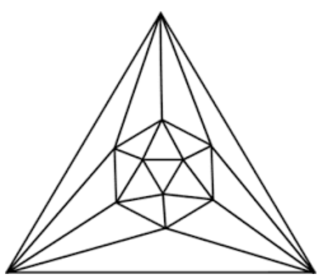
Pajek in Muha
(Pričakovan čas zadetka na razdaljno-regularnih grafih)

Andrej Jočić

1 UVOD

Za začetek predstavimo uganko s pajkovo mrežo, ki povezuje 12 vozlišč v obliki dvanajsterca (glej sliko 1). V enem vozlišču se je ujela muha, ki se ne more več premikati, v antipodnem vozlišču (tj. na maksimalni razdalji od nje) pa se nahaja slepi pajek brez spomina. Pajek se po mreži premika tako, da se v vsakem koraku naključno odloči, v katero sosednje vozlišče se bo podal (lahko tudi v tisto iz katerega je pravkar prišel). Zanima nas pričakovano število korakov, preden doseže muho (t.i. čas zadetka, angl. *hitting time*).



Slika 1. Planarna slika skeleta dvanajsterca

Problem bomo rešili za obsežno skupino grafov (kamor spada tudi zgornji skelet ikozaedra) in za poljubno začetno razdaljo med pajkom in muho.

V 2. delu je predstavljena ideja, kako bi se lahko lotili naloge s simulacijo in pri tem kontrolirali kvaliteto rezultatov. V 3. delu prevedemo problem v jezik verjetnosti, kar nas pripelje do Markovskih verig, ki so na kratko predstavljene v 4. delu. Nato v 5. delu posplošimo problem na razdaljno-regularne grafe, ki so opisani v 6. delu. V 7. delu je predstavljena metoda, s katero lahko najdemo približno rešitev problema na tem razredu grafov. To metodo najprej uporabimo na dvanajstercu, v 8. delu pa še na nekaterih drugih grafih. Na koncu je opisan še alternativni pristop, ki nam da natančno rešitev.

Biggs [1], Devroye in Sbihi [2], Van Slijpe [3] ter Jurišić [4] so sicer že našli učinkovitejše načine za izračun točne rešitve na takšnih grafih, ampak je tukaj predstavljena metoda precej bolj elementarna in poučna.

2 SIMULACIJA

Problema se lahko lotilmo s simulacijo velikega števila naključnih sprehodov in po tej poti izračunamo povprečje števila korakov potrebnih za slepega pajka, da pride do muhe. Zakon velikih števil nam (pod predpostavko končne variance v porazdelitvi dolžine pajkove poti) zagotavlja, da bo povprečje konvergiralo k iskani pričakovani vrednosti. Tu se pojavi vprašanje, koliko simulacij bomo izvedli. Namesto da si samo izmislimo neko veliko število, si lahko pomagamo z intervali zaupanja (približno) doseči neko željeno natančnost pri vnaprej določeni stopnji tveganja. Interval zaupanja za pričakovano vrednost μ normalno porazdeljenih podatkov pri neznani varianci in velikem številu podatkov n je določen s točkama

$$\overline{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}},$$

kjer je α stopnja tveganja, s standardni odklon vzorca, $z_{\alpha/2}$ pa $(1-\alpha/2)$-ni kvantil standardne normalne porazdelitve. Po centralnem limitnem izreku je ta interval dober približek za poljubno (v našem primeru neznano) porazdelitev, če predpostavimo končno varianco in je n dovolj velik. Recimo, da bomo izvajali simulacije, dokler se širina intervala ne skrči pod vrednost w. Ustavimo se, ko je n dovolj velik in

$$2z_{\alpha/2}\frac{s}{\sqrt{n}} < w \quad \text{oz.} \quad n > \left(\frac{2sz_{\alpha/2}}{w}\right)^2.$$

