



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Rozálie Kluvancová

**Coupling, transportní metrika a
aplikace pro přibližné počítání**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Michaela Prokešová, Ph.D.

Studijní program: Obecná matematika

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Ráda bych poděkovala RNDr. Michaelle Prokešové, Ph.D., za podporu a cenné připomínky. Dále bych ráda poděkovala Evě Šmejkalové, Evě Březinové a Marku Bedřichovi za hodiny společného studia v Národní technické knihovně.

Název práce: Coupling, transportní metrika a aplikace pro přibližné počítání

Autor: Rozálie Kluvancová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Michaela Prokešová, Ph.D.

Abstrakt: Důležitou vlastností markovských řetězců s diskrétním časem a konečnou množinou stavů je rychlost konvergence marginálního rozdělení řetězce ke stacionárnímu rozdělení (neboli rychlost mixingu). Pokud zkonstruuujeme coupling dvou markovských řetězců se stejnou maticí pravděpodobností přechodu, kdy jeden startuje ze stacionárního rozdělení a druhý z pevného stavu, můžeme ho použít k odhadu rychlosti mixingu. Cílem práce je popsat, jak můžeme takový coupling sestavit pomocí transportní metriky, a aplikovat tuto metodu při přibližném počítání prvků množiny všech přípustných obarvení grafu.

Klíčová slova: coupling pravděpodobnostních rozdělení, transportní metrika, přibližné počítání

Title: Coupling, transportation metrics and applications to approximate counting

Author: Rozálie Kluvancová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Michaela Prokešová, Ph.D.

Abstract: An important property of discrete-time Markov chains with finite state space is the rate of convergence of the marginal distribution of the chain to the stationary distribution (i.e. mixing rate). If we construct a coupling of two Markov chains with the same transition matrix, where one starts from a stationary distribution and the other from a fixed state, we can use it to estimate the mixing rate. The main goal of this thesis is to describe how we can construct such a coupling using the transportation metric, and to apply this method to approximate counting of all proper colorings of the graph.

Keywords: coupling of probability distributions, transportation metric, approximate counting

Obsah

Značení	2
Úvod	3
1 Základní definice a tvrzení	4
2 Konvergence ke stacionárnímu rozdělení	8
3 Coupling pravděpodobnostních rozdělení	11
4 Transportní metrika, coupling trajektorií a přibližné počítání	13
4.1 Transportní metrika	13
4.2 Coupling trajektorií	14
4.3 Rychlý mixing	17
4.4 Přibližné počítání obarvení grafu	19
4.5 Vylepšený odhad výpočetní náročnosti pro přibližné počítání obarvení grafu	23
Seznam použité literatury	27

Značení

$\mathbf{1}$	Indikátorová funkce
δ_x	Diracova míra v x
$ \Omega $	Mohutnost množiny Ω
\mathbf{P}	Pravděpodobnost
$\mathbf{P}\{A B\}$	Podmíněná pravděpodobnost jevu A za podmínky B
$(X_t)_{t=0}^\infty$	Markovský řetězec s diskrétním časem
\mathcal{X}	Množina stavů
P	Matice pravděpodobností přechodu
$P(x, y)$	Pravděpodobnost přechodu z jevu x do jevu y
$P^t(x, y)$	Pravděpodobnost přechodu ze stavu x do stavu y v t krocích
$P^t(x, \cdot)$	Pravděpodobnostní rozdělení markovského řetězce po t krocích ze stavu x
π	Stacionární rozdělení
μ_t	Rozdělení náhodné veličiny X_t
\mathbf{E}	Střední hodnota
\mathbf{E}_ν	Střední hodnota za podmínky, že počáteční rozdělení je rovno ν
\mathbf{E}_x	Střední hodnota za podmínky, že řetězec startoval ze stavu x
Var	Rozptyl
Var $_\nu$	Rozptyl za podmínky, že počáteční rozdělení je rovno ν
Var $_x$	Rozptyl za podmínky, že řetězec startoval ze stavu x
\mathbf{P}_ν	Pravděpodobnost za podmínky, že počáteční rozdělení je rovno ν
\mathbf{P}_x	Pravděpodobnost za podmínky, že řetězec startoval ze stavu x
t_{mix}	Čas mixingu
$\ \mu - \nu\ _{\text{TV}}$	Vzdálenost μ a ν v totální variaci
ρ_K	Transportní metrika
$\text{diam}(\mathcal{X})$	Průměr prostoru \mathcal{X}
$x \sim y$	Vrchol x sousedí s vrcholem y
$\text{deg}(x)$	Stupeň vrcholu x
Δ	Maximální stupeň grafu
$A_v(x)$	Množina všech přípustných barev vrcholu v při obarvení x
$\mathcal{X}(x, v)$	Množina všech obarvení, která se liší od x pouze ve vrcholu v

Úvod

Klíčovou vlastností markovských řetězců s diskretním časem a konečnou množinou stavů je rychlost konvergence marginálního rozdělení řetězce ke stacionárnímu rozdělení (neboli rychlost *mixingu*). Pokud zkonstruujeme *coupling* dvou markovských řetězců se stejnou maticí pravděpodobností přechodu, kdy jeden startuje ze stacionárního rozdělení a druhý z pevného stavu, můžeme ho použít k odhadu rychlosti mixingů. Jednou z možností, jak takový coupling vytvořit, je pomocí *transportní metriky*. Náplní práce je představení uvedených pojmů a metody, a popsání, jak může být tato metoda aplikována v oblasti přibližného počítání prvků nějaké velké kombinatorické množiny. Speciálně to v práci ukážeme na příkladu množiny všech přípustných obarvení grafu.

Práce je členěna do čtyř kapitol. V první kapitole si připomeneme základní teorii k homogenním markovským řetězcům s diskretním časem a uvedeme potřebné definice z teorie grafů, se kterými budeme pracovat ve čtvrté kapitole. Ve druhé kapitole popíšeme, za jakých podmínek konverguje marginální rozdělení markovského řetězce ke stacionárnímu. Třetí kapitola se věnuje couplingu pravděpodobnostních rozdělení a jeho vlastnostem, které pro nás budou užitečné.

Jádro práce je obsaženo ve čtvrté kapitole. Nejprve zavedeme metriku na prostoru všech pravděpodobnostních rozdělení na množině stavů markovského řetězce. Pokud bude možné ztotožnit množinu stavů řetězce s množinou vrcholů nějakého grafu, coupling zkonstruovaný pomocí této metriky nám dále umožní přesně omezit rychlost mixingů.

Závěr kapitoly se věnuje nastínění problému přibližného počítání prvků velkých kombinatorických množin, které využívá tzv. *zcela polynomiální znáhodněné aproximační schéma*. V případě přibližného počítání všech přípustných obarvení grafu jde o algoritmus, který pro každé $1 > \varepsilon > 0$ dokáže s pravděpodobností minimálně $1 - \frac{1}{\text{poly}(n)}$ a s přesností až na multiplikativní faktor $1 \pm \varepsilon$ odhadnout tento počet v čase, který je polynomiální v $n = |V|$ a ε^{-1} . Ukážeme, že při použití nezávislých realizací markovského řetězce, jakožto nezávislých vzorků z množiny všech přípustných obarvení grafu, přibližné počítání prvků této množiny má asymptotickou složitost $\mathcal{O}(n^3 \log n)$. Dále dokážeme, že pokud místo nezávislých realizací pracujeme s ergodickými průměry tohoto řetězce, tak můžeme podstatným způsobem snížit počet simulací k dosažení stejné přesnosti a dosáhnout asymptotické složitosti $\mathcal{O}(n^3)$. Další výsledky a vylepšení metody využívající markovské řetězce při vzorkování z množiny obarvení daného grafu lze najít například v přehledovém článku Frieze a Vigoda (2007).

Práce je založena na poznatcích z knihy Levin a Peres (2017). Přínosem autorky je přehledné a kompaktní pojednání daného tématu, doplnění mezikroků a oprava drobných chyb v důkazech a vylepšení odhadu na počet kroků simulačního algoritmu z věty 14 v podkapitole 4.5.

1. Základní definice a tvrzení

V první kapitole si uvedeme klíčové definice a tvrzení, které uplatníme v následujících kapitolách. Dodatečnou teorii k homogenním markovským řetězcům lze nalézt v knize Prášková a Lachout (2012).

Základní pojem, se kterým budeme pracovat v následujícím výkladu, je homogenní markovský řetězec s diskretním časem.

Definice 1 (Homogenní markovský řetězec). *Bud' $X = \{X_t, t \in \mathbb{N}_0\}$ náhodný proces s diskretní nejvýše spočetnou množinou stavů \mathcal{X} . Poté X je homogenní markovský řetězec s diskretním časem s maticí pravděpodobností přechodu P , pokud platí*

$$\mathbf{P} \left\{ X_{t+1} = j \mid \left\{ X_t = i \right\} \cap \bigcap_{i=1}^t \left\{ X_{i-1} = i_{i-1} \right\} \right\} = \mathbf{P} \{ X_{t+1} = j \mid X_t = i \} = P(i, j),$$

pro všechna $t \in \mathbb{N}_0$ a všechna $j, i, i_{t-1}, \dots, i_0 \in \mathcal{X}$ taková, že

$$\mathbf{P} \left\{ \left\{ X_t = i \right\} \cap \bigcap_{i=1}^t \left\{ X_{i-1} = i_{i-1} \right\} \right\} > 0.$$

V celé práci se omezíme pouze na případ, kdy \mathcal{X} je diskretní **konečná**.

Rozdělení náhodné veličiny $X_t, t \in \mathbb{N}_0$, budeme značit řádkovým vektorem μ_t , tj. $\mu_t(x) = \mathbf{P}(X_t = x)$ pro $x \in \mathcal{X}$. Poté pro každé $x \in \mathcal{X}$ a $t \in \mathbb{N}_0$ platí

$$\mu_{t+1}(x) = \sum_{y \in \mathcal{X}} \mathbf{P}\{X_t = y\} P(y, x) = \sum_{y \in \mathcal{X}} \mu_t(y) P(y, x)$$

a tedy $\mu_{t+1} = \mu_t P$. Celkově dostáváme platnost $\mu_t = \mu_0 P^t$.

Od této chvíle budeme výrazy *markovský řetězec* či *markovský řetězec s maticí přechodu P a množinou stavů \mathcal{X}* označovat homogenní markovský řetězec s diskretním časem s maticí pravděpodobností přechodu P a s množinou stavů \mathcal{X} .

