МАТЕМАТИЧКИ ОМНИБУС, 5 (2019), 155 – 167
http://im-pmf.weebly.com/matematicki-omnibus-kniga-5.html

ТЕОРИЈА НА ГРАФОВИ И ЗАДАЧАТА НА СОБИРАЊЕ КУПОНИ

Марко Димовски [1]

Гордана Николовска [2]

1. ВОВЕД

Графовите се прекрасна математичка апстракција. Овие математички структури се користат за моделирање на широк дијапазон реални ситуации во кои е потребно да се изразат релациите помеѓу објектите. Графот е севкупност од точки и линии кои поврзуваат парови од точките. Точките на графот се нарекуваат *темиња*, а линиите кои ги поврзуваат темињата се нарекуваат *ребра*. Темињата се поврзуваат врз основа на определена законитост за графот, а темињата кои се поврзани со ребро ги нарекуваме *соседи*. Теоријата на графови наоѓа широка примена во реалниот свет. Од низата примери, би ги издвоиле:

- поврзувањето со пријателите на социјалните мрежи (секој корисник е теме, а кога корисниците се поврзуваат, тие создаваат ребро, [7]);
- користење на GPS/Google Maps/Yahoo Maps (за пронаоѓање најкраток пат, минимална тура, [8]);
- за пребарување на Google (за да пребарувате веб-страници, каде што страниците на интернет се поврзани меѓусебно со хиперлинкови; секоја страница е теме и врската помеѓу две страници е ребро);
- анализи во биологијата и другите науки.

Во овој труд ќе анализираме една друга примена на графовите во реален пример. Нека темињата на графот се аналогија за бројот на видови купони во игрите од типот „Собери ги сите купони и победи“. Имено, ако имаме по неограничен број купони од конечен број категории и по случаен избор извлекуваме купони, тогаш изборот на секој нов купон претставува движење од темето кое ја претставува категоријата на која припаѓа претходно извлечениот купон до темето кое ја претставува категоријата на која припаѓа новоизвлечениот купон. Претпоставуваме дека по извлекувањето на купон од една категорија, можно е да се извлече купон од било која друга категорија (или повторно од истата) т.е. било кои две темиња се соседи, и при тоа со подеднаква веројатност при секое влечење може да се извлече купон од која било категорија. Ваквиот граф во кој секое теме е поврзано со сите други темиња се вика комплетен граф. Целта на задачата на собирање купони е да се собере барем по еден купон од секоја категорија, затоа, во продолжение, очекуваниот број на чекори кој е потребен за посетување на секое теме од графот барем по еднаш, при што изборот на секое соседно теме по претходното е случаен, ќе го поистоветиме со очекуваниот број на извлечени купони за собирање на барем по еден купон од сите категории, т.е. победа во играта на собирање купони.

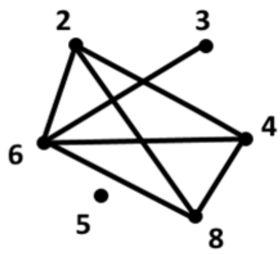
2. ПОИМИ ОД ТЕОРИЈА НА ГРАФОВИ

Граф. *Граф* е пар $G = (V, E)$ што се состои од конечно непразно множество V и множество E од двоелементни подмножества од V. Елементите на V се нарекуваат *темиња* на графот и најчесто ги означуваме со букви: $x, y, z, \dots$. Елементите на E ги нарекуваме *ребра* на графот и се означуваат со xy, што претставува ребро кое ги поврзува темињата x и y. За темињата x и y велиме дека се соседни ако постои ребро кое ги поврзува. Графот е *ненасочен* ако релацијата за соседност е симетрична. Ако ребрата имаат заедничко теме, велиме дека се *соседни ребра*. Графот со p темиња и q ребра го викаме (p, q) граф. Графот $(1,0)$, составен од една точка, се вика *тривијален граф*. За темињата кои се поврзани со ребро велиме дека е *инцидентни* со тоа ребро. *Насочен граф* е граф во кој секое ребро е подреден пар од двете темиња кои се негови крајни точки. Во продолжение, ќе се подразбира дека графот е ненасочен.