Pri tem moramo sproti računati s, kar lahko naredimo z rekurzivno formulo za s^2 in $\overline{x}$:

$$\begin{aligned}
&s_1 = 0\\
&s_n^2 = \left(1-\frac{1}{n-1}\right)s_{n-1}^2 + n\left(\overline{x}_n - \overline{x}_{n-1}\right)^2\\
&\overline{x}_1 = x_1\\
&\overline{x}_n = \frac{(n-1)\overline{x}_{n-1} + x_n}{n}.
\end{aligned}$$

Če si ne moremo privoščiti velikega števila simulacij (npr. če imamo opravka z zelo velikim grafom), bi porazdelitev rezultatov lahko na podlagi majhnega vzorca aproksimirali z metodo bootstrapping (glej npr. Chengov članek o analizi rezultatov simulacij [5]). Namesto tega bomo v tej seminarski nalogi poskusili priti do učinkovitejše in splošnejše rešitve.

3 PREVEDBA PROBLEMA

Naj bo X slučajna spremenljivka za čas zadetka, tj. meri število korakov pajka, preden prispe do muhe. Potem iščemo njeno pričakovano vrednost, ki je enaka

$$E(X) = \sum_{n=r}^{\infty} n \cdot P(X=n),$$

kjer je r začetna razdalja med pajkom in muho.

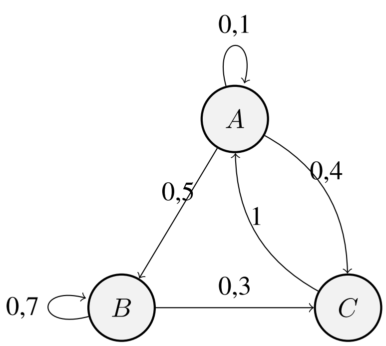
Pajek bo potreboval vsaj r korakov, zgornje meje za dolžino njegove poti pa v splošnem ni. Vrednost $P(X = n)$ je verjetnost, da pajek pride do muhe v natanko n korakih oz. da je tam po n korakih ampak še ni bil po $n-1$ (saj se ustavi, ko enkrat prispe do muhe). Drugače povedano:

$$P(X = n) = P(\text{po } n \text{ korakih je prispel do muhe}) - P(\text{po } n-1 \text{ korakih je na muhi}).$$

Potrebovali bomo verjetnost, da je pajek po določenem številu korakov na določenem vozlišču. Primerno orodje za ta namen so Markovske verige, opisane v naslednjem razdelku.

4 Markovske verige

Markovska veriga je matematični model za sistem, s katerim lahko opišemo zaporedje naključnih dogodkov pod pogojem, da je vsak dogodek odvisen samo od trenutnega stanja. Za nas so relevantni sistemi s končnim številom stanj, ki prehajajo skozi stanja v diskretnih časovnih korakih (sicer se s t.i. zveznimi Markovskimi verigami da modelirati še marsikaj drugega). Tak sistem lahko predstavimo z usmerjenim uteženim grafom, kjer vozlišča predstavljajo stanja in povezava iz vozlišča v v vozlišče u s težo t pomeni, da sistem gre iz stanja v v stanje u z verjetnostjo t.



Slika 2. Sistem s tremi stanji: A, B in C. Če je ta sistem v nekem časovnem koraku v stanju A, bo v naslednjem koraku z verjetnostjo 0,1 ostal v tem stanju, z verjetnostjo 0,5 bo šel v stanje B in z verjetnostjo 0,4 v stanje C (podobno za stanji B in C).
Dogodek je na vsakem koraku odvisen izključno od stanja, v katerem se sistem takrat nahaja in ne od stanj, v katerih je bil prej. Prehodi, ki niso eksplicitno narisani, imajo verjetnost 0, npr. prehoda $C \to B$ in $C \to C$ sta nemogoča, medtem ko je prehod $C \to A$ gotov dogodek.

