

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 neohodnoceného grafu, který reprezentuje interakci mezi uživateli stránky *https://mathoverflow.net/*. Hrana (u, v) vycházející z vrcholu u do vrcholu v , 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

Síť obsahuje 21688 vrcholů a 107581 hran, tedy 21668 uživatelů a 107581 odpovědí na otázky.

2.0.1 Analýza stupňů vrcholů

| | 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 |

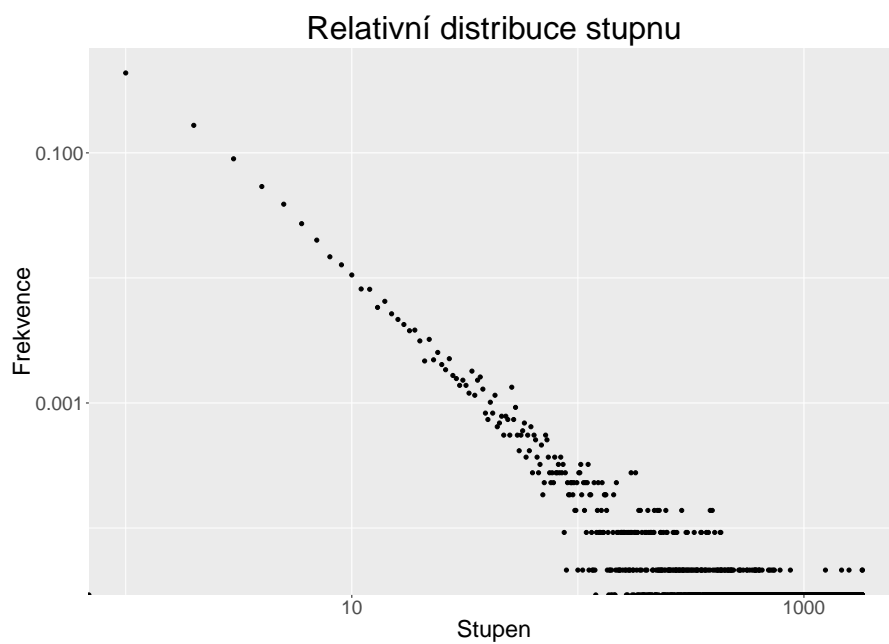
Tabulka 1: Tabulka stupňů vrcholů

Z Tabulky 1 můžeme vyčíst, tyto extrémy:

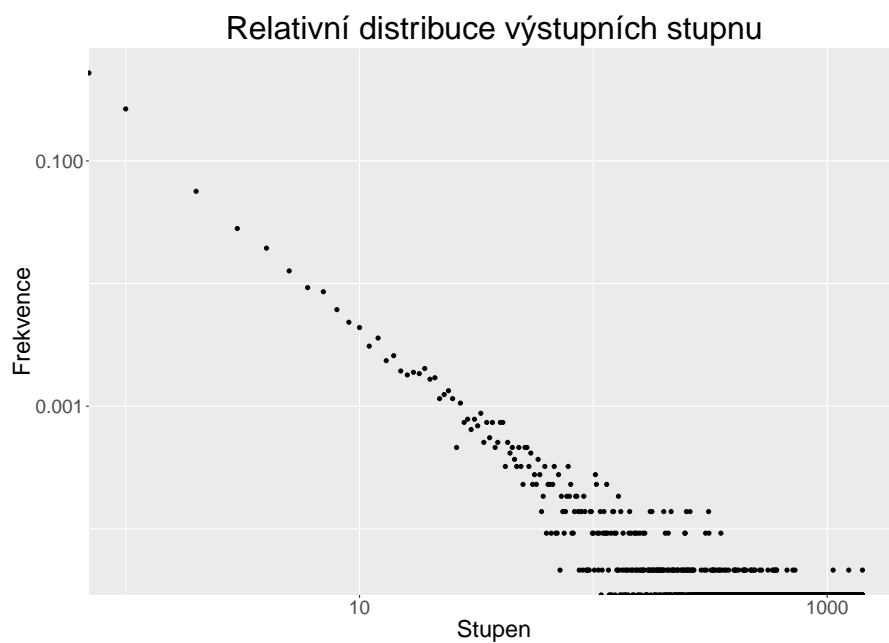
- 5683 uživatelům nepřišla žádná odpověď na jejich otázky.
- Jednomu uživateli přišlo celkem 1102 odpovědí na jeho otázky.
- 11309 uživatelů neposkytlo žádnou odpověď.
- Nejvíce odpovědí napsal jeden uživatel a to 1415 odpovědí.

