



METODY ANALÝZY DAT

ANALÝZA SÍTĚ

MAREK HAJDÍK

HAJ0136

Datová sada

SOCIAL CIRCLES: FACEBOOK

Zdroj: <https://snap.stanford.edu/data/ego-Facebook.html>

Datová sada se skládá ze skupin přátel na sociální síti Facebook. Data byla získána přes dotazník, který byl vytvořen jako aplikace pro Facebook. Všechna jména byla anonymizována a nahrazena identifikačním číslem. Hrany znamenají vzájemné přátelství. Graf je neorientovaný a nevážený.

Vrcholy	4039
Hrany	88234
Počet komponent	1
Průměrný shlukovací koeficient	0,6055
Diametr	8
Průměrný stupeň	43,69
Nejvyšší stupeň	1045
Hustota	0,0108
Průměrná vzdálenost	3,69



Analýza

SAMPLING

Metody:

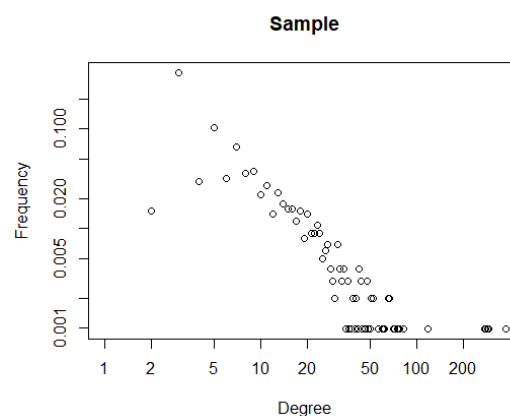
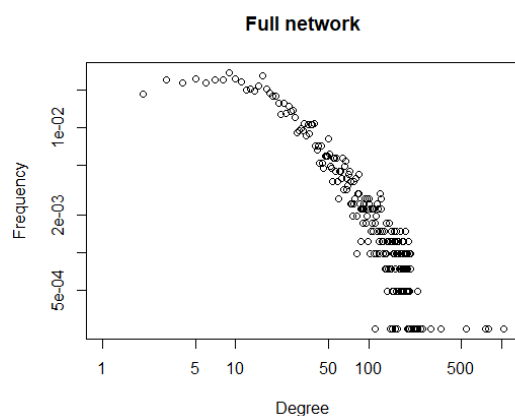
- **Random Walk Sampling** - Využívá principu náhodné procházky, kde se každý další krok volí zcela náhodně. Počáteční vrchol se vybere náhodně a v každém dalším kroku se s pravděpodobností c vrátí algoritmus do počátečního vrcholu a začne novou cestu. Kroky se opakují, dokud nemá algoritmus požadovanou velikost.

Parametry metody:

Velikost: 25%

Pravděpodobnost C : 15%

	Původní síť	Vzorek
Vrcholy	4039	1009
Hrany	88234	5650
Počet komponent	1	1
Průměrný shlukovací koeficient	0,6055	0,3185
Diametr	8	16
Průměrný stupeň	43,69	11,20
Nejvyšší stupeň	1045	375
Hustota	0,0108	0,0111
Průměrná vzdálenost	3,69	3,83





