Pál Balázs
UXB26I

Véletlen fizikai folyamatok
8-9. házifeladat

2019. április 25.

1. FELADAT

K: Szimuláljuk az órán tárgyalt véletlen rekurzív fát (minden lépésben egy új csúcsot adunk a hálózathoz, s az új csúcsot egyenlő valószínűséggel kötjük a meglevő csúcsok egyikéhez).

Feladatok:

(i) Határozzuk meg a csúcsok fokszámeloszlását $P_k = N_k/N$-t, ahol N a csúcsok száma, N_k pedig a k éllel rendelkező csúcsok száma.

(ii) Vizsgájuk mekkora N kell ahhoz, hogy az eloszlásfüggvény P_k hibája kisebb legyen mint 10% minden $k \leq 10$-re. Figyelem, az egzakt eloszlásfüggvényt az órán kiszámoltuk!

(iii) Határozzuk meg az átlagos fokszámot mind elméletileg, mind pedig a szimulációkból!

(iv) Találjuk meg a maximális fokszámú csúcsot a fenti szimulációkban generált hálózatokban. Többször megismételve a szimulációkat $N = 100$, 1000 és 10000 esetére, határozzuk meg a maximális fokszám átlagát, $\langle k_{max} \rangle$-t! Látunk trendet az eredményekben?

2. FELADAT

K: Szimuláljunk egy anti-preferenciális csatolódással növekedő hálózatot, amelyben az új csúcsok nem szeretnek a már sok éllel rendelkező csúcsokhoz kapcsolódni. Egy lépésben egy csúcsot adunk a hálózathoz, s az $N + 1$-edik csúcs csatolásának szabályai legyenek a következők:

(1) Véletlenszerűen kiválasztunk egyet a már meglévő N csúcs közül.

(2) Leszámoljuk a kiválasztott csúcs éleinek számát. Legyen ez a szám k.

(3) A kiválasztott csúcshoz $w_k = k - 2/A$ valószínűséggel csatolunk egy új csúcsot. Itt A a normalizációs állandó $A = \sum_l l^{-2} N_l$, ahol N_l az l éllel rendelkező csúcsok száma.

Feladatok:

(i) Mérjük meg a csúcsok fokszámeloszlását, $P_k = N_k/N$-t!

(ii) Vizsgáljuk a fokszámeloszlás nagy-k aszimptotikáját, s hasonlítsuk össze az eredményt a véletlen rekurzív fára ($w_k \sim 1/N$) kapottakkal!

MEGOLDÁS

2.1. Megvalósítás

A feladatokat megvalósító programok forráskódjának mindegyikét egy Jupyter Notebookban futó Python 3.7 kernel alatt valósítottam meg, valamint az ábrákat is ebben készítettem. Ezeket GitHub-on mind elérhetővé is tettem[1].

Az első feladat során a Véletlen Rekurzív Fát kellett szimuláljuk, tetszőleges N csúcs/élszám darabszámra. Ennek szimulálása a következő lépések elvégzéséből tevődik össze:

(1) Kezdetben lehelyezünk egyetlen pontot az üres térbe. Triviálisan mivel ő az egyetlen, így fokszáma 0, nem csatlakozhat semmilyen másik ponthoz (hisz eleve nincs másik pont).

(2) Első lépésben lehelyezünk még egy pontot a térbe, és véletlenszerűen - azonos valószínűséggel - kiválasztunk egy már meglévőt (jelen esetben ebből egyetlen egy darab van), melyet egy éllel összekötünk az újonnan behozott csúccsal. Minden lehetséges esetben az első két pont szükségszerűen egymással lesz összekötve.

(3) Második körben szintén véletlenszerűen választunk egy pontot - az eddig meglevő 2 közül -, majd egy harmadik pontot kötünk egy éllel hozzá.

(4) A fentieket tetszőleges ideig folytatjuk. A szimulációt esetemben adott élszám elértéig futtattam.

A második feladatban (Anti-Preferenciális Modell) megtett lépések a (2)-es leírásban szereplőkkel megegyezőek.

A szimuláció futtatása mindkét feladat esetében összesen 3 darab adatsorral tért vissza. Ezek egy kezdeti értékkel voltak inicializálva, majd ezek minden lépésben frissültek, végül pedig a végleges értéküket használtam fel a további számításokban és ábrázolásokban. Ezek az adatsorok a következőek voltak, a szimuláció forráskodjában is használt nevük szerint:

- `graph_rrt`: Amikor egy újabb pont érkezik a gráfba, az minden alkalommal pontosan 1 ponthoz csatolódik hozzá. Ez az `array` típusú adatsor sorrendben tartalmazza, hogy az újabb csúcsok melyik másik csúcshoz kapcsolódtak hozzá beérkezésükkor. Hossza egyenlő az összes él darabszámával, mely jelen esetben egyel kevesebb a rendszerben szereplő pontok végleges számánál.

