

Analýza vícevariačních sítí

Tomáš Anlauf, Vedoucí práce: doc. Mgr. Miloš Kudělka, Ph.D.

Fakulta elektrotechniky a informatiky, Vysoká škola Báňská - Technická univerzita Ostrava

Motivace a cíle

Vícevariační sítě jsou tvořeny vrcholy a vztahy ve formě hran, ovšem dále obsahují data o těchto vrcholech a hranách ve formě atributů. Reálné sítě velmi často obsahují několik atributů a mnoho analytických úloh je závislých na analýze jak vztahů, tak i atributů. Většina metod pro analýzu se zaměřuje pouze na síťovou topologii zkonstruovanou z vícevariačních dat. Navíc jsou tyto metody omezené na určité domény a vyžadují odbornou znalost problematiky. Nejčastěji jsou tyto sítě aplikovány v oborech analýzy sociálních sítí, biologických aplikacích a softwarovém inženýrství. Jejich uplatnění se prokázalo i v oborech jako je oceánografie, systémová analýza a strojní inženýrství.

Cílem práce bylo:

- Vyhledat a seznámit se s metodami určenými pro vizualizaci a analýzu vícevariačních sítí.
- Na základě získaných poznatků vytvořit nástroj, který tyto metody implementuje a umožňuje vícevariační sítě analyzovat.

Popis systému

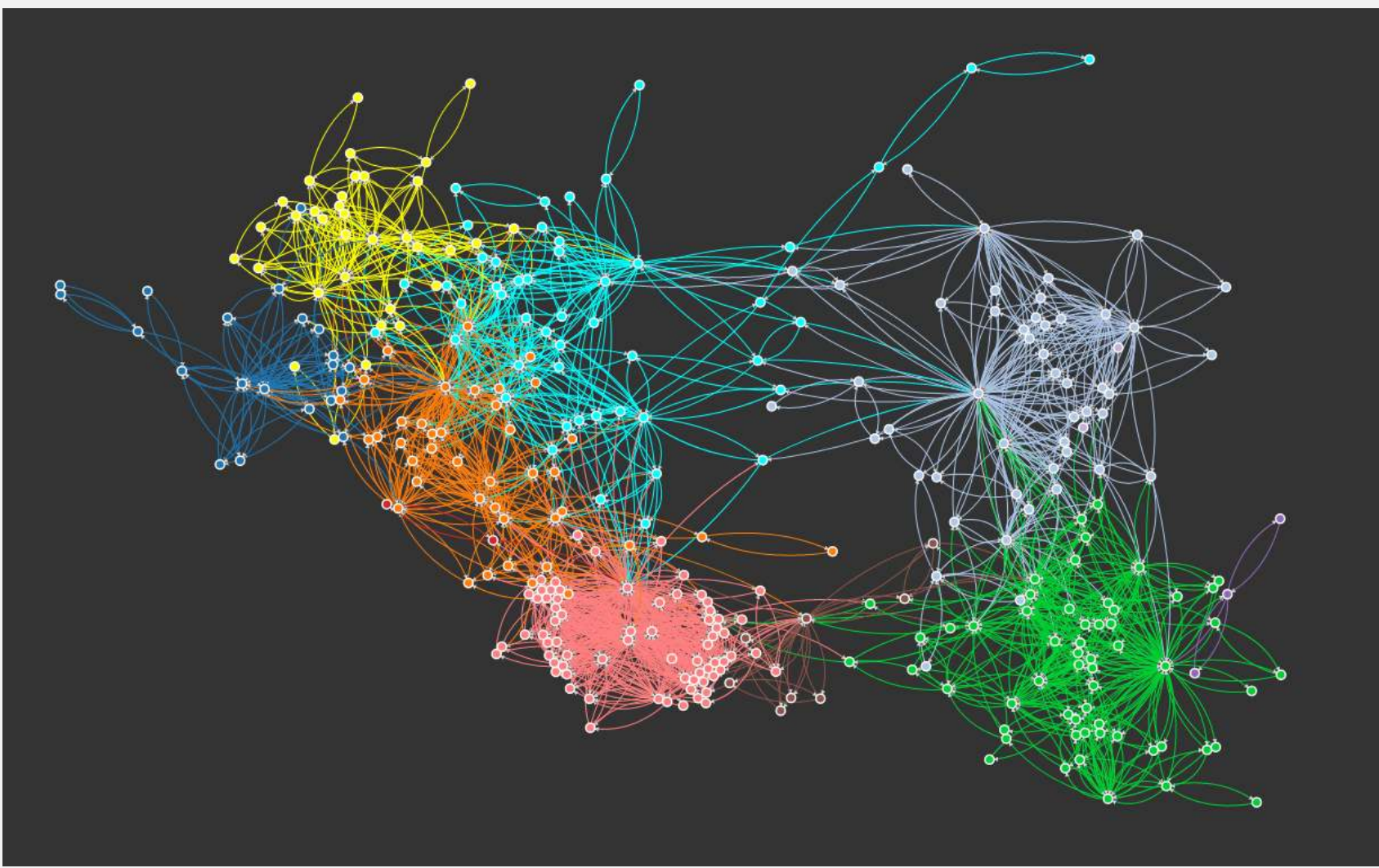
Systém pro analýzu vícevariačních sítí uživateli umožňuje provést vizualizaci vícevariační sítě, kterou si dále může prozkoumávat a upravovat dle svých potřeb. Systém umožňuje vytvořenou vizualizaci modifikovat více způsoby. Uživatel je schopen filtrovat vrcholy dle zadaných mezí atributů. Rozložení sítě není pevné, proto uživatel může vlastnosti grafu sítě modifikovat dle své libosti. V případě, že jsou uživatelem dodána pouze vektorová data, tak je systém schopen na jejich základě vytvořit reprezentativní síť. Další funkcí, kterou lze využít, je rozřazování vrcholů grafu do skupin. Skupiny lze tvořit ručně nebo lze využít funkce detekce komunit.

