IV124 Komplexní sítě

Eva Výtvarová, Jan Fousek, Eva Hladká

Fakulta informatiky, Masarykova univerzita

6. března 2019

Dnešní plán

Začneme s kvantifikací síťové struktury:

- stupně uzlů v náhodném grafu
- délky cest a klastrovací koeficient
- co znamenají konkrétní hodnoty pro danou síť

V návaznosti

- jednoduchý nulový model náhodné sítě

Stupeň uzlu

Stupeň uzlu

- počet hran spojující uzel s ostatními
- počet nenulových prvků v řádku (sloupci) matice sousednosti
- u orientovaných grafů rozlišujeme vstupní/výstupní

Interpretace

- počet přátel
- počet chemických reakcí
- ...

Průměrný stupeň uzlu

Pro neorientovaný graf

- $\overline{k} = \frac{2|E|}{|V|}$
- každá hrana „přispívá“ dvěma uzlům

Distribuce stupně $P(k)$

- pravděpodobnost, že náhodný uzel má stupeň k
- zavádíme, protože v reálných sítích si s průměrem nevystačíme (ukážeme si později)

Cesty a vzdálenosti

Cesta v grafu

- posloupnost hran spojující dva uzly
- délka cesty: počet hran
- u ohodnocených hran záleží na sémantice

Vzdálenost uzlů *d*

- délka nejkratší cesty (také geodetická vzdálenost)
- nejkratších cest mezi uzly může být více
- $d = \infty$ pokud uzly nejsou spojené

Výpočet vzdálenosti

Neohodnocený graf

- prohledávání do šířky

Ohodnocený graf

- Dijkstrův alg.

Průměrná délka cesty

- all-to-all
- Floyd-Warshallův alg.

Cesty a vzdálenosti

Ve souvislých komponentách grafu sledujeme:

- diametr sítě
  - maximální vzdálenost mezi libovolným párem uzlů
- průměrnou délku cesty $\overline{d}$

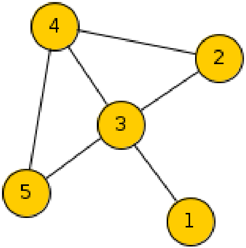
Co vzdálenosti vypovídají o síti

- efektivita šíření např. informace
- okruhy důvěry v sociálních sítích
- ...

Klastrovací koeficient

Klastrovací koeficient C_i uzlu i:

- jak jsou mezi sebou propojeni sousedé uzlu i?
- $C_i = \frac{L_i}{k_i(k_i - 1)}$
- kde L_i jsou hrany mezi sousedy uzlu i



- $C_3 = 1/6$

Klastrovací koeficient

Průměrný klastrovací koeficient $\overline{C}$

- $\overline{C} = \frac{1}{N} \sum_{i=1}^{N} C_i$
- lze též číst jako pravděpodobnost, že dva sousedé náhodného uzlu jsou propojeni

Co vypovídá o síti

- *lokální* tranzitivita: přátelé mých přátel jsou moji přátelé
- pravidelnost ve struktuře sítě: trojúhelníky

Příklady reálných sítí

síť	$\lvert V\rvert$	$\lvert E\rvert$	$\overline{k}$	$\overline{d}$
WWW	192K	609K	6.34	6.98
E. Coli metab.	1K	5.8K	5.84	2.98
citace	450K	4.7M	10.47	11.21
kvasinky proteiny	2K	2.9K	5.84	2.98
distrib. síť	4.9K	6.5K	2.67	18.99

Jaký mají ale konkrétní hodnoty význam?

Model náhodného grafu

Proč modelovat náhodný graf:

- vlastnosti lze matematicky odvodit
- užitečný pro srovnání s reálnou sítí:
  - čím se liší?
  - co nám to o síti říká?

Model náhodného grafu Erdös-Rényi:

- $G(N, p)$, kde N je počet uzlů a p pravděpodobnost spojení dvou uzlů

Erdös-Rényi model: vlastnosti

Počet hran $|E|$

- binomiální distribuce
  $P(|E|) = \binom{E_{max}}{|E|} p^{|E|} (1-p)^{E_{max}-|E|}$
- kde $E_{max} = N(N-1)/2$ je maximální počet hran

Stupeň uzlu

- binomiální distribuce:
  $P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$
  - $\binom{N-1}{k}$ výběr k
  - p^k: pravděpodobnost vzniku k hran
  - $(1-p)^{N-1-k}$: nepřítomnost zbylých hran
  - $\overline{k} = p(N-1)$

Erdös-Rényi model: vlastnosti

Klastrovací koeficient

- $C_i = \frac{L_i}{k_i(k_i-1)}$
- za L_i dosadíme $p\frac{k_i(k_i-1)}{2}$ – pravděpodobnost hrany mezi sousedy
- tedy $C_i = \frac{pk_i(k_i-1)}{k_i(k_i-1)} = p$

Pro reálné řídké sítě je tedy $\overline{C}$ velmi malé.

Erdös-Rényi model: vlastnosti

Odvození

- uvažujme síť s daným $\overline{k}$
- uzel má v průměru $\overline{k}^d$ sousedů ve vzdálenosti d
- tedy počet uzlů ve vzdálenosti d je $N(d) = \frac{\overline{k}^{d+1}-1}{\overline{k}-1}$
- ale $N(d) \leq N$ a tedy $\overline{k}^{d_{max}} \approx N$ a $d_{max} = \frac{log(N)}{log(\overline{k})}$
- pro většinu sítí zároveň dobrá aproximace pro průměrnou délku cesty $\overline{d} \approx \frac{ln(N)}{ln(\overline{k})}$