Poglejmo si primer grafične predstavitve enostavnega sistema (slika 2). Ta sistem lahko opišemo tudi s *prehodno matriko* - poimenujmo jo T (angl. *transition*). Najprej stanjem priredimo zaporedna naravna števila, npr. $A \to 1$, $B \to 2$ in $C \to 3$. Potem sestavimo 3×3 razsežno matriko T, kjer je element T_{ij} enak verjetnosti prehoda iz i-tega v j-to stanje. V našem primeru imamo

$$T = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0 & 0.7 & 0.3 \\ 1 & 0 & 0 \end{bmatrix}.$$

V splošnem bo za sistem z n stanji ta matrika velika $n \times n$. Ker i-ta vrstica predstavlja verjetnosti popolnega sistema dogodkov, je vsota njenih elementov

$$\sum_{j=1}^{n} T_{i,j} = 1.$$

Taki matriki (torej kvadratni, ki ima vse vrstične vsote enake 1) pravimo *(desno) stohastična matrika*. To lastnost lahko zapišemo tudi kot zvezo $T\mathbf{1} = \mathbf{1}$, kjer je $\mathbf{1}$ vektor samih enic ustrezne dolžine.

Recimo, da nas v zgornjem primeru zanima verjetnost, da pridemo v dveh korakih iz stanja A v stanje C. Za to imamo 3 možnosti: $A \to A \to C$, $A \to B \to C$ in $A \to C \to C$. Verjetnost prve je enaka

$$P(A \to A \to C) = P(A \to A) \cdot P(A \to C) = T_{1,1} \cdot T_{1,3},$$

in podobno za ostali dve možnosti. Potem je verjetnost, ki jo iščemo, vsota verjetnosti teh treh poti:

$$P(A \xrightarrow{2} C) = T_{1,1} \cdot T_{1,3} + T_{1,2} \cdot T_{2,3} + T_{1,3} \cdot T_{3,3},$$

kar pa je ravno $(T^2)_{1,3}$. V splošnem je $(T^k)_{ij}$ enako verjetnosti prehoda iz i-tega v j-to stanje v natanko k korakih - označimo jo s $P(i \xrightarrow{k} j)$. To lahko pokažemo to z indukcijo na k z bazo $k = 1$. $(T^1)_{ij} = P(i \xrightarrow{1} j)$ očitno drži, saj smo tako definirali matriko T. Za indukcijski korak moramo pokazati $(T^{k+1})_{ij} = P(i \xrightarrow{k+1} j)$ pod predpostavko $(T^k)_{ij} = P(i \xrightarrow{k} j)$:

$$\begin{aligned} (T^{k+1})_{i,j} &= (T^k T)_{ij} && (1) \\ &= \sum_{\ell=1}^{n} (T^k)_{i\ell} T_{\ell j} && (2) \\ &= \sum_{\ell=1}^{n} P(i \xrightarrow{k} \ell) T_{\ell j} && (3) \\ &= P(i \xrightarrow{k+1} j) && (4) \end{aligned}$$

Pri enačaju v (2) smo uporabili definicijo produkta matrik, v (3) pa indukcijsko predpostavko. Izraz v (3) je vsota verjetnosti vseh (medsebojno nezdružljivih) načinov, kako priti iz i-tega v ℓ-to stanje v k korakih, potem pa v $(k+1)$-vem koraku direktno iz ℓ-tega v j-to stanje. To so natanko vsi načini, kako lahko pridemo iz i-tega v j-to stanje v $k+1$ korakih, torej velja tudi enačaj v (4).

Lahko se še prepričamo, da je tudi potenca T^k desnostohastična. V primeru $k = 2$: $T^2\mathbf{1} = T(T\mathbf{1}) = T\mathbf{1} = \mathbf{1}$, in podobno za $k > 2$.

Markovske verige bomo uporabili v razdelku 7, najprej pa znižamo zahtevnost problema tako, da se omejimo na določeno skupino grafov.

