Nechť $\mathbf{Y}$ je funkce, která každému grafu přiřazuje číslo. Budeme například uvažovat funkci $\mathbf{Y}_m$, která každému grafu přiřaźuje počet klik velikosti mv tomto grafu.

Takové funkci $\mathbf{Y}$ a přirozenému číslu n budeme přiřazovat posloupnost $p_0, p_1, \ldots$, která se často popisuje takto: p_k je pravděpodobnost, že pro náhodně vybraný graf s n vrcholy má vlastnost $\mathbf{Y}$ hodnotu k.

Prvky posloupnosti $p_0, p_1, \ldots$ vyhovují následujícím podmínkám:

$$p_k \geq 0 \text{ pro všechna } k = 0, 1, \ldots \text{ a } \sum_{k=0}^{\infty} = 1.$$

Pro někoho, kdo není zvyklý na jazyk teorie pravděpodobnosti může tato definice být nejasná nebo záhadná. Ale ani pro odborníka v teorii pravděpodobnosti není definice přesná - museli bychom dodat, že náhodný výběr je uniformní, tedy, že každý graf s n vrcholy má stejnou pravděpodobnost býti vybrán.

V tomto článku budeme pravděpodobnost chápat v její nejjednodušší podobě jako podíl počtu případů příznivých ku případům všem. Výše uvedenou definici čísel p_k tedy budeme chápat následujícím způsobem: p_k je podíl počtu grafů s danou množinou n vrcholů, pro které má funkce Y hodnotu k a počtu všech grafů s touto množinou vrcholů. Ale i tak budeme o posloupnosti $p_0, p_1, \ldots$ mluvit jako o posloupnosti pravděpodobností odpovídajících funkci $\mathbf{Y}$.

Výse uvedené vlastnosti jsou pro takto definovanou posloupnost zřejmé. U všech zde studovaných posloupností bude navíc platit, že jen konečné množství jejich prvků je nenulových a proto nebude žádný problém se součty přes celou posloupnost.

Definice 1 *Nechť funkci* $\mathbf{Y}$ *odpovídá posloupnost pravděpodobností. Pak pro každé celé nezáporné číslo* s *budeme výrazu*

$$\sum_{k}^{\infty} p_k k^s$$

říkat s-tý moment *funkce* $\mathbf{Y}$.

Zde budeme uvažovat pouze dva momenty pro $s = 1, 2$. První moment budeme obvykle nazývat *střední hodnota* funkce $\mathbf{Y}$ a budeme jej značit $E(\mathbf{Y})$ a s-tý moment budeme značit $E(\mathbf{Y}^s)$.

Pro střední hodnotu platí velmi důležitá a přitom docela jednoduchá věta, pro jejíž formulaci potřebujeme následující pojem:

Definice 2 *Nechť* $\mathbf{Y}'$ *a* $\mathbf{Y}''$ *jsou dvě číselné funkce definované na grafech. Potom je funkce* $\mathbf{Y}' + \mathbf{Y}''$ *definována jako číselná funkce na grafech, která libovolnému grafu* G *přiřazuje číslo* $\mathbf{Y}'(G) + \mathbf{Y}''(G)$.

Je zřejmé, že funkce $\mathbf{Y}' + \mathbf{Y}''$ přiřazuje grafu G číslo k právě když pro nějaké celé ℓ v rozmezí $0 \leq \ell \leq k$ je $\mathbf{Y}'(G) = \ell$ a $\mathbf{Y}''(G) = k - \ell$.

Jsou-li tedy posloupnosti pravděpodobností odpovídajících funkcím $\mathbf{Y}'$ a $+\mathbf{Y}''$ posloupnosti $p'_0, p'_1, \ldots$ a $p''_0, p''_1, \ldots$, pak součtové funkci $\mathbf{Y}' + \mathbf{Y}''$ odpovídá posloupnost

$$q_0, q_1, \ldots \text{ kde } q_k = \sum_{\ell=0}^{k} p'_\ell p''_{k-\ell}.$$

Nyní tedy ta věta

Věta 1 *Nechť* $\mathbf{Y}'$ *a* $\mathbf{Y}''$ *jsou dvě číselné funkce definované na grafech. Pak*

$$E(\mathbf{Y}' + \mathbf{Y}'') = E(\mathbf{Y}') + E(\mathbf{Y}'').$$

Důkaz: Na základě výše uvedeného je

$$E(\mathbf{Y}' + \mathbf{Y}'') = \sum_{k=0}^{\infty} k \sum_{s=0}^{k} p'_s p''_{k-s}.$$

