scale[j] * scale[k] * probs[i][j] *
29                       probs[j][k] * probs[k][i] ) * int(bool(i!=j and j!=k and k!=i))
30 #szimulacio
31 x = np.array([100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000,
32             7000, 8000, 9000, 10000])
33
34 for i in range(len(x)):
35     n = x[i]
36     G = SBM(n, scale, probs)
37     tri_dens = 2*tri(G) / (len(G)**3)
38     y = np.append(y, tri_dens)
39
40 # grafikon
41 fig, ax = plt.subplots()
42
43 ax.plot(x, y, color = 'blue', marker = 'o')
44 plt.axhline(y = limit, linestyle = '--', color = "red")
45
46 plt.grid(visible = True, which = 'both')
47 ax.loglog()
48
49 ax.set(xlabel = '$n$', ylabel = '$t(\Delta, SBM_n)$')
50 plt.annotate("$\lim_{t \rightarrow \infty} t(\Delta, SBM_n) \approx$", (100, limit
51     +0.012*10**(-3)), color = "red")
52 plt.show()

```

Listing 3.3. Az háromszögsűrűség szimulálása SBM esetén

A 2.3 ábra forráskódja:

```
1 n=1000
2 scale = [0.13, 0.241, 0.145, 0.29, 0.194]
3 sizes = [int(s*n) for s in scale] #community meretek
4 probs = [[0.29, 0.001, 0.001, 0.005, 0.002],
5           [0.001, 0.3, 0.001, 0.001, 0.001],
6           [0.001, 0.001, 0.45, 0.002, 0.0001],
7           [0.005, 0.001, 0.002, 0.25, 0.003],
8           [0.002, 0.001, 0.0001, 0.003, 0.4]]
9
10 G = nx.stochastic_block_model(sizes, probs)
11
12 options = {
13     'node_color': 'red',
14     'node_size': 5,
15     'width': 0.1,
16 }
17
18 nx.draw_spring(G, **options), nx.draw_circular(G, **options)
```

Listing 3.4. SBM-gráf lerajzolása közösségenként és körkörösén

A 2.4 ábra hasonló inicializáció után készült, az alábbi parancsokkal:

```
1 nx.draw_random(G, **options), nx.draw_spring(G, **options)
```

Listing 3.5. SBM-gráf lerajzolása véletlenszerűen és közösségenként

A 3.1 szimuláció forráskódja:

```
1 def SBM(n, scale, probs):
2     sizes = [int(s*n) for s in scale] #community meretek
3     return nx.stochastic_block_model(sizes, probs)
4
5 scale = [1/5, 1/5, 1/5, 1/5, 1/5]
6
7 w = [[1/25, 1/1000, 1/1000, 1/1000, 1/1000],
8       [1/1000, 1/25, 1/1000, 1/1000, 1/1000],
```

```

9     [1/1000, 1/1000, 1/25, 1/1000, 1/1000],
10    [1/1000, 1/1000, 1/1000, 1/25, 1/1000],
11    [1/1000, 1/1000, 1/1000, 1/1000, 1/25]]
12
13 x = np.array([100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000,
14              5000, 6000, 7000, 8000, 9000, 10000])
15
16 y = np.array([])
17
18 for i in range(len(x)):
19     n = x[i]
20     G = SBM(n, scale, w)
21     Gn = nx.barabasi_albert_graph(2*n, n//10, initial_graph = G)
22     edge_dens = 2*len(Gn.edges()) / (len(Gn)**2)
23     y = np.append(y, edge_dens)
24
25 fig, ax = plt.subplots()
26
27 ax.plot(x, y, color = 'blue', marker = 'o')
28
29 plt.grid(visible = True, which = 'both')
30 ax.loglog()
31 ax.set(xlabel = '$n$', ylabel = '$t(-, G_n)$')
32
33 plt.show()

```

Listing 3.6. Az élsűrűség szimulálása a 3.1 szakaszban leírt modellű gráf esetén

A 3.2 szimuláció forráskódja:

```

1 def SBM(n, scale, probs):
2     sizes = [int(s*n) for s in scale] #community meretek
3     return nx.stochastic_block_model(sizes, probs)
4
5 def tri(G): #haromszgek szamanak 3-szorosa
6     return sum(list(nx.triangles(G).values()))
7
8 scale = [1/5, 1/5, 1/5, 1/5, 1/5]

```