5 UČINKOVITEJŠI PRISTOP

Zdaj lahko sprehod pajka predstavimo kot Markovsko verigo, kjer vsako stanje predstavlja eno od n vozlišč grafa. Da dobimo iskano verjetnost, da je pajek po $k \in \mathbb{N}$ korakih pristal na muhi, bomo morali izračunati k-to potenco prehodne matrike. Matrično množenje je računsko precej zahtevno (vsaj $\approx O(n^{2.8074})$ z izjemo algoritmov, ki so hitrejši le za nepraktično velike matrike), zato ta pristop ni bistveno (če sploh) boljši od simulacije.

Bolje bi bilo, če bi lahko vozlišča razdelili na skupine, kjer ena skupina ustreza enemu stanju sistema. Če jih razdelimo glede na razdaljo od muhe, bomo lahko izračunali verjetnost, da je pajek po nekem številu korakov na določeni razdalji od muhe. To pa je v našem primeru vse kar potrebujemo, ker je "biti na muhi" ekvivalentno "biti na razdalji 0 od muhe". Ampak taka delitev vozlišč je smiselna samo, če poznamo verjetnosti prehodov med skupinami. To pa je možno le, če imajo vsa vozlišča na določeni razdalji enake verjetnosti prehodov. Z drugimi besedami: za vsako vozlišče na razdalji r obstajajo tri konstante, ki predstavljajo število sosedov na razdaljah $r-1$, r oz. $r+1$. To lastnost ima razred grafov, ki jim pravimo *razdaljno-regularizirani*. Godsil in Shawe-Taylor [6] sta pokazala, da so taki grafi dveh vrst:

- razdaljno-regularni, kjer so ta števila sosedov enaka ne glede na izbiro izhodiščnega vozlišča (v našem primeru lokacije muhe), in
- razdaljno-biregularni, kjer so vozlišča razdeljena v dve skupini, in so ta števila sosedov odvisna samo od tega, v kateri od dveh skupin je izhodiščno vozlišče.

Tukaj se bomo omejili na razdaljno-regularne grafe, ki jih je leta 1974 vpeljal Biggs [7]. V to skupino spada veliko grafov, ki se pojavijo npr. v teoriji kodiranja in drugih področjih teoretičnega računalništva. Med predstavniki so n-razsežne kocke, Hammingovi grafi (ki jih srečamo v teoriji kodiranja) in Johnsonovi grafi, ki se pojavijo v teoriji načrtov in končnih geometrijah. Za več primerov glej zbirko razdaljno-regularnih grafov [8], ki jo ureja Robert Bailey.

6 RAZDALJNO-REGULARNI GRAFI

Razdaljno-regularni grafi (angl. distance-regular graphs, DRGs) so natanko tisti neusmerjeni grafi brez zank, ki imajo sledečo lastnost. V njih lahko izberemo poljubno vozlišče v in opazujemo množice vozlišč na enaki razdalji od v (množici na razdalji r tudi rečemo $S_r(v)$ kot "sfera" z radijem r in središčem v). Potem mora za vsako sfero $S_i(v)$ ($0 \leq i \leq D$, kjer je D diameter grafa, tj. maksimalna razdalja med pari vozlišč v grafu) veljati naslednje:

- vsa vozlišča v sferi imajo enako število sosedov v tej sferi - recimo temu številu a_i,
- vsa vozlišča v sferi imajo enako število sosedov v naslednji sferi, tj. $S_{i+1}(v)$ - recimo temu številu b_i,
- vsa vozlišča v sferi imajo enako število sosedov v predhodni sferi, tj. $S_{i-1}(v)$ - recimo temu številu c_i,

