

ZÁPADOČESKÁ UNIVERZITA V PLZNI
FAKULTA APLIKOVANÝCH VĚD
KATEDRA MATEMATIKY

NÁHODNÉ PROCHÁZKY NA GRAFU

BAKALÁŘSKÁ PRÁCE

Prohlášení

Předkládám tímto k posouzení a obhajobě bakalářskou práci zpracovanou na závěr bakalářského studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím odborné literatury, jejíž úplný seznam je její součástí.

V Plzni, dne 23. května 2022

.....

vlastnoruční podpis

Poděkování

Děkuji RNDr. Vladimírovi Švíglerovi, Ph.D. za vstřícnost, trpělivost a odborné rady při konzultacích a psaní této práce.

Abstrakt

V této práci se zabýváme očekávaným časem pokrytí náhodné procházky na grafu. Očekávaný čas pokrytí chápeme jako průměrný počet kroků potřebných k navštívení všech vrcholů daného grafu. Ukážeme si odvození očekávaného času pokrytí náhodné procházky na úplném grafu a cestě. Práce je doplněna numerickými výpočty očekávaného času pokrytí pro konkrétní typy grafů a nastínění vlivu minimálního stupně grafu na očekávaný čas pokrytí grafu.

Klíčová slova: náhodná procházka, graf, čas pokrytí, očekávaný čas pokrytí.

Abstract

In the thesis, we study expected cover time of a random walk on a graph. We consider expected cover time as an average number of steps which takes to visit all vertices of the graph. We will show the derivation of an expected cover time formula for the complete graph and the path graph. Furthermore, there are numerical calculations of expected cover time of the random walk on several types of graphs. We will use the numerical calculations to show the relation between the expected cover time and the minimal degree of graph.

Keywords: random walk, graph, cover time, expected cover time.

Obsah

1	Úvod	1
2	Základní pojmy	2
2.1	Teorie grafů	2
2.2	Náhodné procesy a pravděpodobnost	4
3	Očekávaný čas pokrytí úplného grafu a cesty	10
3.1	Metoda odvození	10
3.2	Úplný graf	12
3.3	Cesta	16
3.3.1	Cesta se začátkem na kraji	16
3.3.2	Cesta nezačínající na kraji	18
4	Numerické výpočty	23
4.1	Metoda numerického výpočtu očekávaného času pokrytí	23
4.2	Závislost očekávaného času pokrytí na minimálním stupni grafu	24
4.2.1	Konstrukce posloupnosti grafů	25
4.2.2	Výsledky	26
5	Závěr	28

Kapitola 1

Úvod

Náhodné procházky na grafech našly široké uplatnění v mnoha vědních oborech, jako je fyzika, ekonomie nebo biologie. Dále se náhodné procházky využívají při zkoumání dopravních infrastruktur, sociálních sítí nebo internetu. Náhodná procházka na grafech se používá ke zkoumání reálných struktur, které jsou rozsáhlé a jejichž strukturu lze popsat pomocí grafu.

Jednou z aplikací náhodné procházky při zkoumání je například odhalování komunit v rámci rozsáhlé sítě. Zde se používají tzv. krátké náhodné procházky, které mají tendenci zůstat v komunitách [5].

V této práci se zaměříme na očekávaný čas pokrytí náhodné procházky na grafu a jeho odvození pro úplný graf a cestu. Základní pojmy a definice nutné pro tato odvození jsou uvedeny v kapitole 2, která je rozdělena na dvě části. V první podkapitole zavádíme definice z teorie grafů a druhá podkapitola se věnuje definicím a pojmům z teorie náhodných procesů.

V kapitole 3 se věnujeme metodě odvození očekávaného času pokrytí náhodné procházky, která využívá indukovaný Markovův řetězec, který sleduje množinu již pokrytých vrcholů. Následně si ukážeme odvození očekávaného času pokrytí náhodné procházky na úplném grafu. V tomto případě si uvedeme dva způsoby, jak postupovat. První způsob vychází ze specifické struktury úplného grafu, ve druhém způsobu využíváme postup odvozený z indukovaného Markovova řetězce. Dále si ukážeme odvození očekávaného času pokrytí náhodné procházky na cestě, i zde využijeme postup využívající indukovaný Markovův řetězec. Na rozdíl od úplného grafu bude očekávaný čas pokrytí náhodné procházky na cestě závislý na volbě počátečního vrcholu náhodné procházky.

V kapitole 4 uvedeme numerické výpočty očekávaného času pokrytí náhodné procházky na specifických grafech (úplný graf, cesta, kružnice a wheel graf). Na závěr si nastíníme vztah minimálního stupně grafu a očekávaného času pokrytí náhodné procházky na grafu, vycházející z článku [3].

Kapitola 2

Základní pojmy

V práci budeme používat poznatky zejména ze dvou oblastí matematiky, a to teorie grafů a teorie náhodných procesů. Tato kapitola slouží jako stručný přehled definic a vět, které využíváme v textu.

2.1 Teorie grafů

Práce pojednává o náhodných procházkách na grafech, proto se v této podkapitole věnujeme definicím a pojmům týkajících se grafů. Budeme zde vycházet zejména z materiálu [2].

Definice 2.1.1 ([2], Def. 2.1.1). Graf G je uspořádaná dvojice množin V a H , kde V je konečná množina. Pokud $H \subseteq \binom{V}{2}$, potom graf G nazveme neorientovaným (prostým) grafem, a pokud $H \subseteq V \times V$, pak G nazveme orientovaným grafem.

V reálném světě lze pomocí grafu reprezentovat počítačové, sociální nebo biologické sítě. V praxi jsou tyto struktury velice rozsáhlé, a proto připadá v úvahu zkoumat pouze části těchto sítí. Jako sociální síť chápeme společenskou strukturu jedinců, skupin nebo organizací, které znázorňujeme pomocí vrcholů. Vzájemné interakce, přátelství nebo výměnu informací znázorňujeme hranou.

Definice 2.1.2 ([2], Def. 2.1.10). Necht' G_1, G_2 jsou grafy, potom G_1 je podgrafem grafu G_2 (značíme $G_1 \subseteq G_2$), jestliže $V(G_1) \subseteq V(G_2)$ a současně $H(G_1) \subseteq H(G_2)$.

V nadcházejících definicích se zabýváme okolím vrcholů. V sociálních sítích lze sousedy chápat například jako jedince, se kterými jsme ve vzájemné interakci.

Definice 2.1.3 ([2], Def. 2.2.1.). Necht' G je neorientovaný graf a $v \in V(G)$, potom číslo

$$d_G(v) = |\{u \in V(G) : \{v, u\} \in H(G)\}|$$

se nazývá stupeň vrcholu v v grafu G .

Definice 2.1.4 ([2], Def. 2.2.10). Necht' G je neorientovaný graf. Potom číslo $d_{\min}(G) = \min_{v \in V(G)} \{d_G(v)\}$ nazveme minimálním stupněm grafu G a $d_{\max}(G) = \max_{v \in V(G)} \{d_G(v)\}$ nazveme maximálním stupněm grafu G .

Interakci vrcholů lze reprezentovat pomocí matice sousednosti, ve které sledujeme to, zda spolu dva vrcholy sousedí nebo ne. Vztah vrcholů i a j udává prvek matice na pozici i, j . V případě vzájemné interakce je na dané pozici 1 jinak 0.

Definice 2.1.5 ([2], Def. 2.7.1). Necht' $G = (V, H)$ je neorientovaný graf, potom matice sousednosti je čtvercová matice $A = (a_{ij})_{|V| \times |V|}$ jejíž členy jsou dány předpisem

$$a_{ij} = \begin{cases} 1 & \{i, j\} \in H(G), \\ 0 & \text{jinak.} \end{cases}$$

Definice 2.1.6 ([2], Def. 2.1.4). Necht' G_1 a G_2 jsou neorientované grafy. Zobrazení $f : V(G_1) \rightarrow V(G_2)$ nazveme homomorfismus grafu G_1 a G_2 , pokud platí

$$\{x, y\} \in H(G_1) \Rightarrow \{f(x), f(y)\} \in H(G_2).$$

Definice 2.1.7 ([2], Def. 1.1.11). Zobrazení $f : X \rightarrow Y$ je **prosté**, pokud $x_1, x_2 \in X, x_1 \neq x_2, f(x_1) \neq f(x_2)$. Zobrazení f je z množiny X **na** množinu Y , pokud

$$\forall y \in Y \exists x \in X : f(x) = y.$$

Zobrazení f je vzájemně jednoznačné (bijekcí), pokud je zároveň **prosté a na**.

Definice 2.1.8 ([2], Def 2.1.6). Izomorfismus grafu G do grafu G' je bijekce $f : V(G) \rightarrow V(G')$ a zároveň platí, že f, f^{-1} jsou homomorfismy. Izomorfismus grafu G do sebe sama nazýváme automorfismus.

Poznámka 2.1.9. V rámci práce budeme zkoumat specifické typy grafů např. cesta, úplný graf, kružnice a wheel graf.

- Necht' $G = (V, H)$ je graf, ve kterém jsou vrcholy spojeny vždy s předchozím a následujícím vrcholem. Takovýto graf nazýváme (neorientovanou) cestou a značíme jej \mathcal{P}_N , kde N je počet vrcholů, tj.

$$\mathcal{P}_N = (\{1, 2, \dots, N\}, \{\{i, i+1\}, i = 1, \dots, N-1\}).$$

- Necht' $G = (V, H)$ je graf, ve kterém jsou každé dva různé vrcholy spojené hranou. Takovýto graf nazýváme úplným grafem a značíme jej \mathcal{K}_N , kde N je počet vrcholů, tj.

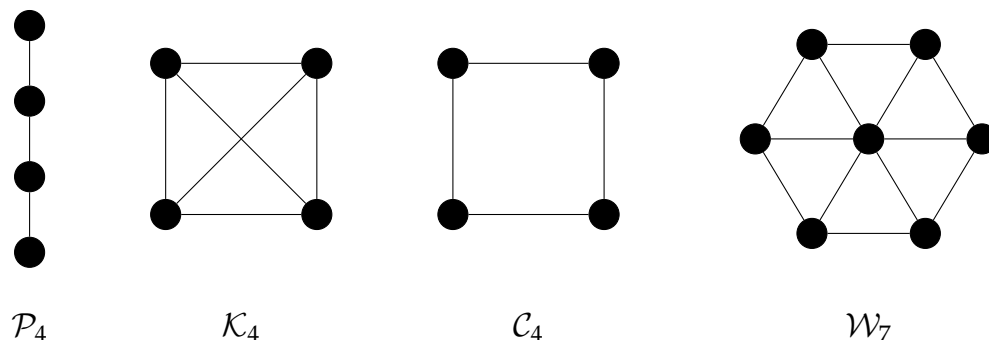
$$\mathcal{K}_N = \left(\{1, 2, \dots, N\}, \binom{\{1, 2, \dots, N\}}{2} \right).$$

- Necht' $G = (V, H)$ je graf, ve kterém jsou vrcholy spojené vždy s předchozím a následujícím vrcholem a poslední vrchol je spojen s prvním. Takovýto graf nazýváme kružnicí a značíme jej \mathcal{C}_N , kde N je počet vrcholů, tj.

$$\mathcal{C}_N = (\{1, 2, \dots, N\}, \{\{i, i+1\}, i = 1, \dots, N-1\} \cup \{\{N, 1\}\}).$$

- Necht' $G = (V, H)$ je graf, který je tvořen kružnicí \mathcal{C}_{N-1} a vrcholem spojeným se všemi vrcholy této kružnice. Takovýto graf nazýváme wheel graf a značíme jej \mathcal{W}_N , kde N je počet vrcholů.