```

9
10 w = [[1/25, 1/1000, 1/1000, 1/1000, 1/1000],
11       [1/1000, 1/25, 1/1000, 1/1000, 1/1000],
12       [1/1000, 1/1000, 1/25, 1/1000, 1/1000],
13       [1/1000, 1/1000, 1/1000, 1/25, 1/1000],
14       [1/1000, 1/1000, 1/1000, 1/1000, 1/25]]
15
16 #szimulacio
17 SBM_tridensities = []
18 Extended_tridensities = []
19
20 x = np.array([500, 750, 1000, 2500, 5000, 7500, 10000, 12500, 15000])
21 for i in range(len(x)):
22     n = x[i]
23     G = SBM(n, scale, w)
24     SBM_tridensities.append(2*tri(G) / (len(G)**3))
25
26     G1 = nx.barabasi_albert_graph(n+1, n//10, initial_graph = G)
27     Extended_tridensities.append(2*tri(G1) / (len(G1)**3))
28
29 y_SBM = np.array(SBM_tridensities)
30 y_BA1 = np.array(Extended_tridensities)
31
32 #limesz
33 #elvileg ez a limesz: (3.3) a dolgozatban
34
35 limit = 0
36 for i in range(5):
37     limit+= scale[i]**3 * w[i][i]**3
38
39 for i in range(5):
40     for j in range(5):
41         limit += ( 3* scale[i]**2 * scale[j] * w[i][i] * w[i][j]**2 ) *
42                 int(bool(i!=j))

```

```

43 for i in range(5):
44     for j in range(5):
45         for k in range(5):
46             limit += ( 6* scale[i] * scale[j] * scale[k] * w[i][j] * w[j][k]
47                 * w[k][i] ) * int(bool(i!=j and j!=k and k!=i))
48
49 #grafikon
50
51 fig, ax = plt.subplots()
52
53 ax.plot(x, y_BA1, color = 'blue', marker = 'o', label = "$G_1$")
54 ax.plot(x, y_SBM, color = 'green', marker = 'o', label = "$G_0$")
55
56 plt.axhline(y = limit, linestyle = '--', color = "red", label = "$\
57     \lim_{\infty} t(\Delta, G_0) = \lim_{\infty} t(\Delta, G_1)$")
58
59 plt.grid(visible = True, which = 'both')
60 ax.loglog()
61 ax.set(xlabel = '$n$', ylabel = "Haromszogsuruseg")
62 ax.legend()
63
64 plt.show()

```

Listing 3.7. Barabási-Albert módszerrel bevett új csúcs háromszögsűrűsége gyakorolt hatás vizsgálata SBM kiindulási gráf esetén

A 3.3 szimuláció forráskódja:

```

1 def SBM(n, scale, probs):
2     sizes = [int(s*n) for s in scale] #community meretek
3     return nx.stochastic_block_model(sizes, probs)
4
5 scale = [1/5, 1/5, 1/5, 1/5, 1/5]
6
7 w = [[1/25, 1/1000, 1/1000, 1/1000, 1/1000],
8     [1/1000, 1/25, 1/1000, 1/1000, 1/1000],
9     [1/1000, 1/1000, 1/25, 1/1000, 1/1000],
10    [1/1000, 1/1000, 1/1000, 1/25, 1/1000],

```

```

11     [1/1000, 1/1000, 1/1000, 1/1000, 1/25]]
12 x = np.array([100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000,
13             5000, 6000])
14
15 for i in range(len(x)):
16     n = x[i]
17     G = SBM(n, scale, w)
18     Gn = nx.barabasi_albert_graph(2*n, n//10, initial_graph = G)
19     edge_dens = 2*tri(Gn) / (len(Gn)**3)
20     y = np.append(y, edge_dens)
21
22 fig, ax = plt.subplots()
23
24 ax.plot(x, y, color = 'blue', marker = 'o')
25
26 plt.grid(visible = True, which = 'both')
27 ax.loglog()
28 ax.set(xlabel = '$n$', ylabel = '$t(-, G_n)$')
29
30 plt.show()

```

Listing 3.8. A 3.1 szakaszban leírt modell háromszögsűrűségének alakulása

A 3.4 szimuláció forráskódja:

```

1 n=5000
2 sizes = [n//5, n//5, n//5, n//5, n//5]
3 w = [[1/25, 1/1000, 1/1000, 1/1000, 1/1000],
4       [1/1000, 1/25, 1/1000, 1/1000, 1/1000],
5       [1/1000, 1/1000, 1/25, 1/1000, 1/1000],
6       [1/1000, 1/1000, 1/1000, 1/25, 1/1000],
7       [1/1000, 1/1000, 1/1000, 1/1000, 1/25]]
8
9 tri = []
10 for i in range(100):
11     G = nx.stochastic_block_model(sizes, w)

```

