

Užduotis: Su turimu duomenų rinkiniu sudaryti tinklą, naudojantis „CytoScape“ programine įranga ir atsakyti į klausimus.

Klausimai:

1. Mokytojų pajėgumui savivaldybėse.
2. Aukštos kvalifikacijos mokytojų pajėgumai savivaldybėse.
3. Mokinių skaičius tenkantis vienam mokytojui savivaldybėse.
4. Bendruomenių radimas.
5. Pastebėjimai apie „Cytoscape“ programinę įrangą.

Duomenys: <https://www.svis.smm.lt/pedagogai/>

Darbo eiga:

1. Duomenų gavyba ir paruošimas.

Gautos 6 „Excel“ formato lentelės:

- a. Aukštos kvalifikacijos mokytojų dalis pagal priklausomybę
- b. Mokinių, tenkančių vienam mokytojui, skaičius pagal priklausomybę
- c. Mokytojų, dirbančių ne vienoje mokykloje, dalis
- d. Mokytojų, mokančių daugiau kaip 1 dalyko, dalis pagal priklausomybę
- e. Pensinio amžiaus mokytojų dalis pagal priklausomybę
- f. Mokytojų, kurių darbo stažas iki 2 metų (0 – 1 metai), dalis (proc.)

Jos konvertuotos į CSV tipo lentelę, pasinaudojant žemiau pateiktu programiniu kodu.

```
for filename in filenames:
    for sheet, skip in zip(sheet_names, row_skips):
        print(sheet)
        ds = pd.DataFrame()
        df = pd.read_excel(filename,
                           sheet_name=sheet, skiprows=skip, nrows=65)
        ds['savivaldybe'] = df['Unnamed: 1']
        ds['stazas_iki_2_metu_proc'] = df['Savivaldybėje']
        ds['mokslo_metai'] = sheet[:9]
        print(ds.shape[0])
        df_2 = pd.concat([df_2, ds])
```

Atliktas perteklinių duomenų valymas pasinaudojant kodu

```
counties = list(df_2['savivaldybe'].unique())
del counties[counties.index('Iš viso (asmenų)')]
```

Duomenų rinkinį sudaro šie atributai:

- a. Savivaldybės pavadinimas.
- b. Mokslo metai.
- c. Mokytojų skaičius
- d. Ekspertų skaičius
- e. Ekspertų procentinė dalis
- f. Mokinių skaičius

- g. Mokinių tenkančių vienam mokytojui skaičius
- h. Mokytojų mokančių daugiau nei vienoje mokykloje skaičius
- i. Mokytojų mokančių daugiau nei vienoje mokykloje procentinė išraiška
- j. Mokytojų, kurie moko daugiau negu vieno dalyko skaičius
- k. Pensijinio amžiaus mokytojų skaičius
- l. Mokytojų su mažesniu, nei 2 m., stažu procentinė išraiška.

2. Ryšių tarp savivaldybių sukūrimas.

Atliktas savivaldybių nepanašumo skaičiavimas pasinaudojant formule $Coef = \frac{|p_1 - p_2|}{\max(p_1, p_2)}$, kur p_1 – pirmo stebėjimo požymis, p_2 – antro stebėjimo požymis, Coef – nepanašumo koeficientas. Kuo koeficientas arčiau 0, tuo du stebėjimai, pagal tiriamą požymį panašesni. Programinis kodas pateikiamas žemiau.

Skirtingumo indeksas pasirinktas, nes jis tiesiogiai matuoja kiekvienos apskrities reikšmės nuokrypio mastą ir efektyviai išryškina skirtumus. Šis metodas leidžia greitai nustatyti statistiškai artimesnes arba labiau besiskiriančias apskritis, todėl galima susidaryti aiškų santykinio panašumo vaizdą. Be to, naudojant vieno parametro nepanašumo indeksą, analizė supaprastėja, todėl ją lengva interpretuoti ir naudinga iš karto palyginant regionus. Papildomai įvestas panašumo matas, kurio vertė lygi vidurkių, visų skaičiuotų panašumo įverčių, tai apjungiant kelių atributų panašumus, kurie leidžia plačiau pažvelgti į analizuojamas savivaldybes.

```
def calculate_similarity(row1, row2):
    mokytk_sk = round(abs(row1['mokytoju_skaicius'] - row2['mokytoju_skaicius']) / max(row1['mokytoju_skaicius'], row2['mokytoju_skaicius']), 4)
    mok_sk = round(abs(row1['mokiniu_skaicius'] - row2['mokiniu_skaicius']) / max(row1['mokiniu_skaicius'], row2['mokiniu_skaicius']), 4)
    ne_1_mok = round(abs(row1['mokytoju_skaicius_ne_vienoje_mokykloje'] - row2['mokytoju_skaicius_ne_vienoje_mokykloje']) / max(row1['mokytoju_skaicius_ne_vienoje_mokykloje'], row2['mokytoju_skaicius_ne_vienoje_mokykloje']), 4)
    avg_all = round(np.mean([mokytk_sk, mok_sk, ne_1_mok]), 4)
    return mokytk_sk, mok_sk, ne_1_mok, avg_all

for year, group in df.groupby('mokslo_metai'):
    for (idx1, row1), (idx2, row2) in itertools.product(group.iterrows(), repeat=2):
        mokytk_sk, mok_sk, ne_1_mok, avg_all = calculate_similarity(row1, row2)
        new_data.append([row1['savivaldybe'] + list(row1[cols]) + row2['savivaldybe'] + list(row2[cols]) + [year, mokytk_sk, mok_sk, ne_1_mok, avg_all]])
        new_data.append([row1['savivaldybe'], row2['savivaldybe'], year, mokytk_sk, mok_sk, ne_1_mok, avg_all] + list(row1[cols]))
new_df = pd.DataFrame(new_data, columns=['savivaldybe_1'] + cols + ['savivaldybe_2'] + ['i'] + ['i_2'] for i in cols + ['mokslo_metai', 'mokytk_simil', 'mok_simil', 'ne_1_mok_simil', 'avg_simil'])
```

Naudojantis panašumo skaičiavimo kodu, atsirado tokių stebėjimų, kur savivaldybė lyginama pati su savimi, todėl pasinaudojant kodu pateiktu žemiau, toki stebėjimai šalinami.

```
df_2[ df_2['savivaldybe_1'] != df_2['savivaldybe_2'] ].to_csv(filename, index=False)
```

Papildomai pridėti atributai: nepanašumas pagal mokytojų skaičių, nepanašumas pagal mokinių skaičių, nepanašumas pagal dirbančių daugiau nei 1 mokykloje skaičius ir visų panašumų vidurkis.