Степен на теме во граф. *Степен на теме* x во граф $G = (V, E)$ е бројот на темиња од G кои се соседни со x. Ознака: $deg(x)$ или $d(x)$.

Пример 1. Ќе разгледаме ненасочен граф со множество темиња $\{2,3,4,5,6,8\}$, каде што две темиња се соседни ако имаат заеднички делител различен од 1 (Слика 1).

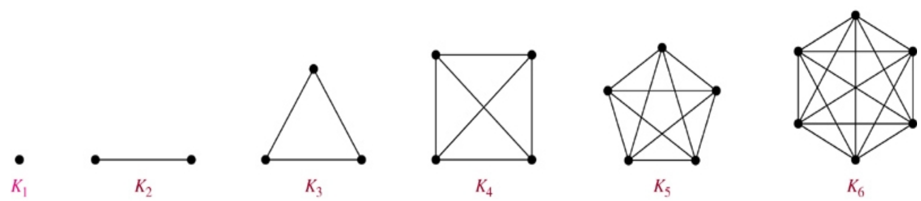
За степените на темињата во графот прикажан на Слика 1, имаме: d(6)=4, d(5)=0, d(8)=3, d(4)=3, d(3)=1 и d(2)=3.



Слика 1. Ненасочен граф со 6 темиња.

Регуларен граф. Граф во кој сите темиња имаат еднаков степен се вика *регуларен граф*.

Комплетен граф. Граф во кој степенот на секое теме е за еден помал од бројот на темиња во графот се вика *комплетен граф*. Тоа значи дека секое теме е поврзано со сите други темиња во тој граф. Нека $|E| = n$, т.е. n е бројот на темиња во графот $G = (V, E)$. Тогаш, за секое теме x е исполнето $d(x) = n - 1$, а вкупниот број на ребра во графот е бројот на комбинации без повторување од n елементи, класа 2 и изнесува $\frac{n(n-1)}{2}$. Комплетен граф кој има n темиња се означува со K_n.



Слика 2. Комплетни графови со различен број темиња.

Сврзан граф. Низа од ребра која поврзува низа од темиња во еден граф се вика *пат*. Графот $G = (V, E)$ е *сврзан граф*, ако секои две темиња од G се поврзани со пат.

Маршрута. *Маршрута* во графот $G = (V, E)$ е конечна наизменична низа од темиња и ребра $x_0, e_1, x_1, e_2, x_2, .., x_n$ која почнува и завршува со теме, каде што e_i е теме коешто ги поврзува темињата x_{i-1} и x_i, $i = \overline{1, n}$. За бројот n велиме дека е должина на маршрутата. Маршрутата е затворена ако $x_0 = x_n$. Маршрутата е *верига* ако ниту едно ребро не се повторува два или повеќе пати во низата $x_0, e_1, x_1, e_2, x_2, .., x_n$. Маршрутата е *циклус* ако е затворена верига.

Дрво. Графот во кој секои две темиња се поврзани со единствен пат се вика *дрво*. Со други зборови, секој сврзан граф кој не содржи циклуси, се вика дрво.

3. ЗАДАЧАТА НА СОБИРАЊЕ КУПОНИ

Постојат разни формулации на задачата на собирање купони, а ние тука ја даваме со следната формулација, [1]: Во секоја кутија со житарици има по еден купон кој носи одредена награда, така што постојат r различни типови награди. Еден човек купува по едно пакување од кутијата со житарици, сѐ додека не собере барем по еден награден купон од сите видови. Колку кутии со житарици треба да купи човекот?