```

12 Gn = nx.barabasi_albert_graph(2*n, n//10, initial_graph = G)
13 t = sum(nx.triangles(Gn).values())
14 tri.append(t)
15
16 t = [e*2 / ((2*n)**3) for e in tri]
17 x = np.array(t)
18
19 # hisztogram
20 fig, ax = plt.subplots()
21
22 ax.hist(x, bins=10, linewidth=0.5, edgecolor="white", zorder=2)
23 ax.set(xlabel = '$t(\Delta, G_n)$', ylabel = 'db.')
```

Listing 3.9. A 3.1 szakaszban leírt modell háromszögsűrűségéből készített hisztogram 100 független sorsolás esetén ($n = 5000$ mellett)

Irodalomjegyzék

- [1] E. ABBE, *Community detection and stochastic block models: recent developments*. <https://arxiv.org/abs/1703.10146>, 2017.
- [2] W. ARVESON, *A Short Course in Spectral Theory*, Springer, New York, Egyesült Államok, 2002.
- [3] N. BERGER, C. BORGS, J. T. CHAYES, AND A. SABERI, *Asymptotic behavior and distributional limits of preferential attachment graphs*, *The Annals of Probability*, 42 (2014), pp. 1–40.
- [4] B. BOLLOBÁS, *Random Graphs*, Cambridge University Press, Cambridge, Egyesült Királyság, 2001.
- [5] B. BOLLOBÁS AND O. RIORDAN, *Mathematical results on scale-free random graphs*, Wiley-VCH, Berlin, Németország, 2003, pp. 1–34.
- [6] C. BORGS, J. CHAYES, L. LOVÁSZ, V. T. SÓS, B. SZEGEDY, AND K. VESZTERGOMBI, *Graph limits and parameter testing*. <https://www.microsoft.com/en-us/research/publication/graph-limits-parameter-testing/>, 2006.
- [7] C. BORGS, J. T. CHAYES, L. LOVÁSZ, V. T. SÓS, AND K. VESZTERGOMBI, *Convergent graph sequences 1: Subgraph frequencies, metric properties, and testing*, *Advances in Mathematics*, 219 (2008), pp. 1801–1851.

- [8] A. DECELLE, F. KRZAKALA, C. MOORE, AND L. ZDEBOROVÁ, *Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications*, Phys. Rev., 84 (2011).
- [9] P. ERDŐS AND A. RÉNYI, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci., 5 (1960), pp. 17–61.
- [10] W. EVANS, C. KENYON, Y. PERES, AND L. J. SCHULMAN, *Broadcasting on trees and the ising model*, The Annals of Applied Probability, 10 (2000), pp. 410–433.
- [11] A. FRIEZE AND R. KANNAN, *Quick approximation to matrices and applications*, Combinatorica, 19 (1999), pp. 175–220.
- [12] A. FRIEZE AND M. KAROŃSKI, *Introduction to Random Graphs*, Cambridge University Press, Cambridge, Egyesült Királyság, 2016.
- [13] E. N. GILBERT, *Random graphs*, Ann. Math. Statist., 30(4) (1959), pp. 1141–1144.
- [14] R. V. D. HOFSTAD, *Random Graphs and Complex Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, Egyesült Királyság, 2016.
- [15] L. LOVÁSZ, *Very Large Graphs*, International Press of Boston, Somerville, Massachusetts, USA, 2009, pp. 67–128.
- [16] ———, *Large Networks and Graph Limits*, American Mathematical Society, Providence, Rhode Island, 2012.
- [17] L. LOVÁSZ AND B. SZEGEDY, *Limits of dense graph sequences*, J. Combin. Theory B, 96 (2006), pp. 933–957.
- [18] ———, *Szemerédi’s lemma for the analyst*, Geom. Func. Anal., 17 (2007), pp. 252–270.
- [19] S. MILGRAM, *The small-world problem*, Psychology Today, 1 (1967), pp. 61–67.

- [20] E. MOSSEL, J. NEEMAN, AND A. SLY, *Reconstruction and estimation in the planted partition model*, Probability Theory and Related Fields, 162 (2015), pp. 431–461.
- [21] S. ROSENGERN, *Random Graph and Growth Models*, PhD thesis, Stockholm University, Stockholm, Svédország, 2020.
- [22] B. RÁTH AND L. SZAKÁCS, *Multigraph limit of the dense configuration model and the preferential attachment graph*, Acta Mathematica Hungarica, 136 (2012), pp. 196–221.
- [23] J. WENGER, *Six degrees of wikipedia*. <https://www.sixdegreesofwikipedia.com/>. Utoljára elérve: 2022-03-07.

NYILATKOZAT

Név: Kovács Bálint

ELTE Természettudományi Kar, szak: Matematika BSc

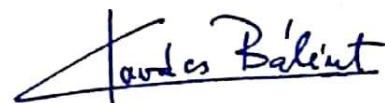
NEPTUN azonosító: UM6A4G

Szakdolgozat címe:

Gráfimeszek és a sztochasztikus blokkmodell kapcsolata

A szakdolgozat szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2022.05.28.



a hallgató aláírása