Galutinis duomenų rinkinys pasiekiamas per nuorodą: https://drive.google.com/file/d/1kAAUZrD_FdT-K6rtfHHyYW0BRkkCozP3/view?usp=sharing

3. Tinklo sudarymas.

Pradžios mazgas – savivaldybės pavadinimas.

Pabaigos mazgas – kitos savivaldybės pavadinimas.

Mazgo atributai – pirmoje dalyje pateikiami duomenys.

Jungties atributai – antroje dalyje gaunami panašumo įverčiai ir mokslo metai.

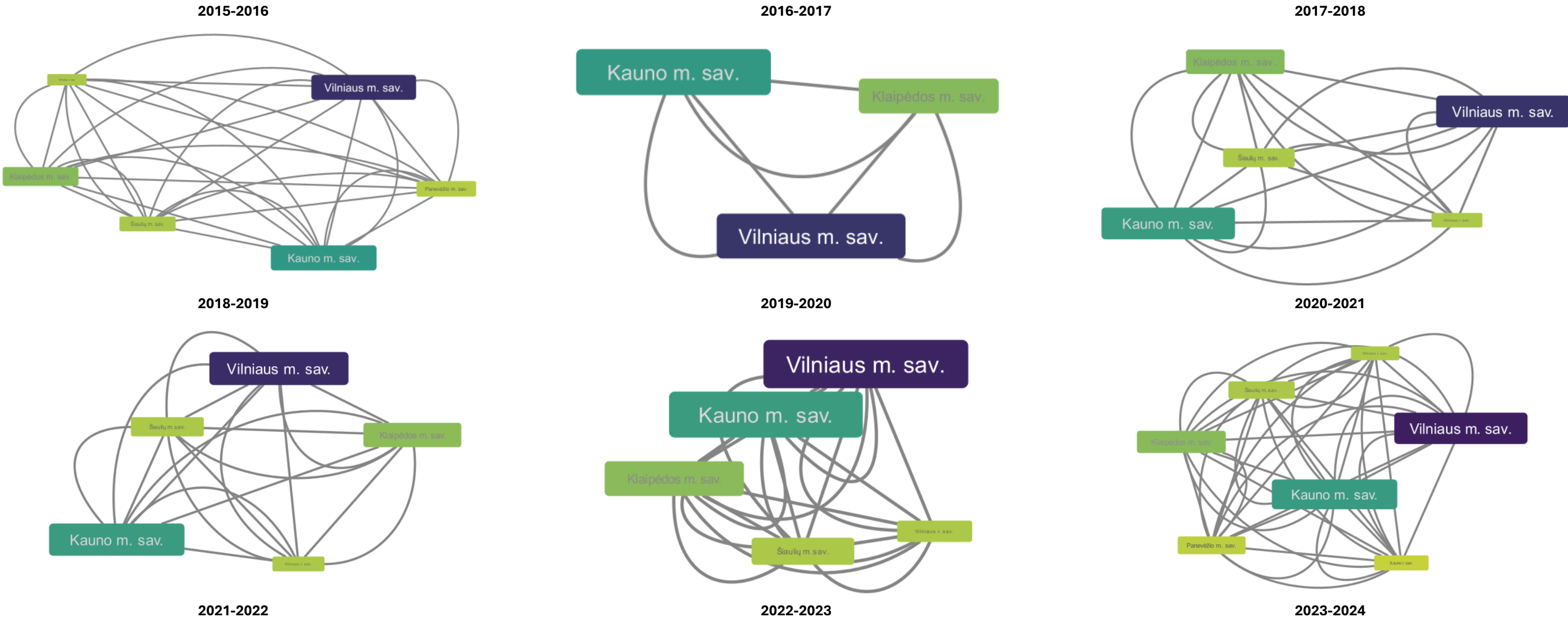
Eimantas Zaranka MIF230003

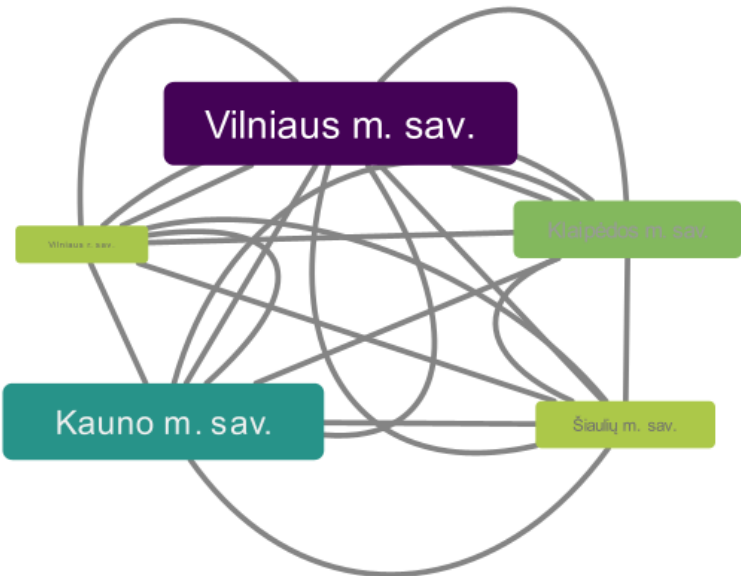
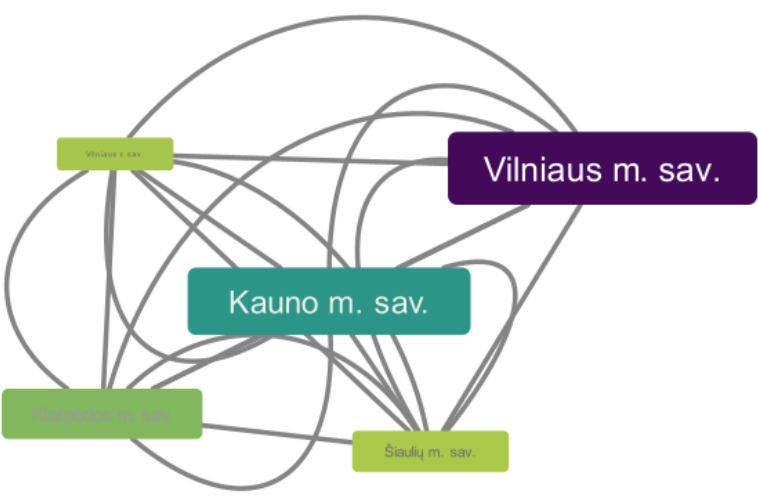
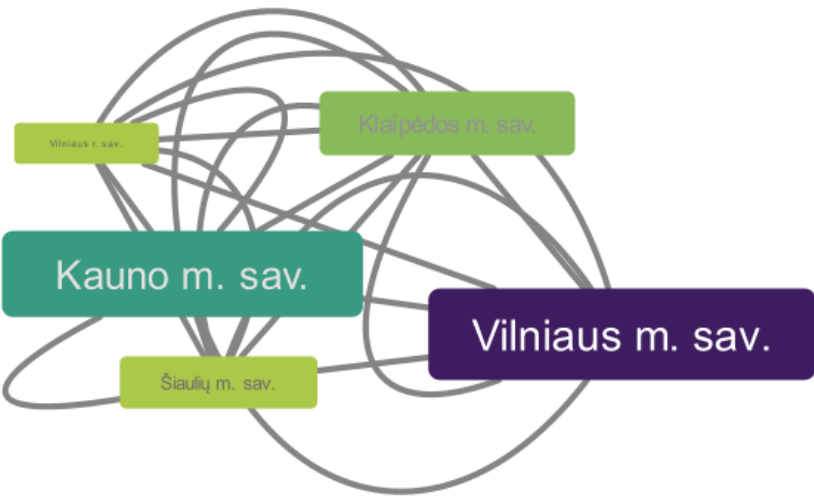
Sudarytas tinklas pasiekiamas per nuorodą: https://drive.google.com/file/d/1Y1VbiF_-UJaElXSOMbs6Qbaknwn84TUE/view?usp=sharing

4. Klausimų atsakymai.

Klausimams 1 ir 2 pasirinktas mazgų stilius, kuriame naudojamos mazgų spalvos ir dydis. Šviesesnės spalvos demonstruoja mažesnį mokytojų skaičių, kai tuo tarpu žalsva spalva mokytojų vidurkj, o violetinė spalva išskiria savivaldybes su daugiausiai mokytojų. Mazgo