

Centralita vrcholov v sociálnych sieťach – centralita vlastného vektora

(slajdy siete02.pdf, od str. 21)

- Základná myšlienka (ako je uvedené aj na str. 22) je, že chceme navrhnúť centralitu, ktorá by závisela od centrality susedných vrcholov. Inak povedané, vysokú centralitu by mal byť taký vrchol, ktorý je spojený v inými veľmi centrálnymi vrcholmi. Áno, je to cyklická definícia, ale uvidíme, že to nevádi. Vektor centralít bude vystupovať aj na ľavej aj na pravej strane rovnosti, ktorá ho definuje – ale to v princípe vadit' nemusí, ak ho z tej rovnosti budeme vedieť vyjadriť. V tomto prípade to pôjde.
- Na str. 22 máme (pracujeme s neorientovanou sieťou):
 - ▶ Ak A je matica susednosti (prípadne aj s váhami) a v je vektor centralít, tak by malo pre najaké kladné číslo c platiť

$$v_i = c \sum_j A_{ij} v_j$$

- Najskôr si zoberme neváženú sieť. V tom prípade má matica A hodnoty 0 a 1 podľa toho, či sú dané vrcholy spojené hranou. To znamená, že suma na pravej strane predstavuje súčet centralít vrcholov, ktoré susedia s vrcholom i . Nejaký násobok tohto súčtu by mal byť centralitou tohto vrchola.
- Teraz nech je sieť vážená. V tom prípade sú prvky matice A váhy. Tým pádom suma vyjadruje vážený súčet centralít susedných vrcholov, pričom váhami v súčte sú váhy hrán. Ak je vrchol i spojený s vrcholom j hranou s vyššou váhou, centralita vrchola j bude do výpočtu centrality vrchola i vstupovať výraznejšie. Čo je presne to, čo intuitívne chceme.

Toto nakoniec zapíšeme maticovo a uvidíme, že dostávame definíciu vlastného vektora. Preto sa tejto centralite hovorí **centralita vlastného vektora** (angl. **eigenvector centrality**).

- Na str. 23, 24, 25 máme príklad:
 - Okrem iného je tu užitočná funkcia **graph_from_literal**, pomocou ktorej môžeme „ručne“ zadať nejaký graf. Môže sa to hodiť pri vytváraní nejakých ilustračných obrázkov, keď sú uvažované siete malé. Presne to je táto situácia.
 - Funkcia **eigen_centrality** počíta centralitu vlastného vektora. Pozrite si celý výstup z tejto funkcie, aby ste videli, prečo potrebujeme cez dolár pristupovať k hodnote **vector**, keď chceme vektor centralít s hodnotami pre jednotlivé vrcholy siete.
 - Na str. 25 je obrázok, ktorý je jednou z možných vizualizácií centralít. Centralita sa vhodne preškáľuje, aby bol boli veľkosti vrcholov primerané, t. j. aby sa nám obrázok páčil ;-).
 - Zopakujte si tento výpočet a grafické znázornenie pre nasledovné siete
 - **g <- graph_from_atlas(123)** (pomocou **?graph_from_atlas** si pozrite, aké siete generuje táto funkcia, tiež môže byť užitočná pri vyrábaní ilustračných obrázkov).
 - karate klub, s ktorým sme už pracovali (z balíka **igraphdata**, ak si chcete zobrazit' všetky siete dostupné v tomto balíku, dosiahnete to pomocou **data(package = "igraphdata")**)
- Teraz, keď už vidíme, že táto centralita dáva rozumné výsledky, poďme sa pozrieť na to, ako sa počíta.
 - Doterajšia úvaha, že by to mal byť vlastný vektor, nestačí:
 - Matica má toľko vlastných čísel, aký je jej rozmer. Ktoré vlastné číslo, resp. ktorý príslušný vlastný vektor zobrať?