Průměrný uživatel napsal skoro 5 odpovědí na otázky ostatních a také průměrně získal 5 odpovědí od ostatních uživatelů. Větší výstupní stupeň v této síti naznačuje, že daný uživatel je vlivným vzhledem k této komunitě.

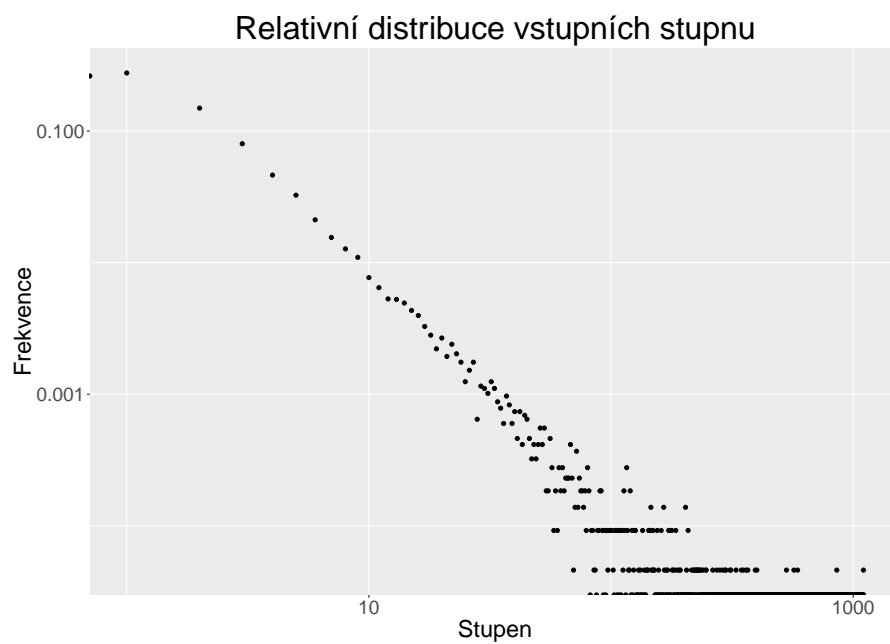
Dále na Obrázcích 1, 2 a 3 můžeme v grafech vidět relativní distribuci stupňů. Tato distribuce odpovídá mocninnému dělení, jak je tomu u obecně reálných sítí. Na Obrázcích 4, 5 a 6 můžeme vidět kumulativní distribuci stupňů.



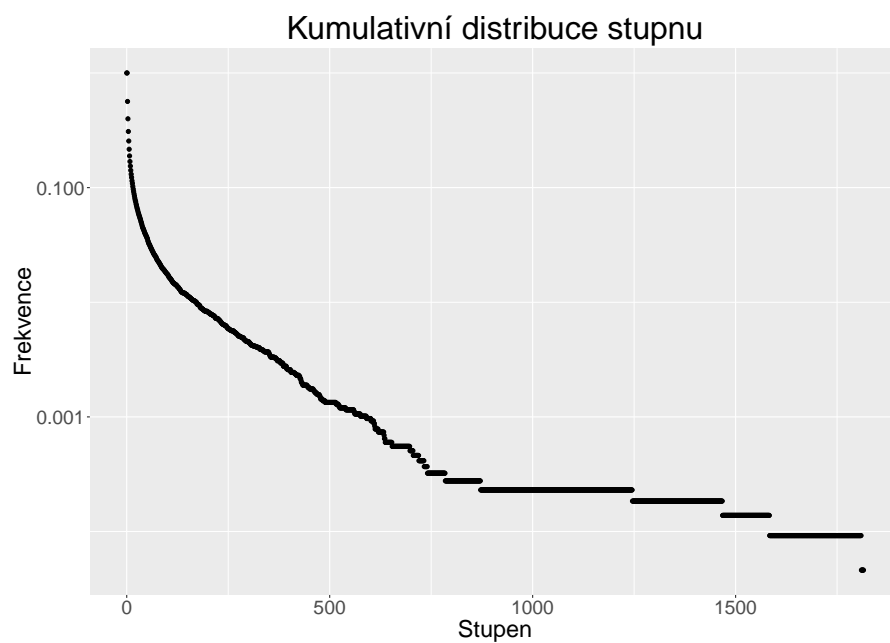
Obrázek 1: Graf relativní distribuce stupňů