Dále definujeme *počáteční rozdělení* markovského řetězce jakožto rozdělení μ_0 náhodné veličiny X_0 . Pro zjednodušení zápisu budeme označovat $\mathbf{P}_\mu(\cdot)$, respektive $\mathbf{E}_\mu(\cdot)$, kde μ je libovolné pravděpodobnostní rozdělení na \mathcal{X} , pravděpodobnost nějakého náhodného jevu, respektive střední hodnotu nějaké náhodné veličiny, za předpokladu, že $\mu_0 = \mu$. Speciálně pro $\mu = \delta_x$, kde $x \in \mathcal{X}$, značíme $\mathbf{P}_x(\cdot)$ a $\mathbf{E}_x(\cdot)$.

Definice 2 (Nerozložitelný řetězec). *Markovský řetězec se nazývá nerozložitelný, pokud pro každé dva stavy $x, y \in \mathcal{X}$ existuje přirozené číslo t , pro které platí $P^t(x, y) > 0$.*

Definice 3 (Aperiodický řetězec). *Pro $x \in \mathcal{X}$ definujeme*

$$\tau(x) = \{t \geq 1 : P^t(x, x) > 0\}.$$

Periodu stavu x definujeme jako největší společný dělitel množiny $\tau(x)$. Řekneme, že markovský řetězec je aperiodický, pokud mají všechny jeho stavy periodu 1.

Nyní bez důkazu uvedeme tvrzení, které říká, že pro aperiodický nerozložitelný řetězec s konečnou množinou stavů existuje minimální počet kroků, které jsou potřebné, abychom se s kladnou pravděpodobností mohli dostat z libovolného stavu do dalšího libovolného stavu.

Tvrzení 1. *Pro každý aperiodický nerozložitelný markovský řetězec s konečnou množinou stavů existuje přirozené číslo r_0 takové, že pro $r \geq r_0$, pro každé $x, y \in \mathcal{X}$ platí $P^r(x, y) > 0$.*

Důkaz. Levin a Peres (2017, tvrzení 1.7, strana 7)

□

Definice 4 (Stacionární rozdělení). *Pravděpodobnostní rozdělení π na množině stavů \mathcal{X} markovského řetězce s maticí přechodu P nazveme stacionárním rozdělením, pokud platí*

$$\pi(y) = \sum_{x \in \mathcal{X}} \pi(x)P(x, y) \quad \text{pro každé } y \in \mathcal{X}.$$

V maticovém zápisu

$$\pi = \pi P.$$

Pokud π je stacionární rozdělení a $\mu_0 = \pi$, poté $\mu_t = \mu_0 P^t = \pi P^t = \pi$ pro každé $t \geq 0$.

Bez důkazu uvedeme klíčové tvrzení o existenci jednoznačně určeného stacionárního rozdělení nerozložitelného markovského řetězce s konečnou množinou stavů.

Tvrzení 2. *Mějme nerozložitelný markovský řetězec s konečnou množinou stavů. Poté existuje jeho stacionární rozdělení a je určené jednoznačně.*

Důkaz. Levin a Peres (2017, kapitola 1.5, rozloženo mezi více tvrzení)

□

Pro připomenutí uvedeme Čebyševovu nerovnost, na kterou se odkážeme v důkazu věty 12.

Tvrzení 3 (Čebyševova nerovnost). *Bud X reálná náhodná veličina s konečným druhým momentem. Poté pro $a > 0$ platí*

$$\mathbf{P}\{|X - \mathbf{E}X| \geq a\} \leq \frac{\text{Var}(X)}{a^2}.$$

Nyní uvedeme základní definice z teorie grafů, se kterými budeme pracovat ve čtvrté kapitole.

Definice 5 (Graf). *Grafem G rozumíme uspořádanou dvojici $G = (V, E)$, kde V nazýváme množina vrcholů a $E \subseteq \{\{x, y\} : x, y \in V, x \neq y\}$ nazýváme množina hran.*

V našem případě bude V vždy *konečná* a hrany definujeme jakožto *neuspořádané* dvojice dvou *různých* vrcholů. Pokud $\{x, y\} \in E$, píšeme $x \sim y$ a říkáme, že x a y jsou *sousedící* vrcholy. *Stupeň* vrcholu x definujeme jako počet vrcholů s ním sousedících a značíme $\deg(x)$. *Cestou* mezi vrcholy x a y rozumíme posloupnost sousedících vrcholů, která začíná v x a končí v y . V této práci se budeme zabývat pouze *souvislými* grafy. To jsou takové grafy, v nichž pro každé dva vrcholy existuje cesta, která je spojuje.

Ve čtvrté kapitole budeme popisovat vlastnosti markovských řetězců, jejichž množina stavů se dá přepsat do tvaru S^V , kde V je množina vrcholů nějakého grafu a S je konečná množina. Prvky S^V jsou zobrazení $f : V \rightarrow S$ a budeme je dále nazývat *konfigurace*.

Definice 6 (Přípustné q -obarvení grafu). *Bud' $q \in \mathbb{N}$. Mějme množinu barev $\{1, 2, \dots, q\}$ a graf $G = (V, E)$. Přípustným q -obarvením grafu G rozumíme libovolnou konfiguraci $x \in \{1, 2, \dots, q\}^V$, pro kterou platí $x(v) \neq x(w)$ kdykoliv $\{v, w\} \in E$.*

Definice 7 (Glauberova dynamika). *Bud' V a S konečné množiny a $\mathcal{X} \subseteq S^V$. Pro $x \in \mathcal{X}$ a $v \in V$, označme*

$$\mathcal{X}(x, v) = \{y \in \mathcal{X} : y(w) = x(w) \forall w \neq v\}$$

množinu všech stavů, které se od x liší pouze ve vrcholu v . Bud' π pravděpodobnostní rozdělení na S^V , pro něž platí $\pi(\mathcal{X}) = 1$. Glauberovou dynamikou pro π rozumíme markovský řetězec s množinou stavů \mathcal{X} , který se vyvíjí následovně:

Nechť se řetězec nachází ve stavu $x \in \mathcal{X}$. Zvolíme rovnoměrně náhodně vrchol v z množiny V a označíme ho v , poté řetězec přejde do stavu y s pravděpodobností $\pi^{x,v}(y)$, kde

$$\pi^{x,v}(y) = \pi(y | \mathcal{X}(x, v)) = \begin{cases} \frac{\pi(y)}{\pi(\mathcal{X}(x, v))}, & \text{pokud } y \in \mathcal{X}(x, v), \\ 0, & \text{pokud } y \notin \mathcal{X}(x, v). \end{cases}$$

Z definice je zřejmé, že pro $x, y \in \mathcal{X}$ platí

$$P(x, y) = \sum_{v \in V} \frac{1}{|V|} \pi^{x,v}(y).$$

Glauberova dynamika pro π má jednu obzvlášť důležitou vlastnost a to, že π je stacionární rozdělení tohoto řetězce, neboť ze spočetnosti \mathcal{X} pro každé $y \in \mathcal{X}$ platí

$$\sum_{x \in \mathcal{X}} \pi(x) \sum_{v \in V} \frac{1}{|V|} \pi^{x,v}(y) = \sum_{v \in V} \frac{1}{|V|} \sum_{x \in \mathcal{X}} \pi(x) \pi^{x,v}(y) = \sum_{v \in V} \frac{1}{|V|} \pi(y) = \pi(y).$$

Ve čtvrté kapitole budeme zkoumat rychlost konvergence marginálního rozdělení Glauberovy dynamiky k jejímu stacionárnímu rozdělení. Glauberova dynamika pro q -obarvení souvislého grafu je aperiodický řetězec, neboť zřejmě pro každé $x \in \mathcal{X}, v \in V$ platí $x \in \mathcal{X}(x, v)$, a je nerozložitelný, neboť se můžeme z každého obarvení dostat konečnou posloupností kroků do libovolného jiného obarvení.

Bud' π rovnoměrné rozdělení na množině všech přípustných q -obarvení grafu. Jedním z cílů této práce je odhadnutí počtu prvků této množiny. Budeme k tomu potřebovat rozšířit definici Glauberovy dynamiky pro π na množinu všech možných q -obarvení grafu.

Definice 8 (Modifikovaná Glauberova dynamika pro obarvení grafu). *Mějme souvislý graf $G = (V, E)$ a $q \in \mathbb{N}$. Bud $\mathcal{X} = \{1, 2, \dots, q\}^V$ množina všech q -obarvení grafu G . Bud π rovnoměrné rozdělení na množině všech přípustných q -obarvení grafu G . Modifikovanou Glauberovou dynamikou pro π rozumíme řetězec, který se vyvíjí následovně:*

Nechť se řetězec nachází ve stavu $x \in \mathcal{X}$, náhodně zvolíme vrchol z množiny V , označíme ho w , a rovnoměrně náhodně mu přiřadíme barvu z množiny všech barev, které sousedící vrcholy w při obarvení x nemají.

Pokud takový řetězec začne v přípustném q -obarvení, poté se již vždy nachází v množině všech přípustných q -obarvení G .

2. Konvergence ke stacionárnímu rozdělení

V této kapitole zavedeme složitější nástroje užitečné k popisu dlouhodobého chování markovských řetězců. Bude nás především zajímat odhad rychlosti konvergence marginálního rozdělení řetězce ke stacionárnímu. Je tedy třeba zavést metriku, vůči níž budeme tuto konvergenci uvažovat.

Definice 9 (Vzdálenost v totální variaci). *Budte μ a ν pravděpodobnostní rozdělení na prostoru \mathcal{X} . Vzdálenost v totální variaci rozdělení μ a ν definujeme předpisem*

$$\|\mu - \nu\|_{\text{TV}} = \max_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)|. \quad (2.1)$$

Při výpočtu vzdálenosti v totální variaci může být použití vztahu (2.1) relativně nepohodlné. Uvedeme nyní alternativní definici, která je výpočetně jednodušší.

Tvrzení 4. *Bud μ a ν pravděpodobnostní rozdělení na prostoru \mathcal{X} , kde množina \mathcal{X} je diskrétní a konečná. Poté platí*

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|. \quad (2.2)$$

Důkaz. Levin a Peres (2017, tvrzení 4.2, strana 48) □

Ze vztahu (2.2) je již zřejmé, že vzdálenost v totální variaci je metrika na množině všech pravděpodobnostních rozdělení na prostoru \mathcal{X} .

Nyní můžeme dokázat větu o konvergenci, která tvrdí, že marginální rozdělení nerozložitelného aperiodického markovského řetězce s konečnou množinou stavů konverguje k jeho stacionárnímu rozdělení v totální variaci.