Tyto specifické grafy jsou uvedeny na obrázku 2.1.



Obrázek 2.1: Přehled specifických grafů uvedených v poznámce 2.1.9. Na obrázku zleva cesta, úplný graf, kružnice a wheel graf.

Poznámka 2.1.10. Uvažujme rotaci vrcholů grafu

$$(v_1, v_2 \dots, v_{N-1}, v_N) \longrightarrow (v_N, v_1 \dots, v_{N-2}, v_{N-1})$$

a reflexi vrcholů grafu

$$(v_1, v_2 \dots, v_{N-1}, v_N) \longrightarrow (v_N, v_{N-1} \dots, v_2, v_1),$$

potom pro specifické grafy platí:

- libovolná permutace vrcholů úplného grafů \mathcal{K}_N je automorfismus,
- libovolná rotace a reflexe vrcholů kružnice \mathcal{C}_N je automorfismus,
- libovolná reflexe vrcholů cesty \mathcal{P}_N je automorfismus.

2.2 Náhodné procesy a pravděpodobnost

V této podkapitole se budeme zabývat základními pojmy z oblasti teorie náhodných procesů a teorie pravděpodobnosti. Pojmy jsou převzaty z materiálů [1], [6], [7] a [8].

Definice 2.2.1 ([1], §1.1). Neprázdnou množinu všech výsledků náhodného pokusu značíme Ω . Náhodným jevem nazveme jakoukoli podmnožinu množiny Ω .

Jako náhodný pokus označujeme pokus, jehož výsledek není jednoznačně určen podmínkami, za kterých je pokus uskutečněn.

Definice 2.2.2 ([1], §1.2). Systém \mathcal{A} podmnožin množiny Ω se nazývá σ – algebrou, pokud

- $\Omega \in \mathcal{A}$,
- $A \in \mathcal{A} \implies \Omega \setminus A \in \mathcal{A}$,
- $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup A_i \in \mathcal{A}$.

Definice 2.2.3 ([1], §1.3). Pravděpodobnostní mírou rozumíme reálnou funkci P definovanou na σ – algebře \mathcal{A} jevů Ω , která splňuje

- P má hodnoty z intervalu $\langle 0, 1 \rangle$,
- $P(\Omega) = 1$,
- pro disjunktní množiny $A_1, A_2, \dots \in \mathcal{A}$ je $P(\cup A_i) = \sum P(A_i)$.

Definice 2.2.4 ([1], §1.3). Trojice (Ω, \mathcal{A}, P) se nazývá pravděpodobnostní prostor.

Definice 2.2.5 ([1], §1.6). Pravděpodobnost jevu C za předpokladu, že nastal jev D (s pravděpodobností $P(D) > 0$) nazýváme podmíněnou pravděpodobností definovanou jako

$$P(C|D) = \frac{P(C \cap D)}{P(D)}.$$

Definice 2.2.6 ([1], §2.1). Funkce $X : \Omega \rightarrow \mathbb{R}$ definovaná na neprázdňém prostoru Ω se σ – algebrou \mathcal{A} se nazývá náhodná veličina, pokud

$$X^{-1}(B) = \{\omega : X(\omega) \in (a, b)\} \in \mathcal{A}$$

pro libovolné $-\infty \leq a < b \leq \infty, B = (a, b)$.

Definice 2.2.7 ([1], §2.2). Mějme náhodnou veličinu X definovanou na prostoru (Ω, \mathcal{A}, P) . Pak rozdělením (distribucí) této veličiny chápeme pravděpodobnostní míru Q

$$Q(B) = P(X \in B), B \in \mathcal{B}(\mathbb{R}),$$

na vhodném systému podmnožin \mathbb{R} , tzv. Borelovské σ – algebře $\mathcal{B}(\mathbb{R})$, která označuje nejmenší σ – algebru v \mathbb{R} obsahující všechny intervaly.

Definice 2.2.8 ([1], §2.3). Distribuční funkce $F : \mathbb{R} \rightarrow \langle 0, 1 \rangle$ je funkce, která splňuje

- F je neklesající a zprava spojitá,
- $\lim_{x \rightarrow -\infty} F(x) = 0$,
- $\lim_{x \rightarrow \infty} F(x) = 1$.

Definice 2.2.9 ([1], §2.3). Mějme (reálnou) náhodnou veličinu X s rozdělením Q na \mathbb{R} . Distribuční funkcí této veličiny rozumíme funkci

$$F(x) = P(X \leq x), x \in \mathbb{R}.$$

Definice 2.2.10 ([1], §2.4). Mějme náhodnou veličinu X definovanou na spočetné množině $M = \{x_1, x_2, x_3 \dots\} \subset \mathbb{R}$. Takovouto náhodnou veličinu budeme označovat jako diskrétní náhodnou veličinu.

V této práci budeme nadále uvažovat pouze diskrétní náhodné veličiny.

Definice 2.2.11 ([1], §2.5). Pravděpodobnostní funkcí rozumíme nezápornou funkci nulovou všude s výjimkou spočetně bodů $x_1, x_2, \dots \in \mathbb{R}$, v nichž nabývá postupně hodnot $p_1, p_2, \dots \in (0, 1)$, přičemž $\sum p_i = 1$.

Definice 2.2.12 ([1], §2.9). Střední hodnotu diskrétní náhodné veličiny X definujeme jako

$$E(X) = \sum_i x_i \cdot P(X = x_i).$$

Lemma 2.2.1 ([1], §2.9). Mějme konstanty $\alpha, \beta \in \mathbb{R}$ a dvě náhodné veličiny X a Y , potom platí

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

Tato rovnost se nazývá linearitou střední hodnoty.

Nadcházející lemma říká, že pokud mají dvě náhodné veličiny stejnou distribuci, potom mají i stejnou střední hodnotu, čehož využijeme v pozdějších výpočtech.

Lemma 2.2.2. Mějme dvě náhodné veličiny X a Y s diskrétním rozdělením, které mají stejnou distribuci, tj.

$$X \stackrel{d}{=} Y,$$

potom tyto náhodné veličiny mají stejné střední hodnoty

$$E(X) = E(Y).$$

Důkaz. Máme diskrétní náhodné proměnné X a Y , které mají stejnou distribuci. Z toho plyne, že mají stejnou pravděpodobnostní funkci. Platí tedy

$$P(X = x_i) = P(Y = y_i) \quad \text{pro } i = 1, 2, \dots$$

Dále bude platit

$$\begin{aligned} \sum_i x_i \cdot P(X = x_i) &= \sum_i y_i \cdot P(Y = y_i), \\ E(X) &= E(Y). \end{aligned}$$

□

Další pojmy a definice se týkají teorie náhodných procesů, a to zejména Markovových řetězců.

Definice 2.2.13 ([7], Def. 1.1). Necht' (Ω, A, P) je pravděpodobnostní prostor a $T \subseteq \mathbb{R}$, potom rodinu náhodných veličin $\{X_t, t \in T\}$ v prostoru (Ω, A, P) nazýváme náhodným procesem.

Z hlediska hodnoty množiny T rozlišujeme náhodné procesy v diskrétním čase ($T = \mathbb{Z}$) a ve spojitém čase ($T = (a, b)$, kde $-\infty \leq a < b \leq \infty$).

Zvláštním případem náhodného procesu je tzv. Markovův řetězec. Mějme opět pravděpodobnostní prostor (Ω, A, P) a uvažujme posloupnost náhodných veličin $\{X_n, n \in \mathbb{N}_0\}$ na tomto prostoru, které nabývají pouze celočíselných hodnot. Necht' S je množina stavů náhodného procesu, tj. množina celých čísel i , pro které platí, že když $n \in \mathbb{N}_0$ tak $P(X_n = i) > 0$. Prvky této množiny budeme označovat jako stavy.

Definice 2.2.14 ([7], Def. 2.1). Posloupnost náhodných veličin $\{X_n, n \in \mathbb{N}_0\}$ s celočíselnými hodnotami se nazývá Markovův řetězec s diskrétním časem a množinou stavů S , pokud

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i), \quad (2.2.1)$$

pro všechna $n \in \mathbb{N}_0$ a pro $i, j, i_{n-1}, \dots, i_0 \in S$, taková, že $P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$. Pokud je množina stavů S konečná, jedná se o tzv. konečný Markovův řetězec.

Rovnost (2.2.1) vyjadřuje tzv. markovskou vlastnost, tj. že stav Markovova řetězce závisí pouze na stavu předchozím.

Definice 2.2.15 ([7], Str. 15). Podmíněné pravděpodobnosti

$$P(X_{n+1} = j | X_n = i) = p_{i,j}$$

nazýváme pravděpodobnostmi přechodu Markovova řetězce.

Definice 2.2.16 ([7], Def. 2.1). Stochastická matice je taková matice, která splňuje

$$p_{i,j} \geq 0, \quad i, j \in S; \quad \sum_{j \in S} p_{i,j} = 1, \quad i \in S. \quad (2.2.2)$$

Pravděpodobnosti přechodu Markovova řetězce můžeme sestavit do čtvercové matice \mathbf{P} .

Definice 2.2.17 ([7], Str. 16). Matice pravděpodobností přechodu je čtvercová matice \mathbf{P} , která je tvořena jednotlivými pravděpodobnostmi přechodu $\mathbf{P} = \{p_{i,j} : i, j \in S\}$.

Je zřejmé, že tato matice splňuje vlastnost (2.2.2), tedy matice pravděpodobností přechodu \mathbf{P} Markovova řetězce je stochastická matice.

Definice 2.2.18 ([7], Str. 16). Pravděpodobnostní rozdělení $\mathbf{p} = \{P(X_0 = i), i \in S\}$ nazýváme počáteční rozdělení.