- `count_rrt`: Ez a - szintén - `array` típusú adatsor tartalmazza, hogy az egyes pontok fokszámát. Az első eleme az első pont fokszáma, a második eleme a második pont fokszáma és így tovább. Hossza megegyezik az összes, rendszerben levő pont számával, ami pontosan egyel több, mint a rendszerben levő élek száma ebben az esetben.

- `dist_rrt`: Ez az `array` tartalmazza sorrendben az egyes fokszámmal rendelkező csúcsok számát. Első eleme a 0 fokszámú csúcsok száma, második eleme az 1 fokszámú csúcsok száma, és így tovább. Hossza egyenlő a maximális lehetséges fokszámmal. Ez megegyezik az összes él számával, ugyanis ebben a kis valószínűségű esetben minden csúcs ugyanahhoz az egyetlen ponthoz csatlakozik. Ekkor a központi csúcshoz csatlakozik az összes N darab él, és így fokszáma is N.

A második modell esetén az eredetileg visszatérő adatsorok típusai az elsővel megegyezőek voltak, csak ott rendre `graph_apm`, `count_apm` és `dist_apm` jelölésekkel.
Az első modell esetén az `rrt` a *Random Recursive Tree*, míg a második esetben az `apm` rövidítés az *Anti-Preferential Model* kifejezés rövidítése.

2.2. Kiértékelés

Mindkét modell egy-egy növekedő hálózat, melynek egyes tulajdonságait (pl. P_k eloszlásfüggvény) már az órán kiszámoltuk. A továbbiakban azokra fogok alapozni.
A (1) és (2) ábrákon sorrendben a Véletlen Rekurzív Fa és az Anti-Preferenciális Modell alapvető karakterisztikáit ábrázoltam, $E = 100$ él esetére. Nővekedő hálózat gyanánt ilyenkor $N = E + 1 = 101$ csúcs lesz a hálózatban, hisz minden alkalommal egy beérkező fa, pontosan 1 másik, már bent levő csúcshoz csatlakozik hozzá.
Mindkét ábrán a felső grafikon az egyes pontok végleges k fokszámait mutatja, a hálózatba történő beérkezési sorrendjükben, 1-től indexelve. A középső grafikonok az adott k fokszámú pontok N_k darabszámát mutatják, kezdve a $k = 0$ fokszámtól egészen az $E = 100$ esetében maximálisan lehetséges $k = 100$-ig. Az alsó grafikonokon a $P_k = \frac{N_k}{N}$ diszkrét eloszlásfüggvényt ábrázoltam, mely tulajdonképpen csak a középső grafikonok normáltja.

Mivel az első feladatban a Véletlen Rekurzív Fa esetében vizsgálnunk kell ennek a diszkrét, mért P_k függvénynek az elméletileg várt folytonos P_k függvényhez képesti hibáját, így itt ezt a függvényt is ábrázoltam az alsó grafikonon, egy szaggatott szürke vonallal. A két eloszlásfüggvény külsőre nagyon hasonló, azonban két fő különbség észrevehető rajta. Az első, hogy a Véletlen Rekurzív Fa esetén a $k = 1$ fokszámú csúcsok magasan a legnagyobb arányban szerepelnek a hálózatban, míg az Anti-Preferenciális Modell esetében minden alkalommal közel azonos a $k = 1$ és $k = 2$ csúcsok darabszáma. Ezek közül is legtöbb esetben a $k = 2$ fokszámú csúcsok vezetnek számosság terén. A második különbség, hogy az Anti-Preferenciális Modell esetén a fokszámeloszlás sokkal „simább”, nincsenek a 0 darabszámú esetek sorozatát elérve apró fluktuációk, ahogy az a Véletlen Rekurzív Fa esetében a (1)-es ábrán is láthatóan történik.

Az órán kiszámoltuk, hogy a nővekedő hálózat elméleti fokszámeloszlása a következő:

$$P_k = e^{-k \ln(2)} \tag{1}$$

A szimuláció hibáját az elméletitől való százalékos eltérésnek vettem. Feladatunk, hogy meghatározzuk azt az N-t, ami felett az első 10 fokszám hibája mindig kisebb, mint 10%. Ezt meghatározandó, vizsgáltam az első 10 elem hibáinak maximumát az N növelésének függvényében. Egy tetszőleges futásra ($N = 40\,000$) esetén kapott adatsorokat a (3)-as ábrán közlöm. A kapott eredmények analízise a képaláiratban olvashatóak részletesebben. Összefoglalóként, az ilyen módon vett hiba csak fluktuációk során csökken 10% alá, máskülönben - feltételezhetően - ahhoz propagál végtelenül hosszú futásidő során.