Во случај кога од i-тиот вид има n_i купони, човекот треба да собере $n_1 + n_2 + \cdots + n_r - \min\{n_1, n_2, \dots, n_r\} + 1$. Но, во случај на бесконечно многу купони од секој вид, овој начин на размислување не одговара, т.е. може да резултира со купување на бесконечно многу кутии со житарици. Затоа се преминува на веројатносно решавање на задачата и се дава одговор на прашањето колку е очекуваниот број на кутии кои треба да ги купи човекот за да собере барем по еден купон од секој вид? Во теоријата на веројатност овој тип на задача се нарекува „Задача на собирање купони“ и со неа се опишани сите игри од типот „собери ги сите купони и победи“. Кон решавањето на овие задачи се пристапува на тој начин што правиме преформулирање на задачата и даваме одговор на прашањето „Кој е очекуваниот број на кутии кои треба да ги купи човекот, за својата колекција од $j - 1$ купон да ја зголеми за еден купон, односно да собере нов j-ти купон?“, [1]. При тоа, напоменуваме дека купоните не се индексирани и не се собираат според некој даден редослед. Напоменуваме дека изборот на кутија со житарици е случаен и подеднаква е веројатноста за избор на кутија од кој било вид.

Нека со N $(N \geq r)$ го означиме бројот на купени кутии, целта ни е да собереме барем по еден од секој од r- те различни наградни купони. Нека $Y_j,\ j = 1, \dots r$ е бројот на кутии кои човекот треба да ги купи за да ја зголеми својата колекција од $j - 1$, до колекција од j наградни купони од различен вид. Тогаш,

$$N = Y_1 + Y_2 + \cdots + Y_r. \tag{1}$$

Математичкото очекување на N, односно очекуваниот број на купени кутии потребни за собирање на r-те различни наградни купони, ќе го пресметаме како збир од математичките очекувања на Y_j за $j = 1, \dots, r$.

Да претпоставиме дека човекот има собрано $j-1$ различни купони и му преостанале уште $r-j+1$ купон за пополнување на колекцијата. Бројот на кутии коишто треба да ги купи со цел да собере нов купон, различен од оние кои ги поседува, има веројатносна распределба

$$P\{Y_j = k\} = (1-p_j)^{k-1} p_j, \qquad k = 1,2, \dots,$$

каде што p_j претставува веројатноста за добивање на нов купон во купената кутија. Со оглед на тоа што човекот веќе собрал $j-1$ купони од вкупно r различни видови на купони, од класичната дефиниција на веројатност имаме дека

$$p_j = \frac{r-(j-1)}{r}.$$

Всушност, заклучуваме дека случајните променливи Y_j, $j = 1, \dots, r$ имаат геометриска распределба, $Y_j \sim Geo(p_j)$. Бидејќи, математичкото очекување на случајна променлива која има геометриска распределба е еднакво на реципрочната вредност од веројатноста за успех при еден независен експеримент, заклучуваме дека

$$E(N) = E\left(\sum_{j=1}^{r} Y_j\right) = \sum_{j=1}^{r} E(Y_j) = \sum_{j=1}^{r} \frac{1}{p_j} =$$

$$= \sum_{j=1}^{r} \frac{r}{r-j+1} = r\left(\frac{1}{r} + \frac{1}{r-1} + \dots + 1\right). \qquad (2)$$

При доволно големи вредности на r, хармонискиот ред добиен на десната страна од равенството (2) се апроксимира со $\ln r + \gamma$, каде што $\gamma \approx 0{,}57722$ е Ојлеровата константа, [1]. Па така, за очекуваната вредност на бројот на купени кутии, сѐ додека не биде собрана целата колекција од r различни наградни купони, добиваме дека $E(N) \approx r(\ln r + \gamma)$, за доволно големи вредности на r.