dydis kinta proporcingai, pagal mokytojų ekspertų skaičių, t. y. kuo savivaldybėje mokytojų ekspertų daugiau, tuo mazgas didesnis. Pirmam klausimui fiksuotas viršutinės 10 procentilio mokytojų skaičiaus vertės, antrajam – viršutinio 10 procentilio mokytojų ekspertų skaičiaus vertės. Klausimui 3, mazgų spalvai ir dydžiui parinktas gradientas, kuriam priskirta reikšmė, vienam mokytojui vidutiniškai tenkančių mokinių skaičius ir kirpimai atlikti pagal šio atributo viršutinio 10 procentilio reikšmes.

Klausimas 1: Mokytojų pajėgumui savivaldybėse

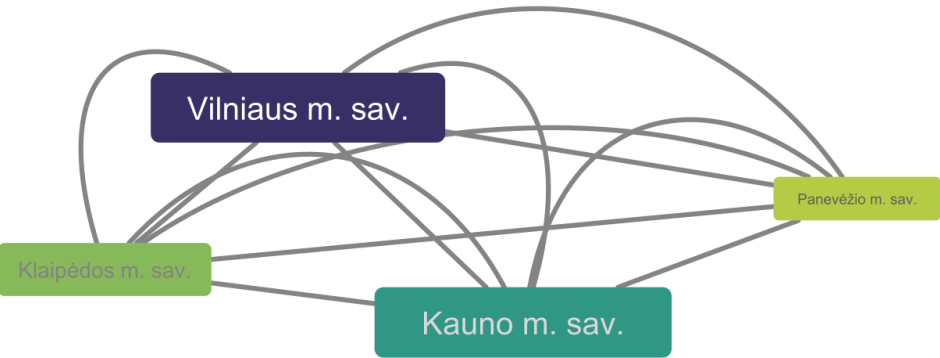




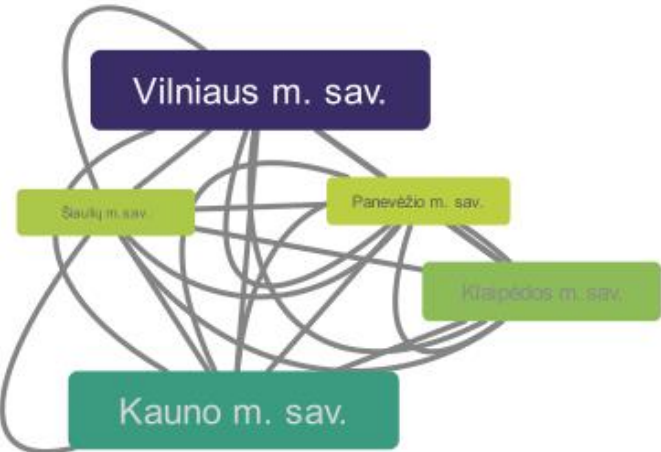
Pagal mokytojų pajėgumus pirmauja didžiosios savivaldybės, t. y. Kaunas m., Klaipėdos m. ir Vilniaus m. Papildomai, metai nuo metų prie lyderių atsiranda ir kitų mažesnių savivaldybių, tokių kaip Panevėžio m., Šiaulių m., Vilniaus r. ir Kauno r., tačiau jos kiekvienais metais kintą ir jų pozicija prie daugiausiai mokytojų turinčių savivaldybių nėra garantuota.

