

Analýza sociálnych sietí 2022/2023 – projekty

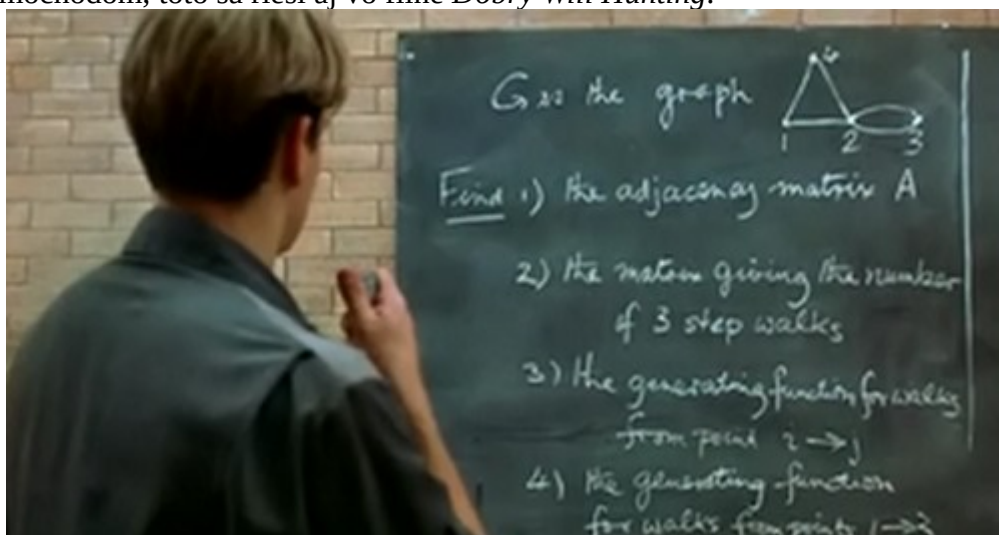
Projekty – časť 1: Centralita

1. Katzova centralita

Centralita vlastného vektora z hodiny berie pri výpočte centrality vrchola do úvahy centrality jeho susedov, teda vrcholov, ktoré sú s ním spojené hranou. Katzova centralita berie do úvahy s nejakou váhou tieto vrcholy tiež, ale s menšou váhou aj vrcholy, do ktorých sa dá dostať v dvoch krokoch, s ešte menšou váhou tie, do ktorých sa dá dostať v troch krokoch atď.

V projekte:

- Súvislosť mocnín matice susednosti siete s takto popísanou dostupnosťou vrcholov. Mimochodom, toto sa rieši aj vo filme *Dobrý Will Hunting*:



- Definícia Katzovej centrality a otázka konvergenie sumy vystupujúcej v definícii, použitie na orientované aj neorientované siete
- Výpočet v R-ku: vplyv parametra, aplikácie

2. Page rank centralita

Ide o centralitu, ktorá stojí za usporiadaním webstránok pri vyhľadávaní. Centralita stránky (vo všeobecnosti vrchola siete) hovorí o tom, s akou pravdepodobnosťou na nej skončíme, ak sa budeme náhodne preklikávať webom cez odkazy na jednotlivých stránkach.

V projekte:

- Súvislosť pravdepodobností a matic.
- Definícia centrality a otázka konvergenie sumy vystupujúcej v definícii,
- Výpočet v R-ku: vplyv parametra, aplikácie

3. Bipartitné siete a centralita ich vrcholov

Bipartitné siete sú také, ktoré majú vrcholy rozdelené do dvoch množín A a B. V sieti existujú len také hrany, ktoré spájajú vrchol z množiny A s vrcholom množiny B. Napríklad z dát, s ktorými sme

robili na hodine pri hľadaní mafiánskych bossov, sa dá zostaviť takáto sieť, kde A sú ľudia a B sú stretnutia.

V projekte:

- Zvolená centralita typu „centrálni ľudia sú tí, ktorí boli na centrálnych stretnutiach a centrálna stretnutia sú tie, na ktorých boli centrálni ľudia“. Definícia, otázka konvergenencie.
- Výpočet v R-ku: bipartitné siete, výpočet centrality, vplyv parametra, aplikácie

Projekty – časť 2: Hľadanie komunit

Po cvičení s touto témou bude doplnená bližšia informácia o týchto témach.