Je to suma dvourozměrného pole čísel $(s+t)p'_s p''_t$ přes $s, t = 0, 1, 2, \ldots$, viz obrázek, kde se ale sčítají postupující diagonální pruhy, indexované číslem k:

Tatáž čísla ale můžeme sečíst také jako

$$\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} (s+t) p'_s p''_t.$$

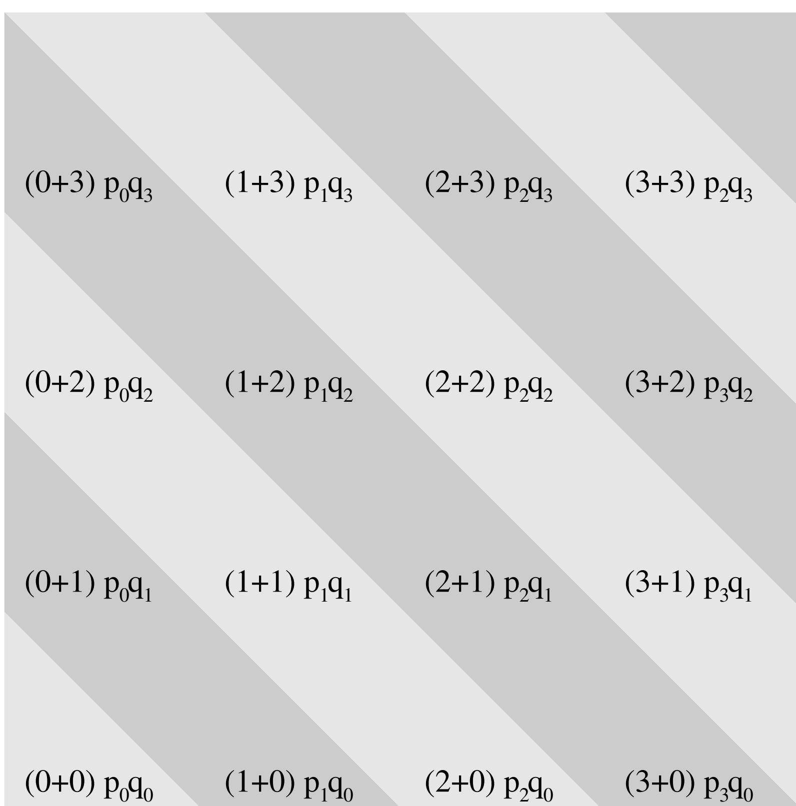
Tuto sumu ale můžeme rozložit na dvě:

$$\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} s p'_s p''_t + \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} t p'_s p''_t.$$

V levém dvojitém součtu vytkneme sp'_s před sumu podle t a využijeme toho, že $p''_0 + p''_1 + \cdots = 1$:

$$\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} s p'_s p''_t = \sum_{s=0}^{\infty} s p'_s \sum_{t=0}^{\infty} p''_t = \sum_{s=0}^{\infty} s p'_s = E(Y'),$$

v pravé sumě prohodíme sčítání podle s a t a obdobně dokážeme, že je rovna $E(Y'')$. ♣



Nechť n a m jsou celá čísla a funkce $\mathbf{Y}$, které odpovídá posloupnost $p_0, p_1, \ldots$, je funkce, která udává, kolik má graf na dané množině vrcholů o velikosti n celkově klik velikosti m. Pak nás bude zajímat hlavně koeficient p_0: když je velmi malý, bude to znamenat, že grafů s n vrcholy, které *nemají žádnou* kliku velikosti m je málo (vzhledem k počtu grafů s n vrcholy. Pokud je hodně blízký 1, znamená to zase, že grafů s n vrcholy, které mají kliku velikosti m je málo.

Je snadné určit, kolik je grafů s danou množinou vrcholů. Má-li ta množina n prvků, pak je dohromady $\binom{n}{2}$ neuspořádaných dvojic různých vrcholů, tedy možných hran takového grafu. Graf je pak dán jak podmnožina této množiny dvojic a takových podmnožin je $2^{\binom{n}{2}}$. Spočítat ale, kolik je grafů, které mají daný počet klik určené velikosti (tedy určit koeficienty p_k dané funkce) je obtížné.

Na druhé straně věta o střední hodnotě součtu dává možnost zcela přesně určit hodnoty momentů funkce. Ukázeme nyní dvě věty, které umožní pomocí prvního a druhého momentu funkce $\mathbf{Y}$ dát docela dobré odhady pro velikost koeficientu p_0.

