Fokszámeloszlás Additív Fitnesszel rendelkező Preferential Attachement Gráfokban

Ferenczi Dávid
Témavezető:Backhausz Ágnes

2020. december 6.

1. Bevezetés

A félév során az úgy nevezett preferential attachement hálózatok egy kitüntetett típusával foglalkoztam, az ún additív fitnesszel rendelkező preferential attachement gráfokkal. Ehhez Bas Lodewijks és Marcel Ortgiese [1] cikkének bizonyos részeit dolgoztam fel.
Az alapmodell a következő: Legyenek n_0, m_0 természetes, nem nulla számok, és vegyünk egy n_0 csúccsal és m_0 éllel rendelkező gráfot. Legyen továbbá minden $1 \leq i \leq n_0$ számra $\deg(i)$ az i-edik csúcs befoka, valamint egy m tetszőleges természetes szám. Vegyünk a gráfhoz egy új pontot, és kössük hozzá a korábbi csúcsokhoz m éllel, úgy hogy az alábbi szabály érvényes:

$$\mathbb{P}((v(n+1), v(i)) \in E(G)) = \frac{\deg(i)}{e_{G_n}},$$

ahol a G gráfnak $E(G)$-vel jelöltük az élhalmazát, valamint a nevezőben e_{G_n} a G gráf éleinek száma az n-edik csúcs hozzávétele után. Az ilyen gráfokat nevezzük preferential attachment gráfoknak.
Azt fontos látni, hogy semmilyen feltételt nem tettünk az élek együttes eloszlására abban az esetben, ha egy lépésben több él is létrejön. Ezt a látszólagos problémát néhány sorral lejjebb orvosoljuk, azáltal, hogy megadjuk a konkrét fejlődési szabályt.

2. Az additív fitnesz esete

Ha az előbbi modellnél egy szemléletes példát keresnénk, lehetne egy iskolai osztályt nézni, ahol minden nap jön egy új gyerek az osztályba, és elkezd a többiekkel barátkozni, mégpedig úgy, hogy minél több barátja van valakinek, az új gyerek annál szívesebben köt vele barátságot. Az additív fitnesz esete abban tér el, hogy a "barátok" száma mellett még minden gyereknek van valamilyen tulajdonsága, ami belejátszik a barátkozásba. Például egy ilyen lehetne a magasság. Ekkor a módosított példánk szerint egy új gyerek az osztályban szeretne olyannal barátkozni, aki lehetőleg minél magasabb és minél több barátja van.
Ez annak felel meg, hogy minden csúcsnak van egy időtől függő és egy állandó paramétere, ami meghatározza, hogy milyen valószínűséggel csatlakoznak hozzá csúcsok.

2.1. Definíció. Legyen $(F_i)_{i\in\mathbb{N}}$ nem negatív,1 valószínűséggel véges, független, azonos eloszlású valószínűségi változó sorozat μ eloszlással. Minden n-re legyen $S_n = \sum_{i=1}^{n} F_i$, valamint $n_0, m_0 \in \mathbb{N}$. Azt mondjuk, hogy a $(G_n)_{n\geq n_0}$ véletlen gráfok egy sorozata preferential attachement additív fitnesszel, ha ez egy irányított gráf, G_{n_0}-nak m_0 éle van, minden j csúcshoz hozzá van rendelve az F_j fitnesz, valamint az élek a nagyobb indexű csúcsoktól a kisebbek irányába tartanak. Meg kell adni még a fejlődési szabályt, vagyis hogyan kapható meg G_n, az egy lépéssel korábbi gráfból, G_{n-1}-ből, amennyiben $n-1\geq n_0$. Három különböző szabályt is megadunk. Legyen m az új létrejövő élek száma, és jelölje $Z_n(i)$ az n lépés utáni gráfunk i-edik csúcsának befokát. Az első szabály szerint az új csúcsot az egyik korábbihoz

$$\frac{Z_n(i)+F_i}{m_0+n-n_0+S_n}$$

valószínűséggel kötjük hozzá.
A második szabály esetén az új csúcsnak adjunk m élet, majd ezeket kössük a többi csúcshoz, egymástól függetlenül

$$\frac{Z_n(i)+F_i}{m_0+m(n-n_0)+S_n}$$

valószínűséggel.
Az utolsó fejlődési szabály lényegileg az, hogy az m új élet egymás után adjuk a gráfhoz, és minden él hozzáadása után változnak a valószínűségek. Legyen $Z_{n,j}(i)$ az i-edik csúcs befoka akkor, amikor az $n+1$-edik csúcsnak már j élét hozzáadtuk a gráfhoz. Ekkor annak a valószínűsége, hogy $n+1$ j-edik élét a gráf i-edik csúcsához kötjük, az alábbi:

$$\frac{Z_{n,j-1}+F_i}{m_0+m(n-n_0)+j-1+S_n}$$

2.2. Megjegyzés.
1. Az első fejlődési szabályban fel kell tenni, hogy a $\Delta Z_n(i) = Z_{n+1}(i)-Z_n(i)$ változók korrelációja nem pozitív.
2. A 2. és 3. szabály között az a különbség, hogy egy új csúcs bekapcsolásakor a többi csúcs foka fixen marad-e, vagy minden új kapcsolódáskor frisstjük őket.

