

# Analýza sítě Gnutella

Lukáš Moravec

11. prosince 2021

## 1 Popis datové sady

[1] Datová sada byla získána ze [2] Standfordské velké kolekce datasetů. Jedná se P2P<sup>1</sup> síť pro sdílení souborů. Samotná datová sada obsahuje 8717 uzlů a 31525 orientovaných hran. Každý uzel reprezentuje účastníka připojeného do sítě Gnutella a hrany reprezentují spojení mezi účastníky. Její kompletní znázornění lze vidět na obrázku 1.

## 2 Statistiky

K analýze sítě byl použit nástroj Gephi. Jelikož síť obsahuje velké množství uzlů, těžce se z ní na první pohled dá vyčíst něco zajímavého. Proto jsem si uzly vyfiltroval podle stupně uzlů  $i$  17. Toto lze vidět na obrázku 2, kde velikost uzlů zachycuje velikost jejich stupně a barevné obarvení příslušnost do určité komunity. Distribuce stupňů lze vidět na obrázcích 3, 4 a 5. Jak jde vidět, distribuce odpovídá spíše mocninnému rozdělení, kde je vysoký počet uzlů s nízkým stupněm a pár uzlů s vysokým stupněm.

Z tohoto už můžeme vyčíst informace. Uzly s vyšším stupněm budou pro tuto síť znamenat pravděpodobně důležitá úložiště, která obsahují buďto velké množství souborů nebo soubory, o které je velký zájem. Mohou to být převážně nějaké větší servery, zatímco uzly s nízkým stupněm mohou představovat účastníky, kteří síť využívají spíše pro své vlastní potřeby a stahují soubory.

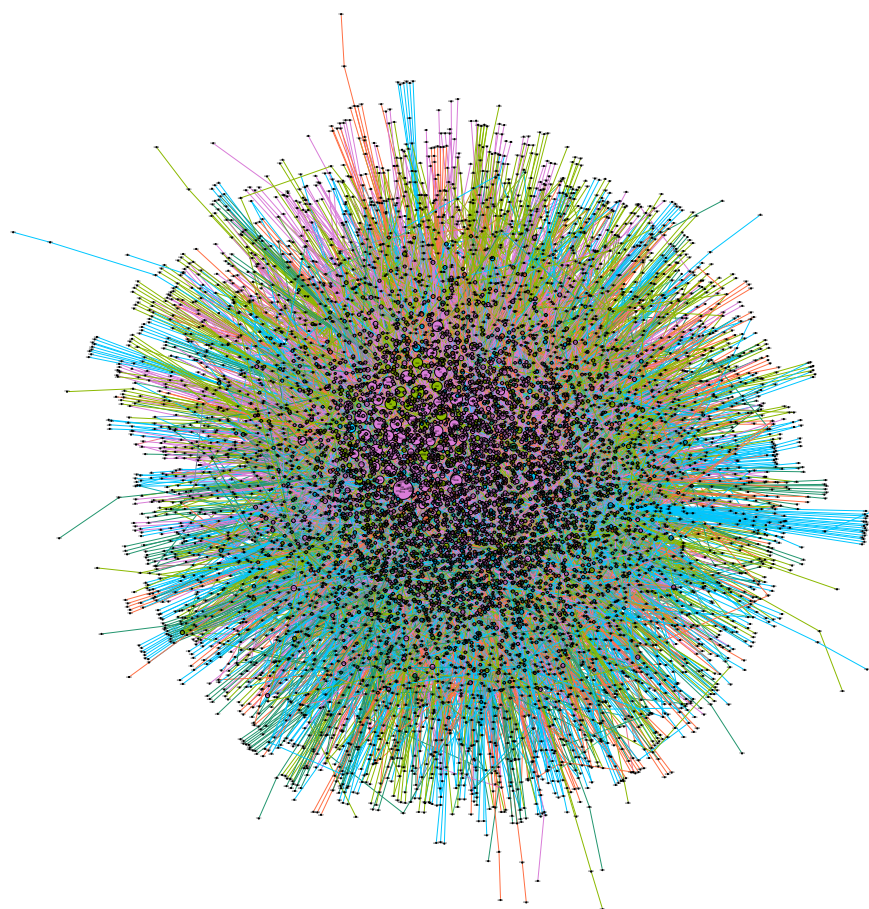
Průměr sítě, jestliže zanedbáme orientaci hran, čítá 10 uzlů. Průměrná délka cesty je však 4,572. To znamená, že pokud se účastník sítě chce podívat na nějaký sdílený soubor, jeho požadavek půjde v průměru přes 4 až 5 dalších zařízení v síti. Síť tvoří jedna velká komponenta. I když síť tvoří několik tisíc uzlů, jedná se o velice řídkou síť, hustota vychází na minimální 0,001.