Definice 2.2.19. Mějme neorientovaný graf $G = (V, H)$. Markovův řetězec s množinou stavů $S = V(G)$ nazveme náhodnou procházkou na grafu G , pokud pro pravděpodobnosti přechodu platí

$$p_{i,j} = \begin{cases} \frac{1}{d_G(i)} & \{i, j\} \in H(G), \\ 0 & \text{jinak.} \end{cases}$$

Náhodná procházka je tedy Markovův řetězec, kde za stavy uvažujeme vrcholy patřící grafu G . Nacházíme-li se ve vrcholu $u \in V(G)$, tak se stejnou pravděpodobností přejdeme do kteréhokoli souseda tohoto vrcholu (např. pokud jsme ve stavu, který sousedí se čtyřmi vrcholy, pak do každého z nich se dostaneme s pravděpodobností $\frac{1}{4}$).

Dále u náhodné procházky uvažujeme, že začneme v určitém vrcholu. Tento vrchol nazveme počátečním vrcholem.

Definice 2.2.20. Mějme neorientovaný graf $G = (V, H)$ a vrchol $v \in V(G)$. Potom tento vrchol nazveme počátečním vrcholem náhodné procházky, pokud bude platit

$$P(X_0 = i) = \begin{cases} 1, & i = v, \\ 0, & i \neq v. \end{cases}$$

Čas pokrytí nám vyjadřuje počet kroků, které jsou potřeba udělat, abychom navštívili (tzn. pokryli) všechny vrcholy daného grafu.

Pro definování očekávaného času pokrytí musíme nejprve definovat náhodnou veličinu času pokrytí, k čemuž použijeme náhodnou veličinu počtu pokrytých vrcholů.

Definice 2.2.21. Mějme náhodnou procházku na grafu, potom Y_i je náhodná veličina vyjadřující počet pokrytých uzlů v čase i .

Definice 2.2.22. Mějme náhodnou procházku na grafu $G = (V, H)$ s počátečním vrcholem $v \in V(G)$. Potom diskrétní náhodnou veličinu času pokrytí τ_v vyjádříme jako

$$\tau_v = \{n > 0 : (Y_n = |V|) \wedge (Y_{n-1} = |V| - 1)\}.$$

Definice 2.2.23. Mějme náhodnou procházku na grafu a diskrétní náhodnou veličinu času pokrytí τ . Potom střední hodnotu této veličiny nazveme očekávaným časem pokrytí (*cover time*) a budeme jej značit

$$C_v(G) = E(\tau_v) = \sum_{i=1}^{\infty} i \cdot P(\tau_v = i), \quad (2.2.3)$$

kde index v vyjadřuje počáteční vrchol náhodné procházky.

Očekávaným časem pokrytí chápeme průměrný počet kroků, které je potřeba udělat, abychom v rámci náhodné procházky navštívili všechny vrcholy grafu $G = (V, H)$, když začneme ve vrcholu $v \in V(G)$.

Dále si uvedeme definice týkající se klasifikaci stavů a jejich vlastností, které budeme potřebovat.

Definice 2.2.24 ([6], Def. 9). Mějme Markovův řetězec s množinou stavů S , pak čas prvního vstupu resp. návratu do stavu $j \in S$ definujeme jako

$$\psi_j(1) = \inf\{n > 0 : X_n = j\}.$$

Definice 2.2.25 ([6], Def. 10). Stav j se nazývá přechodným, jestliže řetězec, který vychází z j , se s kladnou pravděpodobností do j nikdy nevrátí, to znamená

$$P(\psi_j(1) = +\infty | X_0 = j) > 0.$$

Definice 2.2.26 ([6], Def. 12). Stav j nazveme absorpčním právě tehdy, když platí

$$p_{j,j} = 1.$$

To znamená, že jakmile Markovův řetězec dosáhne absorpčního stavu, už ho neopustí.

Nyní zavedeme fundamentální matici, kterou využijeme v pozdějším výpočtu. Pro definici fundamentální matice uvažujeme matici pravděpodobností přechodu ve speciálním tvaru

$$\mathbf{P}_{adj} = \left(\begin{array}{c|c} P^* & 0 \\ \hline Q & R \end{array} \right), \quad (2.2.4)$$

kde uvažujeme \mathcal{T} jako množinu přechodných stavů Markovova řetězce, pak $P^* = \{p_{i,j} : i, j \notin \mathcal{T}\}$, $Q = \{p_{i,j} : i \in \mathcal{T}, j \notin \mathcal{T}\}$ a $R = \{p_{i,j} : i, j \in \mathcal{T}\}$.

Definice 2.2.27 ([6], Def. 17). Fundamentální matice Markovova řetězce je matice, pro kterou platí

$$\mathbf{F} = (\mathbf{I} - \mathbf{R})^{-1}.$$

Věta 2.2.3 ([8]). Mějme Markovův řetězec s fundamentální maticí $\mathbf{F} = (\mathbf{I} - \mathbf{R})^{-1}$, potom suma prvků na řádku i fundamentální matice vyjadřuje průměrný počet kroků do absorpce, pokud začínáme v i -tém přechodném stavu Markovova řetězce.

Nyní si uvedeme definice a pojmy potřebné k numerickým výpočtům.

Definice 2.2.28 ([1], §6.1). Náhodným výběrem rozsahu n budeme nazývat posloupnost stejně rozdělených a nezávislých náhodných veličin (X_1, \dots, X_n) .

Definice 2.2.29 ([1], §6.2). Měřitelnou funkci pozorovaných hodnot (obvykle z náhodného výběru) nazýváme statistikou.

Definice 2.2.30 ([1], §6.3). Odhadem parametrické funkce $g(\xi)$ rozumíme libovolnou statistiku $S(X_1, \dots, X_n)$ na základě pozorování X_1, \dots, X_n nezávislou na ξ .

Kapitola 3

Očekávaný čas pokrytí úplného grafu a cesty

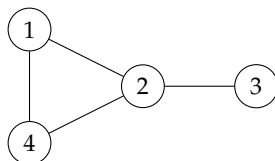
3.1 Metoda odvození

V této podkapitole si nastíníme myšlenku výpočtu očekávaného času pokrytí.

Mějme neorientovaný graf $G = (V, H)$ a náhodnou procházku na tomto grafu podle definice 2.2.18. Náhodná procházka na grafu nám definuje indukovaný Markovův řetězec, jehož stavy se skládají z indexu aktuálně navštíveného vrcholu $j \in V(G)$ (tj. pozice náhodné procházky) a navíc z množiny již pokrytých vrcholů $\mathcal{I} \subseteq V(G)$. Stavy indukovaného Markovova řetězce označíme jako S_{ind} .

Dále si zavedeme náhodnou proměnnou $\chi_{\mathcal{I},j}$, která vyjadřuje čas do pokrytí při pokryté množině vrcholů \mathcal{I} a poloze j . Pro lepší pochopení stavů indukovaného Markovova řetězce si uvedeme následující příklad.

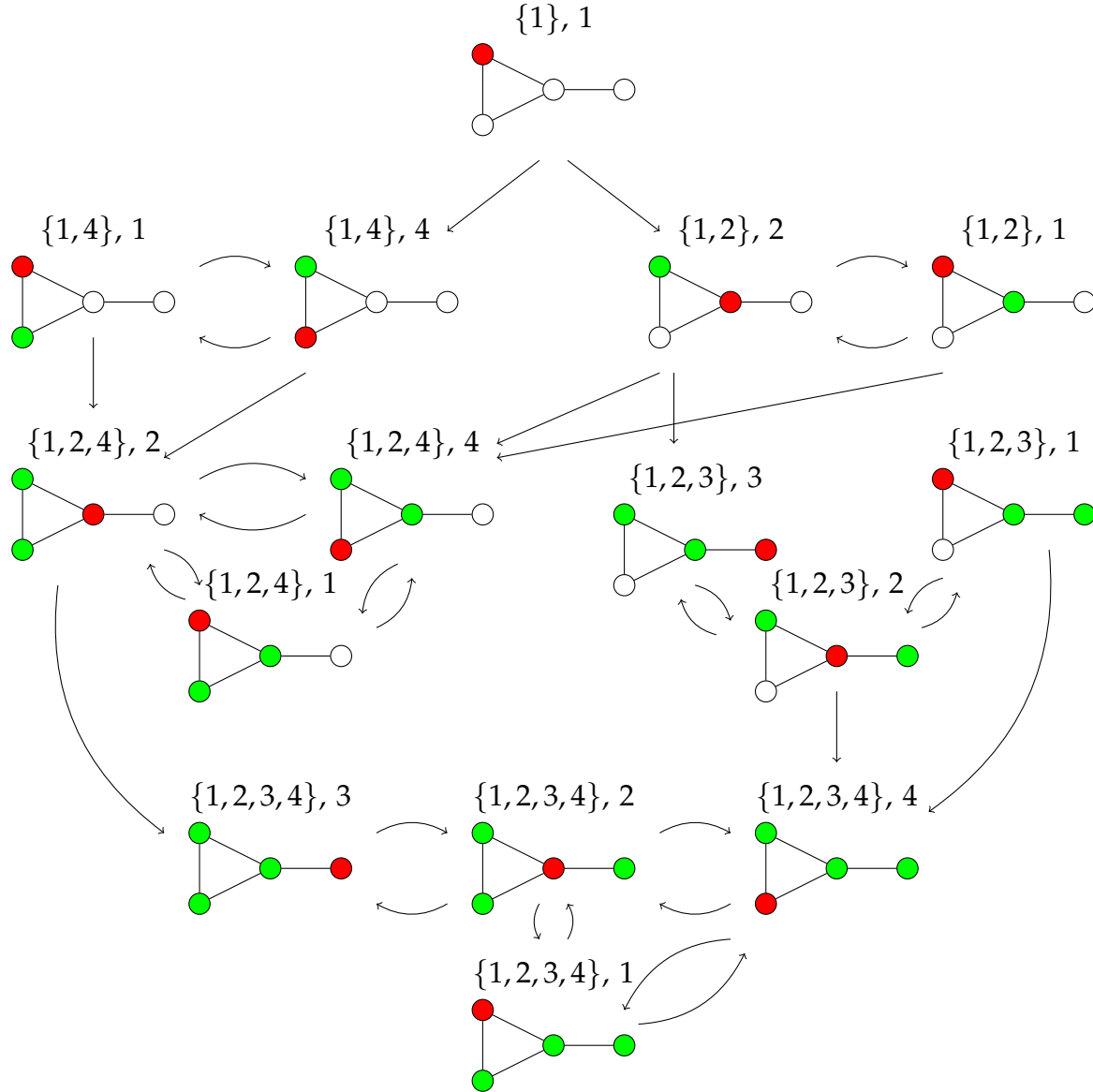
Příklad 1. Uvažujme náhodnou procházku na grafu $G = (V, H)$, kde $V = \{1, 2, 3, 4\}$, $H = \{\{1, 2\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\}$ (viz obrázek 3.1) a dále mějme počáteční vrchol $v = 1$. Z definice 2.2.18 plyne, že: $S = V(G) = \{1, 2, 3, 4\}$.



