

Jan Fousek, Eva Hladká

Fakulta informatiky, Masarykova univerzita

15. března 2018

Důležitost uzlu

Otázky typu

- které osoby jsou klíčové pro šíření nákazy?
- jak cílit útoky proti síti?
- jak zlepšit šíření informací v síti?
- které webové stránky jsou hodnotnější než jiné?
- které osoby mají největší vliv na formování skupinového názoru?
- ...

Centralita jako důležitost uzlu

Důležitost uzlu závisí na

- jeho vlastnostech
- *poloze v síti*

Volba správné metriky závisí na

- původní otázce
- sémantice konkrétní sítě

Stupeň uzlu jako centralita

Uzly s vysokým stupněm jsou

- vysoce propojené se zbytkem sítě
- mají přímý vliv na velké množství uzlů (sousedé)

V orientovaném grafu

- rozlišujeme vstupní a výstupní stupeň
- velmi podstatný rozdíl v interpretaci

Stupeň nic nevyovídá o důležitosti susedů.

Stupeň uzlu: příklad

Síť světového obchodu

- orientovaná síť
- stupeň je počet obchodních partnerů
 - indegree: import
 - outdegree: export

Vývoj nejdůležitějších uzlů podle stupně uzlu odráží proměny struktury světového obchodu.

- vyšší celková propojenost (menší rozdíly)
- změny ve složení nejcentrálnější skupiny

Sít' světového obchodu¹

indegree			outdegree		
1960					
1	0.6438	UK	1	0.5987	USA
2	0.5954	Netherlands	2	0.5861	UK
3	0.5866	France	3	0.5740	France
2000					
1	0.8920	USA	1	0.8636	USA
1	0.8920	Germany	1	0.8636	UK
3	0.8808	UK	1	0.8636	France

¹De Benedictis, L., & Tajoli, L. (2011)

Stupeň uzlu: příklad²

Proteinová síť bakterie *Helicobacter pylori*

- neorientovaná síť, hrana reprezentuje známou fyzickou interakci (katalýza, signalizace, ...)
- známé efekty vyřazení proteinu

Robustnost sítě

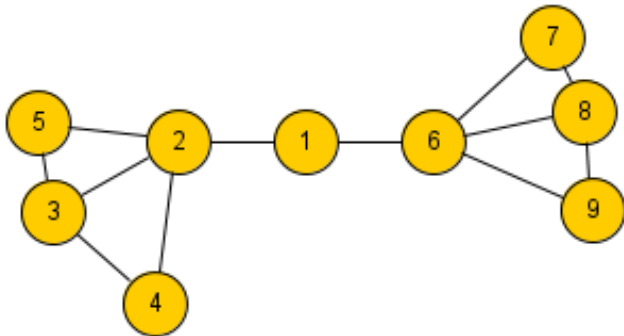
- relativně vysoká tolerance vůči náhodným mutacím
- odstranění proteinů s vysokým stupněm fatální
- korelace závažnosti následků se stupněm uzlu

$$r = 0.75$$

²Jeong, Hawoong, et al. (2001)

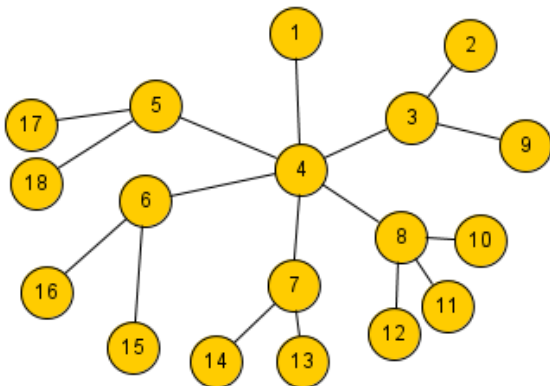
Nejkratší cesty a centralita

I uzly s nízkým stupněm mohou být významné



Nejkratší cesty a centralita

I uzly s nízkým stupněm mohou být významné



Closeness centralita

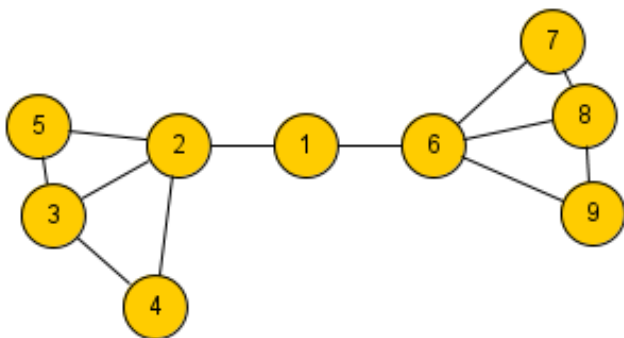
„Být v centru dění“

- nepřímo úměrná průměrné nejkratší cestě do ostatních uzlů
- výhodná pozice pro šíření informace ve smyslu ovlivňování ostatních uzlů