Převod vektorových dat na síť

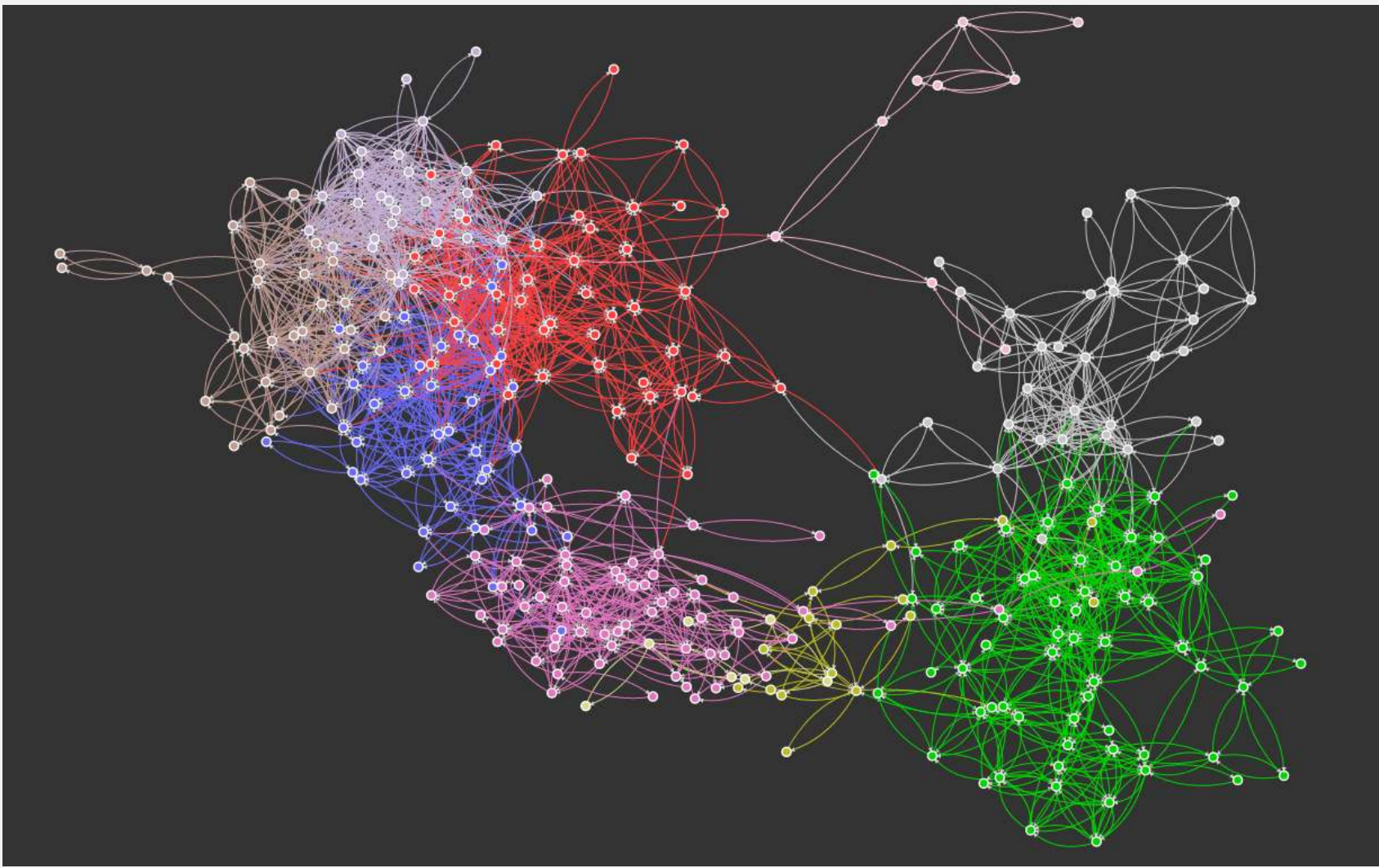
- **Algoritmus LRNet** - vytváří hrany mezi páry nejbližších sousedů a dále mezi individuálními objekty dle jejich vypočtené reprezentativnosti.
- **Algoritmus ϵ a kNN** - vytváří hrany mezi vrcholy na základě prahu použité metricky či podobnosti a minimálního počtu sousedů.