4. Siete so známym rozdelením do komunit a hodnotenie presnosti algoritmov

Príklad siete, ktorá má známe rozdelenie do komunit, sme už videli – karate klub, ktorý je klasickým príkladom. Takýchto sietí je však viac, známe komunity sa označujú ako *ground truth*.

V projekte:

- Príklady takýchto sietí, ich popis a konkrétna práca s vybranými sieťami.
- Použijeme na ne niektoré z algoritmov v balíku *igraph*.
- Vysvetlenie rôznych spôsobov, ako merať presnosť zhľukovania a ako merať podobnosť zhľukovaní (napríklad tak môžeme povedať, ktoré algoritmy dali podobný výsledok).
- Výhody a nevýhody použitých sietí.

5. Naprogramovať si label propagation algoritmus + nejakú jeho modifikáciu

Názov je dosť výstižný. ;)

V projekte:

- Algoritmus, ktorý sa bude programovať, cieľom je mať na konci cvičenia vlastnú implementáciu algoritmu (spoločné programovanie na hodine so spolužiakmi, nie spúšťanie alebo dopĺňanie pripraveného kódu).
- Label propagation je náhodný, zaujíma sa, ako sa správa najčastejšie, inšpirovať sa dá staršou domácou úlohou: <http://www.iam.fmph.uniba.sk/institute/stehlikova/siete21/siete-du2.html>
- Existujú rôzne verzie a modifikácie tohto algoritmu – stručná informácia o nich, alebo naprogramovanie niektorej z nich

6. Hľadanie komunit pomocou vlastného vektora

Ďalšia z metód na hľadanie komunit, referencia: M. E. J. Newman: *Finding community structure in networks using the eigenvectors of matrices*. Phys. Rev. E 74, 036104 (2006), <https://arxiv.org/abs/physics/0605087>

V projekte:

- Na základe časti 4.A článku vysvetlite:
 - ako súvisia vlastné hodnoty a vektory (a akej matice) s modularitou
 - ako z vlastného vektora (a akého) získať rozdelenie vrcholov do dvoch komunit
- Ako cvičenie spoločné naprogramovanie algoritmu (spoločné programovanie na hodine so spolužiakmi, nie spúšťanie alebo dopĺňanie pripraveného kódu).
- Spravte ilustračný príklad (hodí sa karate klub), potom novú zaujímavú aplikáciu
- V R-ku funkcia `cluster_leading_eigen` pokračuje ďalších delení (neobmedzuje sa na dve komunity) – ukážka jej použitia

7. Medzipoloha hrany (*edge betweenness*) a jej využitie pri hľadaní komúnít

Podobne ako medzipoloha vrchola, aj v prípade medzipolohy hrany ide o to, ako často sa vyskytuje v najkratších cestách medzi dvoma vrcholmi.

V projekte:

- Vysvetlite, čo je medzipoloha hrany a ako sa počíta (na malom príklade ručne, potom v R-ku).
- Ako súvisí medzipoloha hrany s hľadaním komúnít.
- Ako konkrétne z informácií o medzipolohy *hrán* vznikne zhlukovanie *vrcholov*?
- Príklady, aplikácie.

Všeobecné poznámky

- Projekt je možné pripraviť samostatne alebo v dvojici. V prípade dvojice sa musia prezentácie aktívne zúčastniť obaja členovia dvojice.
- Cieľom je naštudovať si novú tému, ktorá je blízka tomu, čo sme už preberali a pripraviť si vyučovaciu hodinu, na ktorej tému spolužiakom vysvetlíte a pripravíte si pre nich cvičenia. Štruktúra je teda rovnaká, ako pri ostatných témach, ktoré preberáme. Počas jednej dvojhodinovky sa takto preberú dva projekty.
- Prezentácie musia prebehnúť do začiatku mája, konkrétne rozdelenie termínov si dohodneme.
- Hodnotenie:
 - 20 bodov – priebežná verzia, skonzultovaná osobne aspoň týždeň pred prezentáciou, zvyšok hodnotenia je za finálnu verziu
 - 20 bodov – pripravené slajdy, porozumenie ich obsahu
 - 20 bodov – pripravené výpočty v R-ku, ich porozumenie
 - 20 bodov – prezentovanie