Věta 5 (O konvergenci). *Bud P matice přechodu nerozložitelného aperiodického markovského řetězce s konečnou množinou stavů a π jeho stacionární rozdělení. Poté existuje $\alpha \in (0, 1)$ a $C > 0$ takové, že*

$$\max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq C\alpha^t. \quad (2.3)$$

Důkaz. Z tvrzení 1 víme, že existuje přirozené číslo k takové, že P^k má pouze kladné prvky. Definujme matici Π o rozměrech $|\mathcal{X}| \times |\mathcal{X}|$, jejíž řádky jsou rovny π . Nalezneme $\delta > 0$, které splňuje

$$P^k(x, y) - \delta\pi(y) \geq 0, \quad \text{pro všechna } x, y \in \mathcal{X}.$$

Zřejmě pro každé $x \in \mathcal{X}$ dále platí

$$\sum_{y \in \mathcal{X}} (P^k(x, y) - \delta\pi(y)) = 1 - \delta.$$

Položme $\theta = 1 - \delta$. Poté definujeme stochastickou matici Q jakožto řešení rovnice

$$P^k = (1 - \theta)\Pi + \theta Q. \quad (2.4)$$

Nyní dokážeme indukcí pro $l \geq 1$ platnost vztahu

$$P^{kl} = (1 - \theta^l)\Pi + \theta^l Q^l. \quad (2.5)$$

Pro $l = 1$ vztah platí z (2.4). Předpokládejme, že vztah platí pro l , poté

$$\begin{aligned} P^{k(l+1)} &= P^{kl} P^k = [(1 - \theta^l)\Pi + \theta^l Q^l] P^k \\ &= (1 - \theta^l)\Pi P^k + \theta^l Q^l [(1 - \theta)\Pi + \theta Q] \\ &= (1 - \theta^l)\Pi P^k + (1 - \theta)\theta^l Q^l \Pi + \theta^{l+1} Q^{l+1}. \end{aligned} \quad (2.6)$$

Přímo z definice stochastické matice plyne, že pro každou stochastickou matici M platí $M\Pi = \Pi$. Dále pokud $\pi M = \pi$ pro nějakou matici M , tak i $\Pi M = \Pi$. Zřejmě poté platí $\Pi P^k = \Pi$ a $Q^l \Pi = \Pi$. Můžeme tedy dále upravit (2.6) a dostáváme

$$P^{k(l+1)} = (1 - \theta^{l+1})\Pi + \theta^{l+1} Q^{l+1}.$$

Dokázali jsme tedy platnost vztahu (2.5) pro $l + 1$, a tím i pro všechna $l \geq 1$.

Pokud vynásobíme obě strany rovnosti (2.5) výrazem P^j , $j \in \{1, 2, \dots, k-1\}$, dostáváme

$$\begin{aligned} P^{kl+j} &= [(1 - \theta^l)\Pi + \theta^l Q^l] P^j \\ &= (1 - \theta^l)\Pi P^j + \theta^l Q^l P^j \\ &= (1 - \theta^l)\Pi + \theta^l Q^l P^j \\ &= \Pi + \theta^l (Q^l P^j - \Pi), \end{aligned}$$

a tedy platí

$$P^{kl+j} - \Pi = \theta^l (Q^l P^j - \Pi). \quad (2.7)$$

Buď x_0 libovolný stav z \mathcal{X} . Rozmysleme si nyní, jak vypadají součty absolutních hodnot prvků v řádcích indexovaných x_0 na obou stranách rovnosti (2.7), které jsou navíc vynásobené $\frac{1}{2}$. Na levé straně dostáváme $\|P^{kl+j}(x_0, \cdot) - \pi\|_{\text{TV}}$ ze vztahu (2.2). Na pravé straně je takový řádkový součet výrazu $Q^l P^j - \Pi$ omezen shora maximální vzdáleností v totální variaci dvou rozdělení, tedy 1.

Dostáváme odhad pro libovolné $x_0 \in \mathcal{X}$

$$\|P^{kl+j}(x_0, \cdot) - \pi\|_{\text{TV}} \leq \theta^l = \theta^{\frac{kl+k}{k}-1} \leq \theta^{\frac{kl+j}{k}-1}, \quad (2.8)$$

kde poslední nerovnost platí, neboť $\theta < 1$ a $j < k$.

Pokud označíme $\alpha = \theta^{1/k}$ a $C = \theta^{-1}$, pak z rovnosti (2.8) dostáváme platnost rovnosti (2.3), kterou jsme chtěli dokázat. □

Abychom mohli měřit rychlost konvergence marginálního rozdělení řetězce ke stacionárnímu rozdělení v totální variaci nezávisle na počátečním stavu, zavádíme značení

$$d(t) = \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{\text{TV}},$$

kde \mathcal{X} je konečná množina stavů markovského řetězce s maticí přechodu P a s jednoznačně určeným stacionárním rozdělení π .

Nyní si dokážeme nerostoucnost funkce d . Tedy pokud je vzdálenost v totální variaci stacionárního rozdělení a marginálního rozdělení řetězce po určitém počtu kroků „dostatečně malá“, poté již nikdy nebude větší.

Tvrzení 6. *Bud' P matice pravděpodobností přechodu markovského řetězce s diskrétní konečnou množinou stavů \mathcal{X} a stacionárním rozdělením π . Budte μ a ν libovolná pravděpodobnostní rozdělení na \mathcal{X} . Platí*

$$\|\mu P - \nu P\|_{TV} \leq \|\mu - \nu\|_{TV}. \quad (2.9)$$

Dále pro libovolné $t \geq 0$ platí

$$\|\mu P^{t+1} - \pi\|_{TV} \leq \|\mu P^t - \pi\|_{TV}, \quad (2.10)$$

a tedy zřejmě $d(t+1) \leq d(t)$.

Důkaz. Z (2.2) dostáváme

$$\begin{aligned} \|\mu P - \nu P\|_{TV} &= \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu P(x) - \nu P(x)| \\ &= \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{X}} \mu(y) P(y, x) - \sum_{y \in \mathcal{X}} \nu(y) P(y, x) \right| \\ &= \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{X}} (\mu(y) - \nu(y)) P(y, x) \right| \\ &\leq \frac{1}{2} \sum_{y \in \mathcal{X}} |\mu(y) - \nu(y)| \left(\sum_{x \in \mathcal{X}} P(y, x) \right) \\ &= \frac{1}{2} \sum_{y \in \mathcal{X}} |\mu(y) - \nu(y)| = \|\mu - \nu\|_{TV}. \end{aligned}$$

Druhá část tvrzení plyne přímo z (2.9) pro pravděpodobnostní rozdělení μP^t a π . Nerostoucnost funkce d plyne z platnosti (2.10) pro $\mu = \delta_x$ pro libovolné $x \in \mathcal{X}$. \square

Dále je vhodné zadefinovat parametr, který měří čas potřebný k tomu, aby vzdálenost v totální variaci marginálního rozdělení řetězce a stacionárního rozdělení byla „dostatečně malá“.

Definice 10 (Čas mixingů). *Čas mixingů pro $\varepsilon > 0$ definujeme předpisem*

$$t_{\text{mix}}(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}.$$

Speciálně definujeme $t_{\text{mix}} = t_{\text{mix}}(\frac{1}{4})$.

3. Coupling pravděpodobnostních rozdělání

Nyní zadefinujeme coupling dvou pravděpodobnostních rozdělání a popíšeme jeho vlastnosti. V tvrzení 7 využijeme coupling k určování vzdálenosti mezi pravděpodobnostními rozděláními.

Definice 11 (Coupling dvou pravděpodobnostních rozdělání). *Libovolnou uspořádanou dvojici (X, Y) , kde X a Y jsou náhodné veličiny definované na stejném pravděpodobnostním prostoru, nazveme couplingem rozdělání μ a ν , pokud platí, že marginální rozdělání X je rovno μ a marginální rozdělání Y je rovno ν .*

Poznámka. Coupling dvou pravděpodobnostních rozdělání můžeme definovat pomocí uspořádané dvojice náhodných veličin jako v předchozí definici, nebo pomocí sdruženého rozdělání náhodného vektoru.

Buď (X, Y) coupling pravděpodobnostních rozdělání μ a ν na \mathcal{X} . Označme q sdružené rozdělání náhodného vektoru (X, Y) na $\mathcal{X} \times \mathcal{X}$. Pro rozdělání q definujeme jeho projekci na první souřadnici jako pravděpodobnostní rozdělání na \mathcal{X} , které je rovno

$$q(\cdot \times \mathcal{X}) = \sum_{y \in \mathcal{X}} q(\cdot, y).$$

Analogicky, projekce na druhou souřadnici je rozdělání $q(\mathcal{X} \times \cdot)$.

Z výše uvedeného plyne, že projekce rozdělání (X, Y) na první, respektive druhou, souřadnici je právě μ , respektive ν .

Naopak, pokud q je jako výše definované, poté identická funkce na pravděpodobnostním prostoru $(\mathcal{X} \times \mathcal{X}, q)$ je coupling rozdělání μ a ν .

Nyní popíšeme vztah mezi couplingem dvou pravděpodobnostních rozdělání a jejich vzdáleností v totální variaci.

Pokud nejsou pravděpodobnostní rozdělání identická, není vždy možné, aby X a Y měly stejné hodnoty. Následující tvrzení určuje, jak nejméně se mohou náhodné veličiny X a Y lišit.

Tvrzení 7. *Buď μ a ν pravděpodobnostní rozdělání na prostoru \mathcal{X} . Poté*

$$\|\mu - \nu\|_{\text{TV}} = \inf\{\mathbf{P}\{X \neq Y\} : (X, Y) \text{ je coupling } \mu \text{ a } \nu\}.$$

Důkaz. Levin a Peres (2017, tvrzení 4.7, strana 50)

□

Je zřejmé, že pomocí tohoto tvrzení můžeme jednoduše získat horní odhad vzdálenosti v totální variaci dvou pravděpodobnostních rozdělání.

Analogií couplingu pro náhodné procesy je coupling dvou markovských řetězců.

Definice 12 (Coupling markovských řetězců). *Náhodný proces $(X_t, Y_t)_{t=0}^{\infty}$, pro který platí, že $(X_t)_{t=0}^{\infty}$ a $(Y_t)_{t=0}^{\infty}$ jsou markovské řetězce se stejnou maticí pravděpodobností přechodu P , nazveme coupling markovských řetězců s maticí přechodu P .*

V této práci se budeme zabývat pouze případem, kdy je vytvořený náhodný proces markovským řetězcem.

