

Kā Datorikas fakultātes pasniedzēji liek atzīmes: divi klasteri

Kārlis Podnieks, LU
Pēteris Ručevskis, LU MII

Latvijas Universitātes 75.zinātniskā konference
2017.gada 24.februārī

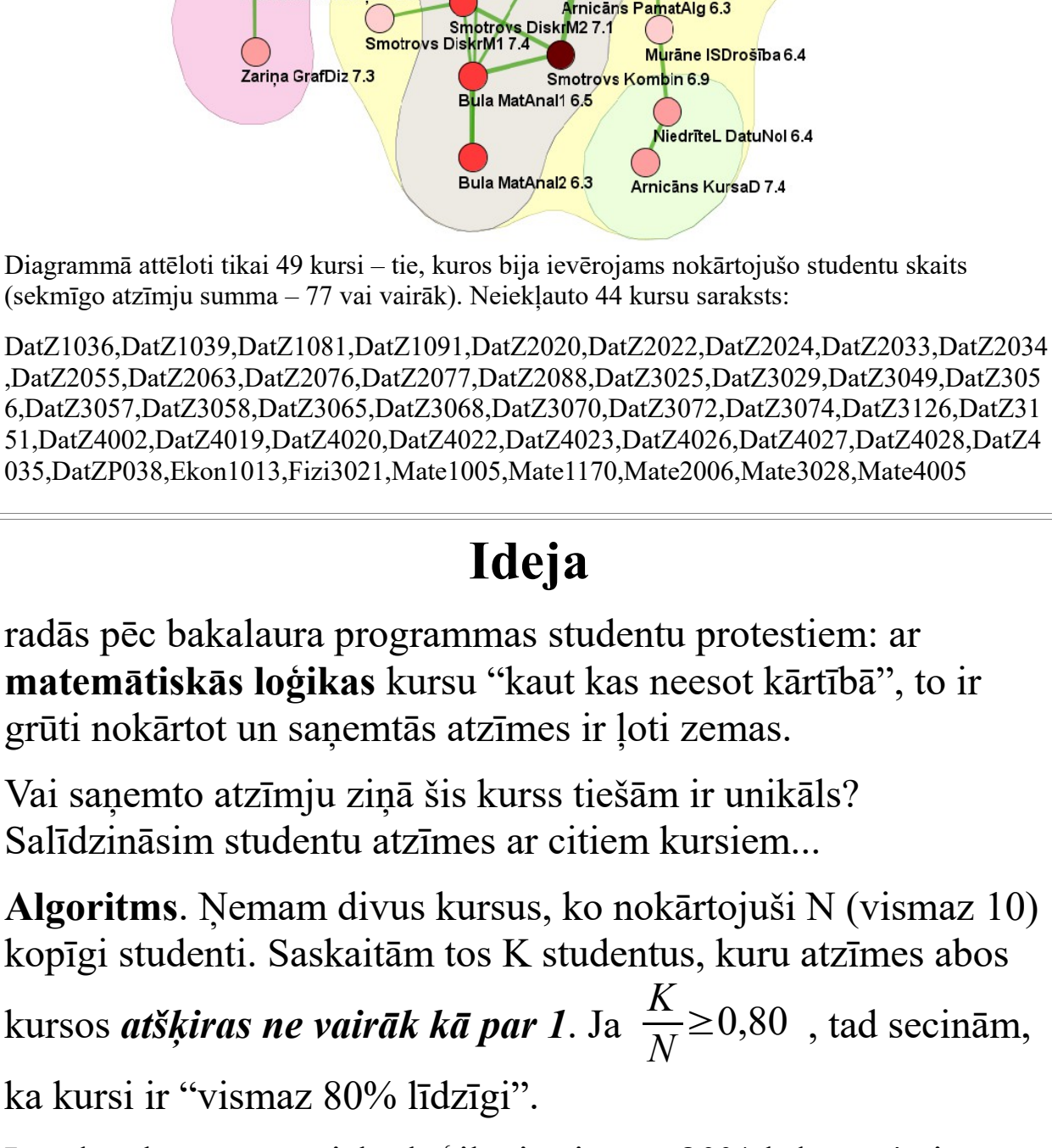
This work is licensed under a [Creative Commons License](#) and is copyrighted © 2017 by us, Karlis Podnieks and Peteris Ručevskis

Pētījuma interesantākais rezultāts

Diagrammā divi kursi ir savienoti, ja studentu saņemtās atzīmes tajos ir "vismaz 80% līdzīgas".