Hodnotenie podrobnejšie:

- **(20 bodov)** Priebežná verzia, ktorú mi pošlete a následne si ju prejdeme, potom bude čas na dopracovanie – toto musí prebehnúť aspoň týždeň pred prezentáciou. Na plný počet netreba mať hotový projekt, ale treba mať naštudovanú teóriu (nie je jej v žiadnom projekte veľa), v R-ku spravené aspoň základné výpočty na nejakých dátach (ostatné výpočty môžete doplniť neskôr, ale malo by vám byť jasné, že viete, ako na to, tiež doplniť ďalšie dát). Konzultácia nemá vyzeráť tak, že vám budem vysvetľovať nejaký dôkaz alebo použitie R-kovskej funkcie, ale napríklad že vám dám námety, ako prezentovať numericky nejaký výsledok, ktorý sa nebude dokazovať, ktoré z nájdených aplikácií by mohli byť vhodné, čo by sa z nich dalo transformovať na cvičenie a pod.
 - Samozrejme, konzultovať sa dá aj skôr a priebežne pred odovzdaním tejto verzie (a je to odporúčané, minimálne na začiatku povedať, do čoho sa idete pustiť a dostať k tomu nejakú spätnú väzbu)
- Vo finálnej verzii potom:
- **(20 bodov)** Pripravené slajdy, porozumenie ich obsahu. Čo nie je dobré: iba vzorce bez vysvetlenia, dlhé výpočty na celý slajd skopírované z nejakého textu, kompletný kód na spustenie (tak ako na ostatných hodinách, práca študentov je aktívna, nespočíva len v spustení kódu alebo dopísaní niekoľkých parametrov). Na konzultácii predbežnej verzie sa o tom dá hovoriť.
 - **(20 bodov)** Pripravené výpočty v R-ku, ich porozumenie. Aspoň snaha o efektívny kód (nemusi byť „dokonalý“), bez „úplne hrozných vecí“ (nebudeme hľadať maximum vektora tak, že v cykle prechádzame prvky alebo vytvárať identickú maticu dvojitém for-cykлом).
 - V súvislosti s predchádzajúcimi dvoma bodmi treba myslieť aj na zmysluplnosť príkladov a aplikácií. To, že mám nejakú sieť neznamená, že sa na ňu dá všetko rozumne použiť. Napríklad „vygenerujme náhodnú sieť a spočítajme centralitu vrcholov“ nie je užitočný a názorný príklad (ak nechcete ukazovať vlastnosti náhodných sietí), ale „vygenerujeme dve nezávislé náhodné siete s veľkým počtom hrán, prepojíme ich niekoľkými hranami a pôjdeme hľadať komunity“ užitočné je,

lebo tým vytvoríte sieť so známym rozdelením vrcholov a môžete tak testovať výsledok. Takisto môžete uviesť naviac ukážky z článkov (napríklad obrázok siete s vyznačenými výsledkami, ktorý vysvetlíte, nie napr. slajd typu „v [1] sa metóda použila v biológii, v [2] pri zostavení portfólia a v [3] pri analýze siete metra” bez toho, aby ste o tom vedeli niečo viac povedať – aspoň 2-3 vety, s akou sieťou robili a čo im vyšlo.)

- **(20 bodov)** Prezentovanie. Prezentácia má byť vysvetľovanie, nie prečítanie slajdov. Nehovorte tvrdenia, ktorým nerozumiete a sú len prebrané z nejakého textu. Treba si premyslieť štruktúru hodiny, nielen spísať tému do slajdov. Počas práce spolužiakov v R-ku treba sledovať, ako to ide a podľa potreby im pomôcť.