Obrázek 3.1: Graf z příkladu 1.

Stavy indukovaného Markovova řetězce S_{ind} a přechody mezi těmito stavy ilustruje následující obrázek 3.2, kde zelené vrcholy vyjadřují již pokryté vrcholy a červený vrchol značí aktuální pozici (tzn. $\mathcal{I} = \{\text{'množina zelených vrcholů'} \cup \{\text{'červený vrchol'}\}$ a $j = \{\text{'červený vrchol'}\}$).

o



Obrázek 3.2: Schéma dosažitelných stavů indukovaného Markovova řetězce S_{ind} a jejich přechodů pro náhodnou procházku s počátečním vrcholem $v = 1$.

Na obrázku 3.2 jsou pouze dosažitelné stavy Markovova řetězce indukovaného z náhodné procházky začínající ve vrcholu 1, množina dosažitelných stavů S_{ind} se většinou liší v závislosti na počátečním vrcholu náhodné procházky. Je tedy patrné, že množina stavů indukovaného Markovova řetězce je rozsáhlejší než stavy náhodné procházky a práce s nimi v rámci obecných grafů je složitá. Avšak u grafů se specifickou strukturou (např. úplný graf a cesta) můžeme sjednotit stavy indukovaného Markovova řetězce, pro které má náhodná veličina $\chi_{I,j}$ stejné pravděpodobnostní rozdělení.

Pozorování 3.1.1. V případě úplného grafu \mathcal{K}_N sjednotíme stavy indukovaného Markovova řetězce se stejnou velikostí množiny \mathcal{I} . Potom získáme náhodnou veličinu χ_i , kde $i = |\mathcal{I}|$ (tj. čas do pokrytí při pokrytých i vrcholech).

Pozorování 3.1.2. Uvažujeme-li náhodnou procházku na cestě \mathcal{P}_N s krajním počátečním vrcholem $v = 1$, sjednotíme stavy indukovaného Markovova řetězce se stejnou pozicí j a získáme náhodnou veličinu χ_j (tj. čas do pokrytí, když se nacházíme ve vrcholu j).

Poznámka 3.1.3. Mějme náhodnou procházku na grafu G s počátečním vrcholem v , potom platí

$$\chi_{\{1\},1} \stackrel{d}{=} \tau_v,$$

a tedy

$$E(\chi_{\{1\},1}) = E(\tau_v) = C_v(G),$$

viz Lemma 2.2.2.

3.2 Úplný graf

Úplný graf je graf, ve kterém je každý vrchol spojen s každým. U úplného grafu je čas pokrytí pro všechny počáteční vrcholy stejný, to plyne z poznámky 2.1.10.

Věta 3.2.1. Mějme úplný graf \mathcal{K}_N , potom očekávaný čas pokrytí platí

$$C_v(\mathcal{K}_N) = (N - 1) \cdot \sum_{n=1}^{N-1} \frac{1}{n} \quad (3.2.1)$$

pro libovolné $v \in V(\mathcal{K}_N)$.

V této práci si ukážeme dva důkazy věty 3.2.1.

Důkaz 1. Mějme úplný graf \mathcal{K}_N . Díky struktuře úplného grafu můžeme z jakéhokoliv vrcholu pokrýt kterýkoli další vrchol, můžeme tedy vždy opustit již pokrytý podgraf grafu \mathcal{K}_N . Každý vrchol úplného grafu $u \in V(\mathcal{K}_N)$ má stupeň $d_{\mathcal{K}_N}(u) = N - 1$. Z každého vrcholu tedy máme $N - 1$ možností, kam jít. Pokud si označíme počet pokrytých uzlů jako n , máme $N - n$ nepokrytých sousedů vrcholu u a $n - 1$ je počet pokrytých sousedů vrcholu u . Získáme tedy pravděpodobnosti

$$\begin{aligned} P(\text{'zůstanu uvnitř již pokrytého grafu'}) &= \frac{n - 1}{N - 1}, \\ P(\text{'pokryji další vrchol'}) &= \frac{N - n}{N - 1}. \end{aligned} \quad (3.2.2)$$

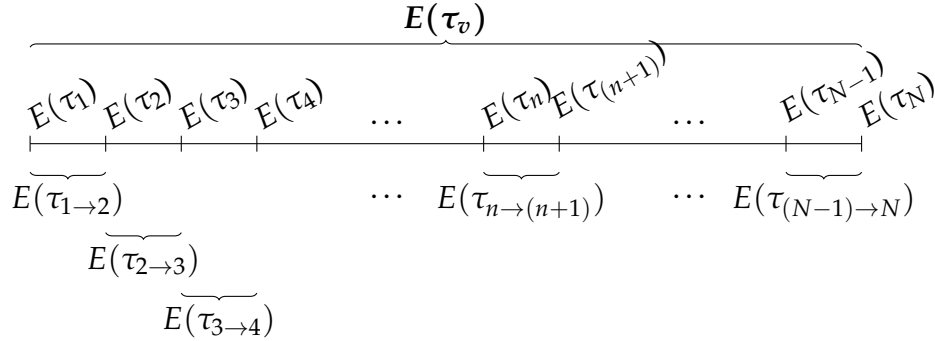
Označme si náhodnou proměnnou času pokrytí pro úplný graf \mathcal{K}_N jako τ_N . Nyní zavedeme pomocnou náhodnou proměnnou $\tau_{n \rightarrow (n+1)}$ pokrytí dalšího vrcholu, když máme pokryto n vrcholů. Pak platí

$$\tau_v = \tau_{1 \rightarrow 2} + \tau_{2 \rightarrow 3} + \cdots + \tau_{(N-2) \rightarrow (N-1)} + \tau_{(N-1) \rightarrow N}. \quad (3.2.3)$$

Použijeme lemma 2.2.1 a získáme očekávaný čas pokrytí $E(\tau_N)$ ve tvaru

$$\begin{aligned} C_v(\mathcal{K}_N) = E(\tau_v) &= E(\tau_{1 \rightarrow 2}) + E(\tau_{2 \rightarrow 3}) + \cdots + E(\tau_{(N-2) \rightarrow (N-1)}) + E(\tau_{(N-1) \rightarrow N}), \\ &= \sum_{n=1}^{N-1} E(\tau_{n \rightarrow (n+1)}) \end{aligned} \quad (3.2.4)$$

znázorněný na obrázku 3.3.



Obrázek 3.3: Očekávaný čas pokrytí jako suma očekávaných mezičasů pokrytí dalšího vrcholu vyjádřené v rovnosti (3.2.4).

Jednotlivé střední hodnoty zapíšeme ve tvaru (2.2.3) jako

$$E(\tau_{n \rightarrow (n+1)}) = \sum_{i=1}^{\infty} i \cdot P(\tau_{n \rightarrow (n+1)} = i). \quad (3.2.5)$$

Nyní vyjádříme pravděpodobnosti $P(\tau_{n \rightarrow (n+1)} = i)$ z výrazu (3.2.5). K tomu použijeme pravděpodobnosti odvozené ve výrazu (3.2.2). Abychom pokryli další vrchol v i -tém kroku, musíme $(i-1)$ -krát zůstat v pokrytém podgrafu a v posledním i -tém kroku pokrýt libovolný stále nepokrytý vrchol. Zapíšeme tedy jednotlivé pravděpodobnosti ve tvaru

$$P(\tau_{n \rightarrow (n+1)} = i) = \left(\frac{n-1}{N-1} \right)^{i-1} \cdot \frac{N-n}{N-1}. \quad (3.2.6)$$

Pravděpodobnost vyjádřenou ve výrazu (3.2.6) dosadíme do rovnosti (3.2.5), to celé zakomponujeme do původní sumy (3.2.4). Tím získáme očekávaný čas pokrytí ve tvaru

$$\begin{aligned} C_v(\mathcal{K}_N) = E(\tau_v) &= \sum_{n=1}^{N-1} \sum_{i=1}^{\infty} i \cdot \left(\frac{n-1}{N-1} \right)^{i-1} \cdot \frac{N-n}{N-1}, \\ &= \sum_{n=1}^{N-1} \frac{N-n}{N-1} \cdot \sum_{i=1}^{\infty} i \cdot \left(\frac{n-1}{N-1} \right)^{i-1}, \\ &= \sum_{n=1}^{N-1} \frac{N-n}{N-1} \cdot \left(\frac{n-1}{N-1} \right)^{-1} \cdot \sum_{i=1}^{\infty} i \cdot \left(\frac{n-1}{N-1} \right)^i. \end{aligned} \quad (3.2.7)$$

Upravíme vnitřní sumu rovnosti (3.2.7) pomocí součtu aritmeticko-geometrické řady, tj.

$$\sum_{k=1}^{\infty} kr^k = \frac{r}{(1-r)^2}$$

pro $0 < r < 1$. Po úpravě získáme

$$\begin{aligned} E(\tau_v) &= \sum_{n=1}^{N-1} \frac{N-n}{N-1} \cdot \left(\frac{n-1}{N-1}\right)^{-1} \cdot \frac{\frac{n-1}{N-1}}{\left(1 - \frac{n-1}{N-1}\right)^2}, \\ &= \sum_{n=1}^{N-1} \frac{\frac{N-n}{N-1} \cdot \frac{n-1}{N-1}}{\frac{n-1}{N-1} \cdot \left(\frac{N-n}{N-1}\right)^2} = \sum_{n=1}^{N-1} \frac{1}{\frac{N-n}{N-1}} = (N-1) \cdot \sum_{n=1}^{N-1} \frac{1}{N-n}. \end{aligned} \quad (3.2.8)$$

V rovnosti (3.2.8) si zavedeme substituci $k = N - n$ a získáme

$$C_v(\mathcal{K}_N) = E(\tau_v) = (N-1) \sum_{k=1}^{N-1} \frac{1}{k},$$

což je výraz (3.2.1). □

Nyní si ukážeme druhý postup, jak dokázat větu 3.2.1. Zde budeme vycházet z myšlenky indukovaných Markovových řetězců uvedené v předchozí kapitole 3.1.

