

Hľadanie komúnít v sieťach - časť I.

(slajdy siete03.pdf)

Téma o hľadaní komúnít bude mať dve časti, v závislosti od toho, čo budeme chápať pod „komunitou“:

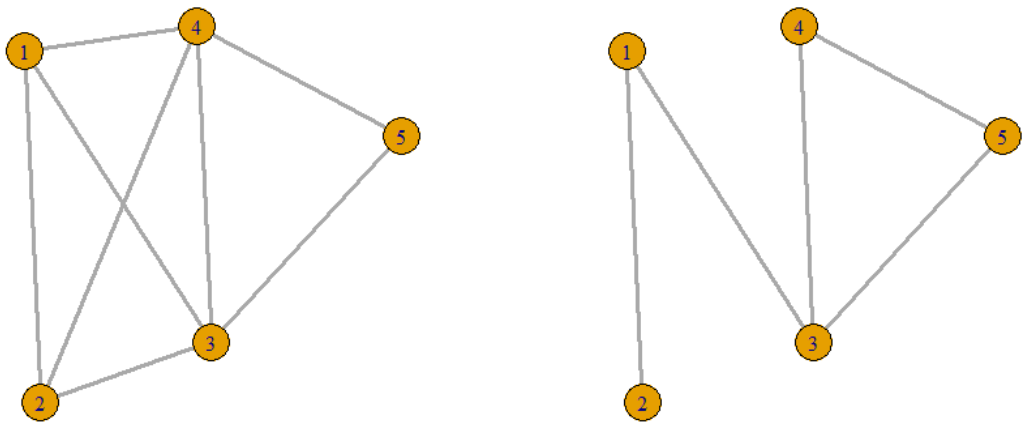
- Môže to byť skupina vrcholov, ktorá je prepojená nejakým konkrétnym spôsobom, napríklad môžeme požadovať, aby každý bol spojený s každým – to bude jedna z vecí, ktoré budeme robiť teraz.
- Alebo to môže byť niečo len intuitívne popísané tým, že vrchol má mať veľa hrán s vrcholmi zo svojej komunity a málo hrán s vrcholmi z ostatných komúnít. Potom nie je úloha definovaná presne a existuje veľa algoritmov na hľadanie takýchto komúnít (*community detection*). Je to podobná situácia ako pri zhlukovaní vektorov, preto sa aj tu niekedy hovorí o zhlukovaní (*clustering*). Toto bude obsahom nasledujúcej témy.

Dnešná téma bude obsahovať dva pojmy:

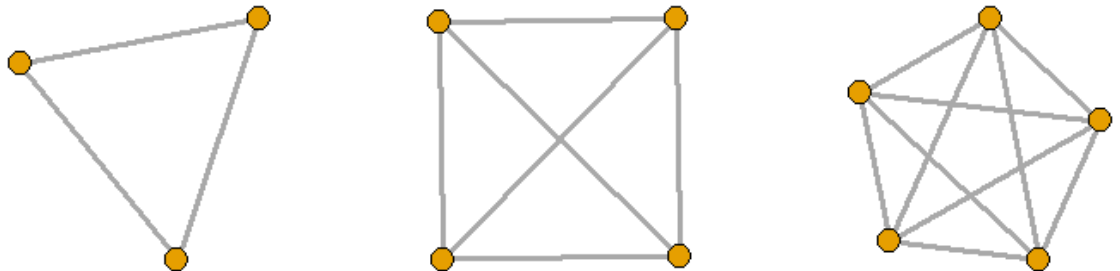
- klika (*clique*)
- jadro (*core*)

1. Klika

- Na str. 5 máme definíciu: *Podgraf, každé dva jeho vrcholy sú spojené hranou*
 - *Podgraf* znamená, že zoberieme nejakú podmnožinu vrcholov a podmnožinu hrán daného grafu. Napríklad zoberme graf na obrázku vľavo. Vpravo je zobrazený jeden jeho podgraf:

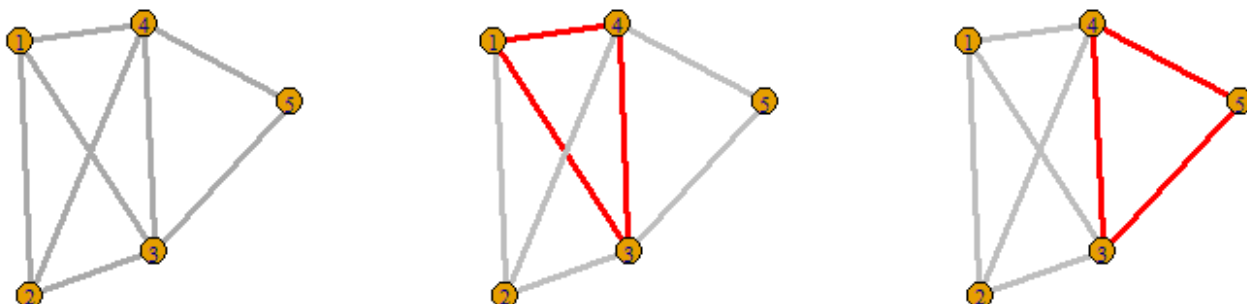


- *Každé dva vrcholy sú spojené hranou* – hľadáme teda v grafe niečo takéto:



To sú kliky veľkosti 3, 4 a 5. (Ako je poznamenané na str. 5, každé dva vrcholy spojené hranou tvoria kliku veľkosti 2. Takisto každý vrchol sám osebe je klikou. Takže prakticky sú zaujímavé kliky, ktoré majú veľkosť aspoň 3)

- Zoberme si znovu graf na obrázku úplne vľavo.
 - Vrcholy 1, 3, 4 (spolu s hranami medzi nimi) tvoria kliku, lebo každé dva z nich sú spojené hranou, ako môžeme vidieť na obrázku v strede
 - Takisto vrcholy 3, 4, 5 (spolu s hranami medzi nimi) tvoria kliku, lebo každé dva z nich sú spojené hranou, ako môžeme vidieť na obrázku vpravo

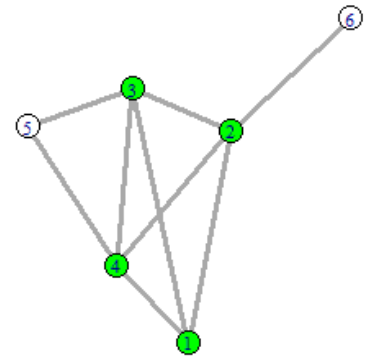
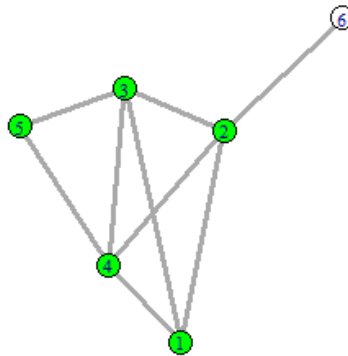
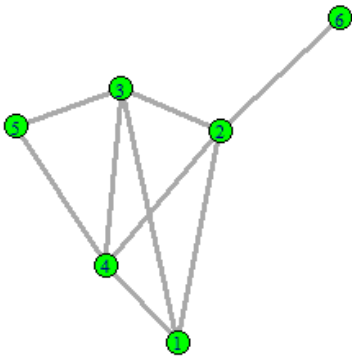


