

Kam dál?

Grafové algoritmy



ISKM89 Organizace dat - sémantický web | podzim 2023
Zuzana Nevěřilová | Centrum zpracování přirozeného jazyka

Znalostní grafy jsou grafy

formálně: Graf $G = (V, E)$, $V(G)$ - **uzly** (vertex, vertices), $E(G)$ - **hrany** (edge)

$E \subseteq V \times V$, což znamená, že $e_i = (v_k, v_l)$

orientovaný, s pojmenovanými hranami

Stupeň (degree) uzlu: počet hran, se kterými je uzel spojen (in-degree, out-degree)

Orientovaná **cesta** (directed path): sekvence vrcholů a hran $(v_0, e_1, v_1, \dots, e_t, v_t)$, kde platí $e_i = \{v_{i-1}, v_i\} \in E(G)$

Souvislý graf - pokud pro každé dva vrcholy existuje cesta

<https://teorie-grafu.cz/>

Znalostní grafy jsou grafy

Grafové algoritmy:

- zjistit, zda je graf spojitý
- najít cestu
- najít nejkratší cestu (co mají společné X a Y?)
- najít uzel s největším stupněm

Síťová analýza

- topologie (modul, skupina, komunita, klastr)
- sociogramy (30. léta 20. stol.), dynamika skupiny

Cílem je najít nejdůležitější uzly (centralita)

- počet uzlů, počet hran, max. stupeň - absolutní veličiny
- mezilehlost (betweenness) - propojení jinak málo propojených částí
- blízkost (closeness) - cesty k ostatním uzlům
- eigencentralita - důležitost sousedních uzlů
- page rank - jak “hodnotný” je uzel v závislosti na svých sousedech
- hustota - kolik uzlů je propojeno (sociální sítě jsou poměrně řídké)
- excentricita - maximální vzdálenost od ostatních uzlů; maximální průměr

Síťová analýza

Graph Centrality measurements

Node Degree

quantify node connectivity

local measure !

Eigen centrality

take in consideration neighbors connectivity (global)

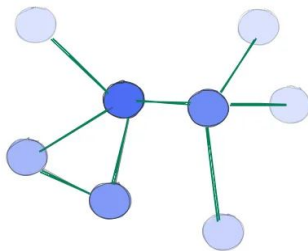
can be fooled by fake hub connectivity

PageRank

can dump or scale hub effect

closeness

easy access to all nodes
(shortest distance from graph nodes)



betweenness

quantify node importance
in information flow

shortest path assumption (not always the case !)

Síťová analýza - software

Gephi, Gephi Lite

Python + NetworkX