

Analýza sociálnych sietí - niekoľko tipov na projekty

1. Zhlukovanie a zhľukovanie v úplných sieťach

Táto téma sa tento školský rok preberať nebude, v slajdoch z minulého roka sú základné pojmy – koeficienty zhľukovania (lokálny a globálny). Treba si o nich ešte niečo asi prečítať, aby ste boli schopní dobre ich vysvetliť a uviesť niekoľko príkladov.

Potom je v týchto slajdoch špeciálna téma – koeficient zhľukovania v úplných vážených sieťach. Pod úplnou sieťou rozumieme takú, že každé dva vrcholy sú spojené hranou. Takéto siete vznikajú napríklad vtedy, keď počítame korelácie medzi nejakými časovými radmi – časové rady predstavujú vrcholy siete a korelácie medzi nimi určujú váhy hrán. V článku *A clustering coefficient for complete weighted networks (2015)* bol navrhnutý postup, ako v takýchto sieťach merať mieru zhľukovania.

Obsah projektu za 60 bodov:

- vysvetliť algoritmus
- ukázať výpočet pomocných funkcií a výsledného zhľukovania pre malú sieť z článku
- implementovať výpočet v R-ku pre ľubovoľnú zadanú sieť
- generovať náhodnú sieť a vytvoriť podobné obrázky ako v článku
- použiť na nejakú reálnu sieť korelácií

Obsah projektu za 80 bodov: Tá reálna aplikácia by mala byť nejaká originálna, nielen „tu je sieť a má takýto koeficient zhľukovania“. Napríklad sa dá porovnať koeficient zhľukovania v rôznych obdobiach a pozrieť sa nejakým spôsobom na to, ktoré vrcholy majú podobný vývoj týchto koeficientov. Niečo takéto som robila s bývalou študentkou, Radkou Litvajovou, tu: https://fmv.euba.sk/www_write/files/veda-vyskum/konferencia-smolenice/2018/Smolenice_2018.pdf

V projekte by malo byť niečo iné, nie takéto 2D zhľukovanie, ale napríklad zhľukovanie pre viacrozmerné vektory charakterizujúce vývoj koeficientu, zhľukovanie časových radov vývoja koeficientu a pod. Alebo aj niečo úplne iné, kde by sa takýto koeficient zhľukovania pre úplné siete dal použiť.

2. Centrality pre orientované siete: Katzova centralita a Page Rank

Téma navrhnutá v slajdoch o centralite vrcholov v sieťach aj s motiváciou.

Obsah projektu za 60 bodov:

- vysvetliť tieto centrality
- ako ich počítať v R-ku
- ilustrovať ich na jednoduchých príkladoch, ktoré by mali byť dobre zvolené tak, aby sa na nich dalo vidieť, čo tieto centrality merajú, čo vyjadrujú
- použiť ich na nejakú reálnu sieť

Obsah projektu za 80 bodov:

Znovu, tak ako v predchádzajúcom nápade (a aj v tých nasledujúcich), aplikácia by mala byť nejakým reálnym prínosom, nie iba ukážkou výpočtu. Môžete si dať hľadať články pomocou scholar.google.com a nájsť nejaké analýzy, kde sa tieto centrality použili. Môžete spraviť niečo podobné a zhodniť, čo vyšlo podobne a čo inak. Môžete si zostaviť vlastnú sieť, ktorú je zaujímavé analyzovať (teda nepoužiť len sieť, ktorá už má na webe pripravené dáta) a nejakou je aktuálna alebo pre vás zaujímavá.

3. Iné vybrané miery centrality

V závislosti od ich jednoduchosti, resp. komplikovanosti si vyberiete jednu alebo viac mier centrality, o ktorých sme pri tejto téme nehovorili. Je ich veľa, stále vznikajú nové. Niekedy sú príliš komplikované pre účely projektu, treba sa vždy pozrieť na algoritmus.

Jeden takýto algoritmus (a v tomto prípade by stačil tento jeden) je napríklad v článku *Ranking the spreading influence of nodes in complex networks - An extended weighted degree centrality based on a remaining minimum degree decomposition (2018)*. Môžete si, samozrejme, vybrať aj inú novú centralitu, resp. niekoľko centralít (ak sú jednoduchšie).

Obsah projektu za 60 bodov:

- vysvetliť výpočet – v tomto prípade ide o „RMD decomposition“, na tom založený „weighted degree“ a „extended weighted degree“ centrality
- naprogramovať výpočet v R-ku, ktorý tieto hodnoty vypočíta pre zadanú sieť
- pomocou týchto funkcií zrekonštruovať hodnoty uvedené v článku pre danú ilustračnú sieť
- použiť výpočet na nejakú reálnu sieť

Obsah projektu za 80 bodov:

Áno, znovu to isté - aplikácia by mala byť nejakým reálnym prínosom, nie iba ukážkou výpočtu.