Ќе ја најдеме и распределбата на веројатности на N. Да го разгледаме експериментот во кој се фрла хомогена r –страна коцка и се забележуваат бројот на паднатите точки на коцката, [1]. Го дефинираме настанот A – „секоја страна на коцката се паднала барем еднаш при n-те повторувања на експериментот". Нека A_i е настанот „i точки не се паднале ниту еднаш при n-те изведени експерименти", $i = 1, \dots, r$. Тогаш,

$$P(A_i) = \left(1 - \frac{1}{r}\right)^n.$$

Јасно е дека

$$A = A_1^c \cap A_2^c \cap \dots \cap A_r^c = (A_1 \cup A_2 \cup \dots \cup A_r)^c,$$

каде што A_i^c е спротивниот настан на настанот A_i, $i = 1, \dots, r$. Аналогно, веројатноста да не се паднале $i_1, i_2, \dots, i_j$ точки при n-те изведени експерименти ќе биде

$$P\left(A_{i_1} \cap \dots \cap A_{i_j}\right) = \left(1 - \frac{\mathrm{j}}{r}\right)^n.$$

Со помош на теоремата на вклучување и исклучување, за веројатноста на унијата на настаните $A_i, i = 1, \dots, r$ ќе добиеме дека е

$$P(A_1 \cup \dots \cup A_r) = S_1 - S_2 + \dots + (-1)^{r-1} S_r,$$

каде што

$$S_j = \sum_{i_1 < \dots < i_j} P\left(A_{i_1} \cap \dots \cap A_{i_j}\right) = \binom{r}{j}\left(1 - \frac{\mathrm{j}}{r}\right)^n.$$

Односно, имаме дека

$$P(A_1 \cup \dots \cup A_r) = \sum_{j=1}^{r} (-1)^{j-1} \binom{r}{j}\left(1 - \frac{\mathrm{j}}{r}\right)^n.$$

За настанот A , можеме да искористиме дека

$$P(A) = 1 - P(A_1 \cup \dots \cup A_r) = \sum_{j=0}^{r} (-1)^{j} \binom{r}{j}\left(1 - \frac{\mathrm{j}}{r}\right)^n. \qquad (3)$$

Во нашата задача, веројатноста $P\{N \leq n\}$ можеме да ја разгледаме како веројатноста на настанот „во n кутии, r-те купони ќе се појават барем по еднаш секој" и оваа веројатност е еднаква со веројатноста (3), т.е. за распределбата на веројатности на N имаме

$$P\{N \leq n\} = \textstyle\sum_{j=0}^{r} (-1)^j \binom{r}{j}\left(1 - \frac{\mathrm{j}}{r}\right)^n.$$

4. ВРЕМЕ НА ЦЕЛОСНО ПОКРИВАЊЕ НА КОМПЛЕТЕН ГРАФ

Нека е даден комплетен граф $G = (V, E)$ и v е едно негово произволно теме. Ова теме го земаме за почеток на низа од темиња во која втор член е било кој од соседите на темето v, одбран по случаен пат при што со еднаква веројатност може да се одбере било кое теме. Трет член е било кој од соседите на второто теме и така натаму, секој следен член во низата е случајно одбран сосед на претходниот. Добиената низа од темиња, одбрани на овој начин, претставува *случајно талкање по темињата на графот*, [2]. Според класичната дефиниција на веројатност, веројатноста за одбирање на било кој сосед во секој следен чекор е еднаква на реципрочната вредност од степенот на неговиот претходник. Талкањето трае онолку колку што е потребно. Очекуваниот број на чекори кој е потребен за по пат на случајно талкање по темињата на графот, секое теме да биде посетено барем еднаш, се вика *очекувано време на целосно покривање на граф*. Ако имаме граф $G = (V, E)$ и случајното талкање го започнуваме во темето v, очекуваното време на целосно покривање на графот го означуваме со $E_v(G)$.

