

Analýza linků, statické a dynamické algoritmy ohodnocení stránek

Petr Šaloun

VŠB-Technická univerzita Ostrava

FEI, katedra informatiky



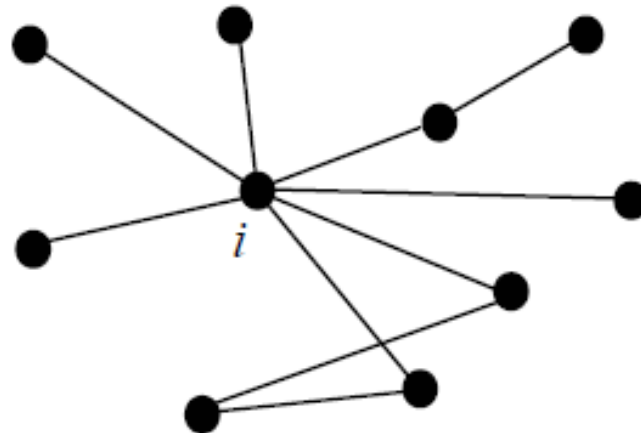
Analýza sociální sítě

- Sociální síť je studium sociálních entit (lidé v organizaci, nazývaní aktoři) a jejich vztahů a interakcí (vzájemného ovlivňování).
- Interakce a vztahy často reprezentovány sítí nebo grafem.
- Každý vrchol (uzel) představuje aktora a každé spojení (hrana) vztah.
- Hledáme i podgrafy – komunity vytvořené skupinami aktorů.



Centrálnost

- Důležití aktoři mají mnoho spojení s jinými
- Spojení nazýváme vazbami.
- Centrální aktor je zahrnutý v mnoha spojeních.
- Aktor i je nejcentrálnější aktor – komunikuje s největším počtem aktorů.



Významnost

- Rozlišujeme spojení od aktora k aktorovi.
- K vypočítání významnosti aktora nás pouze zajímají spojení vedoucí k aktorovi.
- Významnost může být spočítána pouze pokud použijeme orientovaný graf.
- **Centrálnost** – spojení směrem od aktora
- **Významnost** – spojení směrem k aktorovi



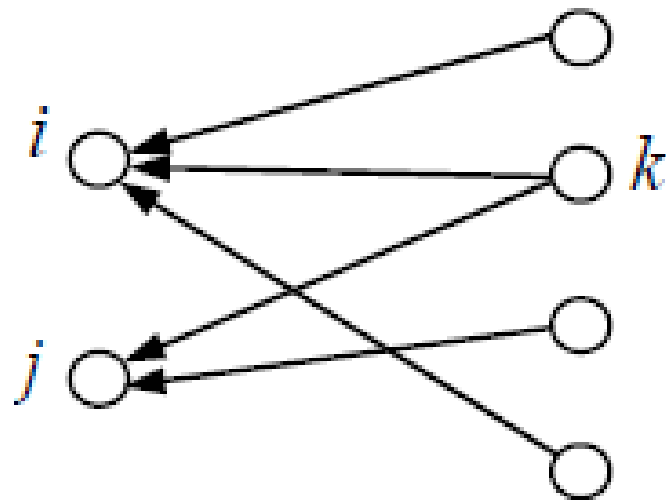
Paralelní citace

- Měření podobnosti dvou dokumentů.
- Pokud jsou články i a j oba citovány v článku k , můžeme říct, že v nějakém smyslu spolu souvisejí, i když navzájem se články i a j necitují.
- Pokud jsou články i a j společně citovány v mnoha člancích, znamená to, že mají silný vztah nebo podobnost.



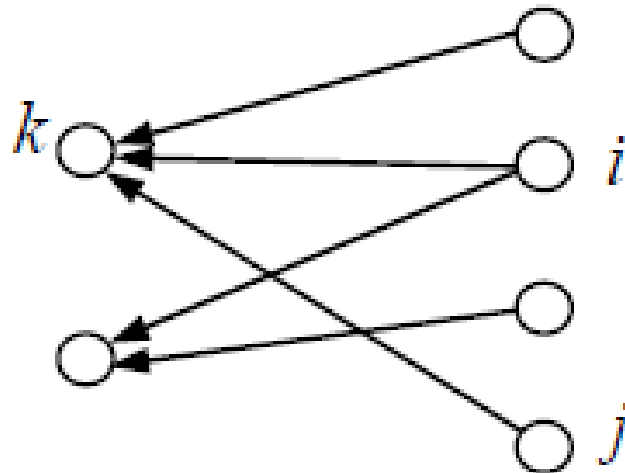
Paralelní citace

- Paralelní citace článků i a j článkem k



Bibliografické spojení

- Funguje zrcadlově k paralelní citaci.
- Bibliografické spojení spojuje články, které citují stejné články.



PageRank

- Spoléhá na demokratickou povahu webu a využívá jeho obrovskou strukturu propojení k určení kvality konkrétní stránky.
- Interpretuje hyperlink ze stránky x na stránku y jako hlas od stránky x pro stránku y .
- Dívá se i na počty hlasů jednotlivých stránek nebo spojení, které stránka přijímá.
- Analyzuje stránku, která dává hlas druhé stránce



HITS

- Hypertext Induced Topic Search
- PageRank je statický algoritmus.
- HITS je závislý na vyhledávaném dotazu.
- Po zadání dotazu vytvoří seznam relevantních stránek od vyhledávače, pak vytvoří dvě řazení sady stránek – autoritativní a středové

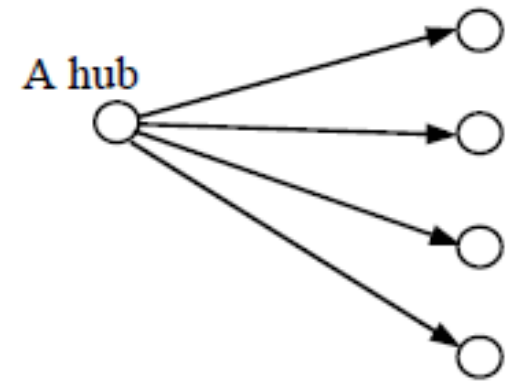
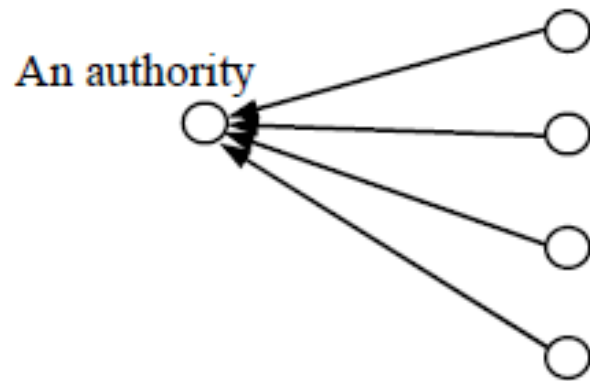


HITS

- Autorita (authority) je stránka s mnoha spojeními směrem k ní.
- Stránka může mít dobrý obsah týkající se nějakého tématu a tedy jí mnoho lidí věří a odkazuje se na ni.
- Střed (hub) je stránka s mnoha spojeními směrem od ní.
- Na středu najdeme mnoho užitečných odkazů, které se týkají daného tématu.



Authority a Hub



Závěr

- V kapitole jste se seznámili s analýzou linků.
- Byla vysvětlena analýza sociální sítě.
- Seznámili jste se s paralelní citací a bibliografickým spojením.
- Byly vysvětleny i algoritmy PageRank a HITS.