Klausimas 2: Aukštos kvalifikacijos mokytojų pajėgumai savivaldybėse

2015-2016

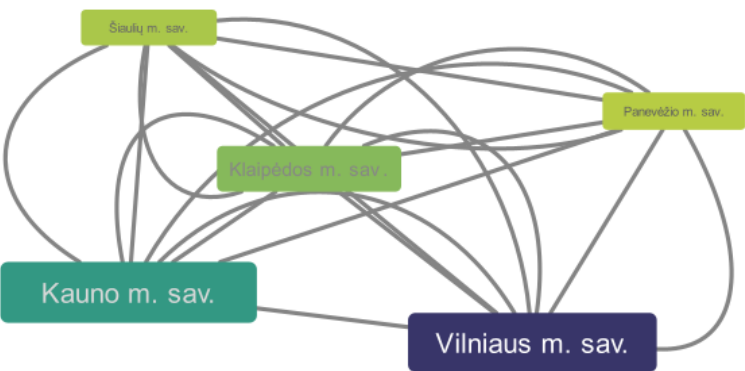


2018-2019

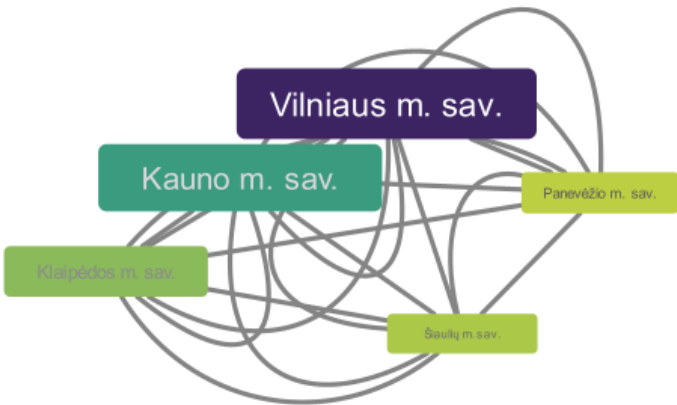


2021-2022

2016-2017

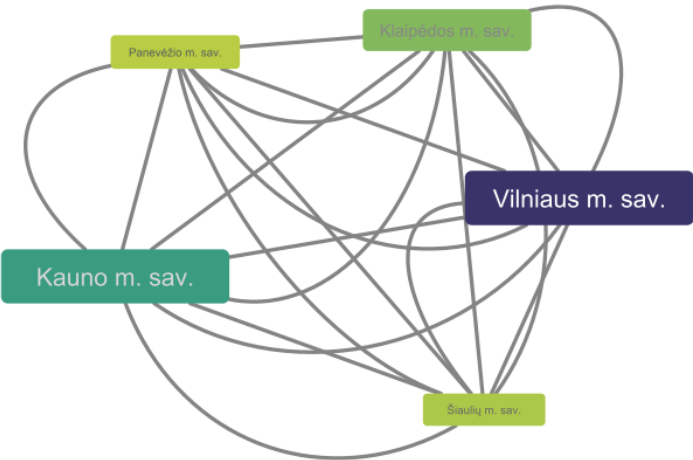


2019-2020

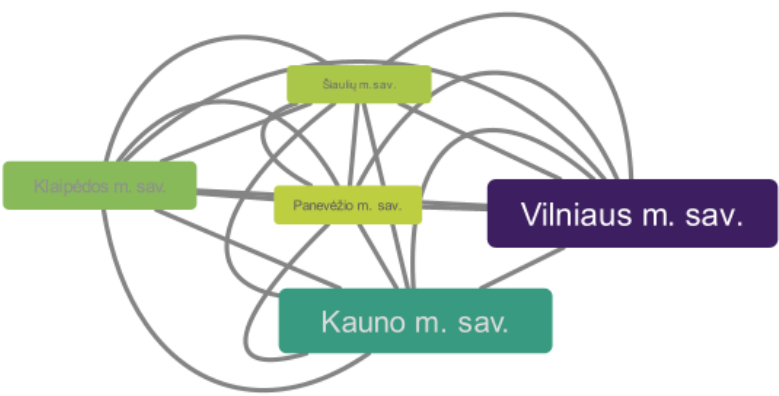


2022-2023

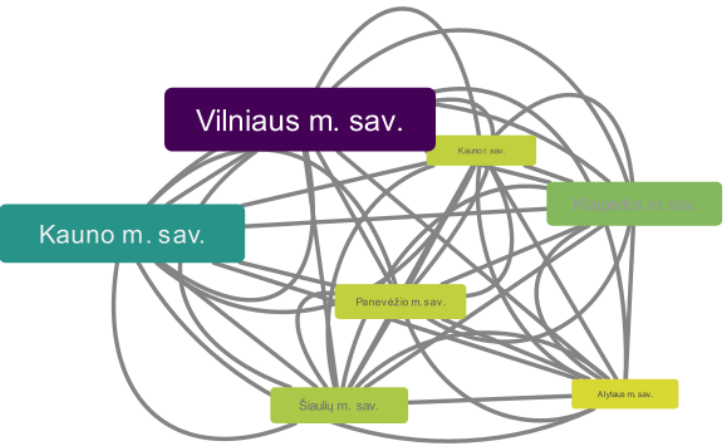
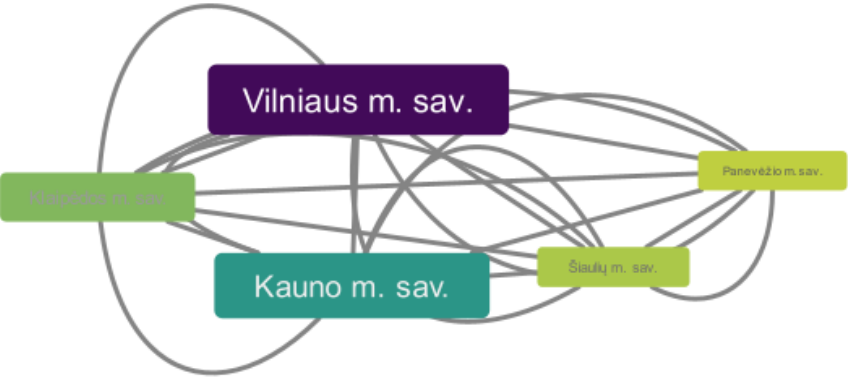
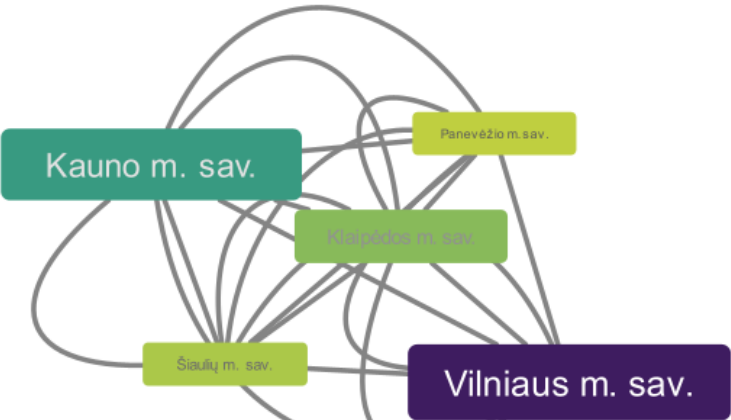
2017-2018