Во продолжение се дадени неколку ограничувања за времето на целосно покривање на граф, кои датираат уште од седумдесетите години од минатиот век [2], [4]:

- За очекуваното време на целосно покривање на граф G и при случајно талкање со почеток во темето v е исполнето:
$$E_v(G) \leq 2|V||E|.$$
- Нека d_{min} е најмалиот степен на теме во графот G. Тогаш:
$$E_v(G) \leq 16\frac{|V||E|}{d_{min}}.$$
- За секое теме v, на регуларен граф G, е исполнето и неравенството $E_v(G) \leq O(|V|^2)$. Ова значи дека очекуваниот број на чекори за покривање на регуларен граф, почнувајќи од темето v, ќе биде помал или еднаков од производот на некој позитивен реален број и квадратот од бројот на темиња во графот.
- Нека G е дрво со n темиња и v е негово произволно теме. Тогаш: $E_v(G) \geq n \ln n - O(n)$.

Да разгледаме комплетен граф со n темиња, т.е. граф во кој секое теме е сврзано со останатите темиња во графот. Очекуваното време за целосно покривање на комплетниот граф може да го поистоветиме со решението на задачата на собирање купони, [2]. Имено, при секое купување на кутија со житарици, секој купон може да биде следен, независно кој бил неговиот претходник, исто како што од било кое теме на комплетниот граф можеме со случајно талкање да се префрлиме на било кое друго или истото теме. Разликата е во тоа што при купувањето кутии, може да ни се падне истиот купон како претходно собраниот, додека при преминувањето на друго теме кај графот, не можеме да „преминеме“ на истото, бидејќи комплетниот граф нема јазли. Но, при доволно големи вредности за n, приближувањето ќе биде еднакво, бидејќи асимптотските резултати се совпаѓаат.

Нека K_n е комплетен граф со n темиња. Темето во коешто го почнуваме случајното талкање е произволно, а во задачата на собирање купони, тоа е претставник од првата класа купони. Веројатноста да се посети претходно непосетено теме во талкањето, откако се посетени i темиња, е еднаква на $p_i = \frac{n-i}{n-1}$. Математичкото очекување на случајната променлива Y_i, која го опишува бројот на чекори кој е потребен за да од i посетени темиња, да преминеме на $i+1$ посетено теме е еднакво на $E(Y_i) = \frac{1}{p_i} = \frac{n-1}{n-i}$. Следствено,

$$E_v(K_n) = \sum_{i=1}^{n-1} \frac{n-1}{n-i} = (n-1)\left(1 + \frac{1}{2} + \cdots + \frac{1}{n-1}\right). \qquad (4)$$

Ако бројот на темиња n е доволно голем, времето на целосно покривање на графот е приближно еднакво на $n \cdot (\ln n + \gamma)$, т.е.

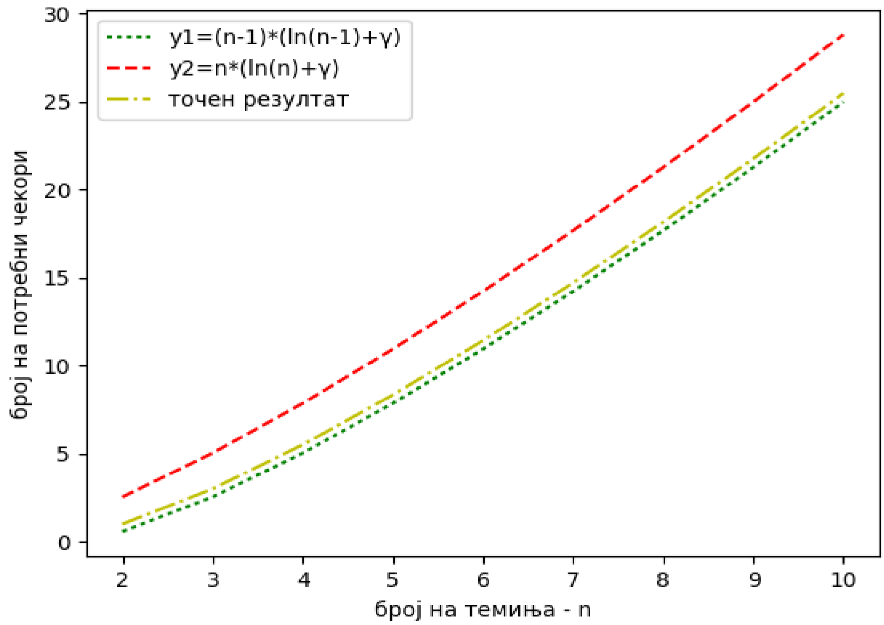
$$E_v(K_n) \approx n \cdot (\ln n + \gamma),$$

каде што γ е Ојлеровата константа.