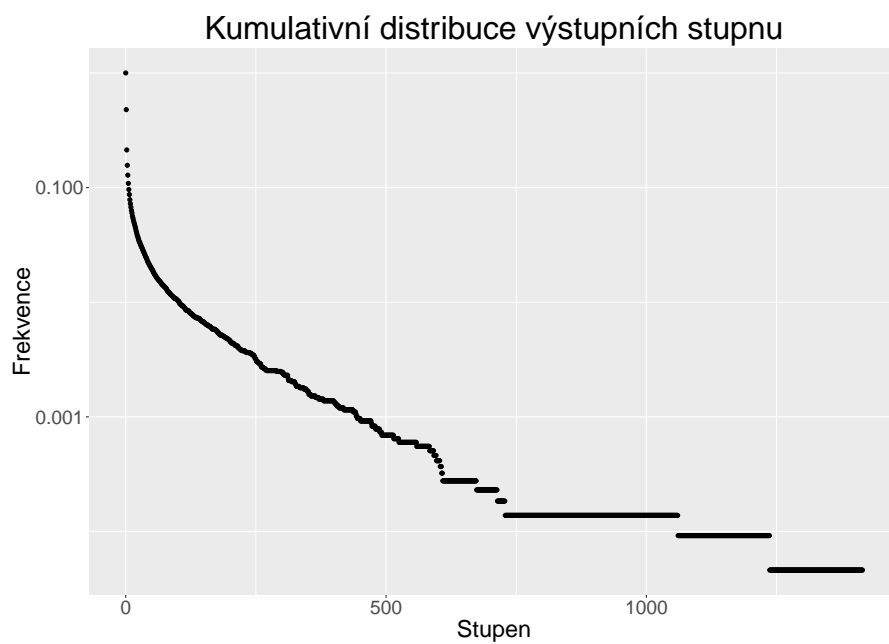
Obrázek 2: Graf relativní distribuce výstupních stupňů



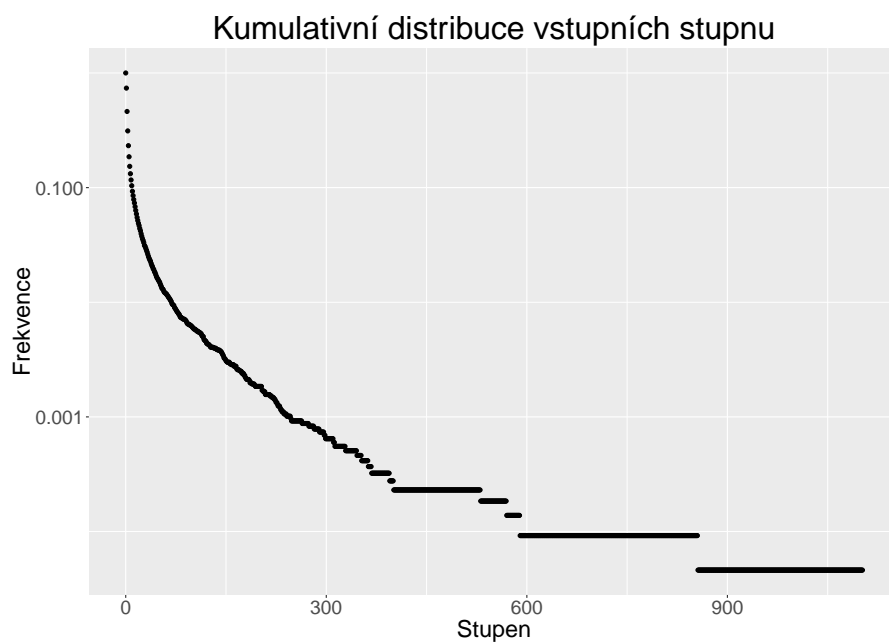
Obrázek 3: Graf relativní distribuce vstupních stupňů