2020-2021

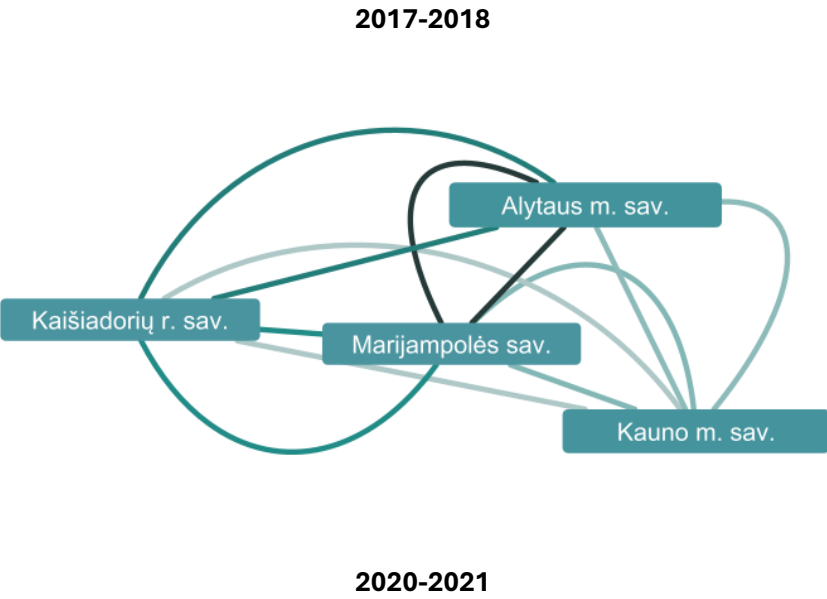
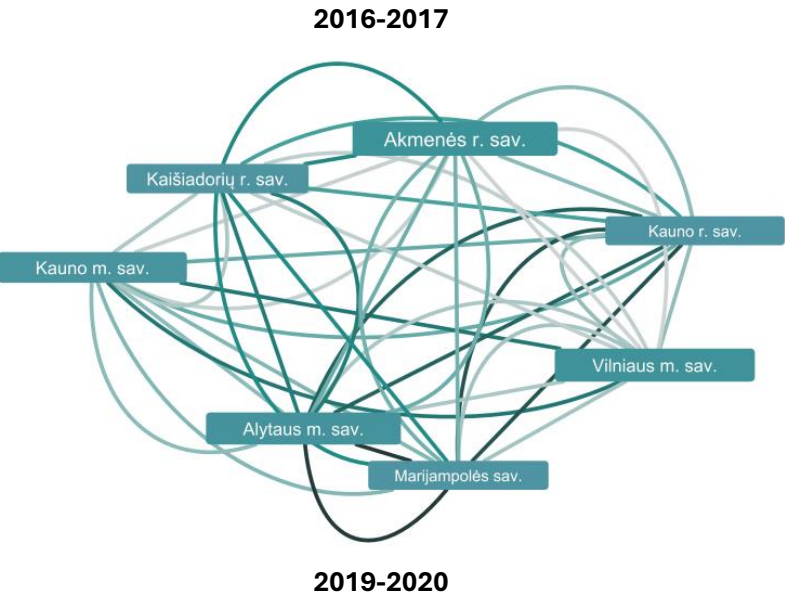
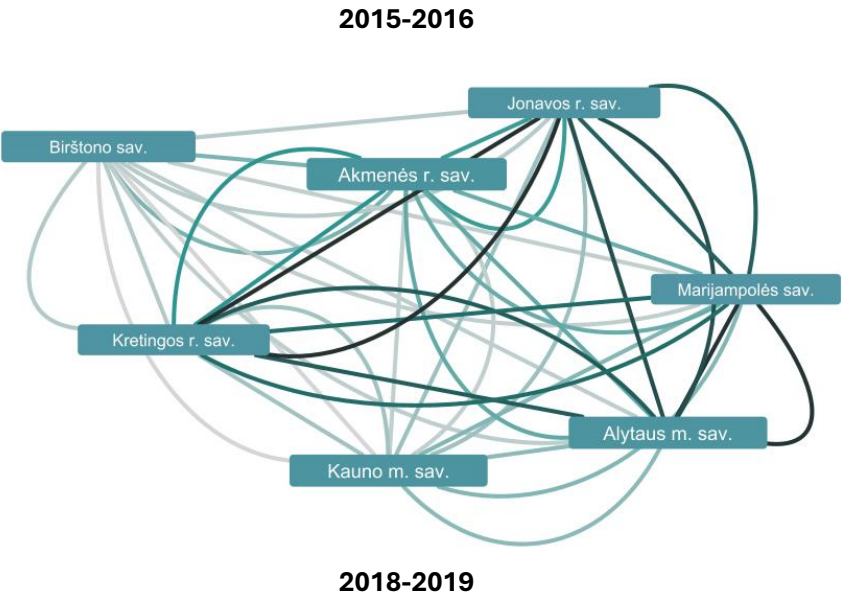