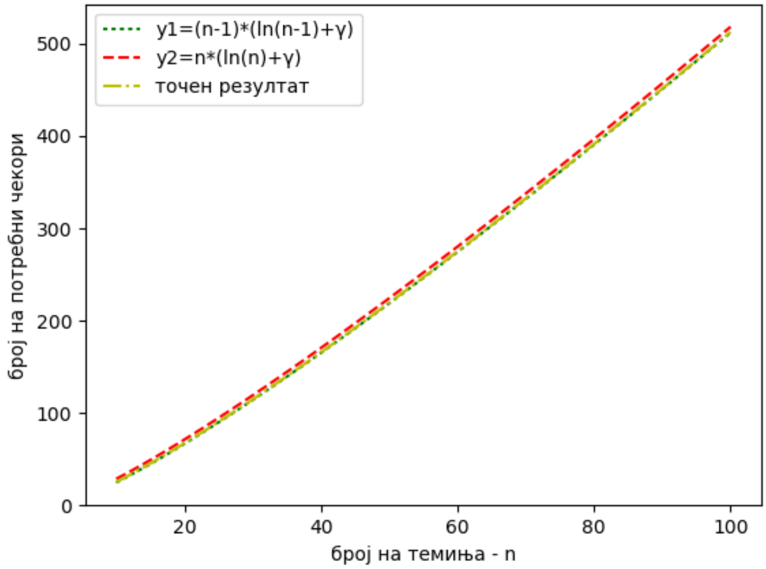
5. СИМУЛАЦИИ

Прво ќе дадеме графички приказ на очекуваниот број на чекори потребни за целосно покривање на комплетен граф со n темиња, каде што $2 \leq n \leq 10000$.

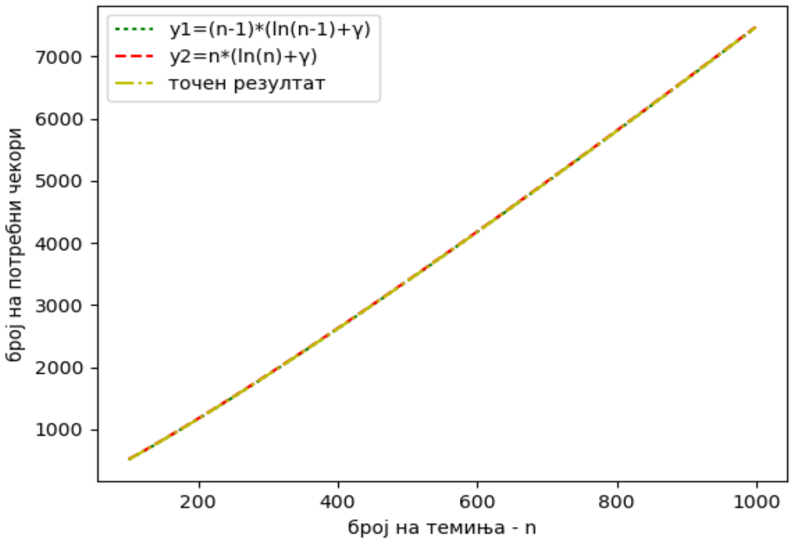
На Слика 3, Слика 4 и Слика 5, се прикажани точните вредности за очекуваното време за целосно покривање на комплетен граф со n темиња, добиени од (4), како и графиците на функцијата $f(n) = n \cdot (\ln n + \gamma)$ и функцијата $g(n) = (n-1) \cdot (\ln(n-1) + \gamma)$, за кои укажавме дека при големи вредности на n претставуваат добра апроксимација на десната страна на (4). При тоа, разгледани се три различни случаи во однос на опсегот на вредности на n. Мала вредност на n $(n \leq 10)$, средни вредности $(10 < n \leq 100)$ и големи вредности во случајот кога $100 < n \leq 1000$.