3. Aszimptotikus fokszámeloszlások

A tétel, aminek a bizonyításának megértése az elsődleges cél volt, azt mondta ki, hogy a tapasztalati fokszámeloszlások konvergensek.

3.1. Tétel ([1]). *Vegyünk egy preferential attachement modellt, amire $\mathbb{E}(F)<\infty$, és $\theta_m = 1+\frac{\mathbb{E}(F)}{m}$, és $\mathbb{P}(t>F)=\mu((t,\infty))$.*
Vegyük az alábbi tapasztalati fokszámeloszlásokat:

$$\Gamma_n = \frac{\sum_{i=1}^{n} Z_n(i)\delta_{F_i}}{n} \qquad \Gamma_n^{(k)} = \frac{\sum_{i=1}^{n} \mathbb{1}_{Z_n(i)=k}\delta_{F_i}}{n} \qquad p_n(k)=\Gamma_n^{(k)}([0,\infty])$$

Ekkor minden k-ra 1 valósíznűséggel teljesülnek az alábbi konvergenciák.

$$\Gamma_n\to\Gamma, \qquad \Gamma_n^{(k)}\to\Gamma^k, \qquad p_n(k)\to p(k),$$

ahol

$$\Gamma(dx) = \frac{x}{\theta_m - 1}\mu(dx) \qquad\qquad \Gamma^{(k)}(dx) = \frac{\theta_m}{x+\theta_m}\prod_{l=1}^{k}\frac{(l-1)+x}{l+x+\theta_m}\mu(dx),$$

valamint

$$p(k) = \int_0^\infty \frac{\theta_m}{x+\theta_m}\prod_{l=1}^{k}\frac{l-1+x}{l+x+\theta_m}\mu(dx)$$

A bizonyítás megtalálható [1]-ben, itt egy rövid összefoglaló látható annak menetéről. A bizonyítás kulcslépése annak megmutatása, hogy a tapasztalati fokszámeloszlásokra megfelelő A_n, B_n, R_n sorozatokkal a preferential attachement dinamika miatt teljesül az alábbi:

$$\Gamma_{n+1}((f,f']) - \Gamma_n((f,f']) \leq \frac{1}{n+1}(A_n - B_n\Gamma_n((f,f'])) + (R_{n+1} - R_n)$$

Ezután egy lemmát bizonyítunk, ami biztosítja, hogy ha az A_n, B_n, R_n sztochasztikus folyamatok egy valószínűséggel konvergálnak, akkor a Γ_n folyamat limeszét is megkaphatjuk. Hasonló jön ki $\Gamma_n^{(k)}$-ra is, amire még egy indukciót kell alkalmazni, a végén pedig integrálni, hogy megkapjuk a $p(k)$-t. Az indukciós lépés a martingáltulajdonságnak köszönhetően hajtható végre. A tételben elég sok az olyan feltétel, amiket érzésre meglehet próbálni gyengíteni. Az első, ami a szembetűnik, hogy mit lehet mondani az új élek sorsolásáról, azonban amivel én a félévben több időt töltöttem, az a fitnessekre tett feltételek esetleges gyengítése. Természetes gondolat lenne, hogy milyen, a függetlenségnél gyengébb feltételek mellett lehet igaz a tétel.

3.1. Stacionárius, ergodikus fitneszek esete

3.2. Lemma. *Ha az alábbi teljesül, ahol a határértékek egyvalószínűséggel értendőek, a fenti tétel akkor is igaz marad:*

$$\lim_{n\to\infty}\frac{\sum_{i=1}^{n}\mathbb{1}_{F_i\in(f,f']}}{n} = \mu_F((f,f']), \tag{1}$$

$$\lim_{n\to\infty}\frac{\sum_{i=1}^{n}F_i}{n} = \mathbb{E}(F), \tag{2}$$

$$F_i \geq 0, \tag{3}$$

ahol f, f' tetszőleges valós számok.

Az eredeti tétel bizonyításában is csak ezt kellett látni.

Tehát az kezd látható lenni, hogy a függetlenség nem kell nekünk feltétlenül, elég az, hogy a számtani közép és a tapasztalati eloszlás is konvergens legyen 1-valószínűséggel. Erre fogunk kimondani egy elégséges feltételt.

3.3. Tétel. *Legyen $(F_i)_{i\in\mathbb{N}}$ egy erősen stacionárius, nem negatív ergodikus sorozat. Ekkor az ebből az eloszlásból kapott fitnessekből generált preferential attachement modell tapasztalati mértékei konvergálni fognak.*

Azt elég látni, hogy az ilyen eloslásra teljesül a lemma. Ez igaz, egyrészt (2) pont az ergodicitás, (1) pedig a Glivenko-Cantelli tétel egy általánosítása[2].

Hivatkozások

[1] Bas Lodewijks, Marcel Ortgiese. A phase transition for preferential attachment models with additive fitness, arXiv:2002.12863,2020

[2] Tucker, Howard G. A Generalization of the Glivenko-Cantelli Theorem. Ann. Math. Statist. 30 (1959), no. 3, 828–830. doi:10.1214/aoms/1177706212. https://projecteuclid.org/euclid.aoms/1177706212