Distribuci velikosti komunit zobrazuje obrázek 6. V této síti toto může znamenat účastníky v síti, kteří si navzájem sdílí soubory, případně mají mají vzájemné připojení a navzájem se preferují, jako zdroj dat. Tato lze lépe pozorovat na obrázku 7, kde teplota barvy reprezentuje velikost shlukovacího koeficientu. Průměrný shlukovací koeficient uzlů je taktéž velmi nízký (0,009), takže žádné velké shluky v síti nelze pozorovat, to je viditelné na obrázku 8. Absence shluků napovídá tomu, že se skutečně jedná o síť typu P2P a nejsou zde nějaké centrální prvky, na kterých síť stojí, ale vše je sdíleno.

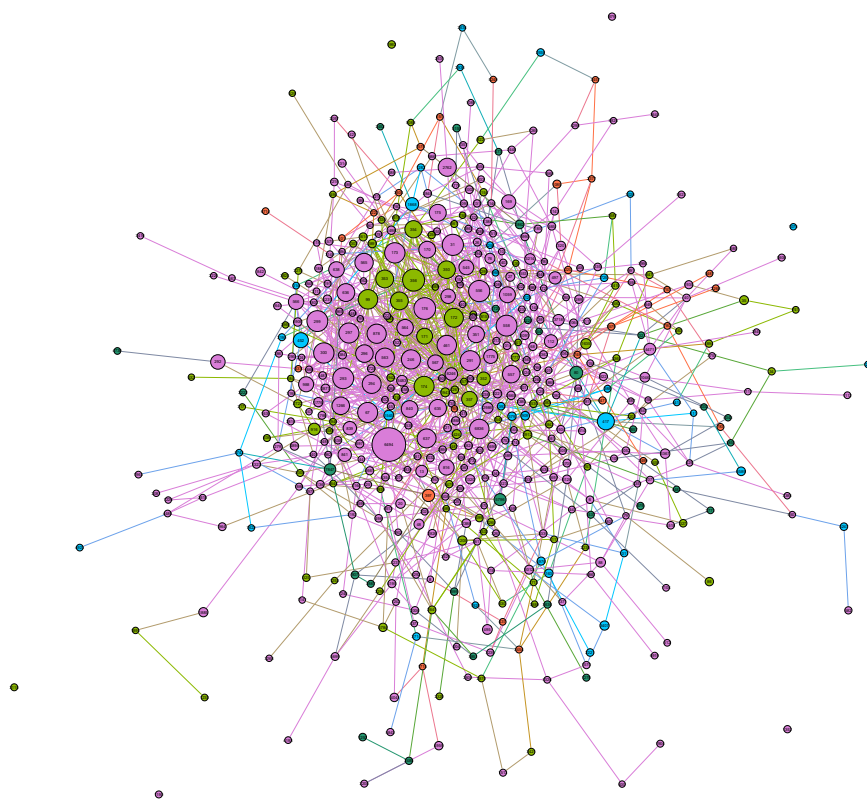
Centrality jsou viditelné na obrázku 9, kde velikost uzlů odpovídá betweenness centralitám a barva (červená = vysoká centralita, zelená = střední centralita, modrá = nízká) closeness centralitám. Pozorovat můžeme zajímavý uzel uprostřed sítě, který má jak vysokou closeness tak betweenness centralitu. Toto bude nějakým způsobem důležitý uzel v síti (např. datové centrum zapojené do této sítě).

