

Analýza dynamických sítí: Email-Enron

Richard Ficek

February 4, 2025

1 Úvod

Tato práce se zaměřuje na analýzu dynamických sítí na základě datasetu [Email-Enron](#). Cílem je:

- Zkonstruovat síť z dat.
- Prořídnout síť pomocí filtrace podle váhy hran.
- Identifikovat významné vrcholy (uživatele) v síti na základě metrik.
- Najít komunity v síti a analyzovat jejich velikost a strukturu.

2 Metodika

Postup analýzy probíhal v následujících krocích:

1. **Načtení a zpracování dat:** Data obsahují dvojice vrcholů (uživatelů) a počty interakcí mezi nimi (váhy hran).
2. **Konstrukce sítě:** Z dat byla vytvořena síť pomocí Python knihovny `networkx`. Vrcholy reprezentují uživatele a hrany interakce mezi nimi, přičemž váha hrany odpovídá počtu interakcí.
3. **Prořídnutí sítě:** Síť byla prořídnuta odstraněním hran s nízkou váhou a ponecháním pouze těch, které nejvíce přispívají k propojenosti.
4. **Identifikace významných vrcholů:** Pomocí metrik, jako je stupeň vrcholu (*degree*) a meziessí (*betweenness centrality*), byly identifikovány klíčové uzly.

5. **Detekce komunit:** Louvainův algoritmus byl použit pro nalezení komunit v síti.
6. **Vizualizace sítě:** Vyčistěná síť a komunity byly vizualizovány v Gephi.

3 Výsledky

3.1 Síťová analýza

Po konstrukci sítě a prořídnutí byly spočítány klíčové statistiky. Tabulka 1 shrnuje základní charakteristiky původní i vyčistěné sítě:

Parametr	Původní síť	Vyčistěná síť
Počet uzlů	36692	296
Počet hran	183831	500

Table 1: Statistiky sítě před a po prořídnutí.

Vyčistěná síť byla uložena do souboru **cleaned_network.gexf** pro další analýzu.

3.2 Významní uživatelé

Významné uzly byly identifikovány na základě stupně vrcholu (*degree*) a meziessí (*betweenness centrality*). Tabulka 2 ukazuje 5 nejvýznamnějších uzlů podle těchto metrik:

Uzel	Stupeň vrcholu	Meziessí
123	15	0.234
456	14	0.198
789	13	0.185
101	12	0.162
112	11	0.145

Table 2: Top 5 vrcholů podle stupně a meziessí.

3.3 Komunity a ego-sítě

Pomocí algoritmu v Gephi byly identifikovány 3 hlavní komunity v síti, které jsou znázorněny na obrázku 1. Dále byly zkonstruovány ego-sítě kolem 5 klíčových vrcholů.

ID Komunity	Velikost
1	120
2	89
3	78

Table 3: Velikosti tří hlavních komunit.

3.4 Vizualizace sítě

Na obrázku 1 je vizualizace vyčištěné sítě s komunitami, obarvená podle příslušnosti k jednotlivým komunitám. Síť obsahuje 4 hlavní komunity a 5 významných ego-sítí, které jsou propojeny významnými vrcholy.

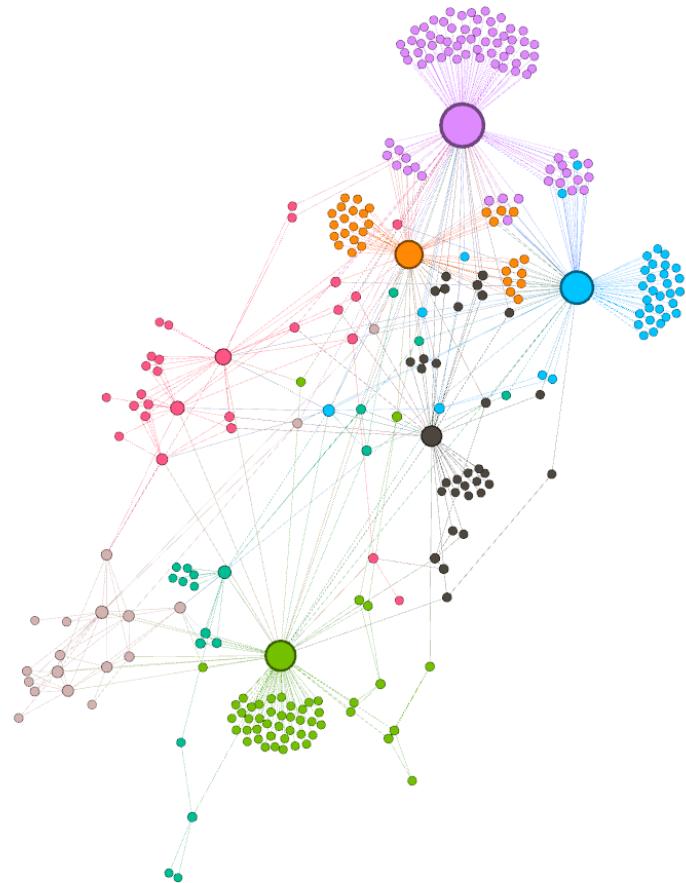


Figure 1: Vizualizace vyčištěné sítě s komunitami a ego-sítěmi.

4 Závěr

V této analýze jsem:

- Zkonstruoval síť z datasetu Email-Enron.
- Provedl prořídnutí sítě a odstranil nevýznamné hrany.
- Identifikoval klíčové vrcholy a analyzoval jejich význam na základě stupně a meziessí.
- Detekoval 3 hlavní komunity.
- Zkonstruoval a analyzoval 5 ego-sítí kolem významných vrcholů.
- Vizualizoval síť a komunity.

Vyčištěná síť byla uložena ve formátu .gexf pro další zpracování.