Věta 2 *Markovova nerovnost.*
Nechť $\mathbf{Y}$ *je číselná funkce definovaná na grafech, které odpovídá posloupnost pravděpodobností* $p_0, p_1, \ldots$. *Pak* $1 - p_0 \leq E(\mathbf{Y})$.

Důkaz:

$$1 - p_0 = \left(\sum_{k=0}^{\infty} p_k \right) - p_0 = \sum_{k=1}^{\infty} p_k \leq \sum_{k=1}^{\infty} p_k \cdot k = \sum_{k=0}^{\infty} p_k \cdot k = E(\mathbf{Y}).$$

♣

Nechť ϕ je booleovská funkce, definovaná na grafech. Můžeme ji také popsat jako *vlastnost* grafu - graf G má tu vlastnost, jestliže $\phi(G)$ je 'true'. Pro dané n budeme výrazem $Prob(\phi)$ označovat podíl počtu grafů na dané n-vrcholové množině, které mají vlastnost ϕ a všech grafů na té množině. Neformálně budeme říkat, že je to pravděpodobnost, že graf s n vrcholy spolňuje ϕ ('Prob' je z anglického 'Probability', tedy pravděpodobnost).

Věta 3 *Čebyševova nerovnost.*
Nechť $\mathbf{Y}$ *je číselná funkce definovaná na grafech, které odpovídá posloupnost pravděpodobností* $p_0, p_1, \ldots$ *a* t *je kladné reálné číslo Pak*

$$Prob(|Y - E(\mathbf{Y})| \geq t) \leq \frac{E(\mathbf{Y}^2) - E^2(\mathbf{Y})}{t^2}.$$

Důkaz: Definujme booleovskou ϕ tak, že pro daný graf G říká, že $|Y(G) - E(\mathbf{Y})| \geq t$. Dále definujme reálnou funkci f tak, že $f(x) = 1$, pokud $|x - E(\mathbf{Y}| \geq t$ a $f(x) = 0$ jinak (černá funkce na obrázku s nespojitostmi v bodech $E(Y) - t$ a $E(Y) + t$). Pak

$$Prob(|Y - E(\mathbf{Y})| \geq t) = \sum_{k=0}^{\infty} p_k f(k).$$

Definujme ještě $g(x) = (x - E(\mathbf{Y}))^2/t^2$ (zelená funkce, jejíž grafem je parabola). Je zřejmé, že $f(x) \leq g(x)$ pro každé x.

Proto tedy

$$Prob(|Y - E(\mathbf{Y})| \geq t) = \sum_{k=0}^{\infty} p_k g(k) = \frac{1}{t^2} \sum_{k=0}^{\infty} p_k (k - E(\mathbf{Y}))^2 =$$

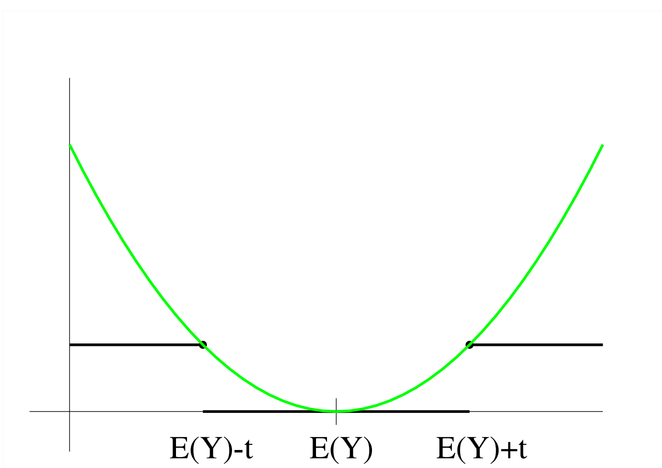
$$\frac{1}{t^2} \sum_{k=0}^{\infty} p_k (k^2 - 2kE(\mathbf{Y}) + E(\mathbf{Y}^2) =$$

$$\frac{1}{t^2} \left(\sum_{k=0}^{\infty} p_k \cdot k^2 - 2E(\mathbf{Y}) \sum_{k=0}^{\infty} p_k \cdot k + E(\mathbf{Y}^2 \sum_{k=0}^{\infty} p_k \right) =$$

$$\frac{1}{t^2} \left(E(\mathbf{Y^2} - 2E(\mathbf{Y})E(\mathbf{Y}) + E(\mathbf{Y}^2) \right) =$$