---

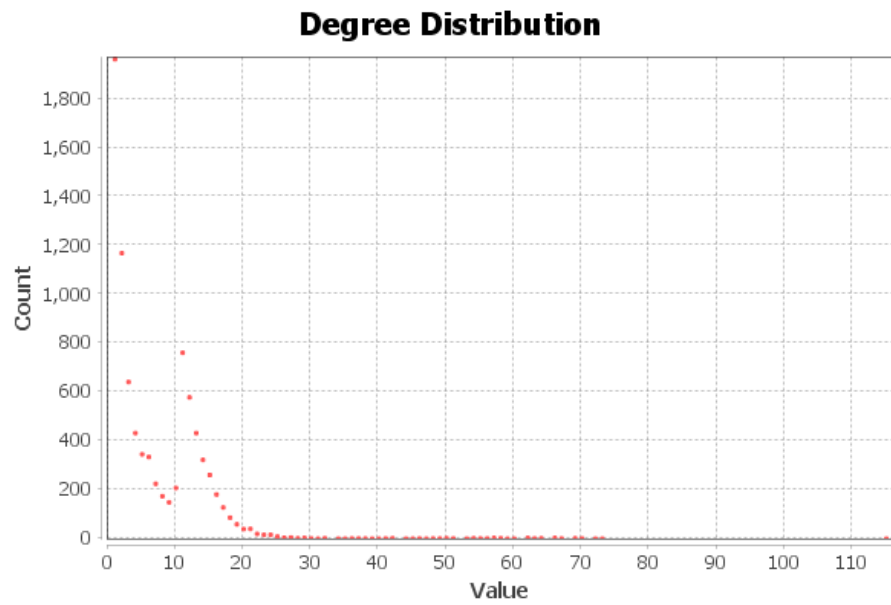
<sup>1</sup>Peer to peer



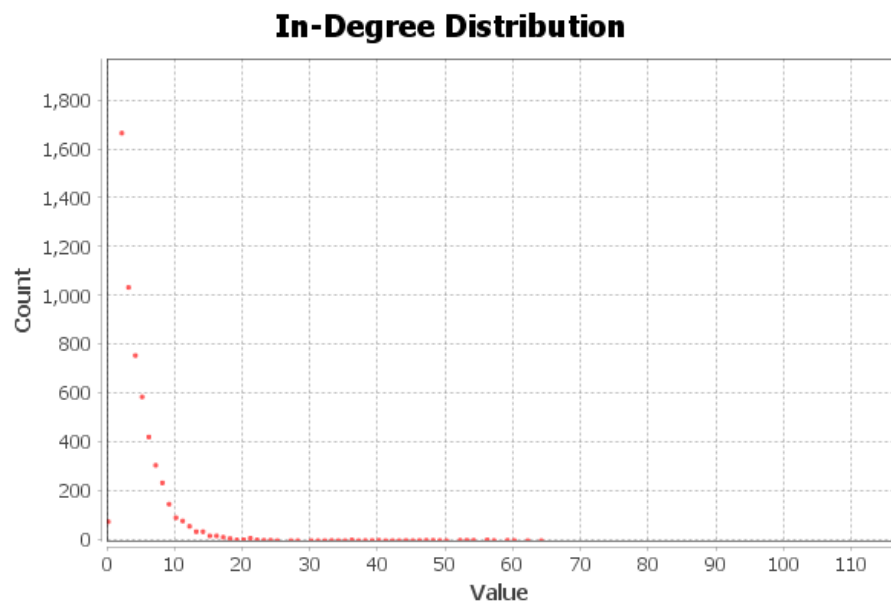
Obrázek 1: Znáznornění kompletní sítě



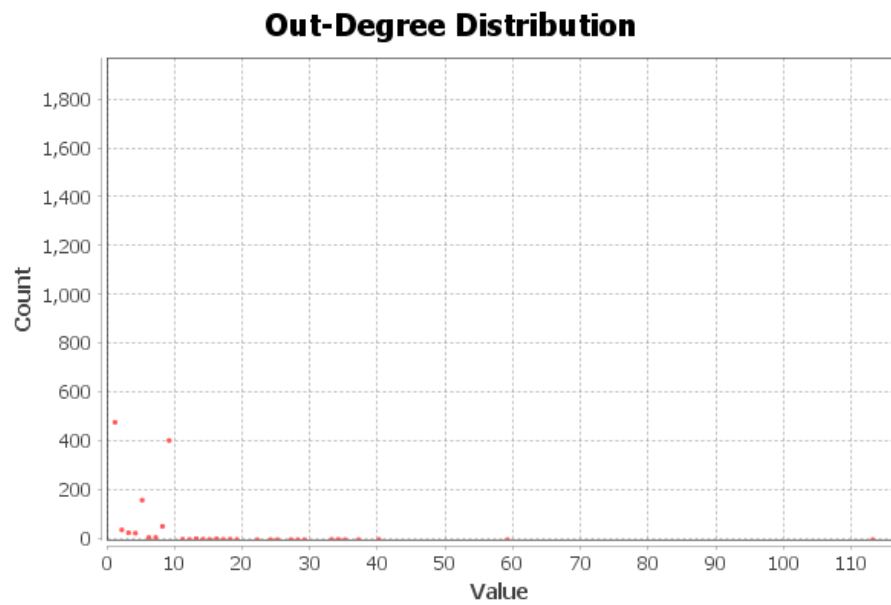
Obrázek 2: Síť s bez uzlů se stupněm nižším než 17