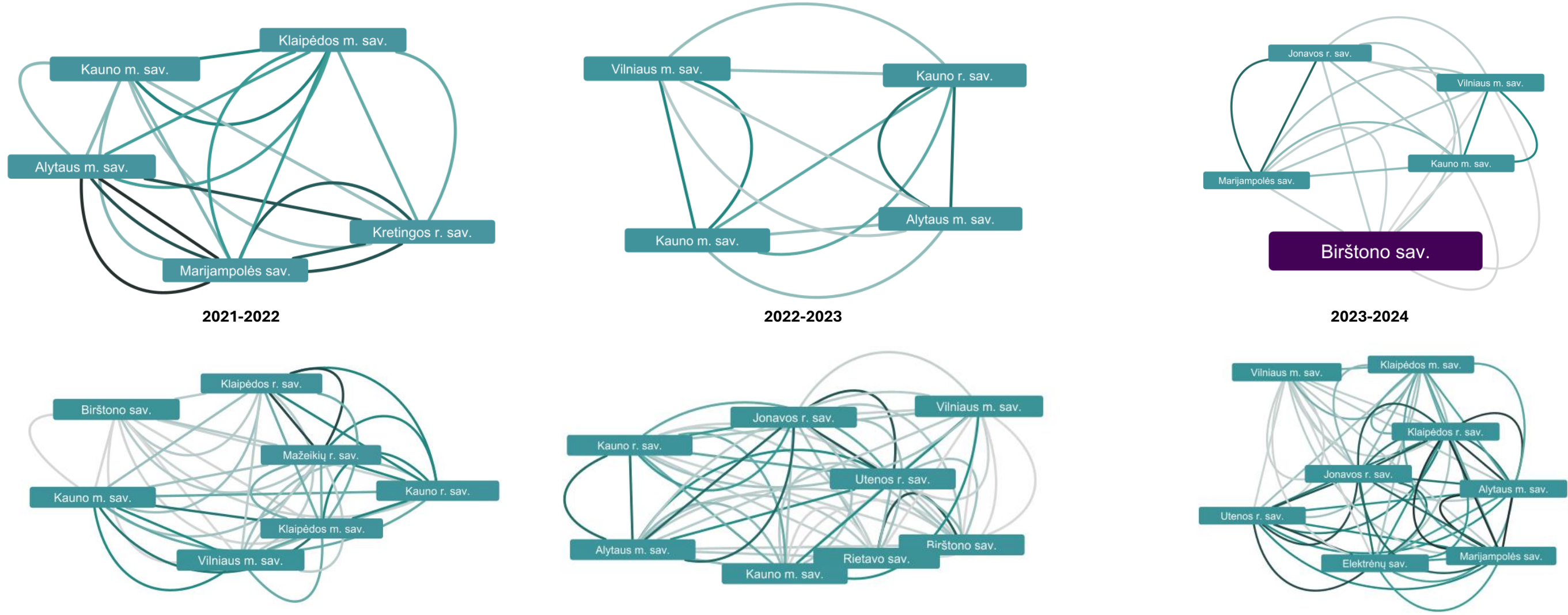
2023-2024



Situacija su mokytojų ekspertų skaičiumi lygiai tokia pati, kaip ir su mokytojų skaičiumi. Didžiosios savivaldybės pastoviai turi daugiau mokytojų specialistų, tačiau pastebima, kad Šiaulių m. ir Panevėžio m. savivaldybės pasižymi mokytojų specialistų gausa ir yra pastoviai tarp lyderiaujančių savivaldybių. Tuo tarpu kitos mažosios savivaldybės, kurios vis atsirasdavo pirmo klausimo metu, tendencingai atsiranda ir antrojo klausimo metu, kuomet jų mokytojų skaičius išauga.

Klausimas 3: Mokinių skaičius tenkantis vienam mokytojui savivaldybėse





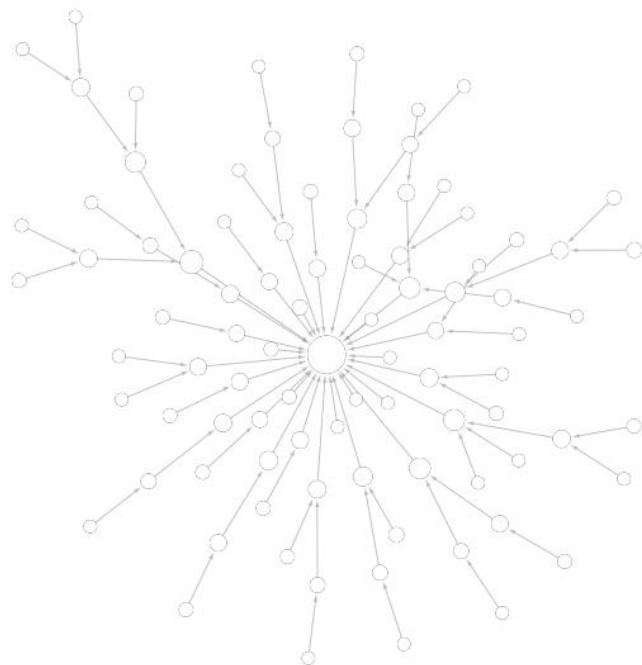
Stebint mokinių tenkančių mokytojų skaičių, pastebima, kad visose savivaldybėse kiekvienam mokytojų tenka daugmaž vienodas mokinių skaičius. Išimtis tik 2020-2021 m. m., prasidėjus COVID-19, Birštone stipriai išaugo mokinių skaičius vienam mokytojų. Ši anomalija pastebima tik vienais mokslo metais, kas reikalautų detalesnio įsigilinimo, kodėl taip atsitiko ir išsiaiškinti ar tai nėra duomenų įvedimo klaida. Šio klausimo metu pastebima, kad neretai didžiosios savivaldybės nėra prie lyderių, tam įtakos turi esamų mokytojų skaičius savivaldybėje, kur mažiau mokytojų turinčiose savivaldybėse darbo krūvis vienam mokytojų išauga.