$$\frac{1}{t^2} \left(E(\mathbf{Y^2} - E(\mathbf{Y}^2) \right).$$

♣



Věta 4 *Důsledek Čebyševovy nerovnosti.*
Nechť $\mathbf{Y}$ *je číselná funkce definovaná na grafech, které odpovídá posloupnost pravděpodobností* $p_0, p_1, \ldots$ *a* t *je kladné reálné číslo Pak*

$$Prob(\mathbf{Y} = 0) \leq \frac{E(\mathbf{Y}^2 - E^2(\mathbf{Y})}{t^2}.$$

Důkaz: Stačí položit $t = E(\mathbf{Y})$, protože když $\mathbf{Y} = 0$, je $|Y - E(\mathbf{Y})| \geq t$. ♣

A nyní se již můžeme pustit do počítání, jak velká je nejčastěji největší klika v grahu, který má n vrcholů.

Označme jako Y_m funkci, která grafu G přiřadí počet klik velikosti m v tomto grafu. Pro použití Markovovy a Čebyševovy nerovnosti potřebujeme spočítat $E(Y_m)$ a $E(Y^2)$.

Použijeme trik, který je postaven na větě o aditivitě střední hodnoty: Nechť A je množina vrcholů grafu G, která má velikost m. Nechť Y_A je funkce, která udává počet klik grafu G, které mají m vrcholů a jsou rovny množině A. Je to trochu krkolomná formulace, která znamená, že $Y_A(G)$ je buď 1, pokud A je klika, nebo 0, pokud není klika.

Ale je zřejmé, že

$$Y_m = \sum_{|A|=m} Y_A,$$

tedy Y_m je součet všech $\binom{n}{m}$ funkcí Y_A přes všechny množiny A vrcholů grafu G velikosti m.

Proto tedy podle věty o aditivitě střední hodnoty je

$$E(Y_m) = \sum_{|A|=m} E(Y_A).$$

Jelikož Y_A má jen hodnoty 0 a 1, je její střední hodnota rovna pravděpodobnosti, že je rovna jedné, což zase je počet grafů, ve kterých A je klikou, dělená počtem všech grafů na dané množině vrcholů.

Počet všech grafů je $2^{\binom{n}{2}}$, protože $\binom{n}{2}$ je neuspořádaných dvojic vrcholů, tedy dvojic vrcholů, které mohou ale nemusejí být spojeny hranou a konkrétní graf je dán podmnožinou této množiny, představující ty dvojice, které jsou spojeny hranou.

Předpokládáme-li, že daná množina A s m vrcholy je klikou, pak dvojic, u kterých můžeme volit, zda budou spojeny nebo ne je o $\binom{m}{2}$ méně, neboli $\binom{n}{2} - \binom{m}{2}$. Grafů, ve kterých A je klika je tedy jen $2^{\binom{n}{2}-\binom{m}{2}}$, takže

$$E(Y_A) = \frac{2^{\binom{n}{2}-\binom{m}{2}}}{2^{\binom{n}{2}}} = 2^{-\binom{m}{2}}.$$

Dosazením do výše uvedeného vztahu tedy dostáváme

$$E(Y_m) = \sum_{|A|=m} 2^{\binom{-m}{2}} = \binom{n}{m} 2^{-\binom{m}{2}}.$$

Teď použijeme Markovovu nerovnost, abychom dokázali, že pro libovolné malé kladné ε a velké n je počet grafů na dané n-prvkové množině vrcholů s klikou velikosti $2(1+\varepsilon)\log_2 n$ velmi malý ve srovnání s počtem všech grafů. Stačí k tomu dokázet, že $E(Y_m)$ je za těchto okolností velmi malé.

$$E(Y_m) = \binom{n}{m} 2^{-\binom{m}{2}} \leq n^m 2^{-m(m-1)/2} = \left(n 2^{-(m-1)/2}\right)^m \leq$$

$$\leq \left(\sqrt{2} n 2^{-(1+\varepsilon)\log_2 n}\right)^m = \left(\sqrt{2} n n^{-(1+\varepsilon)}\right)^m = \left(\sqrt{2} n^{-\varepsilon}\right)^m \to 0.$$

Komplikovanější je dokázet, že pro $m \leq 2(1-\varepsilon)\log_2 n$ je velmi pravděpodobné, že graf s n vrcholy *má* kliku elikosti m. Podle Čebyševovy nerovnosti k tomu stačí ukázat, že $E(Y_m^2)/E^2(Y_m) \to n$ pokud se n blíží k ∞.