Obrázek 3: Distribuce uzlů podle stupně



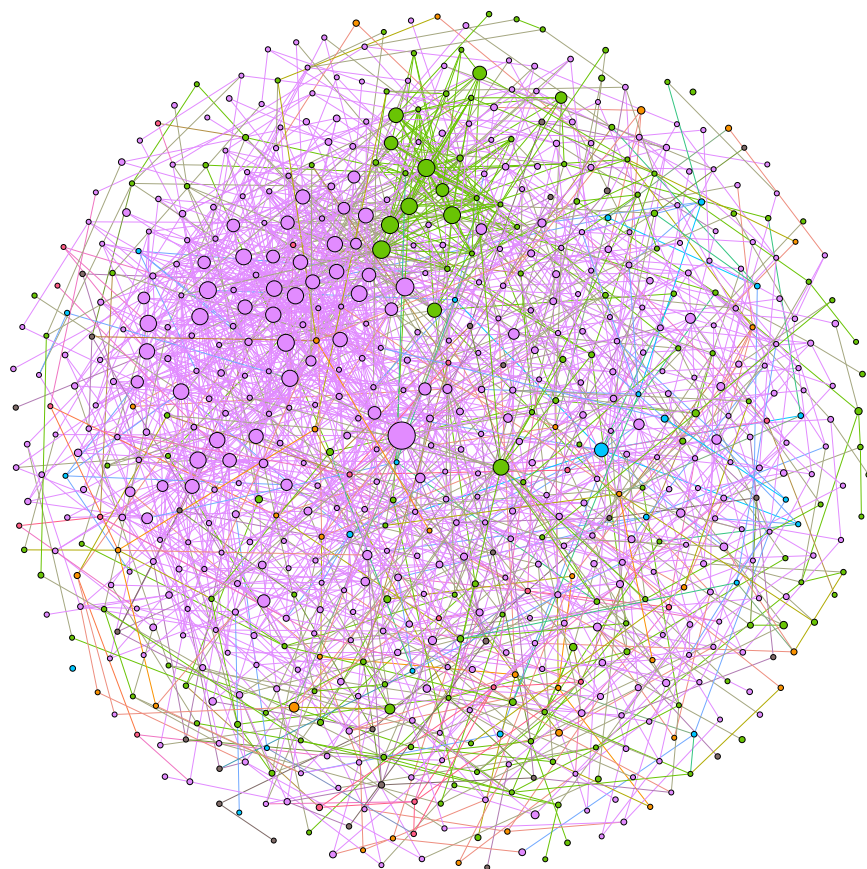
Obrázek 4: Distribuce uzlů podle stupně vstupních hran