A fokszámok átlagos értékének számára már jóval konkrétabb, és az elméletből is várt eredményt kaptam. Mind a Véletlen Rekurzív Fa , mind pedig az Anti-Preferenciális Modell növekedő hálózat mivoltából kiindulva tudjuk, hogy a k fokszám várható értéke $\langle k \rangle = 2$, melyet az alábbi számításból kaphatunk meg:

$$\langle k \rangle = \sum_{i=1}^{\infty} k_i \cdot P_i \equiv \sum_{k=1}^{\infty} k \cdot P_k = \sum_{k=1}^{\infty} k \cdot e^{-k \ln(2)} = \sum_{k=1}^{\infty} k \cdot \left(e^{\ln(2)}\right)^{-k} = \sum_{k=1}^{\infty} k \cdot 2^{-k} \tag{2}$$

Mely egy számelméletből régóta ismert összeg. Írjuk át az alábbi formába a kapottakat:

$$\sum_{k=1}^{\infty} k \cdot 2^{-k} = \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^{k} \tag{3}$$

Melyre alkalmazhatjuk az alábbi ismert összefüggést:

$$|x| < 1: \quad f\left(x\right) = \sum_{n=1}^{\infty} x^n = \frac{x}{1-x} \tag{4}$$

$$x f'\left(x\right) = \sum_{n=1}^{\infty} n \cdot x^n = \frac{x}{(1-x)^2} \tag{5}$$

Ha most $n = k$ és $x = \frac{1}{2}$ értékeket veszünk, akkor felírhatjuk a következőt:

$$\langle k \rangle = \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^{k} = \frac{\frac{1}{2}}{(1-\frac{1}{2})^2} = \frac{\frac{1}{2}}{\frac{1}{4}} = 2 \tag{6}$$

Tehát végtelenül sok él esetén azt várjuk, hogy az átlagos k érték 2-höz tart. Pontosan ezt figyelhetjük meg a szimulációban is, melyet a (5)-ös és (6)-os ábrán illusztráltam, rendre a Véletlen Rekurzív Fa és az Anti-Preferenciális Modell esetében.

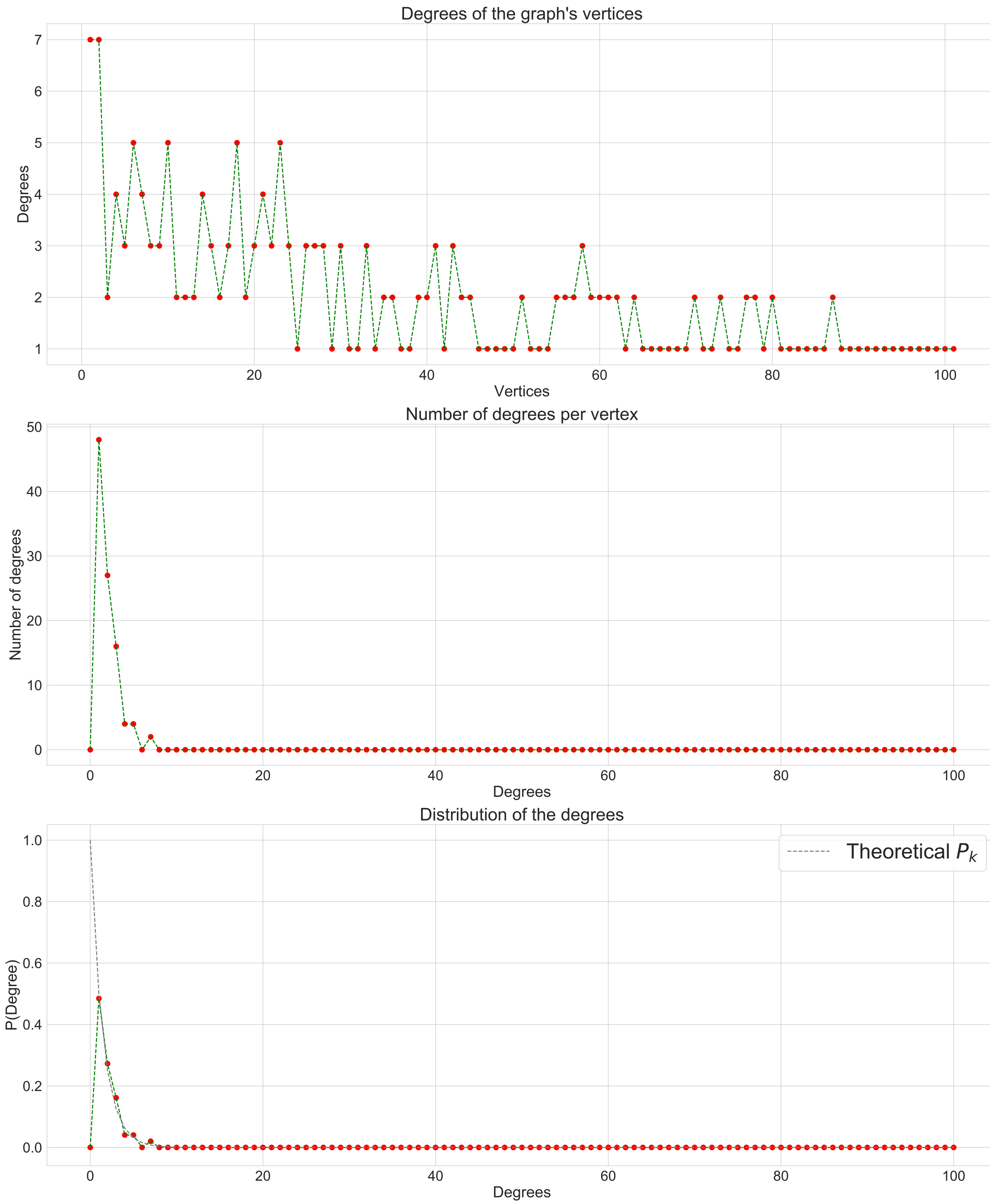
Végül a maximális fokszámok alakulását kellett megvizsgálnunk, az N csúcsok számának növelésének függvényében. Ezt a két modell esetében a (7)-es és (8)-as képeken ábrázoltam. A Véletlen Rekurzív Fa esetén megfigyelhető, hogy a maximális értékek a csúcsok/élszámok függvényében egy exponenciális, vagy esetleg gyökös növekedést mutatnak. Az Anti-Preferenciális Modell esetén a hosszú futásidő miatt csak $E = 2000$ élig tudtam a