Správa komunit

- **Paralelní Louvain algoritmus** - hledá rozdělení do komunit s nejlepší globální modularitou
- **Ruční správa komunit**



(a) LRNet Ecoli



(b) ϵ -kNN Ecoli

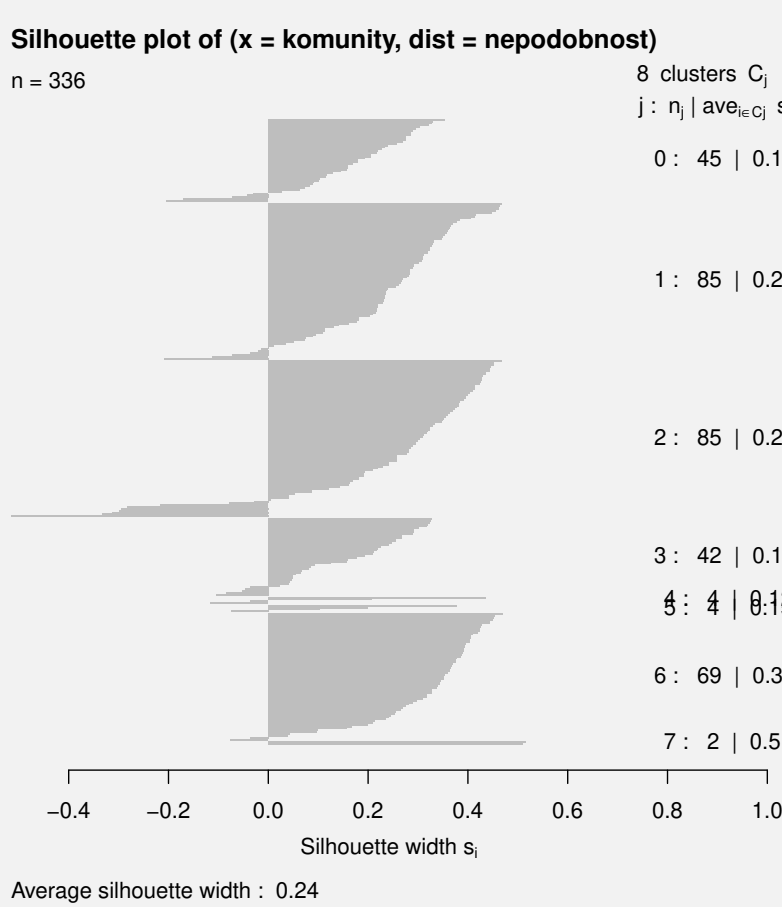
Obrázek: Vytvořené sítě převodem vektorových dat datasetu Ecoli[1] s barevně odlišenými nalezenými komunitami