in so zaporedja $\{a_i\}_{i=0}^D$, $\{b_i\}_{i=0}^D$ in $\{c_i\}_{i=0}^D$ neodvisna od izbire začetnega vozlišča v. To npr. pomeni, da je stopnja vozlišča v, ki jo prestavlja ravno parameter b_0, enaka za vsako vozlišče. Z drugimi besedami, graf mora biti regularen s stopnjo $k = b_0$. Za vsak i velja $a_i + b_i + c_i = k^{(\star)}$, ker vozlišče ne more imeti sosedov nikjer razen v svoji sferi ali eni od sosednjih. S temi tremi zaporedji lahko (ne nujno enolično) opišemo nek DRG, ampak eno od njih lahko zaradi ($\star$) kar spustimo - izberemo $\{a_i\}_{i=0}^D$. Tudi presečnega števila b_D ni potrebno navajati, ker je vedno enako 0 (vozlišče, ki je že na max. razdalji ne more imeti sosedov, ki so še dlje). Podobno je tudi $c_0 = 0$, saj je $S_{-1}(v)$ prazna množica in $a_0 = 0$, ker so vsi sosedi na razdalji 1. Prišli smo do *presečnega zaporedja* $\{b_0, ..., b_{D-1}; c_1, ..., c_D\}$, s katerim lahko izračunamo npr. vse lastne vrednosti grafa (in še marsikaj). Vanj je vključeno tudi število c_1, čeprav je vedno enako 1.

Če vemo, da je nek (dovolj majhen) graf razdaljno-regularen, lahko določimo njegovo presečno zaporedje na roke z naslednjim postopkom. Graf si narišemo, izberemo poljubno vozlišče v in vsa vozlišča označimo glede na razdaljo do v. Torej v označimo z 0, njegove sosede z 1, njihove (neoznačene) sosede z 2 itd. Potem število b_i ($0 \leq i \leq D-1$) razberemo tako, da pogledamo poljubno vozlišče označeno z i in preštejemo njegove sosede, ki so označeni z $i+1$. Podobno je c_i ($1 \leq i \leq D$) število sosedov, označenih z $i-1$. Če ne vemo, ali je graf sploh razdaljno-regularen, moramo pri določanju b_i pogledati vsa vozlišča, označena z i, in preveriti, da imajo vsa enako število sosedov z oznako $i+1$ (ter podobno za c_i). Postopek ponovitmo za vse možne izbire začetnega vozlišča v. Če za vsako vozlišče v dobimo ista presečna števila, je graf razdaljno-regularen.

7 REŠITEV Z MARKOVSKO VERIGO

Naključni sprehod pajka po razdaljno-regularnem grafu premera d lahko opišemo s sistemom, ki ima $d+1$ stanj - eno za vsako možno razdaljo med pajkom in muho. Opisali ga bomo z $(d+1) \times (d+1)$ razsežno prehodno matriko T, za kar moramo izračunati verjetnosti prehodov med stanji. Spomnimo se definicije presečnega zaporedja - tam z v označimo vozlišče, kjer se nahaja muha. Recimo, da se pajek nahaja na razdalji r od muhe. Potem ima v naslednjem koraku k možnosti, na katerega soseda se bo premaknil. Od teh je a_r takih, ki so tudi na razdalji r, b_r takih, ki so na razdalji $r+1$ in c_r takih, ki so na razdalji $r-1$. Naj bo D_i dogodek, da je pajek po premiku na razdalji i. Potem velja

$$P(D_r) = \frac{a_r}{k},\ P(D_{r+1}) = \frac{b_r}{k} \text{ in } P(D_{r-1}) = \frac{c_r}{k}.$$

Označimo stanje "pajek je na razdalji r od muhe" z r. Stolpce in vrstice prehodne matrike bomo označili od 0 naprej (namesto od 1, kot je v navadi). Verjetnosti za i-to vrstico prehodne matrike so potem: c_i/k v $(i-1)$-vem stolpcu, a_i/k v i-tem stolpcu in b_i/k v $(i+1)$-vem stolpcu (z izjemo robnih primerov, kjer $i = 0$ ali $i = d$). Povsod drugje je verjetnost enaka 0, ker se razdalja ne more v enem koraku spremeniti za več kot 1. Dobimo tridiagonalno prehodno matriko

$$T = \frac{1}{k}\begin{bmatrix} a_0 & b_0 & 0 & & \dots & & 0 \\ c_1 & a_1 & b_1 & & & & \\ 0 & c_2 & a_2 & b_2 & & & \vdots \\ & & & \ddots & & & \\ \vdots & & & c_{d-2} & a_{d-2} & b_{d-2} & 0 \\ & & & & c_{d-1} & a_{d-1} & b_{d-1} \\ 0 & & \dots & & 0 & c_d & a_d \end{bmatrix}.$$

