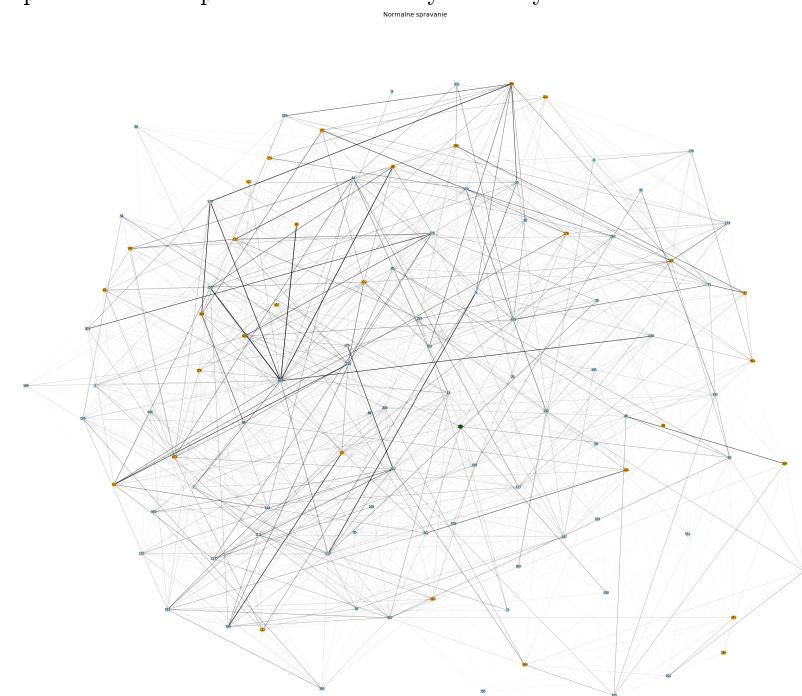


VEDA O SIEŤACH / ALGORITMY NA SIEŤACH 2023 DOMÁCA ÚLOHA 2

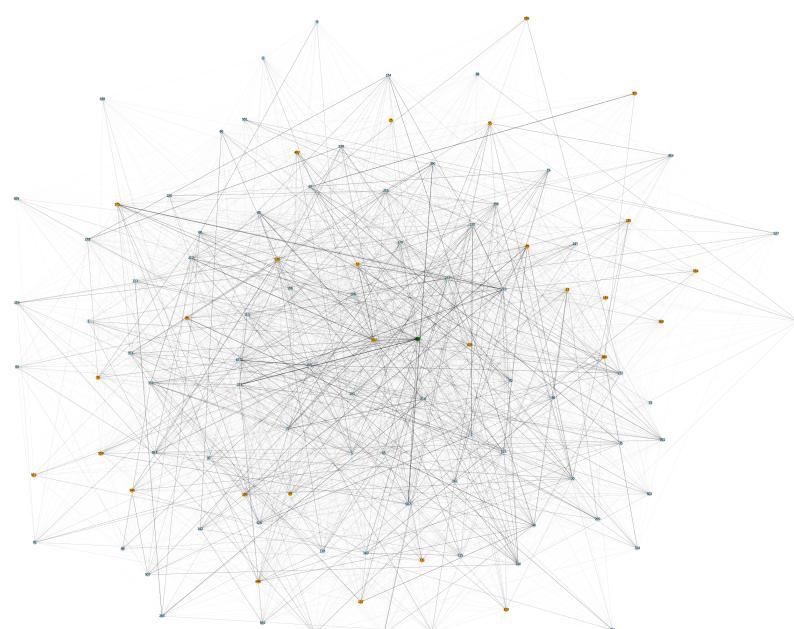
Autor: Marián Kravec

1 Úloha 1

- (a) V prípade pozorovaného správania mravcov vyzerala výsledná sieť následovne:



V prípade náhodného správania mravcov vyzerala výsledná sieť následovne:



V oboch grafoch sú mravce typu "nurse" označené svetlo modrou farbou, "forager" oranžovou a "queen" zelenou. Intenzita (v realite hrúbka) hrany udáva celkový čas kontaktu medzi dvojicou mravcov (čím tmavšia/hrubšia tým dlhší kontakt). (Obrázky vo vyššej kvalite v priloženej zložke "kravec_kod")