Důkaz 2. Mějme náhodnou proměnnou χ_i odvozenou v pozorování 3.1.1. Po prvním kroku jistě pokryjeme další vrchol. Bude tedy platit

$$\chi_1 \stackrel{d}{=} \chi_2 + 1. \quad (3.2.9)$$

To znamená, že náhodná proměnná při pokrytém jednom vrcholu má stejnou distribuci jako náhodná proměnná při pokrytých dvou vrcholech posunutá o jeden krok, který jsme již provedli. Pokud máme pokryté 2 vrcholy a více, mohou nastat v dalším kroku dvě varianty - vrátíme se do již pokrytého vrcholu nebo pokryjeme další vrchol. Zapišeme rovnost pro náhodnou veličinu χ_i pro $i > 1$, kde s pravděpodobností $\frac{i-1}{N-1}$ zůstaneme již v pokrytém podgrafu a s pravděpodobností $\frac{N-i}{N-1}$ pokryjeme další vrchol (to odpovídá pravděpodobnostem ve výrazu (3.2.2), zde však používáme pro pokryté uzly index i namísto n) jako

$$\chi_i \stackrel{d}{=} \frac{i-1}{N-1} \cdot (\chi_i + 1) + \frac{N-i}{N-1} \cdot (\chi_{i+1} + 1). \quad (3.2.10)$$

Jelikož jsou náhodné proměnné na levé i pravé straně rovnosti stejně rozdělené, budou mít stejnou střední hodnotu (viz lemma 2.2.2). Námi hledaný očekávaný čas pokrytí je $E(\chi_1)$, neboli očekávaný čas pokrytí při pokrytém jednom vrcholu (počáteční vrchol). Zajisté platí, že pokud jsme již pokryli celý graf, hodnota pro zbývající kroky na pokrytí je nulová, tj.

$$E(\chi_N) = 0. \quad (3.2.11)$$

K úpravě výrazu (3.2.10) použijeme lemma 2.2.2 a lemma 2.2.1, získáme tedy

$$\begin{aligned}
E(\chi_i) &= E\left(\frac{i-1}{N-1} \cdot (\chi_i + 1) + \frac{N-i}{N-1} \cdot (\chi_{i+1} + 1)\right), \\
&= E\left(\frac{i-1}{N-1} \cdot \chi_i + \frac{N-i}{N-1} \cdot \chi_{i+1} + 1\right), \\
&= \frac{i-1}{N-1} \cdot E(\chi_i) + \frac{N-i}{N-1} \cdot E(\chi_{i+1}) + 1.
\end{aligned} \tag{3.2.12}$$

Nyní upravíme rovnost (3.2.12) a vyjádříme si samotné $E(\chi_i)$ jako

$$\begin{aligned}
E(\chi_i) &= \frac{i-1}{N-1} \cdot E(\chi_i) + \frac{N-i}{N-1} \cdot E(\chi_{i+1}) + 1, \\
E(\chi_i) - \frac{i-1}{N-1} \cdot E(\chi_i) &= \frac{N-i}{N-1} \cdot E(\chi_{i+1}) + 1, \\
E(\chi_i) \cdot \left(1 - \frac{i-1}{N-1}\right) &= \frac{N-i}{N-1} \cdot E(\chi_{i+1}) + 1, \\
E(\chi_i) \cdot \frac{N-i}{N-1} &= \frac{N-i}{N-1} \cdot E(\chi_{i+1}) + 1, \\
E(\chi_i) &= E(\chi_{i+1}) + \frac{N-1}{N-i}.
\end{aligned} \tag{3.2.13}$$

Do výrazu ve tvaru (3.2.13) začneme dosazovat. Díky podmínce (3.2.11) začneme odzadu pro $i = (N-1)$ a postupně se dostaneme až k $i = 1$

$$\begin{aligned}
i = (N-1) : E(\chi_{N-1}) &= E(\chi_N) + \frac{N-1}{N-(N-1)} = (N-1), \\
i = (N-2) : E(\chi_{N-2}) &= E(\chi_{N-1}) + \frac{N-1}{N-(N-2)} = (N-1) + \frac{N-1}{2}, \\
i = (N-3) : E(\chi_{N-3}) &= E(\chi_{N-2}) + \frac{N-1}{N-(N-3)} = (N-1) + \frac{N-1}{2} + \frac{N-1}{3}, \\
i = (N-4) : E(\chi_{N-4}) &= E(\chi_{N-3}) + \frac{N-1}{N-(N-4)} = (N-1) + \frac{N-1}{2} + \frac{N-1}{3} + \frac{N-1}{4}, \\
&\vdots
\end{aligned}$$

Tímto způsobem získáme $E(\chi_1)$ jako

$$i = 1 : C_v(\mathcal{K}_N) = E(\tau_v) = E(\chi_1) = E(\chi_2) + \frac{N-1}{N-1} = (N-1) \cdot \sum_{n=1}^{N-1} \frac{1}{n},$$

což odpovídá výrazu (3.2.1).

□

3.3 Cesta

V této kapitole se budeme věnovat odvození očekávaného času náhodné procházky na cestě. Na rozdíl od úplného grafu se zde bude lišit očekávaný čas pokrytí v závislosti na počátečním vrcholu. Dále z poznámky 2.1.10 plyne, že očekávaný čas pokrytí bude stejný pro vrcholy, které jsou stejně vzdálené od kraje (tzn. první a poslední, druhý a předposlední atd.).

Nejprve si odvodíme případ pro krajní počáteční vrchol. Využijeme zde opět myšlenku indukovaného Markovova řetězce. Pokud začneme v některém z ostatních vrcholů, pokryjeme nejprve s jistotou jeden z krajních vrcholů a až poté celou cestu. Očekávaný čas pokrytí se tedy bude skládat z očekávaného času pokrytí cesty se začátkem na kraji a z času nutného k pokrytí krajního vrcholu.

3.3.1 Cesta se začátkem na kraji



Obrázek 3.4: Ilustrace k větě 3.3.1 týkající se očekávaného času pokrytí náhodné procházky na cestě při počátečním vrcholu $v \in \{1, N\}$, které jsou zvýrazněny na obrázku modře.

Věta 3.3.1. Mějme cestu \mathcal{P}_N , potom pro očekávaný čas pokrytí platí

$$C_v(\mathcal{P}_N) = (N - 1)^2 \quad (3.3.1)$$

pro $v \in \{1, N\}$.

Důkaz. Mějme cestu \mathcal{P}_N z obrázku 3.4 a náhodnou proměnnou χ_j odvozenou v pozorování 3.1.2. Uvažujme-li počáteční vrchol náhodné procházky $v = 1$, potom při pokrytí opačného krajního vrcholu N máme pokrytou celou cestu. Z počátečního krajního vrcholu pokryjeme s jistotou vrchol 2. Jelikož se jedná o cestu, každý nekrajní vrchol má právě dva sousedy. Z toho plyne, že s pravděpodobností $\frac{1}{2}$ navštívíme vrchol $j + 1$ a se stejnou pravděpodobností se vrátíme o krok zpátky do stavu $j - 1$. Získáme tedy následující rovnost distribucí

$$\chi_1 \stackrel{d}{=} \chi_2 + 1, \quad (3.3.2)$$

$$\chi_j \stackrel{d}{=} \frac{1}{2}(\chi_{j-1} + 1) + \frac{1}{2}(\chi_{j+1} + 1). \quad (3.3.3)$$

Zajisté i na cestě bude platit výraz $E(\chi_N) = 0$. Podobnou úpravou rovnosti (3.3.2) a (3.3.3) jako u úplného grafu (viz odvození (3.2.12)) pomocí lemma 2.2.2 a lemma 2.2.1 získáme následující systém rovností:

$$E(\chi_j) = \frac{1}{2}E(\chi_{j-1}) + \frac{1}{2}E(\chi_{j+1}) + 1, \quad (3.3.4)$$

$$E(\chi_1) = E(\chi_2) + 1, \quad (3.3.5)$$

$$E(\chi_N) = 0.$$

Pro vyřešení soustavy zavedeme následující substituci

$$\Delta T_j = E(\chi_j) - E(\chi_{j-1}). \quad (3.3.6)$$

Přepíšeme a upravíme rovnost (3.3.4) a (3.3.5) tak, abychom použili substituci (3.3.6). První rovnost (3.3.4) upravíme a získáme

$$\begin{aligned} -\frac{1}{2}E(\chi_{j-1}) + E(\chi_j) - \frac{1}{2}E(\chi_{j+1}) &= 1, \\ -E(\chi_{j-1}) + E(\chi_j) + E(\chi_j) - E(\chi_{j+1}) &= 2. \end{aligned} \quad (3.3.7)$$

Rovnost (3.3.5) upravíme jako

$$E(\chi_1) - E(\chi_2) = 1. \quad (3.3.8)$$

Nyní máme výraz (3.3.7) a (3.3.8) ve tvaru, pro který lze použít substituce (3.3.6) a získáme soustavu ve tvaru

$$\Delta T_j - \Delta T_{j+1} = 2, \quad (3.3.9)$$

$$\Delta T_2 = -1. \quad (3.3.10)$$

Soustavu rovností (3.3.9) a (3.3.10) můžeme přepsat jako

$$\Delta T_j = -2(j-2) - 1. \quad (3.3.11)$$

Nyní máme vše potřebné k tomu, abychom vyjádřili očekávaný čas pokrytí $E(\chi_1)$ jako

$$\begin{aligned} E(\chi_1) &= E(\chi_1) + (E(\chi_2) - E(\chi_2)) + \dots + (E(\chi_{N-1}) - E(\chi_{N-1})) - E(\chi_N), \\ &= E(\chi_1) - E(\chi_2) + E(\chi_2) - \dots - E(\chi_{N-1}) + E(\chi_{N-1}) - E(\chi_N), \\ &= -\Delta T_2 - \Delta T_3 - \dots - \Delta T_N = \sum_{j=2}^N -\Delta T_j = \sum_{j=2}^N 2 \cdot (j-2) + 1. \end{aligned} \quad (3.3.12)$$

Sumu (3.3.12) upravíme pomocí součtu aritmetické řady, tj.

$$s_n = \frac{n \cdot (a_1 + a_n)}{2},$$

kde a_1 označuje první člen aritmetické posloupnosti a a_n n -tý člen aritmetické posloupnosti.

Očekávaný čas pokrytí je součet aritmetické řady (3.3.12), získáme tedy

$$\begin{aligned} C_v(\mathcal{P}_N) &= E(\chi_1) = \sum_{j=2}^N 2 \cdot (j-3) + 1 = \frac{1 + (2N-4+1)}{2} \cdot (N-1), \\ &= \frac{2N-2}{2} \cdot (N-1) = (N-1) \cdot (N-1) = (N-1)^2 \end{aligned} \quad (3.3.13)$$

pro $v \in \{1, N\}$. To je výraz (3.3.1). □

3.3.2 Cesta nezačínající na kraji



Obrázek 3.5: Ilustrace k větě 3.3.2 týkající se očekávaného času pokrytí náhodné procházky na cestě při počátečním vrcholu $v = \{2, 3, 4, \dots, N-1\}$, které jsou zvýrazněny na obrázku modře.

V této podkapitole se budeme věnovat odvození očekávaného času pokrytí cesty nezačínající na kraji. Budeme předpokládat číslování vrcholů z kraje viz obrázek 3.5.