Obrázek 4: Graf kumulativní distribuce stupňů



Obrázek 5: Graf kumulativní distribuce výstupních stupňů

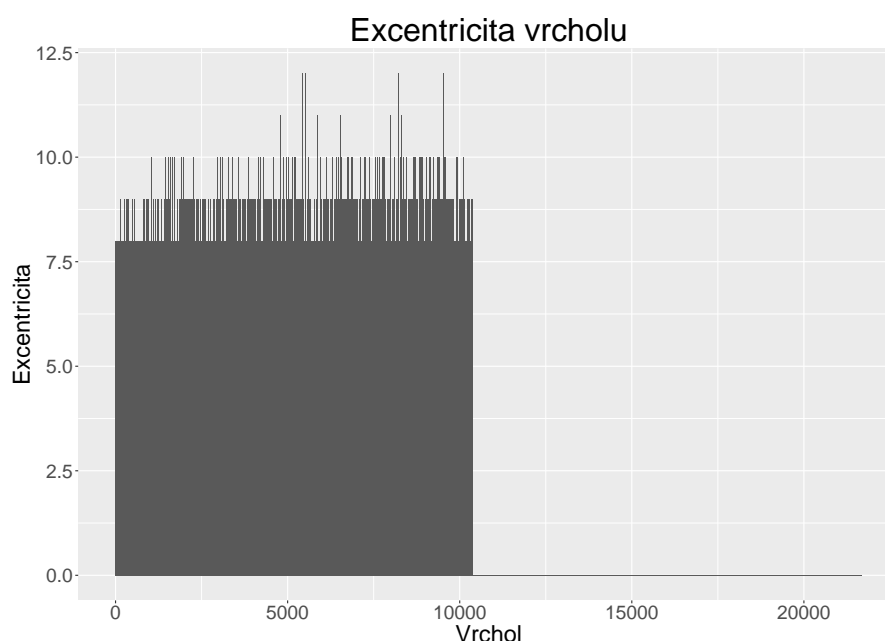


Obrázek 6: Graf kumulativní distribuce vstupních stupňů

2.0.2 Analýza cest v síti

Průměr sítě neboli, nejdelší z nejkratších cest mezi vrcholy, je v naší síti roven 17. Cesta v našem grafu značí posloupnost uživatelů, kteří si odpověděli na otázku. Průměrná délka nejkratší cesty je 4,18. Uživatelé jsou tedy průměrně spojeni přes 4,18 dalších uživatelů.

V grafu na Obrázku 7 můžeme vidět výstupní excentricitu jednotlivých vrcholů. Excentricita je vzdálenost nejkratší cesty k nejvzdálenějšímu vrcholu. 11947 vrcholů z celkového počtu 21688 vrcholů má nulovou excentricitu, což znamená, že jsou izolované, neexistuje z nich cesta do ostatních vrcholů. Jsou to tzv. pasivní uživatelé.