Definice 13 (Markovský coupling). *Bud' $(X_t)_{t=0}^{\infty}$ a $(Y_t)_{t=0}^{\infty}$ markovské řetězce na množině stavů \mathcal{X} s maticí pravděpodobností přechodu P . Markovským couplingem těchto dvou řetězců rozumíme markovský řetězec $(X_t, Y_t)_{t=0}^{\infty}$ s množinou stavů $\mathcal{X} \times \mathcal{X}$, takový, že pro všechna $x, y, x', y' \in \mathcal{X}$ platí*

$$\begin{aligned}\mathbf{P}\{X_{t+1} = x' \mid X_t = x, Y_t = y\} &= P(x, x') \\ \mathbf{P}\{Y_{t+1} = y' \mid X_t = x, Y_t = y\} &= P(y, y').\end{aligned}$$

Každý markovský coupling dvou řetězců s maticí přechodu P můžeme poznamenit tak, aby od chvíle, co se poprvé oba ocitnou ve stejném stavu, měly totožné trajektorie. Neboli aby platilo

$$\text{pokud } X_s = Y_s, \text{ poté nutně } X_t = Y_t \text{ pro každé } t \geq s.$$

Mějme coupling markovských řetězců $(X_t)_{t=0}^{\infty}$ a $(Y_t)_{t=0}^{\infty}$, pro které platí $X_0 = x$ a $Y_0 = y$, poté budeme značit $\mathbf{P}_{x,y}$ pravděpodobnost na stavovém prostoru na němž jsou $(X_t)_{t=0}^{\infty}$ a $(Y_t)_{t=0}^{\infty}$ definované a analogicky budeme střední hodnotu značit $\mathbf{E}_{x,y}$.

4. Transportní metrika, coupling trajektorií a přibližné počítání

V této kapitole zavedeme transportní metriku na prostoru všech pravděpodobnostních rozdělání na množině stavů řetězce. Poté ztotožníme množinu stavů s množinou vrcholů grafu a popíšeme, jak lze zkonstruovat coupling pomocí transportní metriky. Pomocí tohoto couplingu poté přesně omezíme rychlost mixingů modifikované Glauberovy dynamiky. Dále aplikujeme dosavadně popsanou teorii při přibližném počítání prvků množiny všech přípustných obarvení grafu.

4.1 Transportní metrika

V této sekci zavedeme metriku na prostoru všech pravděpodobnostních rozdělání na \mathcal{X} , která bude rozšiřovat definici metriky na \mathcal{X} , a v lemmatu 8 ukážeme, že je to opravdu metrika.

Definice 14. *Mějme danou metriku ρ na množině stavů \mathcal{X} . Transportní metriku dvou pravděpodobnostních rozdělání na \mathcal{X} definujeme předpisem*

$$\rho_K(\mu, \nu) = \inf\{\mathbf{E}\rho(X, Y) : (X, Y) \text{ je coupling } \mu \text{ a } \nu\}. \quad (4.1)$$

Všimněme si, že pro libovolné $x, y \in \mathcal{X}$ platí $\rho_K(\delta_x, \delta_y) = \rho(x, y)$. Transportní metrika ρ_K je tedy rozšířením metriky ρ . Dále pro $\rho(x, y) = \mathbf{1}\{x \neq y\}$ platí $\rho_K(\mu, \nu) = \|\mu - \nu\|_{\text{TV}}$.

Poznámka. Rozepsáním střední hodnoty (dle její definice) a pomocí definice pravděpodobnostního rozdělání q na $\mathcal{X} \times \mathcal{X}$, které je detailně popsáno v poznámce za definicí 11, můžeme transportní metriku ekvivalentně definovat jako

$$\rho_K(\mu, \nu) = \inf\left\{ \sum_{(x,y) \in \mathcal{X} \times \mathcal{X}} \rho(x, y)q(x, y) : q(\cdot \times \mathcal{X}) = \mu, q(\mathcal{X} \times \cdot) = \nu \right\}. \quad (4.2)$$

Poznámka. Množinu všech pravděpodobnostních rozdělání na $\mathcal{X} \times \mathcal{X}$ můžeme ztotožnit s $(|\mathcal{X}|^2 - 1)$ -dimenzionálním simplexem, což je kompaktní podmnožina $\mathbb{R}^{|\mathcal{X}|^2}$. Množina všech pravděpodobnostních rozdělání na $\mathcal{X} \times \mathcal{X}$ jejichž projekce na první, respektive druhou, souřadnici je rovna μ , respektive ν , je uzavřená podmnožina tohoto simplexu a je tedy taktéž kompaktní. Funkce

$$q \mapsto \sum_{(x,y) \in \mathcal{X} \times \mathcal{X}} \rho(x, y)q(x, y)$$

je spojitá na této množině. Existuje tedy pravděpodobnostní rozdělání q_* , ve kterém suma v (4.2) nabývá infima. Takové rozdělání q_* nazýváme ρ -optimální coupling pravděpodobnostních rozdělání μ a ν . Analogicky existuje dvojice náhodných veličin (X_*, Y_*) , které také nazýváme ρ -optimální coupling, splňující

$$\mathbf{E}(\rho(X_*, Y_*)) = \rho_K(\mu, \nu).$$

Lemma 8. *Funkce ρ_K definovaná v (4.1) je metrika na prostoru pravděpodobnostních rozdělání na \mathcal{X} .*

Důkaz. Platnost axiomu nezápornosti a axiomu symetrie plyne přímo z definice (4.1) a z platnosti axiomu nezápornosti a axiomu symetrie pro metriku ρ .

Z nezápornosti metriky ρ platí, že pokud $\rho_K(\mu, \nu) = 0$, pak $\mathbf{E}(\rho(X_*, Y_*)) = \rho_K(\mu, \nu) = 0$, kde $\rho(X_*, Y_*)$ je ρ -optimální coupling μ a ν . Z platnosti axiomu totožnosti pro metriku ρ , musí být náhodné veličiny X_* a Y_* skoro jistě stejné. Pokud jsou náhodné veličiny X_* a Y_* skoro jistě stejné, poté z definice couplingu mají X_* a Y_* stejné rozdělení a tedy $\mu = \nu$. Naopak necht X má rozdělení μ . Poté

$$0 \leq \rho_K(\mu, \mu) \leq \mathbf{E}(\rho(X, X)) = 0,$$

kde poslední rovnost plyne z platnosti axiomu totožnosti pro metriku ρ . Tím jsme dokázali axiom totožnosti pro ρ_K .

Zbývá nám dokázat platnost trojúhelníkové nerovnosti. Necht μ, ν a η jsou pravděpodobnostní rozdělení na \mathcal{X} . Buď p , respektive q , pravděpodobnostní rozdělení na $\mathcal{X} \times \mathcal{X}$ a zároveň coupling μ a ν , respektive ν a η . Definujme pravděpodobnostní rozdělení r na prostoru $\mathcal{X} \times \mathcal{X} \times \mathcal{X}$ předpisem

$$r(x, y, z) = \frac{p(x, y)q(y, z)}{\nu(y)}.$$

Můžeme si všimnout, že projekce r na první dvě souřadnice je rovna p , projekce na poslední dvě souřadnice je rovna q a projekce r na první a poslední souřadnici je coupling μ a η . Předpokládejme nyní, že p je ρ -optimální coupling μ a ν a q je ρ -optimální coupling ν a η . Mějme náhodný vektor (X, Y, Z) s rozdělením r . Z platnosti trojúhelníkové nerovnosti pro metriku ρ , platí

$$\rho(X, Z) \leq \rho(X, Y) + \rho(Y, Z).$$

Nyní z monotonie a linearitě střední hodnoty dostáváme

$$\mathbf{E}\rho(X, Z) \leq \mathbf{E}\rho(X, Y) + \mathbf{E}\rho(Y, Z).$$

Dále z definice optimálního couplingu plyne

$$\mathbf{E}\rho(X, Y) + \mathbf{E}\rho(Y, Z) = \rho_K(\mu, \nu) + \rho_K(\nu, \eta).$$

A nakonec, jelikož (X, Z) je coupling μ a η , platí

$$\rho_K(\mu, \eta) \leq \mathbf{E}\rho(X, Z).$$

□

4.2 Coupling trajektorií

Předpokládejme nyní, že množina stavů \mathcal{X} markovského řetězce $(X_t)_{t=0}^{\infty}$ s maticí pravděpodobností přechodu P je rovna množině vrcholů nějakého souvislého grafu $G = (\mathcal{X}, E_0)$. Necht ℓ je funkce definovaná na E_0 , která každé hraně $\{x, y\} \in E_0$, $x \neq y$, přiřadí délku $\ell(x, y) \geq 1$.

Bud' x_0, x_1, \dots, x_r cesta v G , její *délku* definujeme jako sumu $\sum_{i=1}^r \ell(x_{i-1}, x_i)$. Na prostoru \mathcal{X} dále definujeme *metriku cesty* předpisem

$$\rho(x, y) = \min\{\text{délka } \xi : \xi \text{ je cesta z } x \text{ do } y\}. \quad (4.3)$$

Předpoklad $\ell(x, y) \geq 1$ zřejmě implikuje $\rho(x, y) \geq \mathbf{1}\{x \neq y\}$ a pro každou dvojici náhodných veličin (X, Y) s množinou stavů \mathcal{X} platí

$$\mathbf{P}\{X \neq Y\} = \mathbf{E}(\mathbf{1}\{X \neq Y\}) \leq \mathbf{E}\rho(X, Y). \quad (4.4)$$

Pokud minimalizujeme (4.4) přes všechny couplingy (X, Y) pravděpodobnostních rozdělení μ a ν , dostaneme

$$\|\mu - \nu\|_{\text{TV}} \leq \rho_K(\mu, \nu). \quad (4.5)$$

Tento vztah transportní metriky a vzdálenosti v totální variaci můžeme využít k odvození odhadu rychlosti mixingů řetězce, jak ukážeme v této podkapitole. Myšlenka je následující:

Předpokládejme, že pro každé dva stavy $x, y \in \mathcal{X}$ existuje coupling (X_1, Y_1) pravděpodobnostních rozdělení $P(x, \cdot)$ a $P(y, \cdot)$ takový, že pro nějaké $\alpha > 0$ platí

$$\mathbf{E}_{x,y}\rho(X_1, Y_1) \leq e^{-\alpha}\rho(x, y), \quad \text{pro všechna } x, y \in \mathcal{X}. \quad (4.6)$$