Za potrebe našega vprašanja iz uvoda moramo prvo vrstico nekoliko spremeniti, ker je $i = 0$ končno stanje. Ko enkrat pajek pride do muhe, bo tam tudi ostal z verjetnostjo 1. Dobimo novo matriko

$$R = \frac{1}{k}\begin{bmatrix} k & 0 & 0 & & \dots & & 0 \\ c_1 & a_1 & b_1 & & & & \\ 0 & c_2 & a_2 & b_2 & & & \vdots \\ & & & \ddots & & & \\ \vdots & & & c_{d-2} & a_{d-2} & b_{d-2} & 0 \\ & & & & c_{d-1} & a_{d-1} & b_{d-1} \\ 0 & & \dots & & 0 & c_d & a_d \end{bmatrix}.$$

Naj bo K_i število korakov potrebnih, da iz stanja i prvič pridemo v stanje 0. Potem je $P(K_i = k)$ verjetnost, da smo po k korakih v stanju 0, po $k-1$ korakih pa še nismo bili tam, oz.

$$\begin{aligned} P(K_i = k) &= P(\text{po } k \text{ korakih smo v stanju } 0) - \\ &\quad P(\text{po } k-1 \text{ korakih smo v stanju } 0) \\ &= (R^k)_{i0} - (R^{k-1})_{i0}. \end{aligned}$$

Zdaj lahko izračunamo pričakovano vrednost K_i za poljuben i $(0 \le i \le d)$, torej imamo rešitev ne glede na začetno razdaljo med pajkom in muho. Rešitev naše uganke je $E(K_d)$, ker pajek začne na razdalji d. Torej moramo aproksimirati

$$\begin{aligned} E(K_d) &= \sum_{n=d}^{\infty} n \cdot P(K_d = n) \\ &= \sum_{n=d}^{\infty} n \cdot ((R^n)_{d0} - (R^{n-1})_{d0}). \end{aligned}$$

Recimo da za aproksimacijo seštejemo člene do $n = N$. Potem moramo $(N-1)$-krat množiti neko potenco R^k s prvotno matriko R. Množenje je izvedljivo v času $O(d^2)$, ker ima R veliko ničel, tako da ima izračun časovno zahtevnost $O(N \cdot d^2)$. Za primerjavo, metode omenjene v uvodu so izračunljive v času $O(d)$.

Končno se lahko lotimo uganke iz uvoda. Skelet dvanajsterca, ki je razdaljno-regularen, ima premer 3, tj. $d = 3$ in presečno zaporedje $\{5, 2, 1; 1, 2, 5\}$. Poglejmo, koliko členov moramo sešteti do konvergence:

N	$E(K_3)$
20	7.2065
50	13.6636
100	14.9568
150	14.9988
200	14.9999

Izgleda, da približek konvergira proti 15, kar je tudi prava rešitev (ki jo lahko dobimo npr. z metodo v razdelku 9).

8 Kocke in Hammingovi grafi

Poiščimo še rešitev za n-razsežne kocke, kjer je $n \in \mathbb{N}$ majhen. Te lahko narišemo tako, da za vozlišča vzamemo vse možne binarne n-terice in proglasimo vozlišči za sosednji, če se razlikujeta v natanko enem bitu. Za določanje presečnega zaporedja lahko za začetno vozlišče vzamemo tisto s samimi ničlami. Potem je razdalja vsake točke do začetne kar število enic v n-terici (t.i. *Hammingova razdalja*), in je diameter kocke enak n. Če smo na razdalji r, lahko eno od r enic spremenimo v ničlo in se tako približamo začetni točki - torej $\{c_i\}_{i=1}^{D} = \{1, \dots, n\}$. Ker z eno spremembo ne moremo ostati na isti razdalji $a_i = 0$ $(0 \le i \le D = n)$, zato $b_i = n - c_i$ (stopnja grafa je n). Dobili smo presečno zaporedje $\{n, n-1, \dots, 1; 1, 2, \dots, n\}$.