Слика 3. Графички приказ на функцијата $\boldsymbol{f(n) = n(\ln n + \gamma)}$ и функцијата $\boldsymbol{g(n) = (n-1)(\ln(n-1) + \gamma)}$, како и потребните чекори за целосно покривање на комплетен граф добиени од формула (4) при мали вредности на $\boldsymbol{n}$.

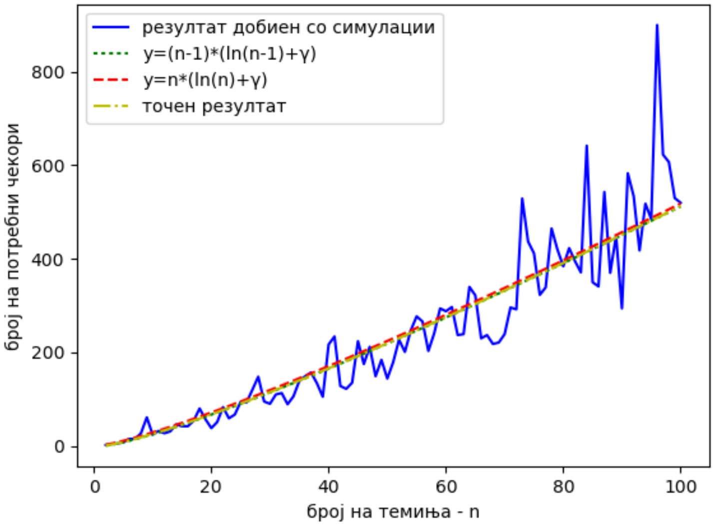


Слика 4. Графички приказ на функцијата $\boldsymbol{f(n) = n(\ln n + \gamma)}$ и функцијата $\boldsymbol{g(n) = (n-1)(\ln(n-1) + \gamma)}$, како и потребните чекори за целосно покривање на комплетен граф добиени од формула (4) при средни вредности на $\boldsymbol{n}$.



Слика 5. Графички приказ на функцијата $\boldsymbol{f(n) = n(\ln n + \gamma)}$ и функцијата $\boldsymbol{g(n) = (n-1)(\ln(n-1) + \gamma)}$, како и потребните чекори за целосно покривање на комплетен граф добиени од формула (4) при големи вредности на $\boldsymbol{n}$.

Извршивме и симулација (во Python) на бројот на чекори, потребни за да се поминат сите темиња на комплетен граф барем по еднаш, при случајно талкање по графот. Имено, за различни вредности на n, се формира низа од првите n природни броеви, кои ги претставуваат купоните кои се собираат. На случаен начин, со помош на рамномерна распределба, во секој чекор се одбира нов број меѓу 1 и n и процесот се повторува додека секој од броевите не се падне барем еднаш. На Слика 6 додадена е и кривата која го опишува симулираниот број на чекори, потребни за целосно покривање на комплетен граф.



Слика 6. Графички приказ на функцијата $\boldsymbol{f(n) = n(\ln n + \gamma)}$ и функцијата $\boldsymbol{g(n) = (n-1)(\ln(n-1) + \gamma)}$, точниот број на чекори од формулата (4) и кривата која го опишува симулираниот број на чекори, потребни за целосно покривање на комплетен граф при средни вредности на n.

Ограничувањата кои важат во теоријата на графовите можат да се применат и на проблемот со купоните, токму поради еднородноста на задачата на собирање купони и очекуваното време на целосно покривање на комплетен граф.