Věta 3.3.2. Mějme cestu \mathcal{P}_N , potom pro očekávaný čas pokrytí platí

$$C_v(\mathcal{P}_N) = (N-1)^2 + \sum_{j=1}^{v-1} \frac{2j \cdot (N-v)}{N-1} + \sum_{j=v}^{N-2} \frac{2(v-1) \cdot (N-j-1)}{N-1} \quad (3.3.14)$$

pro $v = \{2, 3, 4, \dots, N-1\}$.

V důkazu věty 3.3.2 použijeme definice a věty převzaté z článku [9]. Mějme tří-diagonální matici ve tvaru

$$\mathbf{A}_n = \begin{bmatrix} b_1 & c_1 & & & & & 0 \\ a_1 & b_2 & c_2 & & & & \\ & a_2 & b_3 & c_3 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & a_{n-1} & b_{n-1} & c_{n-1} & \\ 0 & & & & a_n & b_n & \end{bmatrix}. \quad (3.3.15)$$

Definice 3.3.1 ([9], §1.3). Vektor θ_n příslušný matici ve tvaru (3.3.15) je vektor splňující

$$\begin{aligned} \theta_n &= b_n \theta_{n-1} - a_n c_{n-1} \theta_{n-2} \quad n \geq 1, \\ \theta_{-1} &= 0, \quad \theta_0 = 1. \end{aligned} \quad (3.3.16)$$

Definice 3.3.2 ([9], §1.4). Vektor ϕ_i příslušný matici ve tvaru (3.3.15) je vektor splňující

$$\begin{aligned}\phi_i &= b_i\phi_{i+1} - a_{i+1}c_i\phi_{i+2} & i = n, n-1, \dots, 2, 1, \\ \phi_{n+1} &= 1, & \phi_{n+2} = 0,\end{aligned}\tag{3.3.17}$$

Věta 3.3.3 ([9], Theorem 1). Mějme symetrickou tří-diagonální matici ve tvaru (3.3.15) o velikosti $n \times n$, kde $a_i = c_i$. Potom členy inverzní matice $\alpha_{i,j}$ jsou dány předpisem

$$\alpha_{i,j} = \begin{cases} (-1)^{i+j} a_{j+1} a_{j+2} \dots a_i \frac{\theta_{j-1}\phi_{i+1}}{\theta_n} & j = 1, 2, \dots, n-1, \\ \frac{\theta_{j-1}\phi_{i+1}}{\theta_n} & j = n. \end{cases}\tag{3.3.18}$$

pro $a_i \neq 0, i > j$ a pro $i = n, n-1, n-2, \dots, 3, 2, 1$.

Důkaz. Důkaz je uveden ve článku [9] na straně 61. □

Nyní provedeme důkaz věty 3.3.2.

Důkaz věty 3.3.2. Očekávaný čas pokrytí náhodné procházky na cestě nezačínající v krajním vrcholu (viz obrázek 3.5) se skládá z očekávaného času pokrytí cesty se začátkem na kraji a času nutného k pokrytí krajního vrcholu. Jelikož z věty 3.3.1 známe očekávaný čas pokrytí cesty začínající z kraje, stačí nám odvodit čas potřebný k pokrytí krajního vrcholu.

Matice přechodu náhodné procházky na cestě \mathbf{P} má tvar

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & & & & & & & \mathbf{0} \\ \frac{1}{2} & 0 & \frac{1}{2} & & & & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & \frac{1}{2} & 0 & \frac{1}{2} & & & \\ \mathbf{0} & & & & \frac{1}{2} & 0 & \frac{1}{2} & & \\ & & & & & 1 & 0 & & \end{bmatrix}.$$

Jelikož nás zajímá průměrný počet kroků potřebný k pokrytí krajního vrcholu, budeme považovat krajní vrcholy za absorpční vrcholy. Tím si převedeme úlohu na výpočet průměrného počtu kroků do absorpce, k čemuž využijeme fundamentální matici. Matice přechodu zůstane až na první a poslední řádek stejná. První řádek matice bude mít hodnotu 1 na pozici $p_{1,1}$ a na ostatních pozicích 0. Obdobně poslední řádek má hodnotu 1 na pozici $p_{N,N}$ a jinak 0. Takovouto matici lze přeuspořádat do tvaru (2.2.4), čímž získáme

$$\mathbf{P}_{adj} = \left[\begin{array}{cc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \hline \frac{1}{2} & 0 & 0 & \frac{1}{2} & & & & \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & & & \mathbf{0} \\ 0 & 0 & & \frac{1}{2} & 0 & \frac{1}{2} & & \\ \vdots & \vdots & & & \ddots & \ddots & \ddots & \\ 0 & 0 & \mathbf{0} & & \frac{1}{2} & 0 & \frac{1}{2} & \\ 0 & \frac{1}{2} & & & & \frac{1}{2} & 0 & \end{array} \right].$$

Pomocí Definice 2.2.27 si vyjádříme fundamentální matici \mathbf{F} jako čtvercovou matici $n \times n$, kde $n = N - 2$, ve tvaru

$$\mathbf{F} = (I - R)^{-1} = \left[\begin{array}{ccccccc} 1 & -\frac{1}{2} & & & & & \\ -\frac{1}{2} & 1 & -\frac{1}{2} & & & & \mathbf{0} \\ & -\frac{1}{2} & 1 & -\frac{1}{2} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\frac{1}{2} & 1 & -\frac{1}{2} & \\ \mathbf{0} & & & & -\frac{1}{2} & 1 & -\frac{1}{2} \\ & & & & & -\frac{1}{2} & 1 \\ & & & & & & -\frac{1}{2} & 1 \end{array} \right]^{-1}.$$

Matice $(I - R)$ je ve tvaru (3.3.15), kde koeficienty $a_n = c_n = -\frac{1}{2}$ a $b_n = 1$. Vyjádříme si vektor θ_n (3.3.16) pomocí diferenční rovnice ve tvaru

$$\theta_n = 1 \cdot \theta_{n-1} - \left(\frac{1}{2}\right)^2 \cdot \theta_{n-2},$$

s počáteční podmínkou

$$\theta_{-1} = 0, \quad \theta_0 = 1.$$

Spočteme kořeny charakteristické rovnice (viz [4], Theorem 3.6, str. 55), čímž získáme

$$\lambda^2 - \lambda + \frac{1}{4} = 0 \implies \lambda_{1,2} = \frac{1}{2},$$

získáme obecné řešení

$$\theta_n = c_1 \cdot \left(\frac{1}{2}\right)^n + c_2 \cdot n \cdot \left(\frac{1}{2}\right)^n.$$

Vyjádříme konstanty c_1 a c_2 z počátečních podmínek jako

$$\begin{aligned} \left(\frac{1}{2}\right)^0 \cdot c_1 + \left(\frac{1}{2}\right)^0 \cdot 0 \cdot c_2 &= 1, & \left(\frac{1}{2}\right)^{-1} \cdot c_1 + \left(\frac{1}{2}\right)^{-1} \cdot (-1) \cdot c_2 &= 0, \\ c_1 &= 1, & 2 \cdot c_1 - 2 \cdot c_2 &= 0, \\ & & c_2 &= 1. \end{aligned}$$

Získáme partikulární řešení ve tvaru

$$\theta_n = \left(\frac{1}{2}\right)^n + \left(\frac{1}{2}\right)^n \cdot n = \left(\frac{1}{2}\right)^n \cdot (1+n). \quad (3.3.19)$$

Z definice 3.3.1 a definice 3.3.2 plyne

$$\phi_i = \theta_{n-i+1}. \quad (3.3.20)$$

Obecný tvar inverze matice $(I - R)$ vyjádříme pomocí výrazu (3.3.18), do kterého dosadíme výrazy (3.3.19), (3.3.20) a koeficienty a_i .

Pro $j < i$ upravíme první část výrazu (3.3.18):

$$\begin{aligned} \alpha_{i,j} &= (-1)^{i+j} \cdot \left(-\frac{1}{2}\right)^{i-j} \cdot \frac{\left(\frac{1}{2}\right)^{j-1} \cdot (1+j-1) \left(\frac{1}{2}\right)^{n-(i+1)+1} \cdot (1+n-(i+1)+1)}{\left(\frac{1}{2}\right)^n \cdot (1+n)}, \\ &= (-1)^{i+j} \cdot (-1)^{i-j} \cdot \left(\frac{1}{2}\right)^{i-j} \cdot \frac{\left(\frac{1}{2}\right)^{j-1} \cdot j \left(\frac{1}{2}\right)^{n-i} \cdot (1+n-i)}{\left(\frac{1}{2}\right)^{j-1} \cdot \left(\frac{1}{2}\right)^{n-j} \cdot \frac{1}{2} \cdot (1+n)}, \\ &= (-1)^{2i} \cdot \frac{\left(\frac{1}{2}\right)^{n-i+i-j} \cdot j \cdot (1+n-i)}{\left(\frac{1}{2}\right)^{n-j} \cdot \frac{1}{2} \cdot (1+n)} = \frac{2j+2nj-2ij}{1+n}. \end{aligned} \quad (3.3.21)$$

Pro $j = i$ (prvky na diagonále) upravíme druhou část a dostaneme

$$\begin{aligned} \alpha_{i,j} &= \frac{\left(\frac{1}{2}\right)^{i-1} \cdot (1+i-1) \cdot \left(\frac{1}{2}\right)^{n-(i+1)+1} \cdot (1+n-(i+1)+1)}{\left(\frac{1}{2}\right)^n \cdot (1+n)}, \\ &= \frac{\left(\frac{1}{2}\right)^{i-1} \cdot \left(\frac{1}{2}\right)^{n-i} \cdot (i+ni-i^2)}{\left(\frac{1}{2}\right)^{i-1} \cdot \left(\frac{1}{2}\right)^{n-i} \cdot \frac{1}{2} \cdot (1+n)} = \frac{2i+2ni-2i^2}{1+n}. \end{aligned} \quad (3.3.22)$$

Dosazením $j = i$ do výrazu (3.3.21) získáme výraz (3.3.22), budeme tedy dále pracovat s výrazem (3.3.21).

Tímto způsobem jsme získali prvky fundamentální matice pod diagonálou a na diagonále. Jelikož je matice symetrická, lze zbývající prvky na rádcích snadno dopočítat. Označme si prvky fundamentální matice jako $f_{i,j}$, potom platí

$$f_{i,j} = \alpha_{i,j} = \begin{cases} \frac{2j \cdot (n-i+1)}{1+n} & \text{pro } j \leq i, \\ \frac{2i \cdot (n-j+1)}{1+n} & \text{pro } j > i. \end{cases} \quad (3.3.23)$$

Vraťme se nyní k náhodné procházce na cestě \mathcal{P}_N nezačínající v krajním vrcholu. Ve fundamentální matici \mathbf{F} suma i -tého řádku odpovídá průměrnému počtu kroků do absorpce z i -tého přechodného stavu (viz věta 2.2.3). Přechodné stavy odpovídají nekrajním

vrcholům cesty, proto vrchol cesty $u \in \{2, 3, \dots, N-1\}$ odpovídá přechodnému stavu $i = u - 1$.