Průměrem prostoru \mathcal{X} rozumíme $\text{diam}(\mathcal{X}) = \max_{x,y \in \mathcal{X}} \rho(x, y)$. Iterováním (4.6) dostaneme

$$\mathbf{E}_{x,y}\rho(X_t, Y_t) \leq e^{-\alpha t} \text{diam}(\mathcal{X}).$$

Celkově pak pomocí tvrzení 7 dostaneme

$$\begin{aligned} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} &\leq \mathbf{P}_{x,y}\{X_t \neq Y_t\} = \\ &= \mathbf{P}_{x,y}\{\rho(X_t, Y_t) \geq 1\} \leq \mathbf{E}_{x,y}\rho(X_t, Y_t) \leq \text{diam}(\mathcal{X})e^{-\alpha t}, \end{aligned}$$

odkud dále platí

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil \frac{1}{\alpha} [\log(\text{diam}(\mathcal{X})) + \log(1/\varepsilon)] \right\rceil.$$

Věta 9. *Předpokládejme, že množina stavů \mathcal{X} markovského řetězce je rovna množině vrcholů grafu. Bud' ℓ jako výše definovaná. Necht' ρ je metrika na \mathcal{X} definovaná v (4.3). Předpokládejme, že pro každou hranu $\{x, y\}$ existuje coupling (X_1, Y_1) pravděpodobnostních rozdělení $P(x, \cdot)$ a $P(y, \cdot)$ takový, že pro nějaké $\alpha > 0$ platí*

$$\mathbf{E}_{x,y}\rho(X_1, Y_1) \leq e^{-\alpha}\rho(x, y). \quad (4.7)$$

Poté pro libovolná pravděpodobnostní rozdělení μ a ν na prostoru \mathcal{X} platí

$$\rho_K(\mu P, \nu P) \leq e^{-\alpha}\rho_K(\mu, \nu). \quad (4.8)$$

Markovský coupling dvou řetězců se stejnou maticí pravděpodobností přechodu sestavený pomocí transportní metriky, kdy jeden startoval v pevném stavu a druhý ze stacionárního rozdělení, nám pomocí věty 9 dá odhad rychlosti mixingů v důsledku 10.

Důkaz věty 9. Nejprve ukážeme, že pro libovolné stavy $x, y \in \mathcal{X}$ platí

$$\rho_K(P(x, \cdot), P(y, \cdot)) \leq e^{-\alpha} \rho(x, y). \quad (4.9)$$

Mějme $x, y \in \mathcal{X}$ pevně zvolené, buď $x = x_0, x_1, \dots, x_{r-1}, x_r = y$ cesta mezi x a y , jejíž délka nabývá minima v (4.3). Z trojúhelníkové nerovnosti pro ρ_K ,

$$\rho_K(P(x, \cdot), P(y, \cdot)) \leq \sum_{k=1}^r \rho_K(P(x_{k-1}, \cdot), P(x_k, \cdot)).$$

Z definice ρ_K a z předpokladu (4.7) plyne, že pro libovolnou hranu $\{a, b\}$ platí

$$\rho_K(P(a, \cdot), P(b, \cdot)) \leq \mathbf{E}\rho(X_a, Y_b) \leq e^{-\alpha} \rho(a, b) = e^{-\alpha} \ell(a, b),$$

kde (X_a, Y_b) je libovolný coupling rozdělení $P(a, \cdot)$ a $P(b, \cdot)$. Z této nerovnosti můžeme odhadnout členy předchozí sumy. Dostáváme

$$\rho_K(P(x, \cdot), P(y, \cdot)) \leq e^{-\alpha} \sum_{k=1}^r \ell(x_{k-1}, x_k) = e^{-\alpha} \rho(x, y).$$

Poslední rovnost platí, jelikož cestu x_0, x_1, \dots, x_r jsme si zvolili tak, aby měla minimální délku. Dokázali jsme platnost (4.9).

Nechť η je ρ -optimální coupling rozdělení ν a μ , neboli platí

$$\rho_K(\mu, \nu) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{X}} \rho(x, y) \eta(x, y). \quad (4.10)$$

Z ekvivalentní definice transportní metriky (4.2), existence optimálního couplingu a nerovnosti (4.9) víme, že pro každé dva stavy $x, y \in \mathcal{X}$ existuje coupling $\theta_{x,y}$ pravděpodobnostních rozdělení $P(x, \cdot)$ a $P(y, \cdot)$ splňující

$$\sum_{u,w \in \mathcal{X}} \rho(u, w) \theta_{x,y}(u, w) = \rho_K(P(x, \cdot), P(y, \cdot)) \leq e^{-\alpha} \rho(x, y). \quad (4.11)$$

Definujme pravděpodobnostní rozdělení $\theta = \sum_{(x,y) \in \mathcal{X} \times \mathcal{X}} \eta(x, y) \theta_{x,y}$ na $\mathcal{X} \times \mathcal{X}$. Takto definované zobrazení je coupling μP a νP , neboť platí

$$\begin{aligned} \sum_{w \in \mathcal{X}} \theta(\cdot, w) &= \sum_{x,y \in \mathcal{X}} \eta(x, y) \left(\sum_{w \in \mathcal{X}} \theta_{x,y}(\cdot, w) \right) = \sum_{x,y \in \mathcal{X}} \eta(x, y) P(x, \cdot) = \\ &= \sum_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{X}} \eta(x, y) \right) P(x, \cdot) = \sum_{x \in \mathcal{X}} \mu(x) P(x, \cdot) = \mu P(\cdot), \end{aligned}$$

$$\begin{aligned} \sum_{u \in \mathcal{X}} \theta(u, \cdot) &= \sum_{x,y \in \mathcal{X}} \eta(x, y) \left(\sum_{u \in \mathcal{X}} \theta_{x,y}(u, \cdot) \right) = \sum_{x,y \in \mathcal{X}} \eta(x, y) P(y, \cdot) = \\ &= \sum_{y \in \mathcal{X}} \left(\sum_{x \in \mathcal{X}} \eta(x, y) \right) P(y, \cdot) = \sum_{y \in \mathcal{X}} \nu(y) P(y, \cdot) = \nu P(\cdot). \end{aligned}$$

Z (4.10) a (4.11) dostáváme

$$\begin{aligned} \sum_{u,w \in \mathcal{X}} \rho(u, w) \theta(u, w) &= \sum_{x,y \in \mathcal{X}} \left(\sum_{u,w \in \mathcal{X}} \rho(u, w) \theta_{x,y}(u, w) \right) \eta(x, y) \\ &\leq e^{-\alpha} \sum_{x,y \in \mathcal{X}} \rho(x, y) \eta(x, y) = e^{-\alpha} \rho_K(\mu, \nu). \end{aligned}$$

Z definice transportní metriky platí $\rho_K(\mu P, \nu P) \leq \sum_{u,w \in \mathcal{X}} \rho(u,w)\theta(u,w)$ a důkaz je tedy hotov. □

Důsledek 10. *Mějme stejné předpoklady jako ve větě 9. Poté platí*

$$d(t) \leq e^{-\alpha t} \text{diam}(\mathcal{X}),$$

tudíž také platí

$$t_{mix}(\varepsilon) \leq \left\lceil \frac{1}{\alpha} [\log(\text{diam}(\mathcal{X})) + \log(1/\varepsilon)] \right\rceil.$$

Důkaz. Ke zkoumání $d(t)$ potřebujeme zjistit, zda existuje jednoznačně určené stacionární rozdělení. Jeho existence je implikována z nerovnosti (4.8). Označme ho π .

Postupným iterováním (4.8) dostaneme

$$\rho_K(\mu P^t, \nu P^t) \leq e^{-\alpha t} \rho_K(\mu, \nu) \leq e^{-\alpha t} \max_{x,y} \rho(x,y),$$

kde platnost druhé nerovnosti plyne z ekvivalentní definice transportní metriky (4.2). Buď $x \in \mathcal{X}$. Pokud do této nerovnosti dosadíme $\mu = \delta_x$ a $\nu = \pi$ a aplikujeme odhad (4.5), získáme

$$\|P^t(x, \cdot) - \pi\|_{\text{TV}} = \|\delta_x P^t - \pi P^t\|_{\text{TV}} \leq \rho_K(\delta_x P^t, \pi P^t) \leq e^{-\alpha t} \max_{x,y} \rho(x,y).$$

Z definice $\text{diam}(\mathcal{X}) = \max_{x,y} \rho(x,y)$ a definice $d(t)$ získáváme platnost

$$d(t) \leq e^{-\alpha t} \text{diam}(\mathcal{X}).$$

Následně z

$$\begin{aligned} \varepsilon &\geq e^{-\alpha t} \text{diam}(\mathcal{X}) \\ \alpha t &\geq \log(\text{diam}(\mathcal{X})) - \log(\varepsilon) \\ t &\geq \frac{1}{\alpha} [\log(\text{diam}(\mathcal{X})) + \log(1/\varepsilon)] \end{aligned}$$

dostáváme platnost druhé části tvrzení. □

4.3 Rychlý mixing

Nechť π je rovnoměrné pravděpodobnostní rozdělení na množině všech přípustných q -obarvení grafu G . V této podkapitole budeme pomocí metody z podkapitoly 4.2 analyzovat rychlost mixingů modifikované Glauberovy dynamiky pro π pro obarvení grafu.

Věta 11. *Mějme modifikovanou Glauberovu dynamiku pro všechna q -obarvení grafu $G = (V, E)$ s n vrcholy a maximálním stupněm Δ . Pokud $q > 2\Delta$, platí*

$$t_{mix}(\varepsilon) \leq \left\lceil \left(\frac{q - \Delta}{q - 2\Delta} \right) n (\log n - \log \varepsilon) \right\rceil. \quad (4.12)$$

Naším cílem bude zavést metriku na množině všech obarvení grafu G , sestavit markovský coupling dvou řetězců modifikované Glauberovy dynamiky a ověřit, že splňuje předpoklady věty 9. Poté již přímo z důsledku 10 získáme odhad rychlosti mixingů.

Důkaz věty 11. Mějme metriku ρ na $\mathcal{X} = \{1, 2, \dots, q\}^V$, která je definovaná předpisem

$$\rho(x, y) = \sum_{v \in V} \mathbf{1}\{x(v) \neq y(v)\}, \quad (4.13)$$

tedy definovanou jakožto počet vrcholů, v níž se daná obarvení liší. Řekneme, že dvě obarvení spolu *sousedí* právě tehdy, když se liší pouze v jediném vrcholu.

Nechť $A_v(x)$ značí množinu všech přípustných barev pro v v konfiguraci x .