Klausimas 4: Bendruomenių radimas

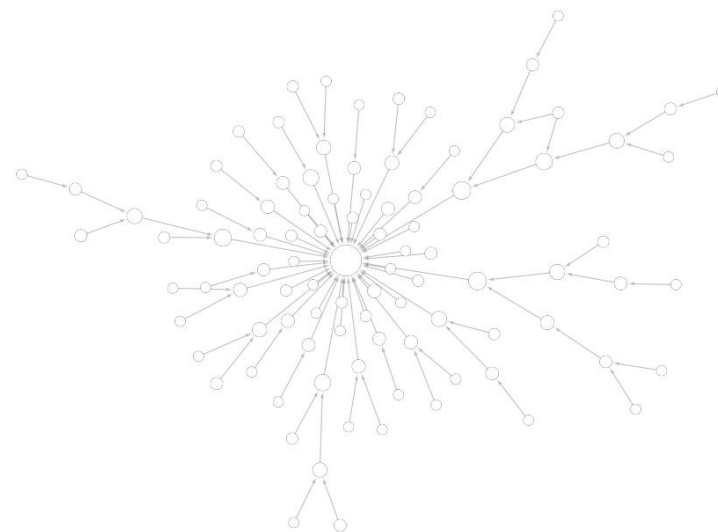
Bendruomenių aptikimui pasirinkti trys tinklo kirpimai, kurie skaičiuojami pagal viršutinį 10 procentilį panašumo įverčių, šie požymiai yra vidutinis vidurkinis panašumas, apimantis visus panašumo matus, pagal mokytojų mokančių daugiau nei 1 mokykloje ir pagal mokinių skaičių savivaldybėje. Pirmasis kirpimas, leidžia pastebėti, kurios Lietuvos savivaldybės yra panašiausios ir kaip bėgant metams, keičiasi situacija, t. y. ar vienai bendruomenei priklausančios mokyklos vystosi vienodai ar bendruomenių viduje atsiranda atskirtis. Antrasis kirpimas parodo, kaip buriasi savivaldybės, kuriose mokytojai moko daugiau nei vienoje mokykloje, šis kirpimas leidžia daryti išvadas apie tai ar mažose savivaldybėse mokytojai labiau linkę dirbti per kelias mokyklas, ar didesnėse. Trečiuoju kirpimu atkreipiamas dėmesys apie moksleivių pokyčius metuose, kas leidžia identifikuoti savivaldybes, kurios yra panašios pagal esamą mokinių skaičių, kas galėtų padėti įvertinti reikalingus sprendimus valstybiniame lygmenyje.

a. Bendruomenių radimas pasinaudojant CytoScape

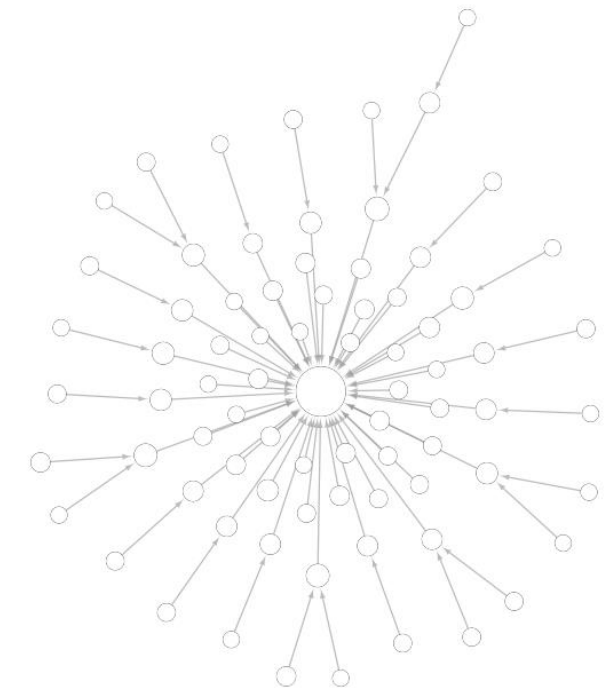
10 proc. vid. panašiausių savivaldybių bendruomenės