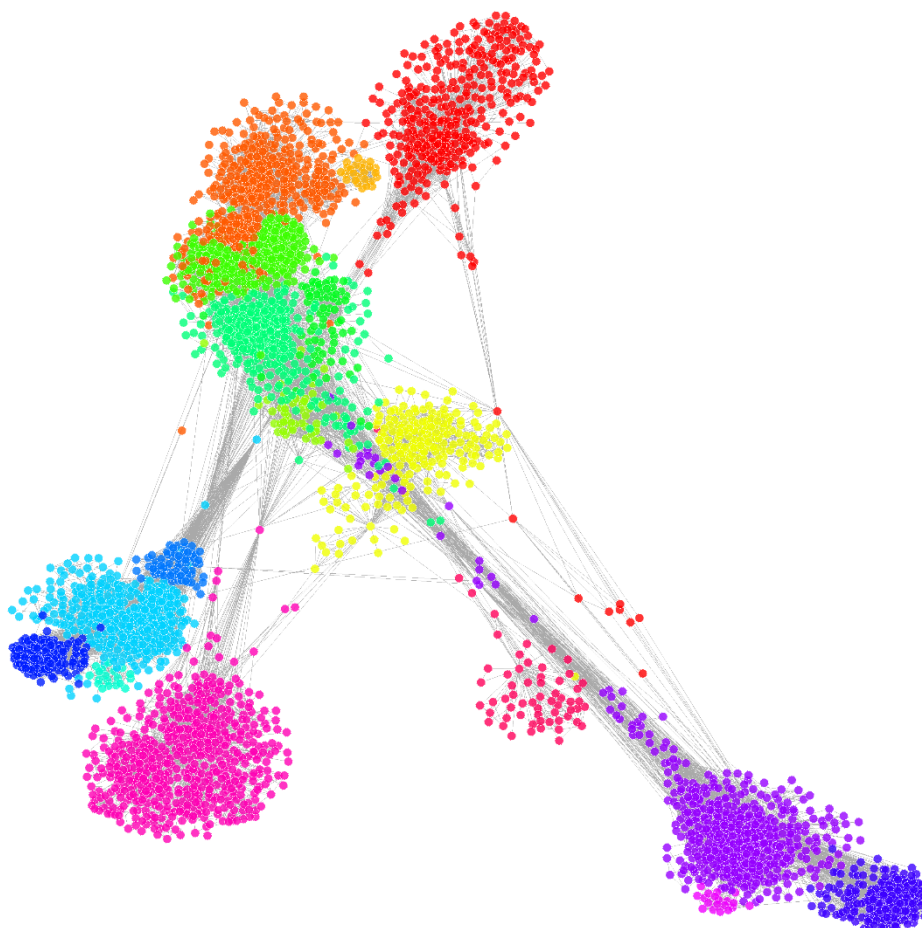
Vzorek zachovává některé charakteristiky původní sítě, skládá se z jedné komponenty, průměrná vzdálenost je velmi podobná. Naopak průměrný shlukovací koeficient je poloviční, takže vrcholy mají menší tendenci se shlukovat a diametr je dvojnásobný. Struktura grafu se vůbec nezachovala. Vzorek se příliš nehodí pro analýzu původní sítě.

DETEKCE KOMUNIT

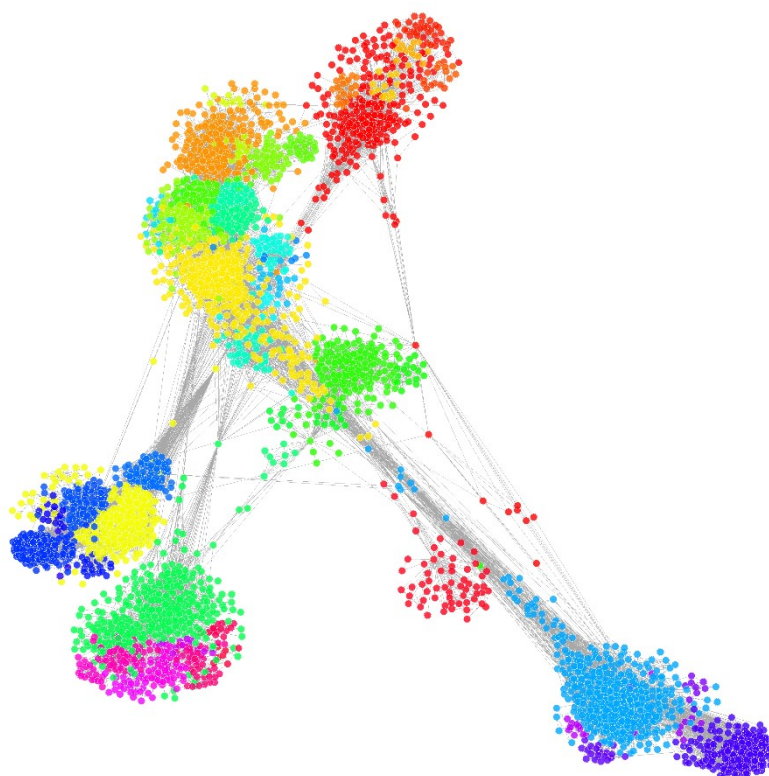
Metody:

- **Louvain method** - metoda pro hledání komunit ve velkých grafech která, používá modularitu a hierarchický přístup. Maximalizuje modularitu každé komunity. Jeden z nejrychlejších algoritmů založený na modularitě. Složitost $O(n \log n)$.

- **Label propagation** – velmi rychlý algoritmus pro detekci komunit v sítích, který má téměř lineární složitost. Funguje na principu označení všech vrcholů unikátními šítky a poté tyto šítky upravuje podle toho jaký šítek převládá v jeho sousedství.