6. ПОВЕЌЕКРАТНО ПОКРИВАЊЕ НА КОМПЛЕТЕН ГРАФ

На сличен начин се пресметува и времето за целосно покривање на комплетен граф, доколку секое теме треба да биде посетено најмалку m пати. Ова *повеќекратно покривање на комплетен граф,* при случајно талкање по темињата на графот, се поистоветува со уште една позната задача, задачата на „*двојниот Дикси куп*“ (Double Dixie cup, [5]). Задачата на двоен Дикси куп е само една задача од типот на задачата на собирање купони. Во [5] е покажано дека очекуваниот број на купени пакетчиња (во кои има по една сликичка) за собирање на сите n сликички, кои се дел од задачата на двојниот Дикси куп, барем m пати секоја, се пресметува како

$$E_m(n) = n \int_0^\infty \left(1 - \left(1 - \sum_{j=0}^{m-1} \frac{e^{-t} t^j}{j!} \right)^n \right) dt.$$

За фиксно m, асимптотскиот резултат кога $n \to \infty$, за очекуваниот број на купени пакетчиња е еднаков на

$$E_m(n) = n(\ln n + (m-1)\ln(\ln n) + C_m + o(1)),$$

при што велиме дека $f(n) = o(1)$, ако $\frac{f(n)}{k} \to 0$, кога $n \to \infty$, за секоја константа k. Во ова равенство, $C_m = \gamma - \ln\big((m-1)!\big)$, каде што γ е Ојлеровата константа. Поради еднородноста на повеќекратното покривање на комплетен граф и задачата на двојниот Дикси куп, асимптотскиот резултат добиен во задачата на двојниот Дикси куп е добра апроксимација и за очекуваното време на повеќекратно целосно покривање на комплетниот граф.

Постојат и други варијанти на дефинираност на задачата на собирање купони. Не секогаш изборот на следната кутија е еднакво веројатен. Исто така, може да се случи некогаш купоните да доаѓаат во групи, како што е случајот во кој собираме албуми со сликички, при што често пати во едно пакетче има барем 5 сликички. Во овие случаи, формулата за очекуваниот број на купени кутии/пакетчиња е поинаква, [3]. Покрај теоријата на графови која може да се искористи при решавање на оваа задача, кон задачата на собирање на купони може да се пристапи и со теоријата на случајни процеси, [6].

ЛИТЕРАТУРА

[1] G. Bloom, L.Holst, D.Sandell, *Problems and snapshots from the world of probability*, Springer-Verlag New York, 1994.

[2] U. Feige, *Collecting coupons on trees, and the cover time of random walks*, Journal of Theoretical Probability, Comput Complexity, December 1996, Volume 6, Issue 4, 341 – 356.

[3] M. Ferrante, M. Saltalamacchia, *The Coupon Collector's Problem*, MATerials MATemàtics, Vol. 2014, No. 2, 35 pp.

[4] J.D. Kahn, N. Linial, N. Nisan, M.E. Saks, *On the cover time of random walks on graphs*, Journal of Theoretical Probability, Vol. 2, No. 1, 1989, 121 – 128.

[5] D.J. Newman, *The Double Dixie Cup Problem*, The American Mathematical Monthly, Vol. 67, No. 1 (Jan., 1960), 58-61.

[6] J. Pitman, *Combinatorial Stochastic Processes*, University of California, Berkeley, 2016.

[7] А. Велкоска, *Математиката на социјалните мрежи*, Математички омнибус 1 (2017), 89 – 99.

[8] И. Стојковска, *Некои алгоритми за решавање на задачата на патувачкиот трговец*, Математички омнибус 1 (2017), 101–113.

[1] Универзитет „Св. Кирил и Методиј“, Скопје
Природно – математички факултет
Архимедова 3, 1000 Скопје, Р. Македонија
e-mail: mdimovski16@gmail.com

[2] СУГС „Георги Димитров“,
Варшавска 3, 1000 Скопје, Р. Македонија
e-mail: gordana.nikolovska.93@gmail.com

Примен: 5.03.2019
Поправен: 6.06.2019
Одобрен: 9.06.2019
Објавен на интернет: 11.06.2019