szimulációt futtatni, itt még nem rajzolódik ki egyértelműen a Véletlen Rekurzív Fa esetében megfigyelhető mintázat.
A maximális fokszám $\langle k_{max} \rangle$ átlagát több, azonos N csúcsszámú futtatásra számítva határoztam meg. Ezek eredményét a (9)-es ábrán közöltem. A maximális fokszámok élszámfüggéséhez hasonló módon itt is egyértelműen egy exponenciális, vagy gyökös növekedés rajzolódik ki. Az érték hibáját itt kiszámolni felesleges, ugyanis a növekedési tendencia bőven az értékek oszcillációjánál nagyobb nagyságrendileg.

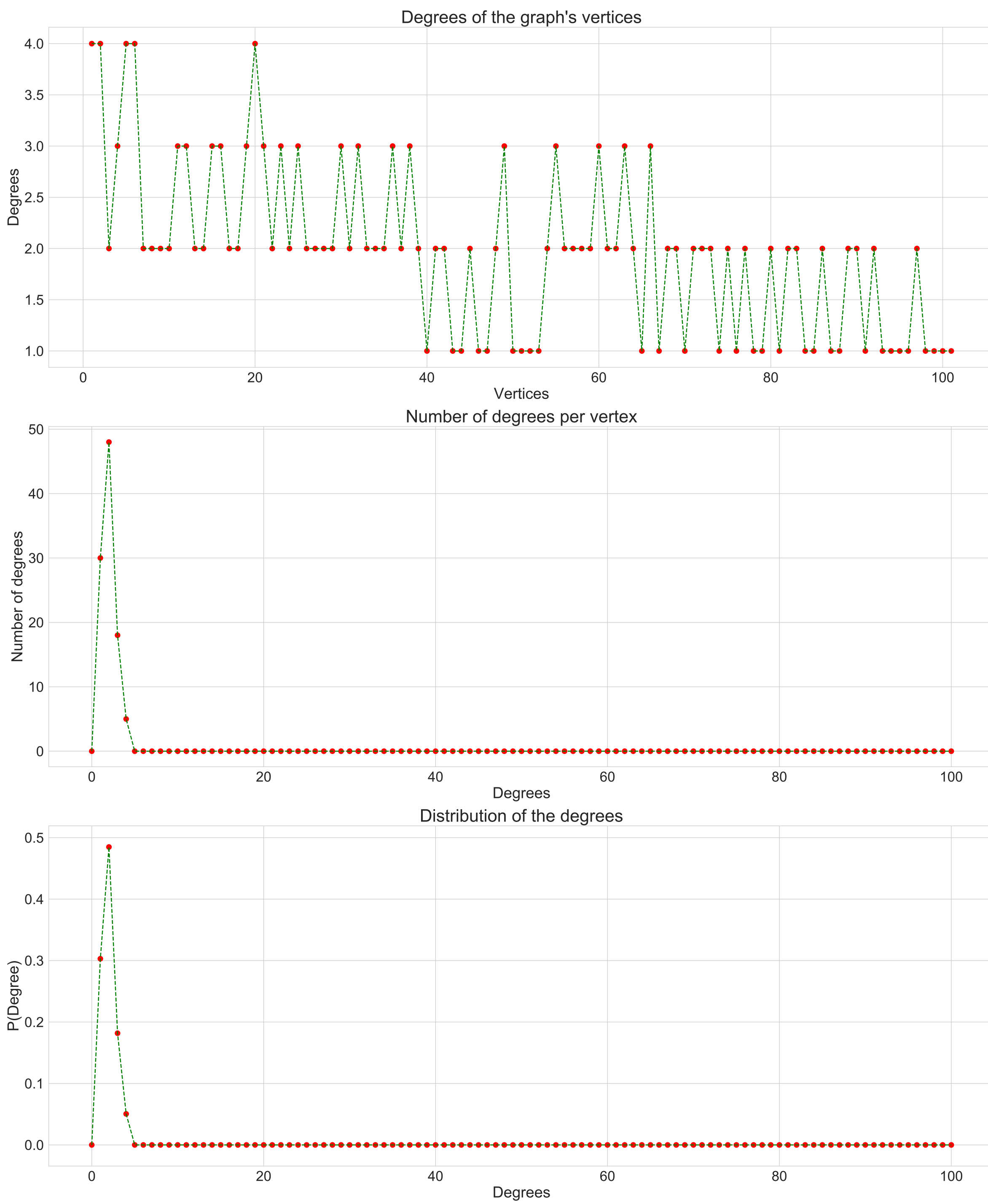
Hivatkozások

[1] Pál Balázs. *ELTE Computer Simulations 2019 — GitHub*. [Online; opened at April 24, 2019]. 2019. URL: https://github.com/masterdesky/ELTE_Comp_Simulations_2019.