Experimenty

Cílem je změřit kvalitu algoritmů pro převod vektorových dat na síť a ukázat, zda Louvain algoritmus pro detekci komunit rozdělí vrcholy sítě do komunit, které budou co nejvíce odpovídat jejich reálným třídám. Pokud ne, tak jestli je možné tuto detekci komunit zlepšit pomocí filtrování vrcholů na základě jejich atributů.



(a) LRNet Ecoli

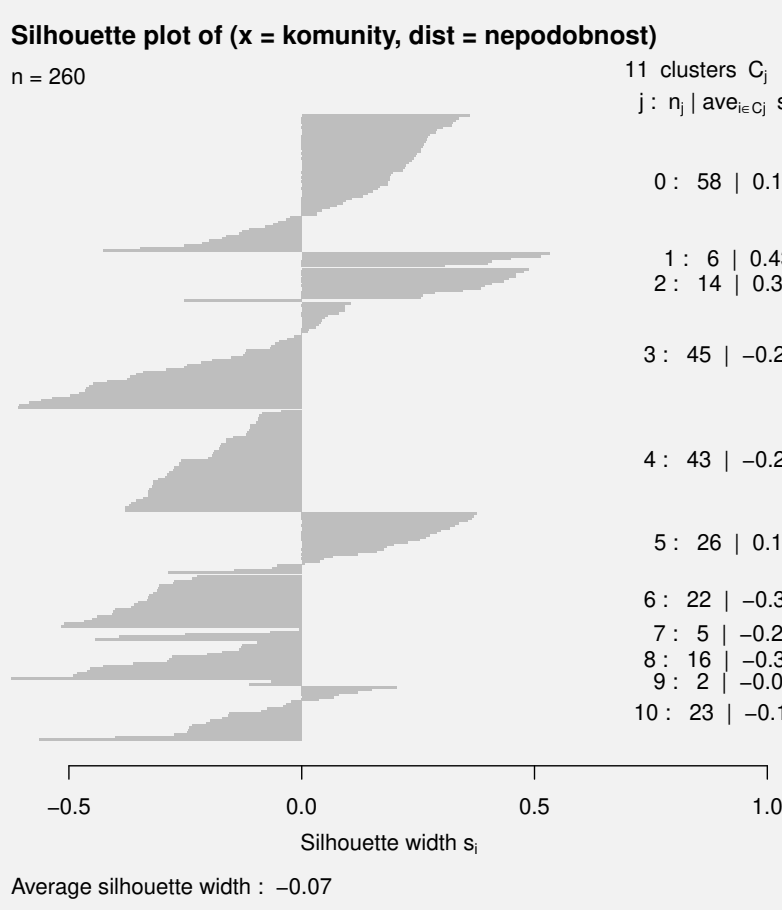


(b) ϵ -kNN Ecoli

Obrázek: Silhouette koeficienty vrcholů detekovaných komunit na celé síti Ecoli



(a) LRNet Ecoli



(b) ϵ -kNN Ecoli

Obrázek: Silhouette koeficienty vrcholů detekovaných komunit na filtrované síti Ecoli

Odkazy

1. DUA, Dheeru; GRAFF, Casey. *UCI Machine Learning Repository*. 2017. Dostupné také z: <http://archive.ics.uci.edu/ml>.