- Centralitu sme použili na definovanie veľkosti vrchola. Vždy teda bola nezáporná, aj keď sme to priamo nekontrolovali. To by aj mala byť – centrality boli doteraz nezáporné, takže takáto by mala byť aj nová centralita. Málo centrálny vrchol by mal mať centralitu blízku nule alebo úplne nulovú, to nám pri niektorých centralitách v minulosti. Vieme, že vlastný vektor môžeme prenásobiť nejakou konštantou a stále zostane vlastným vektorom. To sa dá využiť pri škálovaní. V našej situácii nám to ale nepomôže dosiahnuť kladné zložky, ak má pôvodne získaný vektor aj kladnú aj zápornú zložku. Čo nám zaručí, že vlastný vektor, ktorý budeme chcieť použiť na centralitu, bude mať rovnaké znamienka?
 - „Čo nám zaručí, že...“ z predchádzajúceho bodu – je to Perronova-Frobeniova veta. Uvádza sa v rôznych obmenách (čo je priamo jej tvrdenie, čo je dôsledok, ...), v nejakej podobe si ju môžete pamätať z maticového počtu, ... To, čo z nej budeme potrebovať, je zhrnuté na str. 27. Z nej vyplýva, že treba zobrať vlastný vektor prislúchajúci k vlastnému číslu s najväčšou absolútnou hodnotou a prenásobiť ho tak, aby mal všetky zložky kladné (to sa dá, buď to netreba robiť vôbec, alebo ho stačí prenásobiť mínus jednotkou, ak má zložky záporné), prípadne aby bol preškálovaný tak, ako chceme (napríklad môžeme chcieť, aby maximálna centralita bola 1).
 - Na str. 28 a 29 je postup, ktorým si overíte, že takto (po preškálovaní) dostaneme to isté ako pri priamom výpočte centrality **eigen centrality** – spravte si tento výpočet.
- Na str. 30 je ukážka aplikácie tejto centrality. Mafiánski bossovia majú v priemere vyššiu centralitu vlastného vektora ako ostatní.
- V súvislosti s týmto posledným príkladom sa pozrime ešte na jednu vec: V článku *Identifying Mafia Bosses from Meeting Attendance* sa píše:

on criminal networks [38, 75, 76].¹³ Other measures are eigenvector centrality (calculated on the binary matrix), an adaptation of degree taking into account the degree of a node's contacts (i.e. neighborhood) [38, 77] and the clustering coef-

teda centralita vlastného vektora sa počítala z binárnej matice susednosti. To znamená, že sa nebrali do úvahy váhy hrán, ale len to, či medzi dvoma vrcholmi hrana je alebo nie je. Ak máme v R-ku sieť, nie je problém váhy hrán zrušiť (robili sme to hneď na začiatku slajdov k tejto téme), ale ak to cheme spraviť len jednorazovo pri výpočte centralít, stačí ako parameter funkcie **eigen centrality** zadať **weights = NA**. Porovnajme pre karate klub

- **eigen centrality(karate)**
- **eigen centrality(karate, weights = NA)**

(z týchto výstupov treba ešte vytiahnuť hodnoty vektora, ako to bolo vysvetlené hore).

- Poznámka: Pri centralite blízkosti a medzipolohy boli váhy brané ako vzdialenosti, lebo sa počítali najkratšie cesty medzi vrcholmi. Aby sme teda dostali rozumné výsledky, musíme byť v situácii, že silná väzba medzi vrcholmi je vtedy, keď je váha (t. j. vzdialenosť) malá. Pri centralite vlastného vektora je to naopak, veľká váha hrany musí znamenať silnú väzbu. Ak to v niektorom prípade splnené nie je, treba váhy hrán transformovať alebo uvažovať neváženú sieť.
- Od str. 31 začína námet na projekt, ktorý je motivovaný čudnými výsledkami centrality vlastného vektora použitej pre orientované siete:
 - Vrcholy zo siete zo str. 32 majú všetky rovnakú centralitu, aj keď vrchol D má na rozdiel od ostatných hranu smerujúcu od vrchola E. Jeho centralitu to však nijako nezväčší.
 - Sieť zo str. 34 dáva ešte zvláštnjší výsledok, všetky vrcholy majú centralitu nulovú. R-ko na toto upozorňuje (ak chcete takýto výpočet spraviť). Aké sú dobré miery centrality pre orientované siete, ktoré využívajú podobné myšlienky ako centralita vlastného vektora (ale nespôsobujú takéto problémy)?