Obrázek 1: Louvain method



Obrázek 2: Label propagation method

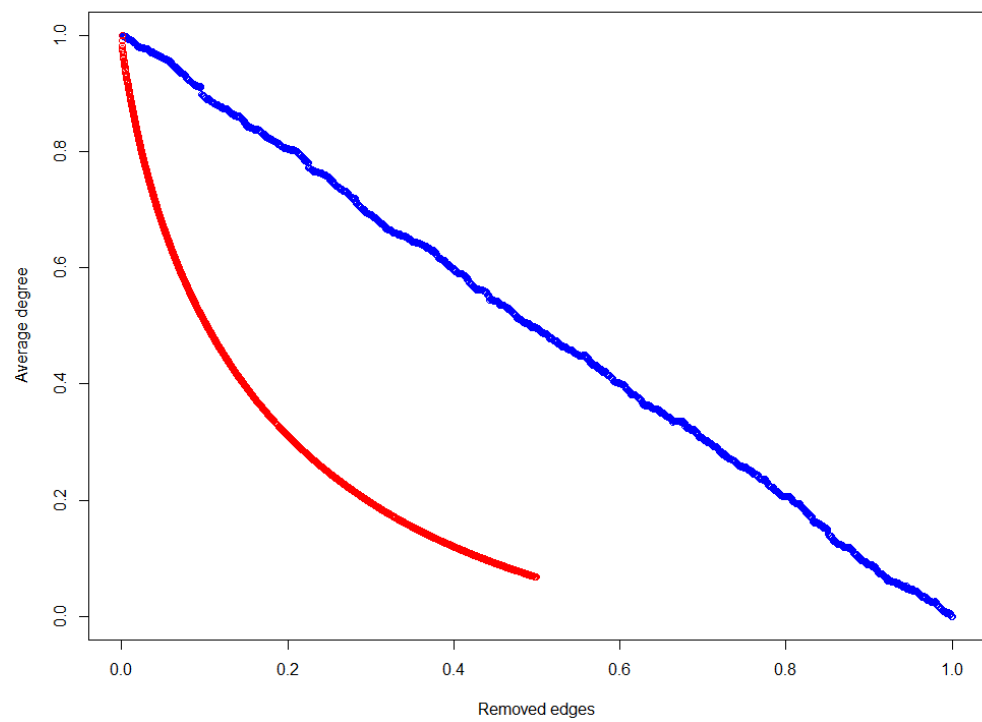
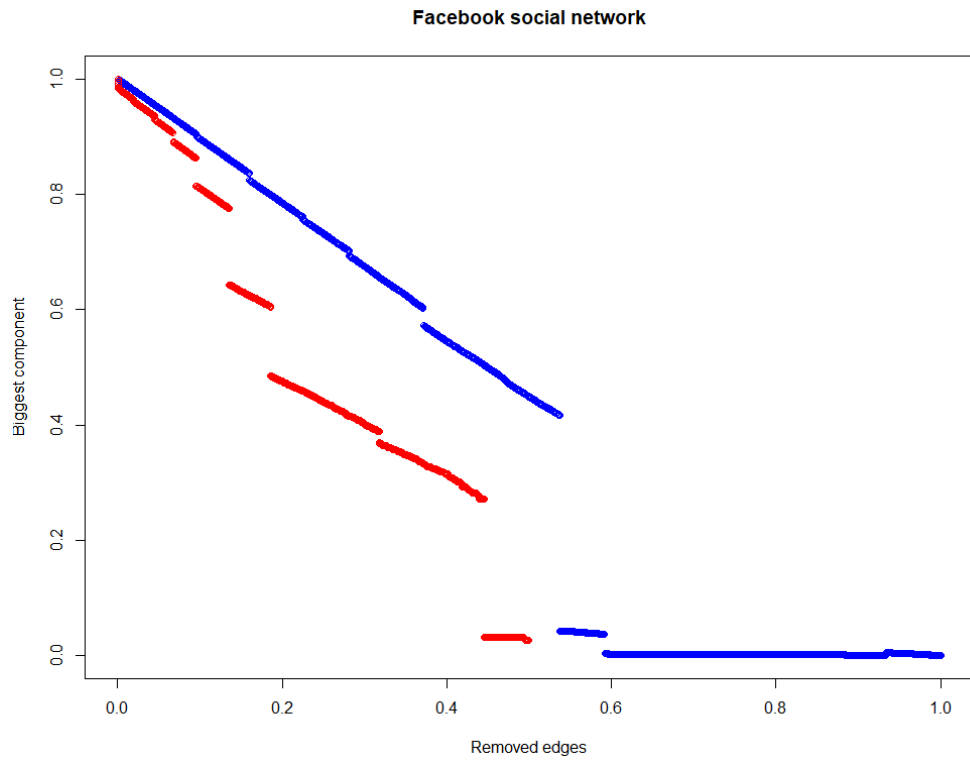
	Louvain	Label propagation
Počet komunit	17	52
Modularita	0,83	0,81
Největší komunita	550	507
Průměrná velikost komunity	238	78