Obrázek 5: Distribuce uzlů podle stupně odchozích hran

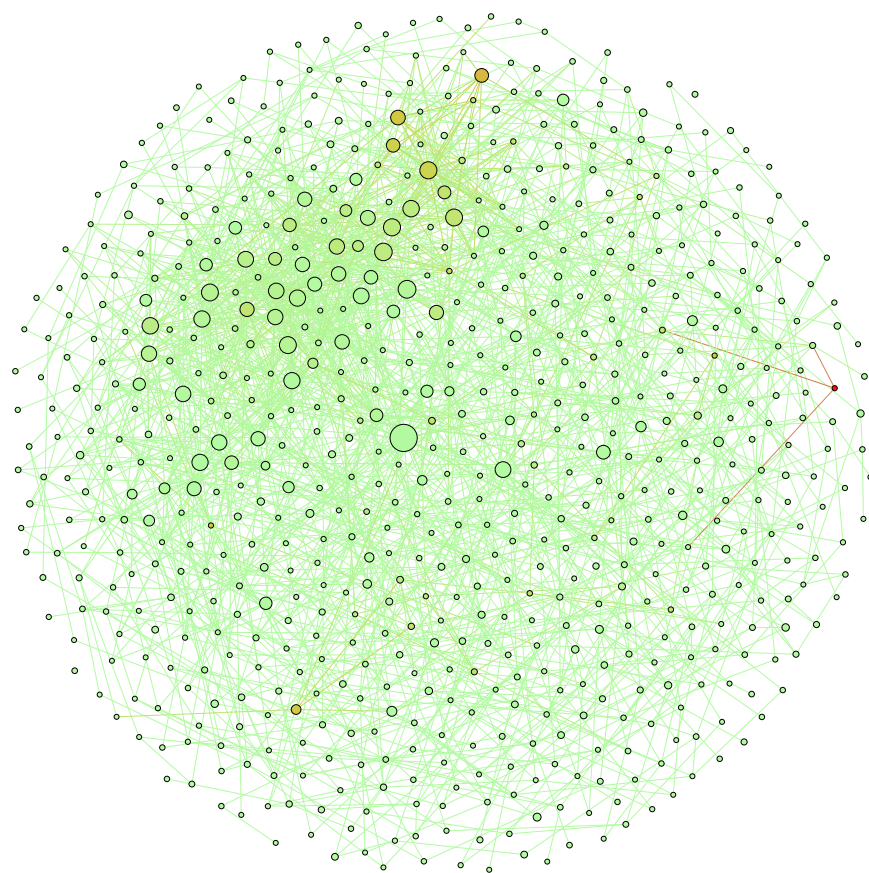


Obrázek 6: Distribuce velikosti komunit



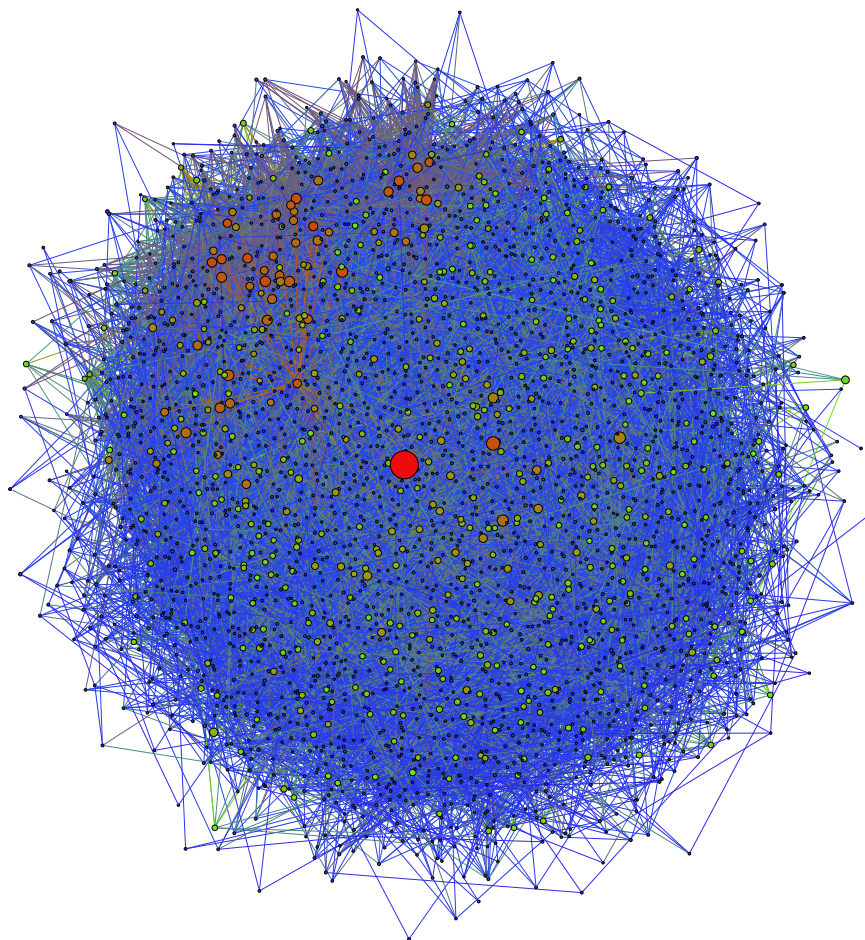
Obrázek 7: Komunity sítě Gnutella





Obrázek 8: Shlukovací koeficienty uzlů sítě Gnutella (zelená = nejnižší, červená = nejvyšší)





Obrázek 9: Centrality, velikost uzlů = betweenness, barva = closeness

## Literatura

1. JURE LESKOVEC, J. Kleinberg; FALOUTSOS, C. *Gnutella dataset* [online]. 2021-12-11 [cit. 2021-12-11]. Dostupné z: <https://snap.stanford.edu/data/p2p-Gnutella04.html>.
2. LESKOVEC, Jure; KREVL, Andrej. *SNAP Datasets: Stanford Large Network Dataset Collection* [<http://snap.stanford.edu/data>]. 2014-06.