Nechť x a y jsou dva sousedící prvky \mathcal{X} a buď v vrchol, ve kterém se tato obarvení liší. Poté jistě platí $A_v(x) = A_v(y)$. Definujme markovský coupling $(X_t, Y_t)_{t=0}^{\infty}$ dvou řetězců modifikované Glauberovy dynamiky, kdy jeden začal ve stavu x a druhý ve stavu y . Nyní popíšeme, jak se budou sdruženě vyvíjet.

Buď w libovolný rovnoměrně náhodně vybraný vrchol z V . Přebarvíme ho v obou řetězcích podle dále popsaných situací.

Pokud v nesusedí s w , poté můžeme přebarvit w stejnou barvou v obou řetězcích. Tuto barvu můžeme vybrat rovnoměrně náhodně z množiny $A_w(x)$, jelikož platí $A_w(x) = A_w(y)$.

Nyní mějme situaci, kdy w sousedí s v . Bez újmy na obecnosti předpokládejme, že $|A_w(x)| \leq |A_w(y)|$. Náhodně vybereme barvu z $A_w(y)$, označíme ji U a položíme $y(w) = U$. Pokud $U \neq x(v)$, poté tedy $U \in A_w(x)$ (z předpokladu $\rho(x, y) = 1$) a položíme $x(w) = U$. Příklad $U = x(v)$ je nutné detailněji rozebrat podle toho, zda platí $|A_w(x)| = |A_w(y)|$. Pokud $|A_w(x)| = |A_w(y)|$, pak položíme $x(w) = y(v)$. Naopak pokud $|A_w(x)| < |A_w(y)|$, poté náhodně vybereme barvu z $A_w(x)$, označíme ji W a položíme $x(w) = W$. Poté tedy rozdělení přebarvení w v x je rovnoměrné na množině $A_w(x)$. Pravděpodobnost, že přebarvení w jsou v obou řetězcích různá je $\frac{1}{|A_w(y)|}$, kde hodnota tohoto výrazu je shora omezená hodnotou $\frac{1}{q-\Delta}$.

Nyní ukážeme, že takto zkonstruovaný coupling (X_1, Y_1) rozdělení $P(x, \cdot)$ a $P(y, \cdot)$ splňuje (4.7) pro ρ definovanou v (4.13).

Vzdálenost $\rho(X_1, Y_1)$ se zvětší v případě, že vybraný vrchol w sousedí s v a jeho nové obarvení je v každém řetězci jiné. Pokud je vybraným vrcholem v , $\rho(X_1, Y_1)$ se zmenší na nulu. V ostatních případech zůstane vzdálenost rovná jedné. Poté platí

$$\begin{aligned} \mathbf{E}_{x,y}(\rho(X_1, Y_1) - 1) &\leq (2 - 1) \frac{\deg(v)}{n} \left(\frac{1}{q - \Delta} \right) + 0 + (0 - 1) \frac{1}{n}, \\ \mathbf{E}_{x,y}(\rho(X_1, Y_1)) &\leq 1 + \frac{\deg(v)}{n} \left(\frac{1}{q - \Delta} \right) - \frac{1}{n}. \end{aligned}$$

Pravá strana poslední rovnice je díky předpokladu $\deg(v) \leq \Delta$ shora omezená výrazem

$$1 - \frac{1}{n} \left(1 - \frac{\Delta}{q - \Delta} \right).$$

Jelikož předpokládáme $q > 2\Delta$, poté je tento výraz menší než 1. Označme $c(q, \Delta) = 1 - \frac{\Delta}{q-\Delta}$, poté

$$\mathbf{E}_{x,y}(\rho(X_1, Y_1)) \leq \exp\left(-\frac{c(q, \Delta)}{n}\right) = \exp\left(-\frac{c(q, \Delta)}{n}\right)\rho(x, y). \quad (4.14)$$

Z předpokladu $q > 2\Delta$ jistě platí $c(q, \Delta) > 0$. Poté jsou splněny předpoklady věty 9 a použitím důsledku 10 dostaneme

$$\begin{aligned} \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{TV} &\leq \text{diam}(\mathcal{X}) \exp\left(-\frac{c(q, \Delta)}{n}t\right) = n \exp\left(-\frac{c(q, \Delta)}{n}t\right), \\ t_{mix}(\varepsilon) &\leq \left\lceil \frac{n}{c(q, \Delta)} (\log n + \log \varepsilon^{-1}) \right\rceil. \end{aligned} \quad (4.15)$$

Konečně, z

$$\frac{1}{c(q, \Delta)} = \frac{1}{1 - \frac{\Delta}{q-\Delta}} = \frac{q - \Delta}{q - 2\Delta}$$

dostáváme platnost (4.12). □

4.4 Přibližné počítání obarvení grafu

V této podkapitole popíšeme, jak lze využít teorii popsanou v předešlých sekcích k přibližnému počítání počtu všech přípustných obarvení grafu a dokážeme přitom následující větu.

Věta 12. *Nechť \mathcal{X} je množina všech přípustných q -obarvení grafu G s n vrcholy a maximálním stupněm Δ . Buď $q > 2\Delta$ a definujme $c(q, \Delta) = 1 - \frac{\Delta}{q-\Delta}$. Pro každé dvě konstanty $\eta, \varepsilon \in (0, 1)$ existuje náhodná veličina W , která může být simulována pomocí nejvýše*

$$n \left\lceil \frac{n \log n + n \log(6eqn/\varepsilon)}{c(q, \Delta)} \right\rceil \left\lceil \frac{27qn}{\eta\varepsilon^2} \right\rceil \quad (4.16)$$

kroků Glauberovy dynamiky a splňuje

$$\mathbf{P}\{(1 - \varepsilon)|\mathcal{X}|^{-1} \leq W \leq (1 + \varepsilon)|\mathcal{X}|^{-1}\} \geq 1 - \eta. \quad (4.17)$$

Poznámka. Toto je příklad tzv. *zcela polynomiálního znáhodněného aproximačního schématu*. Jde o algoritmus, který pro předem dané $\eta, \varepsilon \in (0, 1)$ s pravděpodobností minimálně $1 - \eta$ a s přesností až na multiplikativní faktor $1 \pm \varepsilon$ dokáže odhadnout $|\mathcal{X}|$ v čase, který je polynomiální v n a ε^{-1} .

Nejprve zavedeme potřebné značení. Buď x_0 libovolné přípustné obarvení grafu G . Označme si vrcholy G jako $\{v_1, v_2, \dots, v_n\}$. Dále definujme pro $k \leq n$ množiny

$$\mathcal{X}_k = \{x \in \mathcal{X} : x(v_j) = x_0(v_j) \text{ pro } j > k\}. \quad (4.18)$$

Množina \mathcal{X}_k obsahuje přípustná q -obarvení, kde vrcholy v_1, \dots, v_k mají libovolné obarvení a vrcholy v_{k+1}, \dots, v_n mají stejné obarvení jako v konfiguraci x_0 . Zřejmě platí $\mathcal{X}_{k-1} \subseteq \mathcal{X}_k$, $|\mathcal{X}_0| = 1$ a $|\mathcal{X}_n| = |\mathcal{X}|$.

Následující lemma nám dává spodní odhad hodnot poměrů $|\mathcal{X}_{k-1}|/|\mathcal{X}_k|$, který bude užitečný při důkazu věty 12.

Lemma 13. *Mějme stejné předpoklady jako ve větě 12. Necht \mathcal{X}_k definováno v (4.18). Poté platí*

$$\frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|} \geq \frac{1}{qe}.$$

Důkaz. Vrcholy z množiny v_1, \dots, v_{k-1} , které sousedí s v_k nazveme *volnými* sousedícími vrcholy v_k . Označme r počet volných sousedících vrcholů v_k .

Mějme náhodný proces s počátečním rozdělením rovnoměrným na množině \mathcal{X}_k , který s předem daným pořadím postupně přebarví přípustnými barvami volné sousedící vrcholy v_k a poté i samotný vrchol v_k . Přípustná barva je vždy vybrána rovnoměrně ze všech přípustných barev pro daný vrchol.

Označme stav tohoto procesu po $(r + 1)$ krocích jako Y . Necht A označuje jev, kdy každý z volných sousedících vrcholů v_k je přebarven jinou barvou, než $x_0(v_k)$ a v_k je obarven přímo barvou $x_0(v_k)$. Poté tedy $Y \in \mathcal{X}_{k-1}$ právě tehdy, když nastane jev A , a platí

$$\begin{aligned} \frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|} &= \mathbf{P}\{Y \in \mathcal{X}_{k-1}\} = \mathbf{P}\{A\} \geq \left(\frac{q - \Delta - 1}{q - \Delta}\right)^r \frac{1}{q} \\ &\geq \left(\frac{q - \Delta - 1}{q - \Delta}\right)^\Delta \frac{1}{q} \geq \left(\frac{\Delta}{\Delta + 1}\right)^\Delta \frac{1}{q} = \frac{1}{\left(1 + \frac{1}{\Delta}\right)^\Delta} \frac{1}{q} \geq \frac{1}{eq}. \end{aligned}$$

□

Nyní již můžeme dokázat větu 12.

Důkaz věty 12. Buď $k \in \{1, 2, \dots, n\}$. Buď x_0 libovolné přípustné obarvení a necht \mathcal{X}_k je jako výše definované.

Pomocí úpravy Glauberovy dynamiky můžeme náhodně generovat prvky \mathcal{X}_k . Tato modifikace spočívá v tom, že můžeme přebarvovat pouze vrcholy z množiny $\{v_1, \dots, v_k\}$. Obarvení zbylých vrcholů je dáno konfigurací x_0 . Označme P_k matici pravděpodobností přechodu takto upravené Glauberovy dynamiky na \mathcal{X}_k .

Pokud n ve větě 11 nahradíme k , poté tento horní odhad hodnoty $t_{\text{mix}}(\varepsilon)$ platí i v našem případě. Speciálně, jelikož $k \leq n$, platí i odhad (4.15). Položme

$$t(n, \varepsilon) = \left\lceil \frac{n \log n + n \log(6eqn/\varepsilon)}{c(q, \Delta)} \right\rceil.$$

Označme $\tilde{\varepsilon} = \varepsilon/6eqn$. Poté postupně z věty 11 a z $k \leq n$ platí

$$t_{\text{mix}}(\tilde{\varepsilon}) \leq \left\lceil \frac{k \log k + k \log(\tilde{\varepsilon}^{-1})}{c(q, \Delta)} \right\rceil \leq t(n, \varepsilon).$$

Dále postupně z $t_{\text{mix}}(\tilde{\varepsilon}) \leq t(n, \varepsilon)$, definice vzdálenosti v totální variaci a definice $d(t)$ platí

$$\begin{aligned} \left\| P_k^{t(n, \varepsilon)}(x_0, \cdot) - \pi_k \right\|_{TV} &\leq \left\| P_k^{t_{\text{mix}}(\tilde{\varepsilon})}(x_0, \cdot) - \pi_k \right\|_{TV} \\ &\leq d(t_{\text{mix}}(\tilde{\varepsilon})) < \tilde{\varepsilon} = \frac{\varepsilon}{6eqn}, \end{aligned} \quad (4.19)$$

kde π_k značí rovnoměrné rozdělení na \mathcal{X}_k .