Především je třeba si uvědomit, že Y_m^2 je funkce, která grafu G přiřazuje počet uspořádaných dvojic (A, B), kde A i B jsou kliky velikosti m.

Další krok spočívá v definici funkcí $Y_{A,B}$, kde A i B jsou množiny m vrcholů grafu, která je je rovna počtu dvojic klik v grafu, které jsou rovny dvojici (A, B). Takže zase hodnota této funkce je 1, jestliže A i B jsou kliky a je nula jinak.

Stejně jako výše,

$$E(Y_m^2) = \sum_{|A|=|B|=m} E(Y_{A,B}).$$

$E(Y_{A,B})$ je zase rovno počtu grafů, ve kterých A i B jsou kliky ku počtu grafů všech. To druhé číslo známe, kolik je to první. Je-li Q počet dvojic vrcholů, které leží v klice A nebo klice B, pak je to $2^{\binom{m}{2}-Q}$, tedy $E(Y_{A,B} = 2^{-Q}$.

Kolik dvojic vrcholů leží v klice A nebo klice B. To závisí na velikosti průniku $A \cap B$. Označíme-li si jako j velikost toho průniku, pak takových dvojic je $2\binom{m}{2} - \binom{j}{2}$.

Proto

$$E(Y_{A,B}) = 2^{2\binom{m}{2}-\binom{j}{2}},$$

kde $j = |A \cap B|$.

Pro pevné n a m tedy $E(Y_{A,B})$ závisí jen na velikosti průniku $|A \cap B|$.

Označme si tedy jako $Q_{m,j}$ počet dvojic klik velikosti m, majících průnik velikosti j.

Pak

$$E(Y_m^2) = \sum_{j=0}^{m} E(Q_{m,j}.$$

Uvážíme-li, že množinu A můžeme volit $\binom{n}{m}$ způsoby, její průnik s množinou B celkově $\binom{m}{j}$ způsoby a množinu $B - A$ celkově $\binom{n-m}{m-j}$ způsoby, pak

$$E(Q_{m,j} = \binom{n}{m}\binom{m}{j}\binom{n-m}{m-j}2^{2\binom{m}{2}-\binom{j}{2}}.$$

Vydělíme-li tyto hodnoty druhou mocninou $E(Y_m)$, dostaneme, že

$$\frac{E(Y_m^2)}{E^2(Y_m)} = \sum_{j=0}^{m} c_j,$$

kde

$$c_j = \frac{\binom{m}{j}\binom{n-m}{m-j}}{\binom{n}{m}} 2^{\binom{j}{2}}.$$

Spočítáme hodnoty pro $j = 0, m$:

$$c_0 = \frac{\binom{n-m}{m}}{\binom{n}{m}} \to 1,$$

$$c_m = \frac{1}{\binom{n}{m}} 2^{-\binom{j}{2}} = \frac{1}{E(Y_m)}.$$

Předpokládejme nyní, že pro nějaké kladné ε je $m \le 2(1 - varepsilon) \log_2 n$. Pracným, ale zcela přímočarým způsobem se dá dokázat, že hodnoty $c_1 \to 0$ a c_j nejprve pro j menší než zhruba $\log_2 n$ prudce klesají a pak zase pro j blížící se k m dosti prudce rostou, takže $c_2 + \cdots + c_{m-1}$ jsou majorizovány součtem $c_0 + c_1 + c_m$.

Pokud $m \le 2(1 - \varepsilon) \log_2 n$, pak

$$E(Y_m) = \binom{n}{m} 2^{-\binom{m}{2}} \ge (n-m)^m 2^{-m(m-1)/2} = \left(\sqrt{2}(n-m)2^{-m/2}\right)^m =$$

$$= \left(\sqrt{2}(n-m)2^{-(1-\varepsilon)\log_2 n}\right)^m = \left(nn^{-1+\varepsilon}\right)^m = (n^\varepsilon)^m = n^{\varepsilon m} \to \infty.$$

To tedy znamená, že $c_1 + \cdots + c_m \to 0$ a jelikož $c_0 \to 1$, je $E(Y_m^2)/E^2(Y_m) \to 0$ a proto podle Čebyševovy nerovnosti je počet grafů *bez klik* velikosti menší než $2(1 - \varepsilon) \log_2 n$ velmi malý vzhledem k počtu všech grafů na stejné množině vrcholů.

Naprostá většina grafů s danou množinou n vrcholů má tedy největší kliku v rozmezí $2(1 \pm \varepsilon) \log_2 n$.