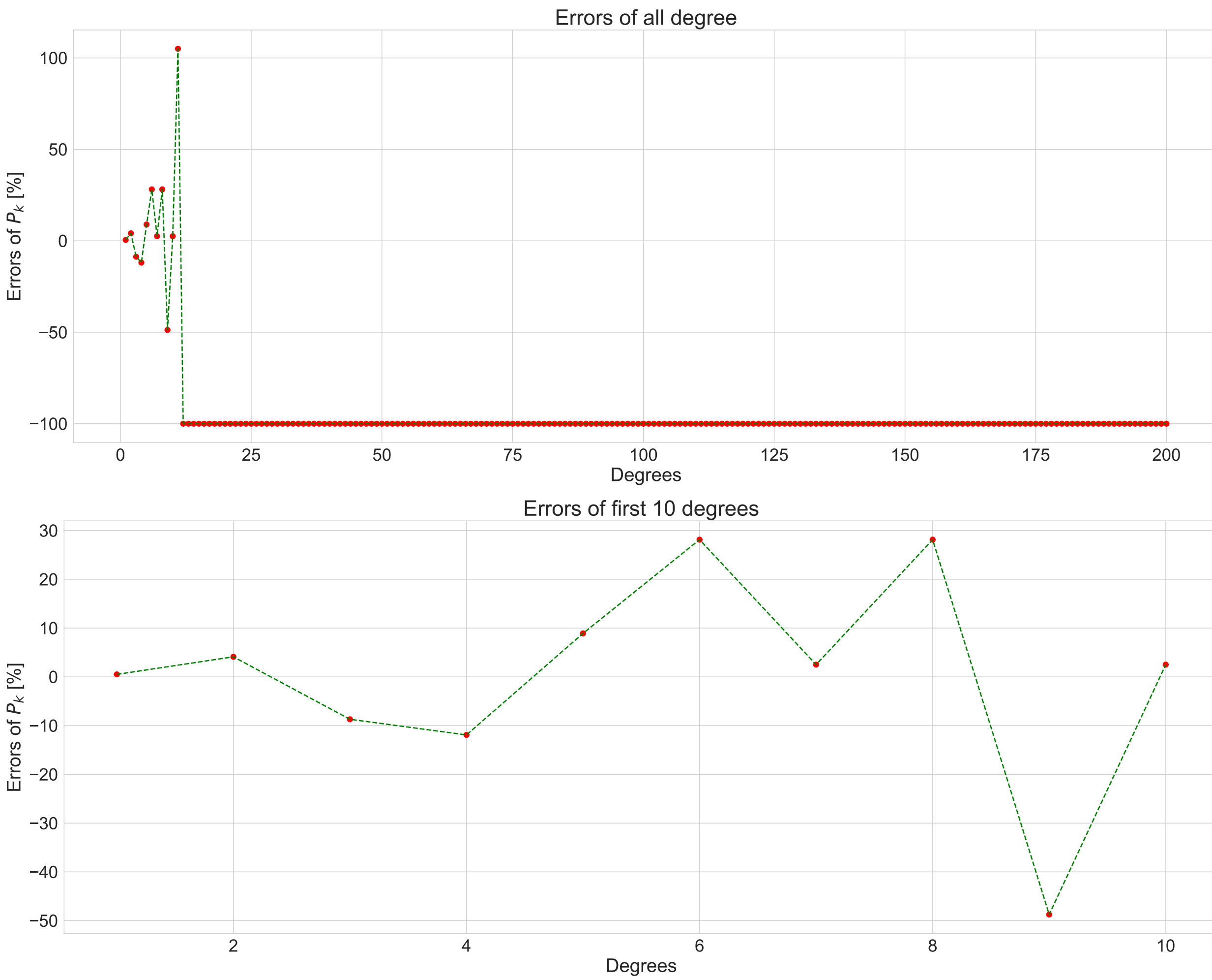
APPENDIX A



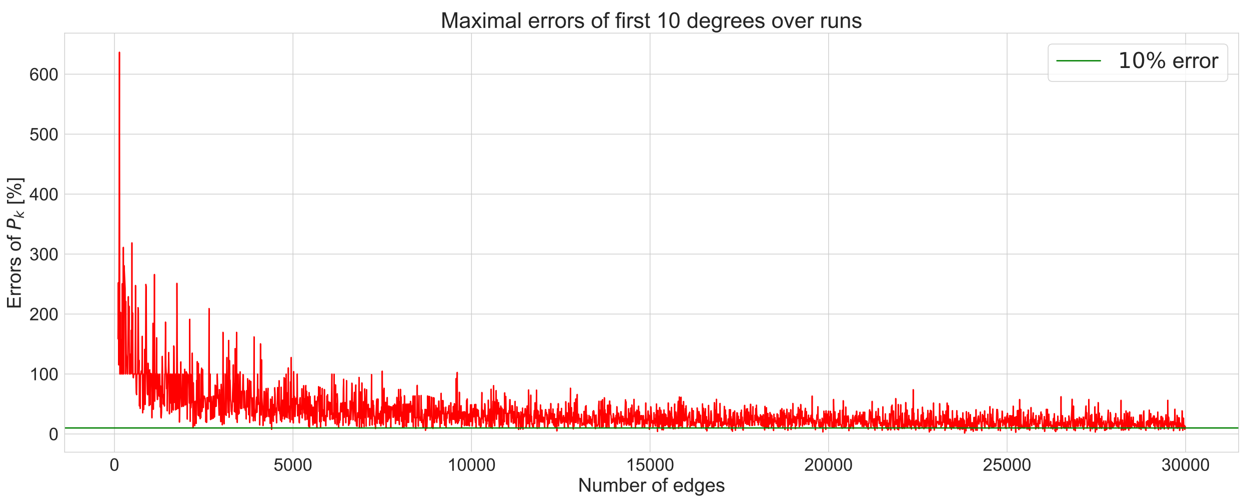
1. ábra. A Véletlen Rekurzív Fa $E = 100$ élre
Fent: Az egyes csúcsok fokszáma
Középen: Adott fokszámú csúcsok darabszáma