Očekávaný čas pokrytí náhodné procházky na cestě nezačínající na kraji získáme jako součet výrazu (3.3.1) a sumy prvků $f_{i,j}$ vyjádřené v rovnosti (3.3.23). Získáme tedy

$$\begin{aligned} C_v(\mathcal{P}_N) &= (N-1)^2 + \sum_{j=1}^i \frac{2j \cdot (n-i+1)}{1+n} + \sum_{j=i+1}^n \frac{2i \cdot (n-j+1)}{1+n}, \\ &= (N-1)^2 + \sum_{j=1}^{v-1} \frac{2j \cdot (N-v)}{N-1} + \sum_{j=v}^{N-2} \frac{2(v-1) \cdot (N-j-1)}{N-1} \end{aligned}$$

pro $v = 2, 3, \dots, N-1$, což je výraz (3.3.14). □

Poznámka 3.3.3. Pro počáteční vrchol $v = N-1$ odpovídajícímu poslednímu přechodnému stavu lze výraz (3.3.14) zjednodušit. Vynecháme druhou sumu, která je pro dopočet řádku pro prvky $j > i$ (v posledním řádku matice jsou všechny prvky $i \leq j$). Dosadíme za $v = N-1$ a upravíme pomocí součtu aritmetické řady, čímž získáme

$$\begin{aligned} C_v(\mathcal{P}_N) &= (N-1)^2 + \sum_{j=1}^{N-2} \frac{2j \cdot (N-N+1)}{N-1}, \\ &= (N-1)^2 + \sum_{j=1}^{N-2} \frac{2j}{N-1}, \\ &= (N-1)^2 + \frac{1}{(N-1)} \cdot \frac{2+2N-4}{2} \cdot (N-2), \\ &= (N-1)^2 + (N-2). \end{aligned} \tag{3.3.24}$$

Jelikož je očekávaný čas pokrytí pro předposlední vrchol stejný jako očekávaný čas pokrytí pro druhý vrchol (viz poznámka 2.1.10), platí výraz pro $v \in \{2, N-1\}$.

Kapitola 4

Numerické výpočty

Numerické výpočty byly provedeny v programu Matlab.

4.1 Metoda numerického výpočtu očekávaného času pokrytí

Metoda numerického výpočtu očekávaného času pokrytí je založena na opakovaném generování realizace náhodné procházky (tzn. trajektorie náhodné procházky). Generování realizace náhodné procházky se ukončí poté, co trajektorie náhodné procházky zahrnuje všechny vrcholy daného grafu. V takovém případě je délka trajektorie námi hledaný očekávaný čas pokrytí.

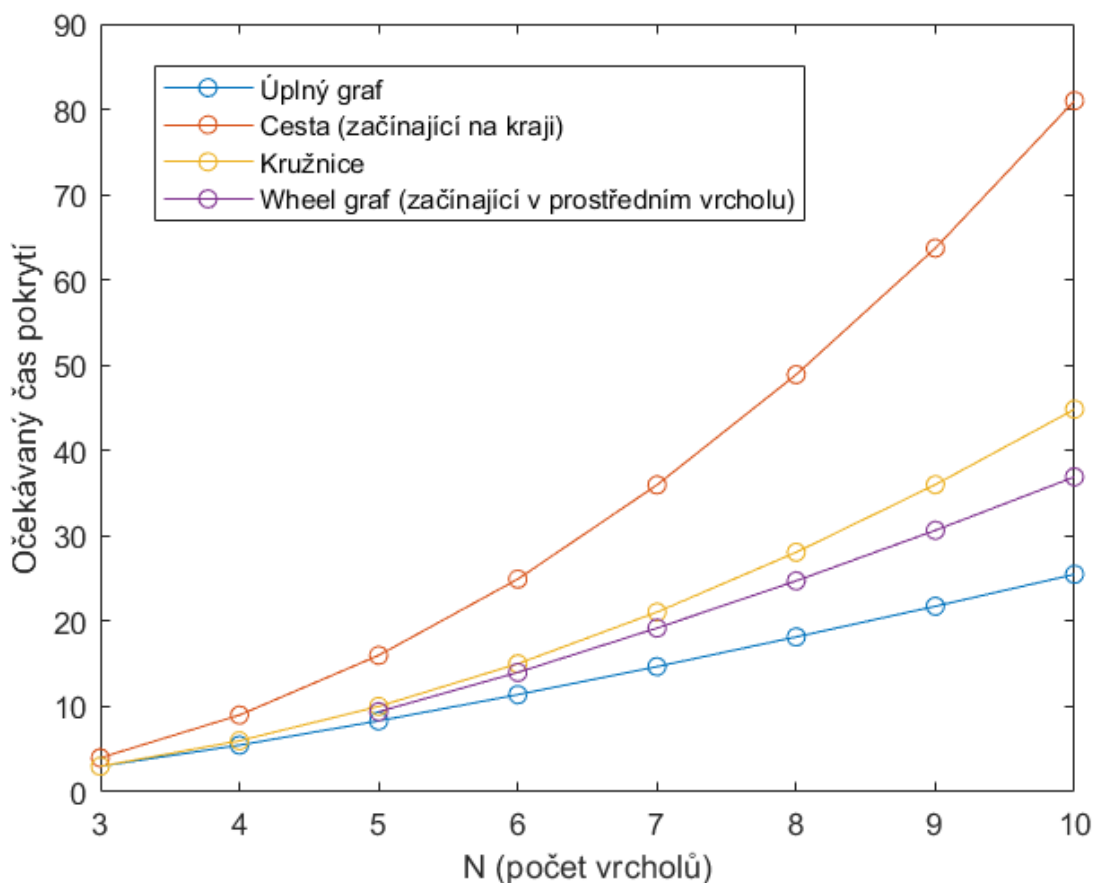
Soubor simulovaných očekávaných časů pokrytí náhodné procházky je náhodným výběrem $\{x_1, x_2, \dots, x_n\}$. Očekávaný čas pokrytí má distribuci se střední hodnotou μ , potom odhadem střední hodnoty bude výběrový průměr pozorování \bar{x} ve tvaru

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Pro zastavení této metody použijeme podmínku pro rozdíl dvou po sobě očekávaných časů pokrytí (tzn. výběrových průměrů).

Tato simulace splňuje všechny vlastnosti metody Monte Carlo, která spočívá v nalezení řešení pomocí opakovaného generování realizací náhodného pokusu.

Na obrázku 4.1 je použita tato metoda pro výpočet očekávaných časů pokrytí specifických typů grafů (úplný graf, cesta, kružnice a wheel graf) s rostoucím počtem vrcholů.



Obrázek 4.1: Numerické očekávané časy pokrytí specifických grafů s rostoucím počtem vrcholů N .

4.2 Závislost očekávaného času pokrytí na minimálním stupni grafu

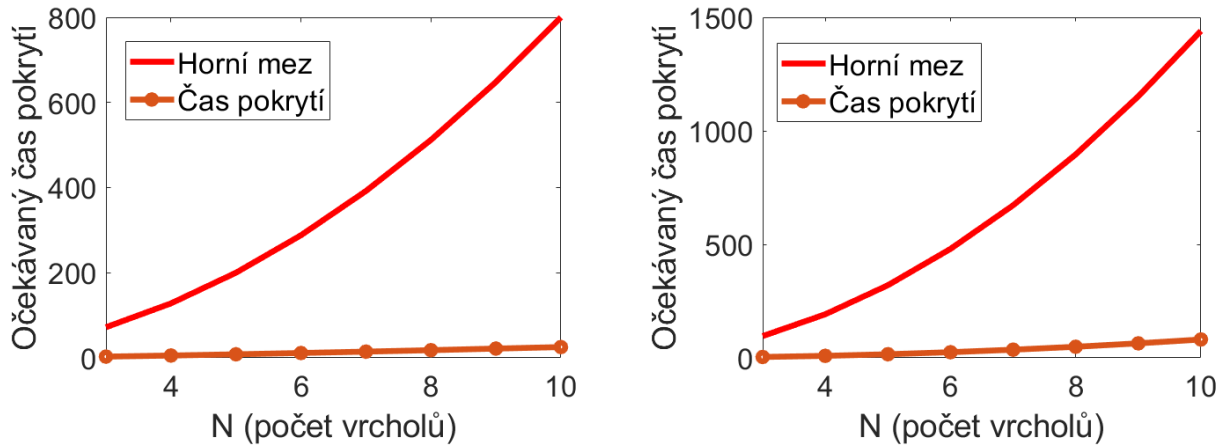
Tato kapitola pojednává o numerickém výpočtu očekávaného času pokrytí grafu $G = (V, H)$ v závislosti na jeho minimálním stupni. Pro libovolný počáteční vrchol $v \in V(G)$ platí následující odhad ([3], Theorem 2.)

$$C_v(G) \leq 16 \cdot \frac{|V| \cdot |H|}{d_{\min}(G)}. \quad (4.2.1)$$

Z této nerovnosti lze usuzovat, že s rostoucím minimálním stupněm grafu G bude klesat očekávaný čas pokrytí. Toto tvrzení se pokusíme ověřit pomocí speciální konstrukce série grafů, ve které bude při neměnném počtu vrcholů růst minimální stupeň grafu G přidáním dodatečné hrany.

Na obrázku 4.2 je znázorněna kvalita tohoto odhadu pro úplný graf a cestu, kterou zkoumáme porovnáním numerických očekávaných časů pokrytí a horního odhadu očekávaného času pokrytí (4.2.1) náhodné procházky pro úplný graf a cestu. Z obrázků je

patrné, že horní mez očekávaného času pokrytí je velmi vzdálená od numerických hodnot očekávaného času pokrytí.



Obrázek 4.2: Porovnání numerických očekávaných časů pokrytí úplného grafu \mathcal{K}_N (v levém bloku) a cesty \mathcal{P}_N (v pravém bloku) s odhadem horní meze očekávaného času pokrytí (4.2.1) při rostoucím počtu vrcholů N .

4.2.1 Konstrukce posloupnosti grafů

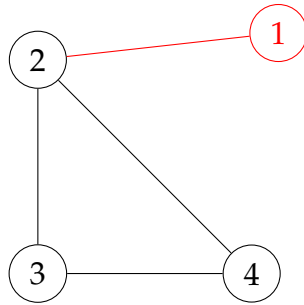
Pro simulaci budeme konstruovat posloupnost grafů $\{G_{d,N}\}_{d=1}^{N-1}$, kde N je fixním počtem vrcholů pro danou posloupnost a d je minimálním stupněm grafu. Postup pro konstrukci posloupnosti:

1. Spojíme $(N - 1)$ vrcholů mezi sebou hranami tak, aby vznikl úplný graf jakožto podgraf grafu G a poslední N -tý vrchol spojíme libovolně pouze s jedním vrcholem, tím vznikne první graf posloupnosti, a to $G_{1,N}$.
2. V každém kroku propojíme N -tý vrchol hranou s dalším vrcholem z podgrafu.
3. Opakujeme, dokud nevznikne úplný graf o N vrcholech.

