

Analýza sítě odpovědí na stránce  
*MathOverflow*

Moravec Vojtěch

Zimní semestr 2018

## 1 Popis datasetu

Analyzovaný dataset byl získán ze stránky *snap.stanford.edu* [1]. Dataset obsahuje seznam hran orientovaného grafu, který reprezentuje interakci mezi uživateli stránky *https://mathoverflow.net/*. Hrana  $(u,v)$  vycházející z vrcholu  $v$  do vrcholu  $u$ , znamená, že uživatel  $u$  odpověděl na otázku uživatele  $v$ . Originální dataset obsahuje také časové razítka vzniku hrany, ale pro naše účely je budeme ignorovat a budeme pracovat se všemi hranami.

## 2 Analýza sítě

Sít' obsahuje 21688 vrcholů a 107581 hran, tedy mezi 21668 uživateli bylo zodpovězeno 107581 otázek.

	Minimum	Výskyt	Maximum	Výskyt	Průměr
Vstupní stupeň	0	5683×	1102	1×	4,9604
Výstupní stupeň	0	11309×	1415	1×	4,9603
Celkový stupeň	1	9445×	1815	1×	9,9207

Tabulka 1: Tabulka stupňů vrcholů

Z Tabulky 1 můžeme vyčíst, tyto extrémy:

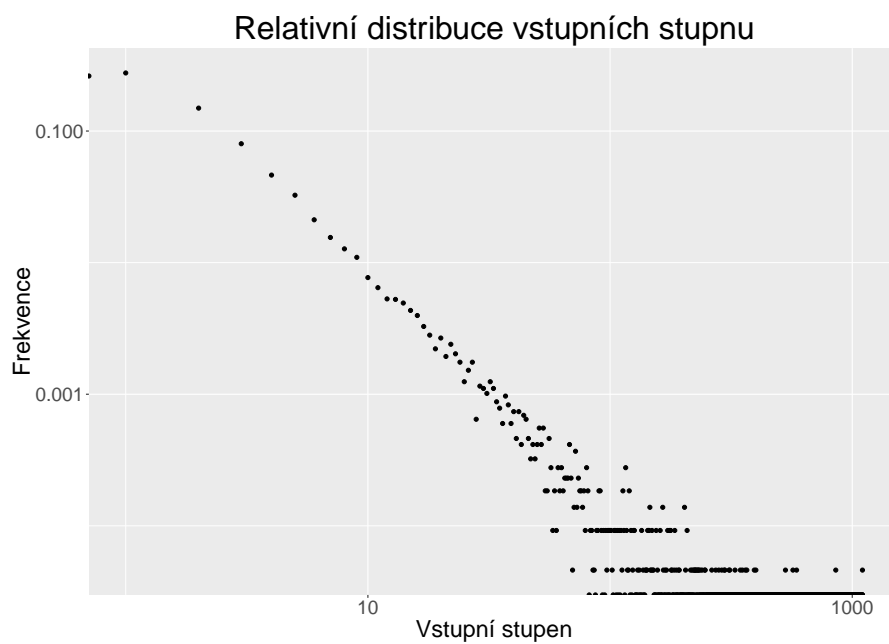
- 5683 uživatelům nepřišla žádná odpověď.
- Jednomu uživateli přišlo celkem 1102 odpovědí na jeho otázky.
- 11309 uživatelů neodpovědělo na žádnou otázku.
- Nejvíce odpovědí napsal jeden uživatel a to 1415 odpovědí.

Průměrný uživatel napsal skoro 5 odpovědí na otázky ostatních a také průměrně získal 5 odpovědí od ostatních uživatelů.

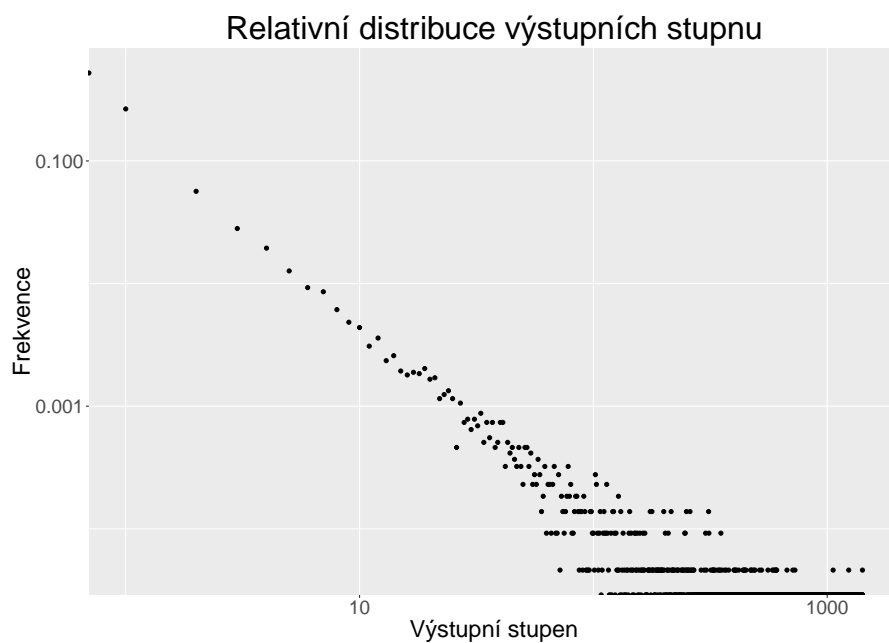
Dále na Obrázcích 1 a 2 můžeme v grafech vidět relativní distribuci stupňů. Tato distribuce odpovídá mocninnému dělení, jak je tomu u obecně u reálných sítí.

## Reference

- [1] J. Leskovec and A. Krevl, “SNAP Datasets: Stanford large network dataset collection.” <http://snap.stanford.edu/data>, June 2014.



Obrázek 1: Graf relativní distribuce vstupních stupňů



Obrázek 2: Graf relativní distribuce výstupních stupňů