Lent: A csúcsok fokszámának eloszlása



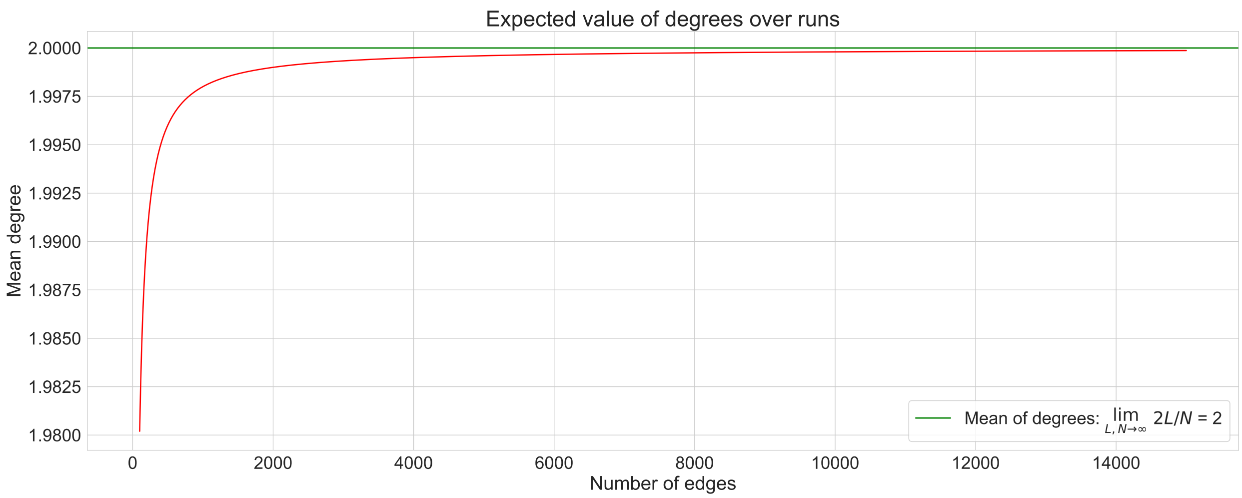
2. ábra. Az Anti-Preferenciális Modell $E = 100$ élre
Fent: Az egyes csúcsok fokszáma
Középen: Adott fokszámú csúcsok darabszáma
Lent: A csúcsok fokszámának eloszlása