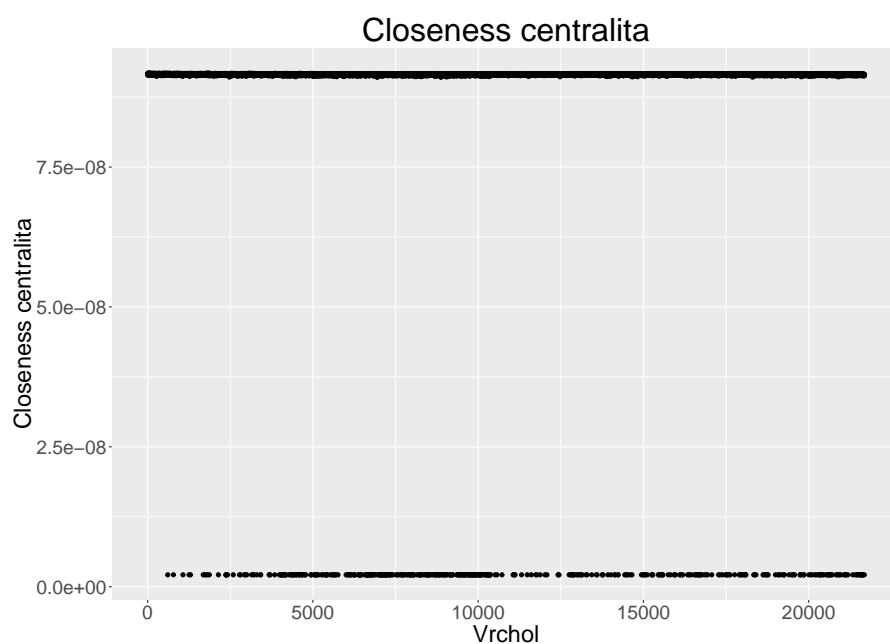


Obrázek 7: Excentricita vrcholů grafu

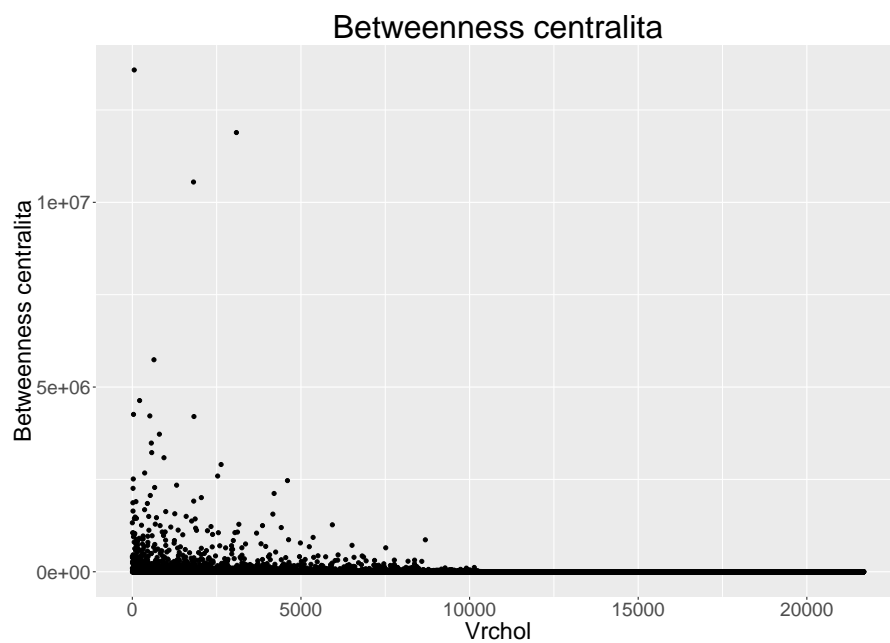
2.0.3 Centrality v síti

Jak už jsme pozorovali u excentricity, uživatelé se dělí na aktivní a pasivní. V grafu na Obrázku 8 můžeme vidět closeness centralitu. Vrcholy jsou zde rozděleny do dvou různých skupin. Na Obrázcích 9 a 10 vidíme betweenness centrality. Více toho vyčteme na grafu s logaritmickou osou y . Betweenness a Closeness centralita mezi sebou korelují, neboť obě pozorují 2 velmi rozdílné skupiny vrcholů.

Jako další dvě centrality zde máme Kleinbergovu hub centralitu 11, která přiřazuje vyšší hodnoty těm vrcholům, s více výstupními hranami, tedy