Poglejmo primer, ko pajek začne na vozlišču, ki je sosednje muhi za nekaj različnih n-kock.

n	$E(K_1)$ konvergira proti..
3	7
4	15
6	63
10	1023

Vidite vzorec? Izkaže se, da je v primeru n-kocke $E(K_1) = 2^n - 1$. Za izpeljavo glej [4], alternativna izpeljava je v (C. Godsil, personal communication, February 11, 2021).

Izračunajmo še $E(K_D) = E(K_4)$ za 4-kocko:

N	aproksimacija $E(K_4)$
100	20.79878
300	21.33331
500	21.33333

Do zdaj smo vedno dobili konvergenco proti naravnemu številu, ampak v resnici ni nobenega razloga, da bi točna rešitev bila celoštevilska. V tem konkretnem primeru izgleda, da je rešitev $21+\frac{1}{3}$, ampak v spošnem ne bo tako lahko prepoznati ulomka (rešitve so vedno racionalna števila - glej razdelek 9 ali [9], str. 96). Ampak enkrat bo računanje treba ustaviti. Recimo, da se ustavimo, ko se vsota ne poveča niti za ε v k zaporednih korakih (vsota strogo narašča, ker prištevamo produkt naravnega števila in verjetnosti). Kriterij začnemo preverjati šele, ko zaporedne verjetnosti začnejo padati (da se ne ustavimo prezgodaj, ko vsota lahko narašča zelo počasi). Intuitivno si lahko predstavljamo, da bodo verjetnosti striktno padale, ko začnemo prištevati $n \cdot P(K_i = n)$ za $n > E(K_i)$. Pri preverjanju padanja zaporedja verjetnosti moramo tudi ignorirati ničle - lahko se npr. zgodi, da pot določene dolžine sploh ne obstaja. V primeru kock se pajkova razdalja od muhe sigurno spremeni v vsakem koraku (ker $a_i = 0$), zato ne more priti do muhe v sodem številu korakov, če začne na lihi razdalji. Vrednosti k in ε bomo določili eksperimentalno. Sicer so tukaj možni bolj sofisticirani pristopi, ampak se v to ne bomo spuščali. Poskusimo $k = 10$ in $\varepsilon = 0.05$ za aproksimacijo $E(K_D)$ za nekaj kock:

n	$E(K_D)$	rel. napaka	N
5	42.47	0.440%	287
9	603.73	0.552%	4355
13	9015.93	0.567%	65929

Za izračun relativne napake približka smo seveda potrebovali točne rešitve, ki jih načeloma vnaprej ne poznamo.

Za konec poglejmo še bolj splošne grafe. *Hammingov graf* $H(n, q)$ dobimo tako, da za vozlišča vzamemo vse n-terice elementov množice velikosti q (pri $q = 2$ dobimo n-kocko) in proglasimo dve za sosednji, če sta na Hammingovi razdalji 1 (torej se razlikujeta v natanko enem mestu). Recimo, da za množico vzamemo $\mathbb{Z}_q$, torej cela števila od 0 do $q - 1$. Spet za določanje presečnega zaporedja za izhodišče vzamemo n-terico samih ničel. Zdaj je razdalja do izhodišča enaka številu neničelnih elementov n-terice. Če smo na razdalji r, imamo $n - r$ ničel, ki jih lahko spremenimo v eno od $q - 1$ drugih števil, da se oddaljimo od izhodišča, torej $b_r = (n - r)(q - 1)$. Prav tako imamo r neničelnih števil, ki jih lahko spremenimo v 0, da se mu pribižamo, zato $c_r = r$. Torej je presečno zaporedje enako $\{n(q - 1), (n-1)(q-1), \ldots, q-1; 1, 2, \ldots, n\}$. Poglejmo, kako natančne so aproksimacije pri $k = 20$ in $\varepsilon = 0.01$ za nekaj različnih Hammingovih grafov:

(n, q)	$E(K_D)$	rel. napaka	N
$(3, 3)$	32.9	0.036%	312
$(6, 3)$	826.9	0.053%	8117
$(4, 7)$	2556.6	0.054%	25259
$(5, 8)$	34054.9	0.055%	337008

9 Rešitev s sistemom enačb

Avtor: Aleksandar Jurišić

Pri vsakem koraku naključnega sprehoda, ko je pajek na razdalji i ($1 \leq i \leq D$), se lahko zgodi en od treh dogodkov:

1) A_i: pajek ostane na enaki razdalji od muhe,
2) B_i: pajek se premakne dlje od muhe,
3) C_i: pajek se premakne bližje muhi.

Naj bo X_i število korakov, če pajek začne na razdalji i od muhe. Označimo $E_i = E(X_i)$. Potem $E_0 = 0$, za ostale pa

$$E_i = 1 + E_{i-1} \cdot P(C_i) + E_i \cdot P(A_i) + E_{i+1} \cdot P(B_i),$$

kjer vzamemo $E_{D+1} = 0$. Konstanta 1 na začetku enačbe predstavlja en korak pajka. Če imamo presečno zaporedje grafa, lahko tudi izračunamo verjetnost dogodkov A_i, B_i, C_i. $P(A_i) = a_i/k$, ker ima pajek a_i od k možnosti, da izbere povezavo v isti sferi okrog muhe. Podobno $P(B_i) = b_i/k$ in $P(C_i) = c_i/k$. Splošna enačba za i ($1 \leq i \leq D$) torej postane

$$E_i = 1 + E_{i-1} \cdot \frac{c_i}{k} + E_i \cdot \frac{a_i}{k} + E_{i+1} \cdot \frac{b_i}{k}.$$

Imamo torej linearen sistem D neodvisnih enačb za D spremenljivk, ki je rešljiv s splošnim postopkom v času $O(D^3)$. Sicer se da izpeljati formulo za E_i, ki je izračunljiva že v času $O(D)$ (glej [4]).

Literatura

[1] N. Biggs, "Potential theory on distance-regular graphs," *Comb. Probab. Comput.*, no. 2, pp. 243–255, 1993.
[2] L. Devroye and A. Sbihi, "Random walks on highly symmetric graphs," *J. Theoret. Probab.*, no. 3, pp. 497–514, 1990.
[3] A. Van Slijpe, "Random walks on regular polyhedra and other distance-regular graphs," *Stat. Neerl.*, no. 38, pp. 273–292, 1984.
[4] A. Jurišić, "pajk+muha." [Online]. Available: https://ucilnica.fri.uni-lj.si/pluginfile.php/130283/mod_label/intro/pajk%2Bmuha.pdf?time=1608260723452
[5] R. Cheng, "Analysis of simulation output by resampling," *International Journal of Simulation*, vol. 1, no. 1, 2002.
[6] C. D. Godsil and J. Shawe-Taylor, "Distance-regularised graphs are distance-regular or distance-biregular," *Journal of Combinatorial Theory, Series B*, vol. 43, pp. 14–24, 1987.
[7] N. L. Biggs, *Algebraic Graph Theory*. Cambridge University Press, 1974.
[8] R. Bailey. [Online]. Available: https://www.distanceregular.org
[9] E. R. van Dam, J. H. Koolen, and H. Tanaka, "Distance-regular graphs," *Electonic Journal of Combinatorics*, 2016.
[10] E. W. Weisstein, "Sample variance computation." [Online]. Available: https://mathworld.wolfram.com/SampleVarianceComputation.html
[11] A. Jurišić, *Verjetnost in statistika (skripta)*. Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, 2021.