Poměry $\frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|}$ můžeme odhadnout podle následujícího postupu. Náhodný prvek z \mathcal{X}_k můžeme vygenerovat pomocí upravené Glauberovy dynamiky na \mathcal{X}_k , kterou necháme běžet $t(n, \varepsilon)$ kroků. Pokud nezávisle na sobě necháme takto běžet $a_n = \lceil 27qn/\eta\varepsilon^2 \rceil$ řetězců, získáme a_n prvků \mathcal{X}_k . Nechť $Z_{k,i}$, pro $i = 1, \dots, a_n$, značí indikátor jevu, kdy i -tý takto generovaný prvek je dokonce prvkem \mathcal{X}_{k-1} . Postupně z definice $Z_{k,i}$ a definice vzdálenosti v totální variaci (jelikož platí $\mathcal{X}_{k-1} \subseteq \mathcal{X}_k$) dostáváme

$$\begin{aligned} |\mathbf{E}Z_{k,i} - \pi_k(\mathcal{X}_{k-1})| &= |P_k^{t(n, \varepsilon)}(x_0, \mathcal{X}_{k-1}) - \pi_k(\mathcal{X}_{k-1})| \\ &\leq \left\| P_k^{t(n, \varepsilon)}(x_0, \cdot) - \pi_k \right\|_{TV} \leq \frac{\varepsilon}{6eqn}, \end{aligned}$$

kde platnost poslední nerovnosti plyne z odhadu (4.19). Pro náhodné veličiny $W_k = a_n^{-1} \sum_{i=1}^{a_n} Z_{k,i}$ tedy platí

$$\left| \mathbf{E}W_k - \frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|} \right| = \left| \mathbf{E}Z_{k,1} - \frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|} \right| = |\mathbf{E}Z_{k,1} - \pi_k(\mathcal{X}_{k-1})| \leq \frac{\varepsilon}{6eqn}. \quad (4.20)$$

Pro náhodnou veličinu Y s alternativním rozdělením platí

$$\text{Var}(Y) = \mathbf{E}Y(1 - \mathbf{E}Y).$$

Poté z nezávislosti náhodných veličin $Z_{k,i}$ platí

$$\text{Var}(W_k) = \frac{1}{a_n^2} \sum_{i=1}^{a_n} \mathbf{E}Z_{k,i}[1 - \mathbf{E}Z_{k,i}] \leq \frac{1}{a_n^2} \sum_{i=1}^{a_n} \mathbf{E}Z_{k,i} = \frac{\mathbf{E}W_k}{a_n}.$$

Poté dostáváme

$$\frac{\text{Var}(W_k)}{\mathbf{E}^2(W_k)} \leq \frac{1}{a_n \mathbf{E}(W_k)}. \quad (4.21)$$

Z lemmatu 13 a aplikací odhadu (4.20) pro dostatečně malé ε získáváme

$$\mathbf{E}(W_k) \geq \frac{1}{eq} - \frac{\varepsilon}{6eqn} \geq \frac{1}{3q}.$$

Použitím toho odhadu v nerovnosti (4.21) dostáváme

$$\frac{\text{Var}(W_k)}{\mathbf{E}^2(W_k)} \leq \frac{3q}{a_n} \leq \frac{\eta\varepsilon^2}{9n}. \quad (4.22)$$

Z (4.20) a lemmatu 13 platí

$$\begin{aligned} \frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|} - \frac{\varepsilon}{6eqn} &\leq \mathbf{E}W_k \leq \frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|} + \frac{\varepsilon}{6eqn}, \\ 1 - \frac{\varepsilon}{6n} &\leq 1 - \frac{|\mathcal{X}_k|}{|\mathcal{X}_{k-1}|} \frac{\varepsilon}{6eqn} \leq \frac{|\mathcal{X}_k|}{|\mathcal{X}_{k-1}|} \mathbf{E}W_k \leq 1 + \frac{|\mathcal{X}_k|}{|\mathcal{X}_{k-1}|} \frac{\varepsilon}{6eqn} \leq 1 + \frac{\varepsilon}{6n}. \end{aligned}$$

Zřejmě platí

$$\prod_{k=1}^n \frac{|\mathcal{X}_k|}{|\mathcal{X}_{k-1}|} = \frac{|\mathcal{X}_n|}{|\mathcal{X}_0|} = |\mathcal{X}_n| = |\mathcal{X}|.$$

Bud $W = \prod_{i=1}^n W_i$. Poté

$$1 - \frac{\varepsilon}{6} \leq \left(1 - \frac{\varepsilon}{6n}\right)^n \leq |\mathcal{X}| \mathbf{E}W \leq \left(1 + \frac{\varepsilon}{6n}\right)^n \leq e^{\varepsilon/6} \leq 1 + \frac{\varepsilon}{3},$$

kde jsme v poslední nerovnosti použili platnost $e^x \leq 1 + 2x$ pro $x \in [0, 1]$. Poté platí

$$\left| \mathbf{E}(W) - \frac{1}{|\mathcal{X}|} \right| \leq \frac{\varepsilon}{3|\mathcal{X}|}. \quad (4.23)$$

Dále platí

$$\mathbf{E} \left(\frac{W}{\mathbf{E}W} \right)^2 = \mathbf{E} \prod_{i=1}^n \left(\frac{W_i}{\mathbf{E}W_i} \right)^2 = \prod_{i=1}^n \frac{\mathbf{E}W_i^2}{(\mathbf{E}W_i)^2},$$

kde jsme v druhé rovnosti využili nezávislost náhodných veličin W_i . Pokud od obou stran této rovnosti odečteme 1, získáváme

$$\frac{\text{Var}(W)}{\mathbf{E}^2(W)} = \prod_{k=1}^n \left[1 + \frac{\text{Var}(W_k)}{\mathbf{E}^2(W_k)} \right] - 1.$$

Tato rovnost nám společně s (4.22) dává platnost

$$\frac{\text{Var}(W)}{\mathbf{E}^2(W)} \leq \prod_{k=1}^n \left[1 + \frac{\eta\varepsilon^2}{9n} \right] - 1 \leq e^{\eta\varepsilon^2/9} - 1 \leq \frac{2\eta\varepsilon^2}{9},$$

kde v poslední nerovnosti jsme opět použili platnost $e^x \leq 1 + 2x$ pro $x \in [0, 1]$. Z Čebyševovy nerovnosti 3 dále platí

$$\mathbf{P}\{|W - \mathbf{E}(W)| \geq \mathbf{E}(W)\varepsilon/2\} \leq \frac{\text{Var}(W)}{(\mathbf{E}(W)\varepsilon/2)^2} = \frac{4\text{Var}(W)}{\varepsilon^2 \mathbf{E}^2(W)} \leq \frac{8\eta}{9} \leq \eta.$$

Použitím trojúhelníkové nerovnosti pro reálná čísla a odhadu (4.23) dostáváme pro každé $\omega \in \{\omega : |W(\omega) - \mathbf{E}W| < \mathbf{E}W\varepsilon/2\}$ odhad

$$\begin{aligned} \left| W(\omega) - \frac{1}{|\mathcal{X}|} \right| &\leq |W(\omega) - \mathbf{E}(W)| + \left| \mathbf{E}(W) - \frac{1}{|\mathcal{X}|} \right| \leq \frac{\varepsilon}{2} |\mathbf{E}(W)| + \frac{\varepsilon}{3|\mathcal{X}|} \\ &\leq \frac{\varepsilon}{2} \left(\frac{1}{|\mathcal{X}|} + \frac{1}{3|\mathcal{X}|} \right) + \frac{\varepsilon}{3|\mathcal{X}|} = \frac{\varepsilon}{|\mathcal{X}|}. \end{aligned}$$

Tedy dokázali jsme platnost (4.17).

Pro každé \mathcal{X}_k , $k = 1, \dots, n$, potřebujeme a_n generovaných prvků a tedy je k simulaci náhodné veličiny W potřeba nejvíce (4.16) kroků Glauberovy dynamiky. \square

4.5 Vylepšený odhad výpočetní náročnosti pro přibližné počítání obarvení grafu

V této kapitole ukážeme, že tvrzení věty 12 lze zesílit. Dokážeme totiž větu 14.

Věta 14. *Nechť \mathcal{X} je množina všech přípustných q -obarvení grafu G s n vrcholy a maximálním stupněm Δ . Buď $q > 2\Delta$ a definujme $c(q, \Delta) = 1 - \frac{\Delta}{q-\Delta}$. Pro každé dvě konstanty $\eta, \varepsilon \in (0, 1)$ existuje náhodná veličina W , která může být simulována pomocí nejvýše*

$$n \left(\left\lceil \frac{n \log n + n \log(6eqn/\varepsilon)}{c(q, \Delta)} \right\rceil + \left\lceil \frac{81q^2n^2}{2\eta\varepsilon^2c(q, \Delta)} \right\rceil \right) \quad (4.24)$$

kroků Glauberovy dynamiky a splňuje

$$\mathbf{P}\{(1 - \varepsilon)|\mathcal{X}|^{-1} \leq W \leq (1 + \varepsilon)|\mathcal{X}|^{-1}\} \geq 1 - \eta. \quad (4.25)$$

Na rozdíl od důkazu věty 12 již nebudeme používat nezávislé realizace markovského řetězce $(X_t)_{t=1}^\infty$, místo nich použijeme ergodický průměr $\frac{1}{s-r} \sum_{t=r+1}^s X_t$. Tak budeme moci podstatným způsobem snížit počet simulací k dosažení stejné přesnosti.

Připomeňme z důkazu věty 12, že při používání nezávislých vzorků je počet iterací Glauberovy dynamiky pro každé $k \in \{1, 2, \dots, n\}$ nejvýše

$$t(n, \varepsilon) \cdot a_n = \left\lceil \frac{n \log n + n \log(6eqn/\varepsilon)}{c(q, \Delta)} \right\rceil \left\lceil \frac{27qn}{\eta\varepsilon^2} \right\rceil.$$

Asymptotická složitost zcela polynomiálního znáhodněného aproximačního schématu z důkazu věty 12 je tedy $\mathcal{O}(n^3 \log n)$.