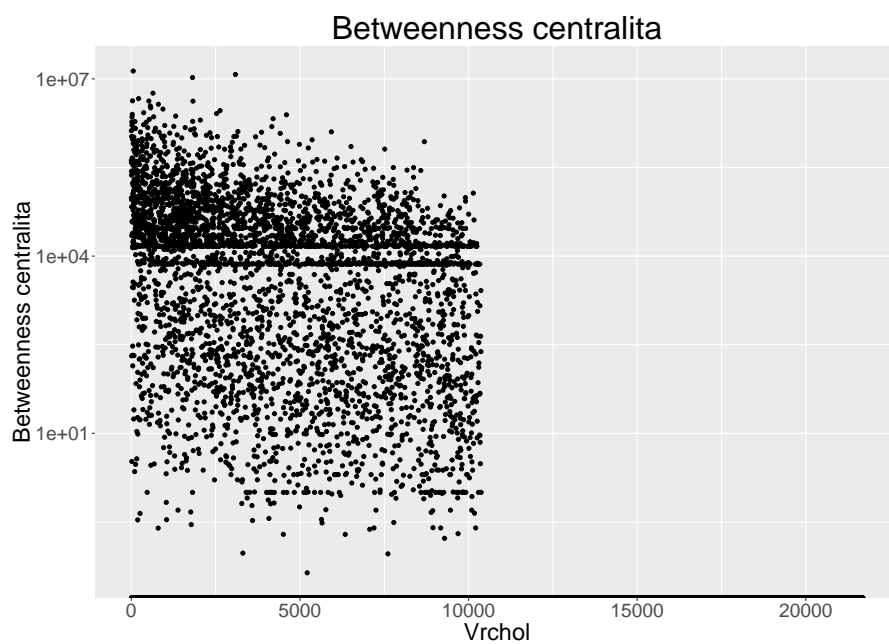


Obrázek 8: Closeness centralita



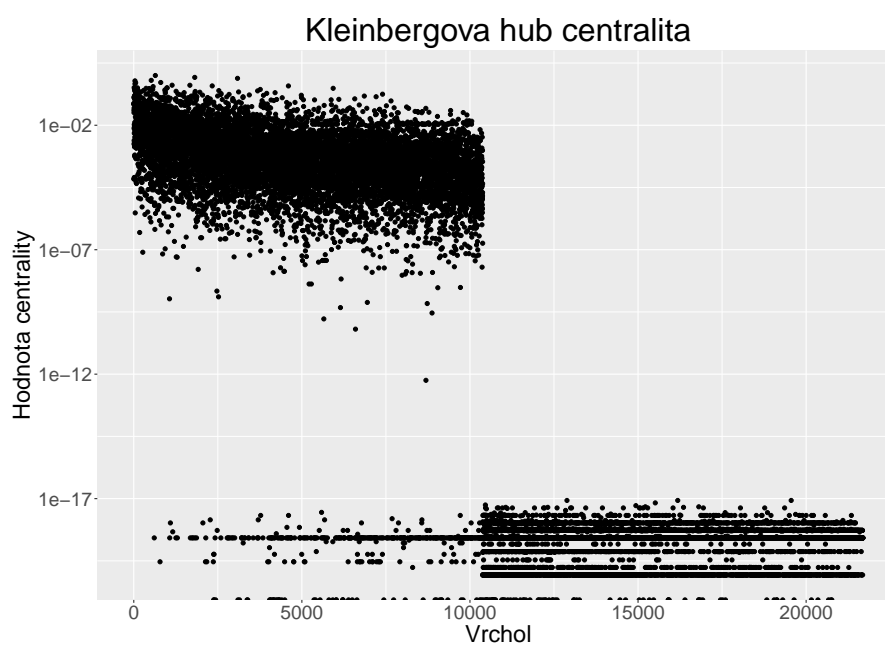
Obrázek 9: Betweenness centralita

těm uživatelům, kteří více odpovídají na otázky. Zase zde vidíme rozdělení na aktivnější uživatele v odpovídání a ty pasivnější. Druhou centralitou je