4. Aplikácia sietí pri konštrukcii portfólia

Nasledujúci obrázok je z článku *A network approach to portfolio selection (2016)*. Čo je to MST, tým sa budeme zaoberať v práve v súvislosti s takýmito sieťami vytvorenými z korelácií. Stručne: MST (minimum spanning tree, v slovenčine minimálna kostra) obsahuje všetky vrcholy pôvodného grafu, je to súvislý graf a neobsahuje cyklus (v tom zmysle je to „kosta“, že spája všetky vrcholy najjednoduchším spôsobom, bez zbytočných cyklov, ktoré k tomu nie sú potrebné). Na obrázku sa dáva do súvislosti centralita vrcholov s váhami pri Markowitzovej diverzifikácii portfólia.

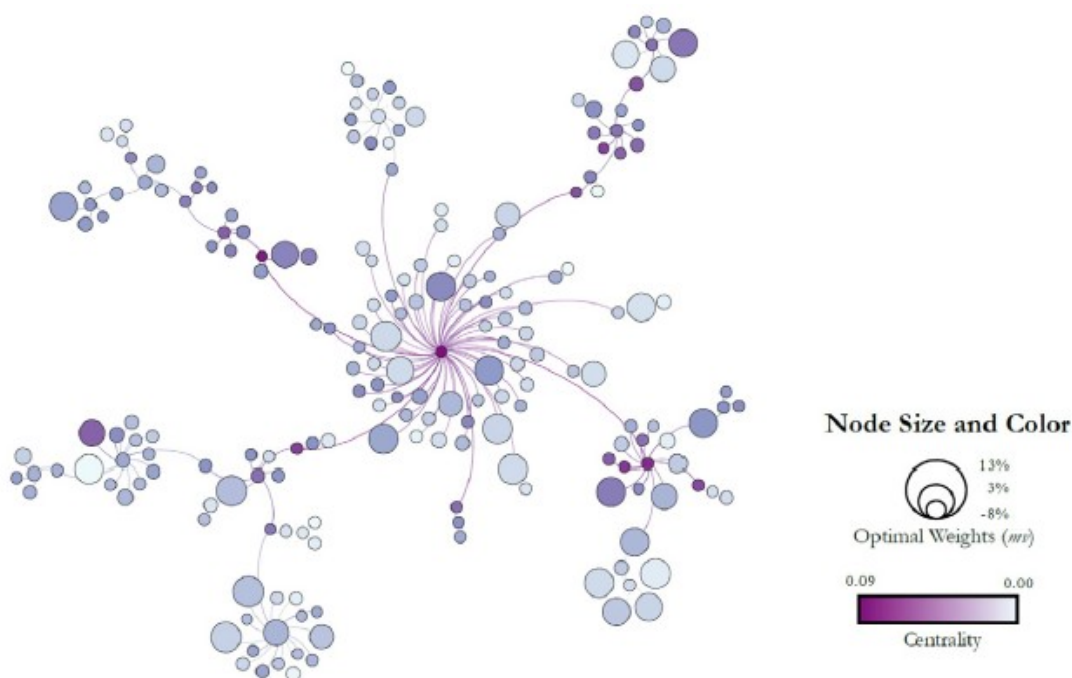


Fig. 2. MST stock market network for the d-S&P dataset. The size of nodes corresponds to the optimal weights for the minimum-variance strategy (see Eq. 3) and the intensity of the colour accounts for the corresponding security's centrality.

Myšlienka je tá, že centrálné vrcholy sa kvôli diverzifikácii dostanú do portfólia s malou váhou.

Obsah projektu za 60 bodov:

- Vytvoriť podobný obrázok pre vlastné dáta.
- Pre vlastné dáta zrekonštruovať aj niektoré z ďalších obrázkov v článku alebo regresii (na ukážku: napríklad je tam lineárna regresia, ktorá modeluje váhy v portfóliu pomocou centrality a štandardnej odchýlky)

Obsah projektu za 80 bodov:

Napríklad nájsť iné články dávajúce do súvislosti siete a optimalizáciu portfólia a otestovať nejakú stratégiu výberu váh, ktorá z takýchto úvah vychádza.

5. Porovnanie centrality z vlastného pohľadu a skutočnej centrality

Obsah projektu za 60 bodov:

V článku *Assessing the Political Landscape - Structure, Cognition, and Power in Organizations* (1990) sú zaujímavé dáta: Každý mal vyplniť tabuľku o svojich spolupracovníkoch a pre každého z nich napísať, ku komu by sa podľa neho išiel poradiť, keby potreboval pomoc, resp. konzultáciu. Tým dostávame možnosť vytvoriť nasledovné dve siete:

- sieť z pohľadu „ku komu by som sa išiel ja poradiť“
- sieť z pohľadu „o kom si myslím, že by sa išiel poradiť so mnou“

Pre nejaké rozumné miery centrality pre orientované siete potom vzniká otázka: Líši sa centralita vypočítaná z týchto dvoch sietí? Vidia sa ľudia reálne? Alebo sa niekto preceňuje (myslí si, že by sa s ním išli mnohí poradiť, ale oni by to nespravili), resp. podceňuje (nemyslí si, že by sa s ním mnohí radili, ale oni ho napriek tomu často označia) – ak áno, ako to súvisí s jeho centralitou? Kto sa podceňuje a kto sa preceňuje?

Obsah projektu za 80 bodov:

Z takýchto odpovedí by sa toho dalo určite zisťovať viac, je tu priestor pre vlastné nápady.