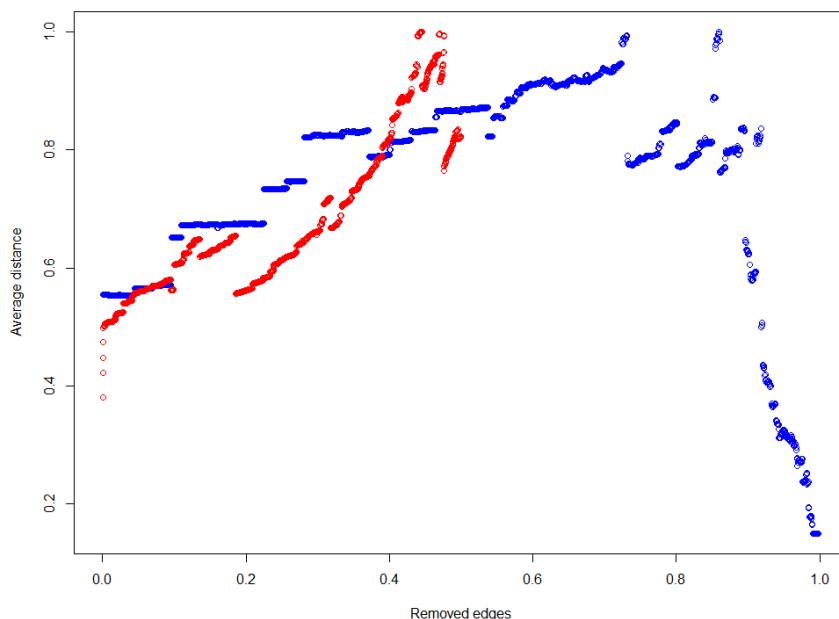
Sít má komunitní strukturu, jednotlivé komunity představují skupiny přátel, které jsou mezi sebou propojené. Louvain metoda našla 17 komunit s průměrnou velikostí 238 přátel, zatímco metoda Label propagation rozdělila síť na 52 komunit s průměrnou velikostí 78 přátel. Obě metody rozdělily komunity tak aby se nepřekrývaly a obě mají přibližně stejně velkou největší komunitu, tato komunita má pravděpodobně velkou hustotu hran.

ODOLNOST SÍTĚ

Metody:

- Náhodná selhání – ze sítě odebíráme náhodné vrcholy a sledujeme, jak se mění charakteristiky sítě
- Cílený útok – ze sítě odebíráme vrcholy s aktuálním nejvyšším stupněm a sledujeme, jak se mění charakteristiky sítě





Červené body = cílený útok

Modré body = náhodná selhání

Na prvním grafu vidíme že při náhodném selhání se největší komponenta zmenšuje lineárně dokud se nedostane na poloviční velikost, kde selhal důležitý vrchol a komponenta se rozpadla na malé části. Při cíleném útoku se velikost největší komponenty snižuje daleko rychleji, při odstranění 20 % vrcholů má největší komponenta už jen poloviční velikost.

Na druhém grafu sledujeme, jak se mění průměrný stupeň grafu. Zatímco při náhodném selhání se průměrný stupeň snižuje lineárně až do konce, tak při cíleném útoku se pokles exponenciálně zpomaluje.

Podle posledního grafu vidíme, jak se mění průměrná vzdálenost. Dokud síť drží pohromadě průměrná vzdálenost narůstá až do doby, kdy se začne rozpadat na malé části.

Tato síť má poměrně vysokou hustotu hran, takže náhodné selhání nemá moc velký vliv a síť dlouho odolává. Při cílených útocích se síť začne poměrně rychle rozpadat, což značí že se síť skládá z hubů vysokých stupňů, které jsou propojeny mezi sebou a při útoku na tyto huby se síť rozpadá na menší části.