- (b) Na výpočet modularity sme využili funkciu z knižnice "networkx" konkrétnie z triedy "community" metódu "modularity" (zdroj). Táto funkcia používa na výpočet následovný vzorec:

$$Q = \sum_{c=1}^n \left[\frac{L_c}{m} - \gamma \left(\frac{k_c}{2m} \right)^2 \right]$$

V tomto vzorci suma prechádza cez všetky komunity c pričom L_c je počet vnútro-komunitných prepojení, k_c je suma stupňov vrcholov v komuniti a γ je "resolution parameter" ktorým je možné nastavovať či má funkcia preferovať veľké komunity ($\gamma < 1$) alebo malé komunity ($\gamma > 1$).

Ide o redukovaná verzia vzorca:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \gamma \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

Výsledná modularita v sieti s reálnymi dátami je 0.0603 a v náhodnej sieti je -0.0023 . Takže modularita v reálnej sieti je kladná zatiaľ čo v náhodnej je záporná preto môžeme tvrdiť, že v reálnej sieti je výraznejšia klustrovanie vrcholov do komunit.

- (c) Na výpočet hustoty siete sme využili vzorec:

$$\rho = \frac{2m}{n(n-1)}$$

Výsledná hodnota reálnej siete vyšla 0.4007 a náhodnej siete 0.6689. Z toho vidíme, že náhodne vygenerovaná sieť je hustejšia.

- (d) Na výpočet hustoty siete sme využili vzorec:

$$\mu = \frac{m}{n}$$

Priemerný stupeň vrchola reálnej siete vyšiel 20.8381 a náhodnej siete 34.7810. Z toho vidíme, že náhodne vygenerovaná sieť má výrazne väčší stupeň vrcholu.

- (e) Vírus by sa mal lepšie šíriť v náhodne generovanej a to z dvoch dôvodov. Poprvé táto sieť má menšiu (zápornú) modularitu čo znamená, že sieť je tvorená jednou komunitou respektíve, že medzi komunitami je pomerne veľké množstvo interakcie vďaka čomu nie je šírenie vírusu nijak obmedzené, zatiaľ čo v reálnej sieti ktorá má väčšiu (kladnú) modularitu existujú komunity medzi v rámci ktorých šírenie vírusu nie je obmedzené ale je výrazne obmedzené šírenie medzi jednotlivými komunitami. Druhým dôvodom je vyššia hustota a priemerný stupeň vrchola náhodnej siete vďaka v priemere existuje viac jedincov (priamych kontaktov) ktorých môže chorý jedinec nakazit.
- (f) Na spracovanie vstupných dát, vytvorenie sietí a výpočet štatistik siete sme použili programovací jazyk Python konkrétnie knižnice "pandas" (na načítanie súboru) a "networkx" (na vytvorenie sietí a výpočet modularity), zároveň sme naprogramovali funkciu "df_matrix_cleaner" ktorá úlohou je vytvorenie matice susedností obsahujúcej súčty stretnutí. V priloženej zložke "kravec_kod" sa nachádza súbor "dataDU.xlsx" obsahujúci vstupné dátá a súbor "kravec_siete.ipynb" obsahujúci samotný kód na spracovanie grafov, ich vykreslenie a výpočet štatistik.

2 Úloha 2

- (a) Vieme, že gigantický komponent zahŕňa presne polovicu vrcholov čiže $S = 0.5$, chceme zistiť priemerný stupeň vrcholu c na čo môžeme použiť následujúcu rovnicu:

$$S = 1 - e^{-cS}$$

Po dosadení S dostaneme:

$$\begin{aligned} 0.5 &= 1 - e^{-0.5c} \\ e^{-0.5c} &= 0.5 \\ -0.5c &= \ln(0.5) \\ c &= -2\ln(0.5) \\ c &= 1.3863 \end{aligned}$$

Takže priemerný stupeň vrcholu v takejto Poissonovej sieti je 1.3863.