Definice

- $C_c(i) = \left[\sum_{j=1}^N d(i, j) \right]^{-1}$
- normalizovaná $C'_c(i) = \frac{C_c(i)}{N-1}$

Mezilehlost



Betweenness centrality

Zachycuje zprostředkování

- uzly spojující klastry
- výhodná pozice pro kontrolu šíření informace

Definice

- $C_b(i) = \sum_{j < k} \frac{g_{jk}(i)}{g_{jk}}$
- g_{jk} je počet nejkratších cest mezi j a k
- $g_{jk}(i)$ je počet nejkratších cest mezi j a k , na kterých leží i

Betweenness příklad ³

Spoluautorská síť (library and information science)

- uzly autoři, hrana značí společně napsaný článek
- analýza impaktu: počtu citací všech prací

Betweenness koreluje s impaktem

- stupeň značí množství spoluautorů
- betweenness odpovídá interdisciplinárním projektům

³Yan, E., & Ding, Y. (2009)

Betweenness příklad⁴

Síť transferu pacientů mezi nemocnicemi

- uzly nemocnice USA, hrany přesuny mezi JIP
- scénář šíření rezistentní infekce

Problém alokace omezených prostředků pro karanténu

- náhodné, podle stupně, podle betweenness, iterativně podle kapacity vystavené nákaze
- betweenness nejlepší ze statických (preventivních) alokací

⁴Karkada, Umanka H., et al. (2011)

Centrality: rozdíly

nízká / vysoká	stupeň	blížkost	mezilehlost
stupeň		vprostřed klastru vzdáleného od zbytku sítě	hrany uzlu jsou pro síť redundantní
blížkost	uzel v bezprostřední blízkosti důležitého uzlu		alternativní nejkratší cesty, množství uzlů je si vzájemně blízké
mezilehlost	most mezi klastry, udržuje významné vazby	spojuje vzdálenou komunitu se zbytkem sítě	

Eigenvector centralita

Důležitost uzlu závisí na důležitosti sousedů

- uvažuje globální topologii sítě
- rekurentní definice
- více variant např. PageRank

Co je to eigenvector (vlastní vektor)

- $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$
- A je matice, u je vektor, λ je číslo
- jak to souvisí s centralitou?

Eigenvector centralita: odvození

Vydeme z

- $C_{eig}(i) \propto \sum_{i \neq j} A_{ij} C_{eig}(j)$
 - jako výchozí hodnotu C_{eig}^0 použijeme např. stupeň
- Iterace pro $x_i = C_{eig}(i)$
- $x_i(t+1) = \sum_{j \neq i} A_{ij} x_j(t)$
 - což je v podstatě násobení vektoru maticí
 - $\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t)$ a tedy $\mathbf{x}(t) = \mathbf{A}^t \mathbf{x}(0)$
 - tím dostáváme mocninou metodu jejímž řešením je dominantní vlastní vektor

Eigenvector centralita: varianty

PageRank

- založený na náhodných procházkách v síti
- vhodný i na orientované grafy (teleportace)
- (C_{eig} selhává na uzlech mimo silně souvislé komponenty)
- A_{ij} modifikována: reprezentuje pravděpodobnost přechodu mezi uzly (suma přes sloupce rovna 1)

Eigenvector příklad⁵

retweet network během prezidentských debat

- uzly účty, hrany @ user zmínky a # témata
- jak identifikovat významné uzly a jaká je struktura komunikace?

Důležité uzly

- stupeň nestačí: protěžuje zpravodajské entity
- EC správně označí debatující
- této znalosti lze využít při analýze sítí, u kterých neznáme předem odpověď

⁵Shamma et al. (2009)

PageRank příklad⁶

Citační síť

- časopisy Physical Review
- uzly články, hrany citace

Význam článku

- běžně podle počtu citací (stupeň)
- podhodnocuje klíčové práce, které umožnily přelomová díla, PageRank ne
- PageRank a stupeň uzlu pozitivně korelují
- outliers: zapadlé poklady

⁶Chen, Peng, et al. (2007)