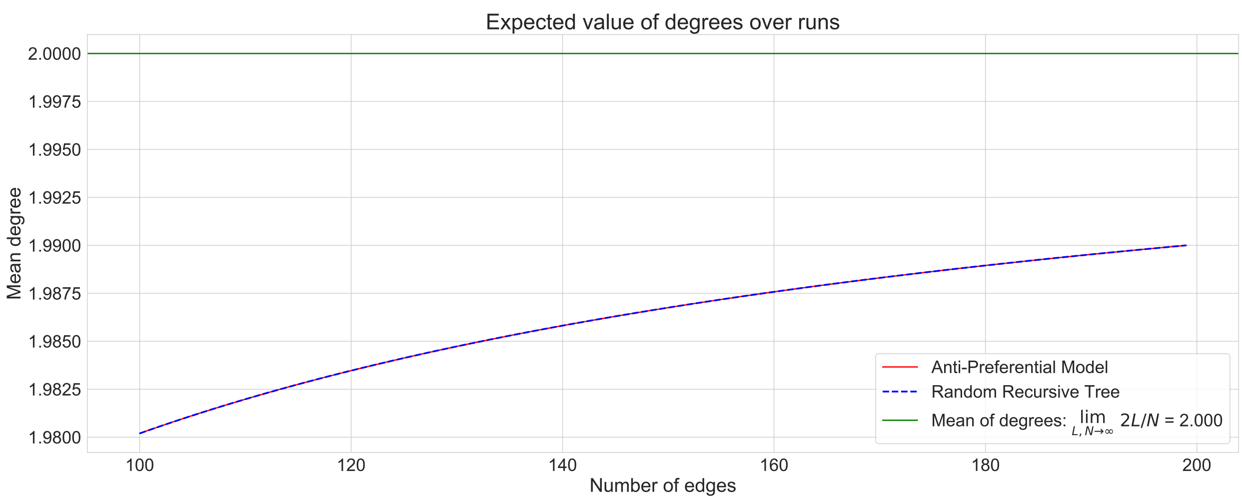
3. ábra. A Véletlen Rekurzív Fa szimulációjának hibája a fokszámeloszlásra vonatkozóan
Fent: A fokszámeloszlás teljes adatsorra vett százalékos eltérése. A -100%-os értékek azt jelzik, hogy azokon a pontokon a szimuláció $k = 0$-t adott eredményül, azonban az elméleti görbe egy exponenciális lecsengő függvény. Emiatt itt az eltérés az elméleti érték 100%-a.
Lent: A fokszámeloszlás csak első 10 értékéhez tartozó hiba ábrázolása.



4. ábra. A Véletlen Rekurzív Fa különböző élszámokra mért P_k élszámeloszlásának maximális eltérései az elméletileg számolt $P_k = e^{-k \cdot \ln(2)}$ értéktől, az adatsor első 10, legnagyobb fokszámú értékére. A görbe egyértelműen exponenciálisan csökkenő karakterisztikát mutat, mely a 10%-os értékhez propagál láthatólag. Mivel a futásidő már nagyon hosszú lett volna a későbbiekben, így nem tudtam a szimulációval nagyonn E értékeket kipróbálni.