Je patrné, že tímto postupem budeme vytvářet grafy, které budou mít minimální stupeň grafu vždy o 1 větší při neměnném počtu vrcholů N a s rostoucím počtem hran. Poslední graf vytvořený tímto postupem je úplný graf $\mathcal{K}_N = G_{N-1,N}$.

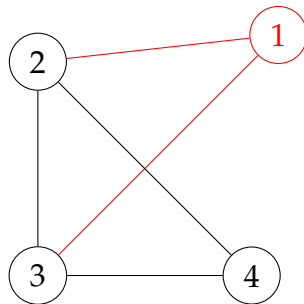
Pro lepší představu si postup ukážeme na obrázcích konstrukce posloupnosti grafů $\{G_{d,4}\}_{d=1}^{N-1}$.

Zvolíme si vrcholy, které mezi sebou spojíme hranami tak, aby vznikl úplný podgraf na 3 vrcholech (v našem případě budeme uvažovat vrcholy 2, 3 a 4). Zbývající vrchol 1 spojíme libovolně s jedním z těchto vrcholů úplného podgrafu (v našem případě spojíme vrchol 1 a 2). Tento krok je uveden na obrázku 4.3.



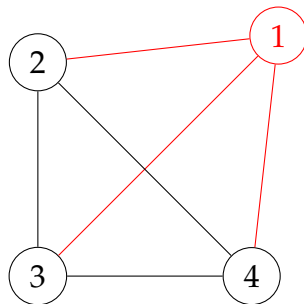
Obrázek 4.3: $G_{1,4}$: první graf posloupnosti vzniklý přidáním hrany $\{1, 2\}$.

V dalším kroku spojíme vrchol 1 s dalším vrcholem našeho úplného podgrafu (viz obrázek 4.4).



Obrázek 4.4: $G_{2,4}$: druhý graf posloupnosti vzniklý přidáním hrany $\{1, 3\}$.

Tento postup opakujeme, dokud nevznikne úplný graf. V našem případě stačí doplnit poslední chybějící hranu a konstrukce celé série grafů je hotová (viz obrázek 4.5).



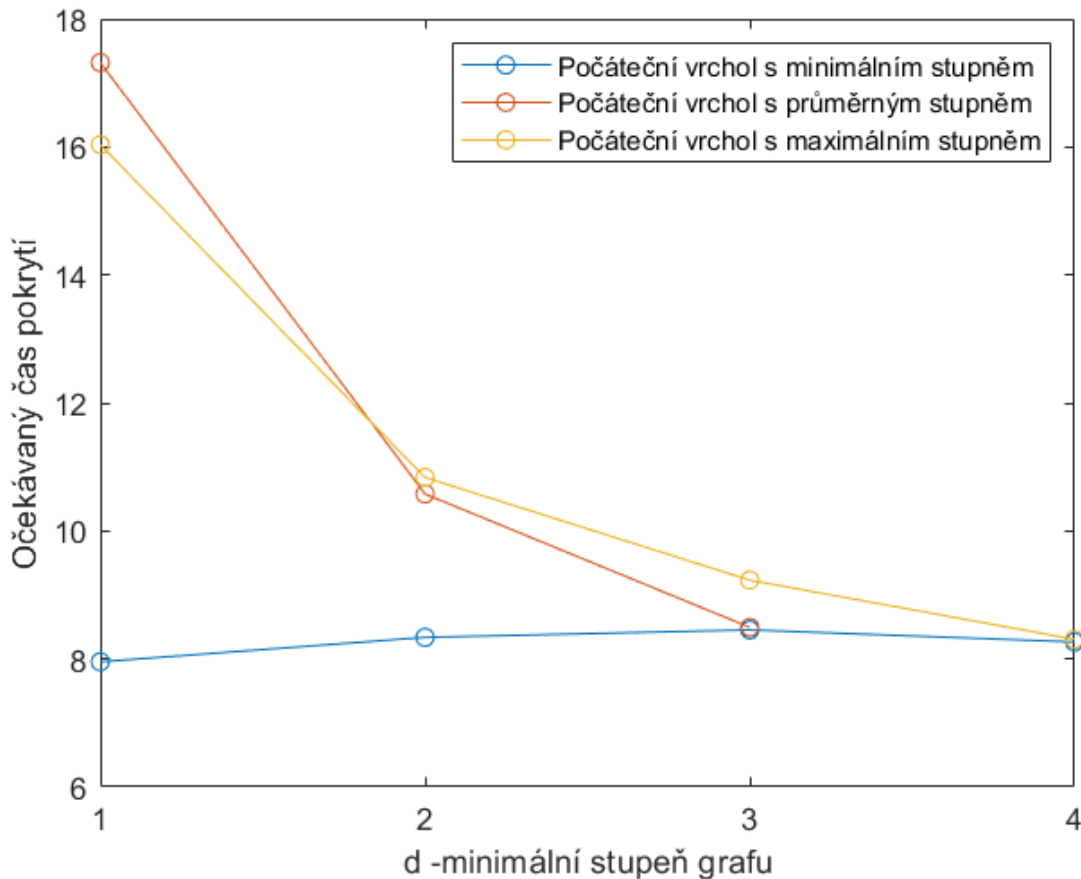
Obrázek 4.5: $G_{3,4}$: třetí a poslední graf posloupnosti vzniklý přidáním hrany $\{1, 4\}$.

4.2.2 Výsledky

Nyní si ukážeme očekávané časy pokrytí pro sérii grafů na 5 vrcholech. Z níže uvedeného obrázku 4.6 je patrné, že při nízkém minimálním stupni grafu je důležitá volba počátečního vrcholu náhodné procházky. V případě, že začneme ve vrcholu s minimálním stupněm, pokryjeme graf nejrychleji. Další zajímavostí je, že v tomto případě dochází

s rostoucím minimálním stupněm k navyšování očekávaného času pokrytí. Naopak pozorujeme, že pro počáteční vrchol s vyšším stupněm je nutno udělat více kroků k pokrytí celého grafu.

Z obrázku 4.6 je dále patrné, že očekávaný čas pokrytí klesá s rostoucím minimálním stupněm grafu (kromě případu, kdy začínáme ve vrcholu s minimálním stupněm). Tím jsme ověřili naši domněnku vycházející z výrazu (4.2.1).



Obrázek 4.6: Na tomto obrázku znázorňujeme numerický očekávaný čas pokrytí v případě náhodné procházky na posloupnosti grafů $\{G_{d,5}\}_{d=1}^4$, ve kterém rozlišujeme 3 případy podle stupně počátečního vrcholu. Modrá křivka znázorňuje numerické očekávané časy pokrytí náhodné procházky, když začneme ve vrcholu s minimálním stupněm. Oranžová křivka znázorňuje případ, kdy počáteční vrchol má průměrný stupeň (v našem případě stupeň 3). Žlutá křivka znázorňuje případ, kdy počáteční vrchol má maximální stupeň. Pro $d = 3$ nastane situace, kdy minimální stupeň grafu bude stejný jako průměrný. Nakonec pro $d = 4$, dojde k rovnosti minimálního a maximálního stupně grafu.

Kapitola 5

Závěr

V této práci jsme si představili očekávaný čas pokrytí grafu, který jsme dále odvodili pro úplný graf a cestu. U úplného grafu jsme ukázali dva způsoby odvození věty 3.2.1. První důkaz vycházel z myšlenky řetězení očekávaných mezičasů pokrytí, které vyjadřují očekávaný čas na pokrytí dalšího vrcholu (tento přístup ilustruje obrázek 3.3). Druhý postup odvození je založen na myšlence indukovaného Markovova řetězce uvedené v podkapitole 3.1.

Tento postup byl použit i pro odvození očekávaného času pokrytí cesty začínající na kraji (viz věta 3.3.1). V případě nekrajových vrcholů jsme zjišťovali očekávanou dobu pokrytí jednoho z krajních vrcholů, jelikož očekávaný čas pokrytí cesty z krajního vrcholu již známe (viz věta 3.3.2).

Poslední kapitola 4 byla věnována numerickému výpočtu očekávaného času pokrytí pomocí simulace v programu Matlab. Uvedli jsme numerickou metodu pro výpočet očekávaného času pokrytí. Výsledky této metody jsou ilustrovány na obrázku 4.1, ve kterém porovnáváme očekávané časy pokrytí pro úplný graf, cestu, kružnici a wheel graf. Dále jsme zkoumali, jak minimální stupeň grafu ovlivní očekávaný čas pokrytí náhodné procházky na grafu.

Pomocí podobných postupů prezentovaných v této práci by mohly být dopočteny očekávané časy pokrytí pro další typické grafy, jako je kružnice nebo wheel graf.

V neposlední řadě vyvstávají otázky ohledně dalších vlastností grafu, které by mohly ovlivnit očekávaný čas pokrytí. Za tímto účelem by mohl být použit numerický výpočet očekávaného času pokrytí (viz kapitola 4).

Literatura

- [1] FRIESL, M. Hypertextové přednášky z předmětu KMA/PSA. <https://courseware.zcu.cz/portal/studium/courseware/kma/dma>, 2020.
- [2] HOLUB, P. Přednášky z předmětu KMA/DMA. <https://courseware.zcu.cz/portal/studium/courseware/kma/dma>, 2020.
- [3] KAHN, J. D., LINIAL, N., NISAN, N., AND SAKS, M. E. On the cover time of random walks on graphs. *Journal of Theoretical Probability* 2, 1 (leden 1989), 121–128.
- [4] KELLEY, W. G., AND PETERSON, A. C. *Difference Equations: An Introduction with Applications*. Academic Press, 1991.
- [5] MOHAMED, S., AND ABDUL, R. Short random walks for community discovery in social networks. In *International Conference on Innovative Computing, Information and Communication Technology* (India, 2009), Sri Sairam Engineering College, pp. 1–4.
- [6] POSPÍŠIL, J. Přednášky z předmětu KMA/ZNP. <https://almamather.zcu.cz/2021/ZNP/prednasky>, 2021.
- [7] PRÁŠKOVÁ, Z., AND LACHOUT, P. *Základy náhodných procesů*. Univerzita Karlova v Praze - Nakladatelství Karolinum, 2001.
- [8] SEHKON, R., AND BLOOM, R. *Applied finite mathematics*. De Anza College, 2021.
- [9] USMANI, R. A. Inversion of Jacobi's tridiagonal matrix. *Computers Math. Applic.* 27, 8 (duben 1994), 59–66.