- Str. 5 – pojem *maximálnej kliky*
 - Čo znamená pojem „vlastný podgraf“ - tak ako je každá množina podmnožinou samej seba, tak aj každý graf je podgrafom samého seba, keďže spĺňa definíciu. Preto sa definuje vlastná podmnožina, kde sa táto možnosť vylúči. Analogicky vlastný podgraf
 - Pri maximálnej kliky sa teda požaduje, aby daná klika nebola súčasťou nejakej väčšej kliky.
 - Napríklad v príklade hore:
 - 1, 3, 4 nie je maximálna klika, lebo je súčasťou väčšej (1, 2, 3, 4)
 - 3, 4, 5 je maximálna klika, nie je súčasťou žiadnej väčšej
- Funkcie v R-ku sú na str. 5, na str. 6 je príklad siete a na str. 7 – 10 je použitie týchto funkcií – **príklad 1** – sieť **kite**. Mali by ste pomocou nich vedieť odpovedať na otázky:
 - Koľko je všetkých klík?
 - Koľko je takých, ktoré majú veľkosť aspoň 3?
 - Koľko je maximálnych klík, ktoré majú veľkosť aspoň 3? (Str. 9 – na kontrolu pochopenia definícií – skúste si ich najskôr vypísať na základe obrázka.)
 - Ktoré kliky sú najväčšie?
- **Príklad 2.** na str. 12 – odpovedzte na predchádzajúce otázky pre sieť Zacharyho karate klubu (samozrejme, okrem vypisovania klík na základe obrázku).
- **Príklad 3** so začiatkom na str. 13
 - Na str. 14 máme užitočnú funkciu, pomocou ktorej vieme do R-ka načítať sieť, ak máme zadaný zoznam hrán
 - Podľa inštrukcií:
 - načítajte sieť,
 - vynechajte násobné hrany (je tam aj 1-2 aj 2-1),
 - priradte k vrcholom mená a nakreslite obrázok siete s menami žien
 - nájdite najväčšiu kliku a interpretujte (čo znamená pre túto sieť klika a maximálna klika?)

2. Jadro

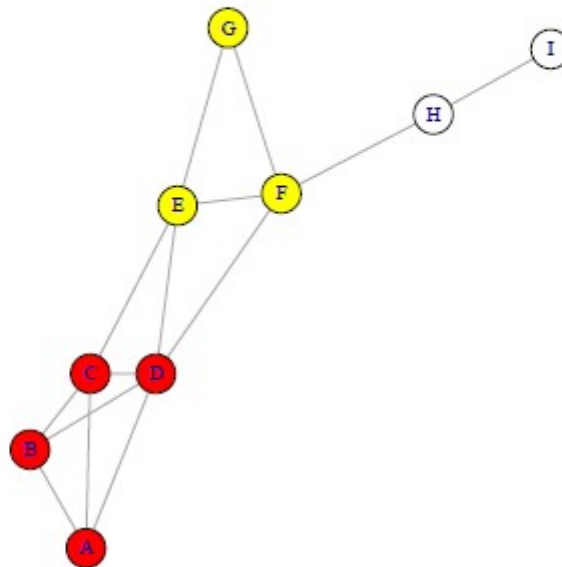
- **K-jadro** - definícia na str. 17: „Maximálny podgraf, ktorého každý vrchol je spojený s aspoň K vrcholmi tohto podgrafu“

- „Maximálny podgraf“ znovu v tom zmysle ako pre kliky – nedá sa zväčšiť
- Zoberme si graf na obrázku vľavo:



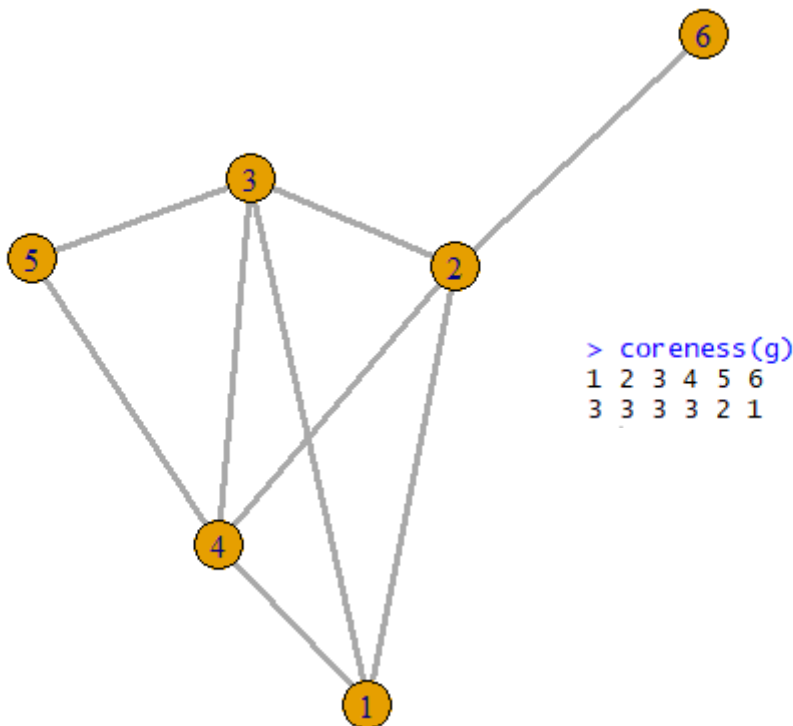
- Graf je súvislý, takže všetky jeho vrcholy patria do 1-jadra. Splňajú totiž podmienku, že sú spojené s aspoň jedným vrcholom z tejto množiny.
 - Na obrázku v strede sú zelenou farbou označené vrcholy, ktoré patria do 2-jadra. Presvedčte sa, že každý z nich je spojený s aspoň dvoma zelenými vrcholmi. Táto množina sa zväčšiť nedá, lebo keby sme do nej pridali ešte vrchol 6, podmienka by už splnená nebola, vrchol 6 by ju nespĺňal.
 - Podobne odôvodnite, prečo 3-jadro obsahuje práve zelené vrcholy na obrázku vpravo.
- Zopakujte si určovanie jadier na príklade zo slajdu 18:

► 3-jadro: A, B, C, D; 2-jadro: A, B, C, D, E, F, G; 1-jadro: všetky



Tu môžeme vidieť, že vrchol H má síce stupeň 2 (je spojený hranou s dvoma vrcholmi), ale napriek tomu sa nenachádza v 2-jadre. Dôvod je ten, že jedna z týchto hrán ho spája s vrcholom I, ktorý je iba v 1-jadre.

- Str. 17: v R-ku máme funkciu **coreness**, ktorá dáva vrcholu hodnotu K v prípade, že tento vrchol je v k-jadre, ale nie je v (k+1)-jadre.
- Napríklad graf z príkladu na predchádzajúcej strane:



Porovnajte s tým, ako sme počítali jadrá.

- Ďalší malý príklad na str. 18 v slajdoch.
- Str. 19 v slajdoch: Zacharyho karate klub, vypočítajte hodnoty *coreness* pre túto sieť a na základe nej zafarbte vrcholy (teda vrcholy s rovnakou hodnotou *coreness* budú mať rovnakú farbu).
- Str. 20, 21 v slajdoch: príklady aplikácií v publikovaných článkoch
- Vrchol, ktorý má vysokú hodnotu *coreness* je v určitom zmysle centrálnejší ako ten, ktorý má túto hodnotu nižšiu.
- Preto môžeme sieť nakresliť tak, že
 - vrcholy budú ležať na sústredných kružniciach
 - na jednej kružnici budú vrcholy s rovnakou hodnotou *coreness*
 - bližšie k stredu budú tie vrcholy, ktoré majú túto hodnotu vyššiu
 - ako je to spravené na str. 22.
- Je tam aj odkaz na kód, ktorý treba ešte trochu upraviť:
 - treba zmeniť stupne na radiány, aby boli body rozmiestnené rovnomerne po celej kružnici
 - nastaviť začiatok – aby nezačínali všetky pod rovnakým uhlom, ale na náhodnom mieste