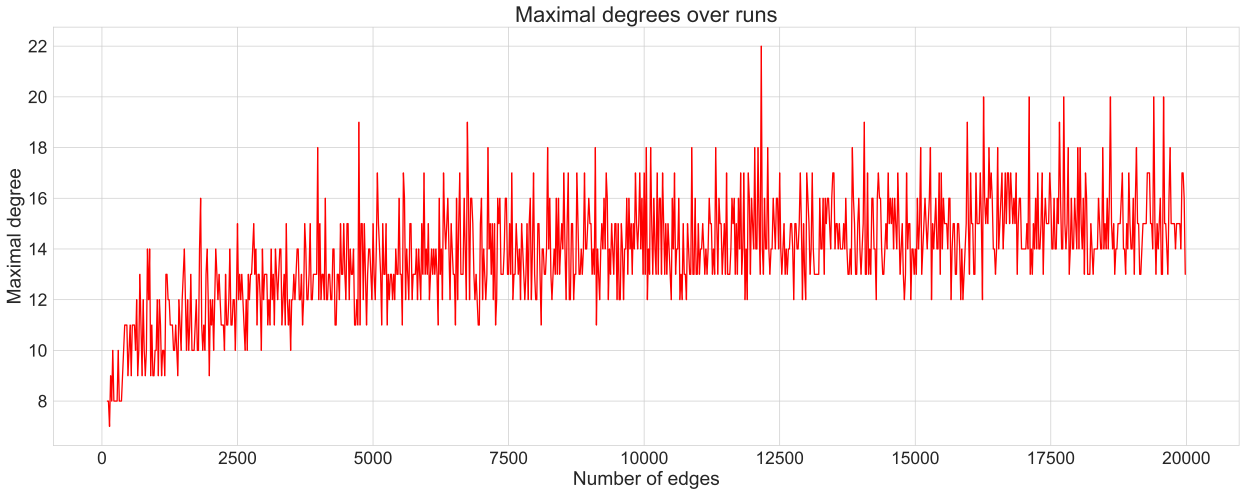


5. ábra. A Véletlen Rekurzív Fa különböző élszámokra mért átlagos fokszáma.
A szimuláció $E = 100$ és $E = 5000$ élszámértékek között futott.



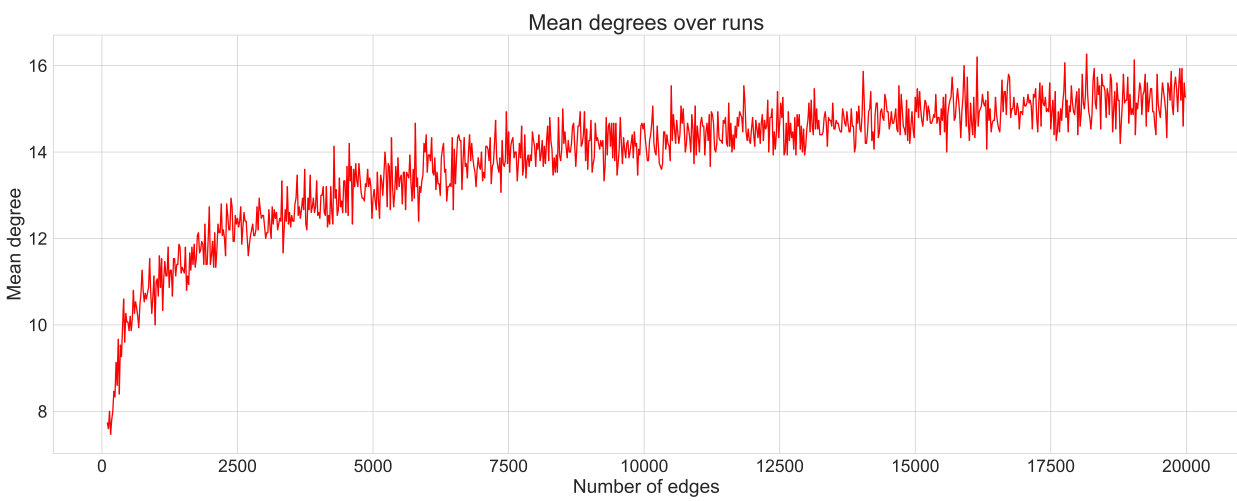
6. ábra. Az Anti-Preferenciális Modell különböző élszámokra mért átlagos fokszáma.
A szimuláció $E = 100$ és $E = 200$ élszámértékek között futott[1].

[1] A két ábra megtévesztőnek tűnhet elsőre. Az átlagos fokszámok mind a Véletlen Rekurzív Fa , mint az Anti-Preferenciális Modell esetén hasonlóan alakultak, azonban az Anti-Preferenciális Modell esetén a futásidő jóval hosszabb volt. Emiatt az élszámok csupán szűk intervallumában tudtam lemérni az egyes átlagos értékeket. Hogy a hasonlóságot szemléltethessem, az Anti-Preferenciális Modell ábráján a Véletlen Rekurzív Fa görbéjét is ábrázoltam, az előbbivel azonos intervallumban. Ezt az ábrán látható jelmagyarázatban jelöltem is.



7. ábra. A Véletlen Rekurzív Fa különböző E élszámok esetén megjelenő maximális fokszáma.
A szimuláció $E = 100$ és $E = 10000$ élszámértékek között futott.

8. ábra. Az Anti-Preferenciális Modell különböző E élszámok esetén megjelenő maximális fokszáma.
A szimuláció $E = 100$ és $E = 2000$ élszámértékek között futott.



9. ábra. A Véletlen Rekurzív Fa modell fokszámeloszlásának átlaga az E élszám függvényében. Minden átlag $n = 15$ db azonos élszámú futtatás eredményeiből lett számolva.