Dokážeme, že pokud místo toho postupujeme pouze s částí trajektorie markovského řetězce, poté pro každé $k \in \{1, 2, \dots, n\}$ potřebujeme nejvýše

$$t(n, \varepsilon) + s = \left\lceil \frac{n \log n + n \log(6eqn/\varepsilon)}{c(q, \Delta)} \right\rceil + \left\lceil \frac{81q^2n^2}{2\eta\varepsilon^2c(q, \Delta)} \right\rceil$$

iterací. Asymptotická složitost upraveného aproximačního schématu je poté pouze $\mathcal{O}(n^3)$.

K důkazu budeme potřebovat umět shora odhadnout rozptyl ergodického průměru $\frac{1}{s-r} \sum_{t=r+1}^s X_t$. K tomu použijeme metodu vlastních čísel a funkcí matice pravděpodobností přechodu P . Uvedeme si tedy nejprve nezbytné definice a tvrzení.

Definice 15 (Vlastní funkce a vlastní číslo matice pravděpodobností přechodu). *Buď P matice pravděpodobností přechodu markovského řetězce s množinou stavů \mathcal{X} . Nenulovou funkci $f : \mathcal{X} \rightarrow \mathbb{R}$ nazveme vlastní funkcí P s příslušným vlastním číslem λ , pokud platí $Pf = \lambda f$.*

Definice 16 (Reverzibilní řetězec). *Buď π pravděpodobnostní rozdělení na množině stavů markovského řetězce s maticí pravděpodobností přechodu P . Řekneme, že řetězec je reverzibilní vzhledem k rozdělení π , pokud pro všechna $x, y \in \mathcal{X}$ platí*

$$\pi(x)P(x, y) = \pi(y)P(y, x).$$

Pokud je řetězec reverzibilní, jsou všechna vlastní čísla P reálná a můžeme je seřadit sestupně $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\mathcal{X}|} \geq -1$. Zavádíme následující značení

$$\begin{aligned}\lambda_* &= \max\{|\lambda| : \lambda \text{ je vlastní číslo } P, \lambda \neq 1\}, \\ \gamma_* &= 1 - \lambda_*, \\ \gamma &= 1 - \lambda_2.\end{aligned}$$

Zřejmě platí $\gamma \geq \gamma_*$.

Lemma 15. *Bud' $(X_t)_{t=1}^\infty$ reverzibilní markovský řetězec. Bud' f reálná funkce definovaná na množině stavů \mathcal{X} taková, že $\mathbf{E}_\pi(f) = 0$. Pak platí*

$$\mathbf{E}_\pi \left[\left(\sum_{t=0}^{s-1} f(X_t) \right)^2 \right] \leq \frac{2s\mathbf{E}_\pi(f^2)}{\gamma}.$$

Důkaz. Levin a Peres (2017, lemma 12.22, strana 174) □

Věta 16. *Bud' (\mathcal{X}, ρ) metrický prostor a bud' P matice pravděpodobností přechodu markovského řetězce s množinou stavů \mathcal{X} . Předpokládejme, že existuje kladná konstanta $\theta < 1$ taková, že pro každé dva stavy $x, y \in \mathcal{X}$ existuje coupling (X_1, Y_1) rozdělení $P(x, \cdot)$ a $P(y, \cdot)$ splňující*

$$\mathbf{E}_{x,y}\rho(X_1, Y_1) \leq \theta\rho(x, y).$$

Pokud $\lambda \neq 1$ je vlastní číslo P , poté $|\lambda| \leq \theta$. Speciálně platí $\gamma_ \geq 1 - \theta$.*

Důkaz. Levin a Peres (2017, věta 13.1, strana 180) □

Abychom mohli použít lemma 15, je nutné ověřit, že je Glauberova dynamika pro π reverzibilní řetězec.

Tvrzení 17. *Glauberova dynamika pro π je reverzibilní vzhledem k π .*

Důkaz. Připomeňme $\mathcal{X}(x, v) = \{y \in \mathcal{X} : y(w) = x(w) \forall w \neq v\}$ z definice Glauberovy dynamiky. Pokud platí $y \in \mathcal{X}(x, v)$, pak $\mathcal{X}(x, v) = \mathcal{X}(y, v)$ a jistě $x \in \mathcal{X}(y, v)$. Poté pro všechna $x, y \in \mathcal{X}$ platí

$$\begin{aligned}\pi(x)P(x, y) &= \pi(x) \sum_{v \in V} \frac{1}{|V|} \pi^{x,v}(y) = \sum_{v \in V} \frac{1}{|V|} \pi(x) \frac{\pi(y)}{\pi(\mathcal{X}(x, v))} \mathbf{1}\{y \in \mathcal{X}(x, v)\} \\ &= \sum_{v \in V} \frac{1}{|V|} \pi(y) \frac{\pi(x)}{\pi(\mathcal{X}(y, v))} \mathbf{1}\{x \in \mathcal{X}(y, v)\} = \pi(y) \sum_{v \in V} \frac{1}{|V|} \pi^{y,v}(x) \\ &= \pi(y)P(y, x).\end{aligned}$$

□

Nyní již přímo k důkazu věty 14.

Důkaz věty 14. Mějme stejné předpoklady jako ve větě 12. Buď $k \in \{1, 2, \dots, n\}$, x_0 libovolné přípustné q -obarvení grafu G . Necht π_k značí rovnoměrné rozdělení na \mathcal{X}_k . Buď $(X_t)_{t=0}^\infty$ Glauberova dynamika pro π_k s množinou stavů \mathcal{X}_k popsaná na začátku důkazu věty 12. Označme P_k matici pravděpodobností přechodu tohoto řetězce.

Naším cílem je odhadnout poměry $\frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|}$ pomocí části trajektorie $(X_t)_{t=0}^\infty$.

Definujme $\theta = \exp\left(-\frac{c(q, \Delta)}{n}\right) < 1$. Z důkazu věty 11, odhadu (4.14) a věty 16 plyne

$$\gamma \geq \gamma_* \geq 1 - \theta \geq \frac{c(q, \Delta)}{n}. \quad (4.26)$$

Buď $f : \mathcal{X}_k \rightarrow \mathbb{R}$ funkce daná předpisem $f(x) = \mathbf{1}\{x \in \mathcal{X}_{k-1}\}$ pro $x \in \mathcal{X}_k$. Poté jistě platí

$$\begin{aligned} \mathbf{E}_{\pi_k}(f) &= \sum_{x \in \mathcal{X}_k} f(x) \pi_k(x) = \frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|}, \\ \text{Var}_{\pi_k}(f) &= \mathbf{E}_{\pi_k}(f)[1 - \mathbf{E}_{\pi_k}(f)] \leq \frac{1}{4}. \end{aligned}$$

Připomeňme značení

$$t(n, \varepsilon) = \left\lceil \frac{n \log n + n \log(6eqn/\varepsilon)}{c(q, \Delta)} \right\rceil.$$

Položme $r = t(n, \varepsilon)$. Označme $Z_{k,i} = f(X_{r+i})$, pro $s \geq 1, i \in \{0, 1, \dots, s-1\}$, a $W_k = \frac{1}{s} \sum_{i=0}^{s-1} Z_{k,i}$.

Potřebujeme opět dokázat nerovnost (4.20). V našem případě to uděláme takto

$$\left| \mathbf{E}W_k - \frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|} \right| \leq \frac{1}{s} \sum_{i=0}^{s-1} \left| \mathbf{E}Z_{k,i} - \frac{|\mathcal{X}_{k-1}|}{|\mathcal{X}_k|} \right| \leq |\mathbf{E}Z_{k,0} - \pi_k(\mathcal{X}_{k-1})| \leq \frac{\varepsilon}{6eqn},$$

kde jsme ve druhé nerovnosti použili tvrzení 6 a ve třetí nerovnosti použili (4.19). Dostáváme odhad střední hodnoty pro dostatečně malé ε

$$\mathbf{E}(W_k) \geq \frac{1}{eq} - \frac{\varepsilon}{6eqn} \geq \frac{1}{3q}. \quad (4.27)$$

Použitím lemmatu 15 a odhadu (4.26) získáváme

$$\text{Var}(W_k) \leq \frac{2\text{Var}_{\pi_k}(f)}{s\gamma} \leq \frac{1}{2s(1-\theta)} \leq \frac{n}{2sc(q, \Delta)}. \quad (4.28)$$

Celkově z (4.27) a (4.28) platí

$$\frac{\text{Var}(W_k)}{\mathbf{E}^2(W_k)} \leq \frac{9q^2n}{2sc(q, \Delta)}.$$

Předchozí výpočty jistě platí pro každé $s \geq 1$. Chceme najít s takové, že platí

$$\frac{\text{Var}(W_k)}{\mathbf{E}^2(W_k)} \leq \frac{9q^2n}{2sc(q, \Delta)} \leq \frac{\eta\varepsilon^2}{9n}. \quad (4.29)$$

Stačí tedy vzít s splňující

$$\frac{81q^2n^2}{2\eta\varepsilon^2c(q, \Delta)} \leq s.$$

Položme

$$s = \left\lceil \frac{81q^2n^2}{2\eta\varepsilon^2c(q, \Delta)} \right\rceil.$$

Poté platí odhad (4.29) a odtud dál běží důkaz stejně jako v důkazu věty 12. □

Poznámka. Formulace a důkaz věty 14 byly obsahem cvičení 14.13 z knihy Levin a Peres (2017, strana 214), které autorka samostatně vyřešila.

Seznam použité literatury

- FRIEZE, A. a VIGODA, E. (2007). A Survey on the Use of Markov Chains to Randomly Sample Colourings. In GRIMMETT, G. a MCDIARMID, C., editors, *Combinatorics, Complexity, and Chance: A Tribute to Dominic Welsh*, pages 53–71. Oxford University Press. ISBN 9780198571278. doi: 10.1093/acprof:oso/9780198571278.003.0004.
- LEVIN, D. A. a PERES, Y. (2017). *Markov Chains and Mixing Times*. Second Edition. American Mathematical Society, Providence, Rhode Island. ISBN 978-1-4704-2962-1.
- PRÁŠKOVÁ, Z. a LACHOUT, P. (2012). *Základy náhodných procesů I*. Druhé vydání. Matfyzpress, Praha. ISBN 978-80-7378-210-8.