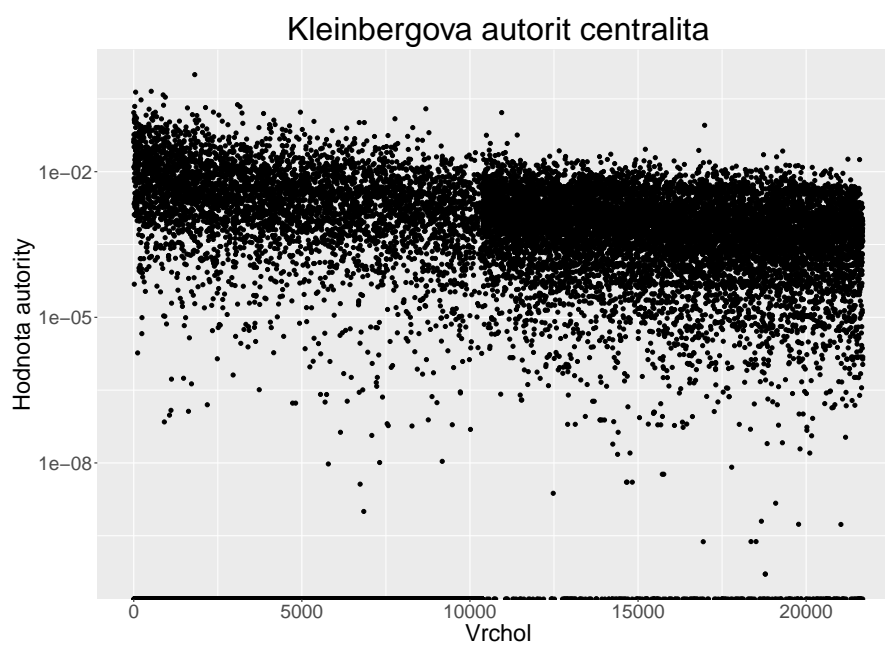


Obrázek 10: Betweenness centralita v logaritmickém měřítku

Kleinbergova autority centralita 12, ta přiřazuje vyšší hodnoty naopak těm s vrcholům s více vstupním hranami.



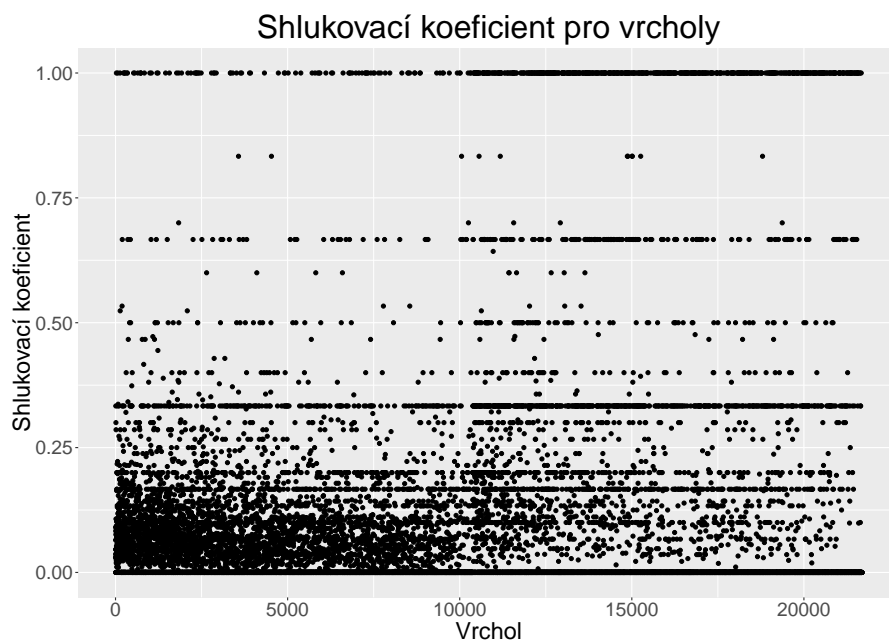
Obrázek 11: Kleinbergova hubs centralita



Obrázek 12: Kleinbergova autorit centralita

2.0.4 Shlukování

V naší síti je globální shlukovací roven 0,0501. Shlukovací koeficient jednotlivých vrcholů je už velmi různorodý jak můžeme vidět na Obrázku v grafu 13. Izolované vrcholy mají v tomto grafu hodnotu 0.



Obrázek 13: Shlukovací koeficient vrcholů

| | Hodnota | Výskyt |
|---------|---------|--------|
| Minimum | 0 | 14847× |
| Maximum | 1 | 767× |
| Průměr | 0,0821 | |

Tabulka 2: Shlukovací koeficient

3 Závěr

V námi analyzované síti jsme objevili dvě skupiny uživatel. Pasivní uživatele a aktivní uživatele. U aktivních je větší pravděpodobnost, že budou více odpovídat na otázky, u pasivních je to přesně naopak.

Reference

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