- (b) V prípade určenia pravdepodobnosti vrchola stupňa 5 môžeme postupovať dvomi spôsobmi. Prvý spôsob je, že budeme uvažovať, že rozdelenie stupňa vrchola je binomické ($k \sim Bin(n - 1, p)$) a v takom prípade pravdepodobnosť vrchola stupňa 5 je $P(k = 5) = p^5(1-p)^{n-1-5} \binom{n-1}{5}$ avšak keďže nepoznáme p ani n a nie je možné ich z nami známych informácií vypočítať je táto pravdepodobnosť nevyčísliteľná. V prípade druhého postupu uvažujeme, že informácia "s dostatočne veľkým n " znamená, že môžeme uvažovať, že rozdelenie stupňa vrcholov je Poissonovské ($k \sim Poi(c)$) a v takomto prípade je pravdepodobnosť, že vrchol má stupeň 5 nasledovná $P(k = 5) = \frac{c^5 e^{-c}}{5!}$, keďže hodnotu c už poznáme, môžeme túto hodnotu vyčísiť a dostaneme:

$$P(k = 5) = \frac{1.3863^5 e^{-1.3863}}{5!} = 0.0107$$

Čiže pravdepodobnosť, že vrchol má stupeň 5 je približne 11%.

- (c) Úprimne netuším ako správne postupovať v tejto úlohe. Logickou úvahou, ktorou som sa rozhodol túto úlohu riešiť je, že ak aspoň jeden zo susedov nášho vrcholu patrí do gigantického komponentu tak aj samotný vrchol musí do gigantického komponentu patrí. Podmienka "aspoň jeden" sa však komplikované počíta a je jednoduchšie vypočítať inverznú úlohu "ani jeden nepatrí" a výsledná pravdepodobnosť je komplementom nami hľadanej pravdepodobnosti. Úlohu ani jeden nepatrí do gigantického komponentu môžeme počítať ako u^k kde $u = 1 - S$ čiže pravdepodobnosť, že vrchol nepatrí do gigantického komponentu a k je počet susedov čiže stupeň vrcholu. Keďže nás zaujíma komplement tejto úlohy dostaneme, že hľadaná pravdepodobnosť pre $k = 5$ je $1 - u^5 = 1 - 0.5^5 = 0.96875$ takže takmer 97%.
- (d) Tento problém budeme riešiť následovne, poznáme pravdepodobnosť, že vrchol má stupeň 5 a pravdepodobnosť, že je v gigantickom komponente, vynásobením týchto pravdepodobností dostaneme pravdepodobnosť, že vrchol je stupňa 5 a je v gigantickom komponente. Na to aby sme zistili, aká je proporcia v gigantickom komponente máme 2 možnosti buď to jednoducho predelíme pravdepodobnosťou, že hociktorý vrchol patrí do gigantického komponentu čiže S alebo súčtom pravdepodobností, že vrchol má stupeň k a patrí do gigantického komponentu. Obe tieto spôsoby by mali dať rovnaký výsledok, čiže si môžeme skontrolovať, či náš vzorec na pravdepodobnosť, že vrchol stupňa k patrí do gigantického komponentu. V ukážke nižšie

použijeme verziu s predelením sumou.

$$\begin{aligned} prop(k) &= \frac{c^k e^{-c}(1-u^k)}{\sum_{i=0}^{n-1} c^i e^{-c}(1-u^i)} \\ prop(5) &= \frac{1.3863^5 e^{-1.3863}(1-u^5)}{\sum_{i=0}^{n-1} 1.3863^i e^{-1.3863}(1-u^i)} \\ &= \frac{0.0107 * 0.96875}{0.5} = \frac{0.0103}{0.5} = 0.0207 \end{aligned}$$

Vidíme, že proporcia vrcholov stupňa 5 v gigantickom komponente je takmer 21% zároveň si môžeme všimnúť, že menovateľ vyšiel presne 0.5 čo je naše S čo hovorí, že náš vzorec na určenie pravdepodobnosti, že vrchol stupňa k patrí do gigantického komponentu môže byť správny.