10 proc. panašiausių sav. pagal mokančių ne 1 mokykloje



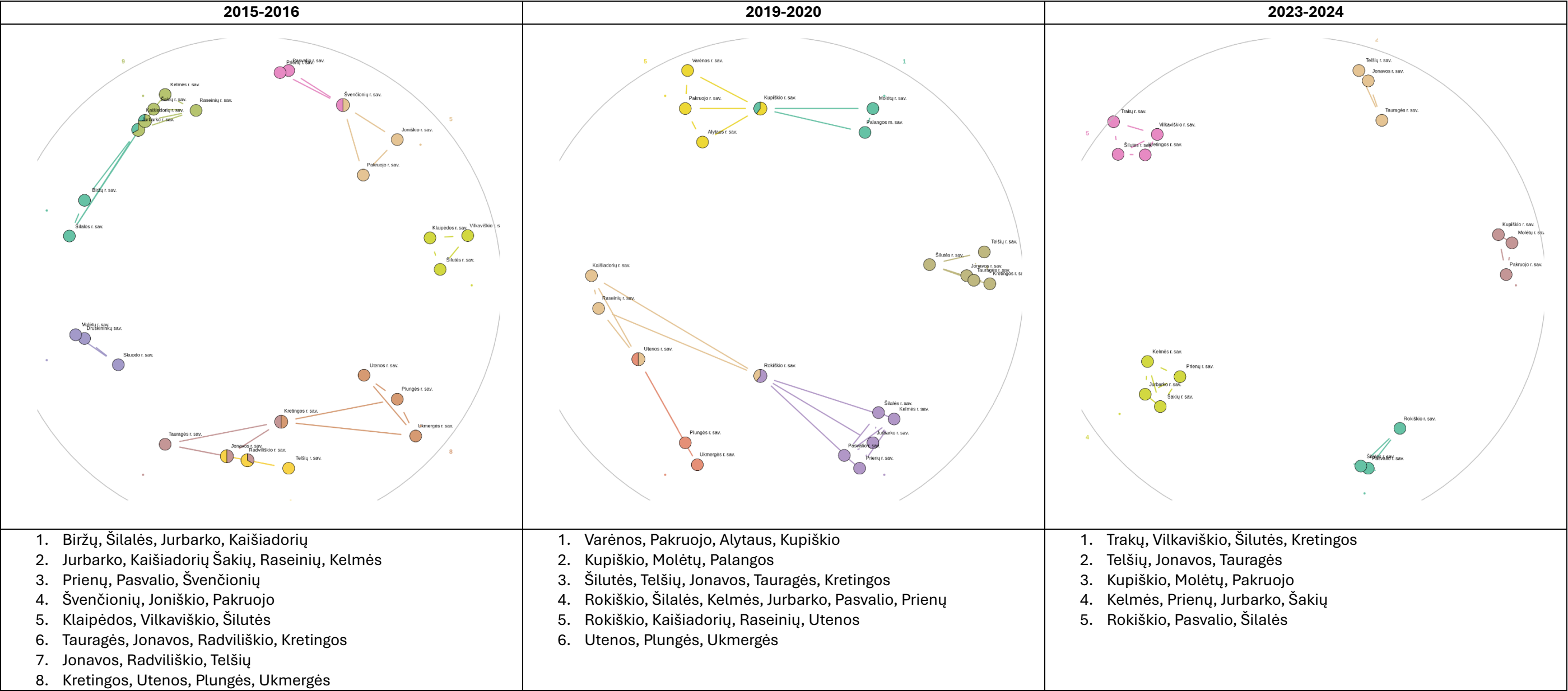
10 proc. panašiausių sav. pagal mokinių skaičių



Atlikus bendruomenių gavimą, pastebėta, kad bendruomenės išsiskaidžiusios pagal metus, t. y. tos pačios savivaldybės skirtingais metais priklauso vienai bendruomenei. Kilo įtarimas, kad metų atributas galėjo turėti poveikį, tačiau jį pašalinus rezultatai nesikeitė ir bendruomenės toliau priklauso nuo metų. Buvo bandyti skirtingi bendruomenių sudarymo algoritmai, tačiau ir šis sprendimas poveikio bendruomenių sudarymui neturėjo. Todėl daroma išvada, kad savivaldybėse statistikos keitėsi tendencingai, kas ir turėjo įtakos būtent tokiam bendruomenių susidarymui.

b. Bendruomenių aptikimas pasitelkiant R kalbos biblioteką linkcomm

Iš grafiko pašalintos bendruomenės, kurias sudaro tik dvi savivaldybės.



Pastebima, kad pirmaisiais duomenų rinkimo metais, bendruomenės buvo mažiau tarpusavyje atsiskyrusius, t. y. yra pastebimos tarpinės savivaldybės, kurios jungia kelias bendruomenes. 2015-2016 m. m. tokios savivaldybės buvo Kaišiadorių, Jurbarko, Švenčionių, Kretingos, Jonavos ir Radviliškio savivaldybės. 2019-2020 m. m. pastebimas bendruomenių grynėjimas, kur mažėja, tarpinių savivaldybių, t. y. šiais metais tokios buvo tik 3: Kupiškio, Utenos ir Rokiškio. Naujausiais stebėjimo metais pastebima, kad bendruomenės yra pilnai atsiskyrusios ir tarpinių savivaldybių tarp bendruomenių neaptikta.

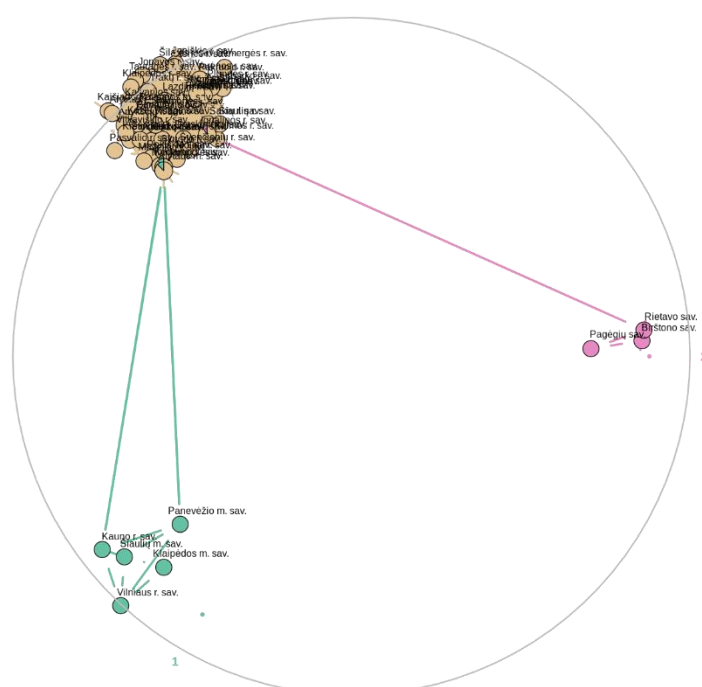
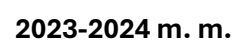
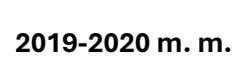
Programinis kodas:

```
sel <- select(edges, savivaldybe_1, savivaldybe_2, mokslo_metai, avg_simil)
info <- subset(sel, avg_simil < 0.11 & mokslo_metai == "2015-2016")

lm <- getLinkCommunities(select(info, savivaldybe_1, savivaldybe_2, avg_simil))

plot(lm, type = "graph", layout = "spencer.circle")
```

2015-2016 m. m.





6. Programinis kodas

<https://github.com/EimantasZaranka/SAM-pedagogai>