3 Úloha 3

- (a) V sieti kde každý vrchol má stupeň 1 nastáva situácia, že vrcholy tvoria dvojice keďže vo chvíli keď prepojíme 2 vrcholy hranou obe budú "nasýtené", čiže nebudú mať žiadnu voľnú polhranu. Z toho vyplýva, že všetky komponenty v grafe sú tvorené presne dvomi vrcholmi, čo je zanedbateľná veľkosť pre grafy ktoré nie sú veľmi malá (2, 4 alebo 6 vrcholov) a zároveň graf obsahuje veľa rovnako veľkých komponentov preto môžeme tvrdiť, že takýto graf nemá gigantický komponent.
- (b) Pre grafy ktoré majú iba vrcholy jedného stupňa a to stupňa k môžeme pravdepodobnostnú funkciu stupňa zapísť nasledovne:

$$\begin{aligned} P(x) &= \begin{cases} 1 & \text{ak } x = k \\ 0 & \text{ak } x \neq k \end{cases} \\ p_x &= \begin{cases} 1 & \text{ak } x = k \\ 0 & \text{ak } x \neq k \end{cases} \end{aligned}$$

Vieme, že vytvárajúca funkcia pre rozdelenie stupňa vrchola má následovný vzorec:

$$g_0(u) = \sum_{i=0}^{\inf} p_i u^i$$

Ale keďže mi vieme, že pre všetky $i \neq k$ je $p_i = 0$ a pre $i = k$ je $p_k = 1$ tak sa nám vzorec zjednoduchší na nasledovný:

$$g_0(u) = u^k$$

Vytvárajúcu funkciu pre rozdelenie stupňa susedného vrchola vieme vypočítať následovne:

$$g_1(u) = \frac{g'_0(u)}{g'_0(1)}$$

Najskôr si vypočítame deriváciu našej vytvárajúcej funkcie g_0 :

$$\begin{aligned} g_0(u) &= u^k \\ g_0(u) &= k u^{k-1} \end{aligned}$$

Ked' to teraz dosadíme, do vzorca pre g_1 dostaneme:

$$\begin{aligned} g_1(u) &= \frac{g'_0(u)}{g'_0(1)} \\ g_1(u) &= \frac{ku^{k-1}}{k1^{k-1}} \\ g_1(u) &= \frac{ku^{k-1}}{k} \\ g_1(u) &= u^{k-1} \end{aligned}$$

Takže vytvárajúca funkcia pre rozdelenie stupňa vrchola je $g_0(u) = u^k$ a vytvárajúca funkcia pre rozdelenie stupňa susedného vrchola je $g_1(u) = u^{k-1}$.

- (c) Vieme, že veľkosť gigantického komponentu vieme v konfiguračnom modely vypočítať ako:

$$S = 1 - g_0(u)$$

Z predchádzajúcej časti sme vypočítali, že $g_0(u) = u^k$, keďže v tomto prípade máme $k = 3$ dostaneme, že veľkosť gigantického komponentu je:

$$S = 1 - u^3$$

teraz potrebujem vypočítať hodnotu u .

Pre u vieme, že platí rovnica:

$$g_1(u) = u$$

V našom prípade kde $k = 3$ vieme, že $g_1(u) = u^2$ po dosadení do predchádzajúcej rovnice dostaneme:

$$\begin{aligned} u^2 &= u \\ u^2 - u &= 0 \\ u(u - 1) &= 0 \\ u &= \begin{cases} 0 \\ 1 \end{cases} \end{aligned}$$

Ak tento výsledok dosadíme do našej rovnice pre S dostaneme následovné výsledky:

$$\begin{aligned} S &= 1 - u^3 \\ S &= \begin{cases} 1 - 0^3 \\ 1 - 1^3 \end{cases} \\ S &= \begin{cases} 1 \\ 0 \end{cases} \end{aligned}$$

Z výsledkov vyplýva, že v grafe vytvoreným konfiguračným modelom kde všetky vrcholy majú stupeň 3 buď gigantický komponent neexistuje alebo pokrýva celý graf. Čomu úplne nerozumiem ale netuším kde som spravil chybu.