Atzīmju dati ņemti par bakalaura programmas 7.semestra 89 studentiem dažus gadus atpakaļ.

Uzrakstu beigās ir kursa vidējā atzīme.



Diagrammā attēloti tikai 49 kursi – tie, kuros bija ievērojams nokārtojumu studentu skaits (sekmīgo atzīmju summa – 77 vai vairāk). Neiekļauto 44 kursu saraksts:
DatI036,DatZ1039,DatZ1081,DatZ1091,DatZ2020,DatZ2022,DatZ2024,DatZ2033,DatZ2034,DatZ2035,DatZ2063,DatZ2076,DatZ2077,DatZ2088,DatZ3025,DatZ3029,DatZ3049,DatZ3056,DatZ3057,DatZ3058,DatZ3065,DatZ3068,DatZ3070,DatZ3072,DatZ3074,DatZ3126,DatZ3151,DatZ4002,DatZ4019,DatZ4020,DatZ4022,DatZ4023,DatZ4026,DatZ4027,DatZ4028,DatZ4035,DatZP038,Ekon1013,Fizi3021,Mate1005,Mate1170,Mate2006,Mate3028,Mate4005

Ideja

radās pēc bakalaura programmas studentu protestiem: ar matemātiskās loģikas kursu "kaut kas neesot kārtībā", to ir grūti nokārtot un saņemtās atzīmes ir ļoti zemas.

Vai saņemto atzīmju ziņā šis kurss tiešām ir unikāls? Salīdzināsim studentu atzīmes ar citiem kursiem...

Algoritms. Ņemam divus kursus, ko nokārtojuši N (vismaz 10) kopīgi studenti. Saskaitām tos K studentus, kuru atzīmes abosursos *atšķiras ne vairāk kā par 1*. Ja $\frac{K}{N} \geq 0,80$, tad secinām, ka kursi ir "vismaz 80% līdzīgi".

Izrādās, ka matemātiskā loģika ir vismaz 80% līdzīga 4 citu pasniedzēju kursiem. Apvairojums kanibālismā tātad nav korekts...

Vispārīgājs

Pēc tam radās doma šādu salīdzināšanu izveikt visiem bakalaura programmas kursiem, un rezultātus attēlot grafiski: kursi būtu grafa virsotnes, bet līdzīgos kursus savienotu grafa šķautnes. Un tad "uzlaist" diagrammai klasteru meklētāju.

Tā mēs arī izdarījām ar tiem pašiem 7.semestra 89 studentu atzīmju datiem. **Protams, pamēģinājām ne tikai 80% līdzību, bet arī citus procentu skaitļus.**

Tehnoloģija

1. Atzīmju datus mēs saņēmām no LUIS, tabulu izskatā: S rindīņas atbilst studentiem, K kolonas – kursiem. Kurss ar pārāk mazu atzīmju skaitu atmetām (tā tika atņemta aptuveni puse kursu). Tad ar Excel *macro* palīdzību ieguvām KxK tabulu ar kursu līdzības procentiem.

2. Šos datus apstrādājām ar rīku *Smart Network Analyzer*, kas noformēts kā *SNA.jar*, un kura autori ir mūsu kolēģi Valdis Krebs, Viesturs Zariņš, Pēteris Ručevskis, Kārlis Freivalds un Mārtiņš Opmanis. Klasterizācijas moduli ir uzrakstījis Rihards Opmanis saziņā ar Pauli Ķikustu.

Rīks prot grafiski un gudri attēlot datus, kas sastāv no objektiem (ar īpašībām) un saitēm (ar īpašībām) objektu starpā.

Ko varējām sagaidīt a priori?

Pirmā galējība: ja kursu līdzības prasība būs 0%, tad diagrammā ik divi kursi tiks savienoti, t.i. mēs iegūsim pilnu grafu un vienu lielu klasteri.

Otrā galējība: ja kursu līdzības prasība būs 100%, tad diagrammā tikai reti kursi tiks savienoti, t.i. mēs iegūsim gandrīz tukšu grafu un aptuveni "cik kursu, tik klasteru".

Kas notiks šīm galējībām pa vidu – ja kursu līdzības prasību pakāpeniski palielināsim no 0 līdz 100? Protams, notiks pakāpeniska dalīšanās arvien mazākos klasteros.

Bet kāds būs pirmais dalījums – divos klasteros vai vairākos?

[Principā šis process ir līdzīgs vienai no datizāces metodēm – t.s. hierarhiskajai klasterizācijai, kuras rezultāts ir klasteru koks.]

Kas mums sanāca?

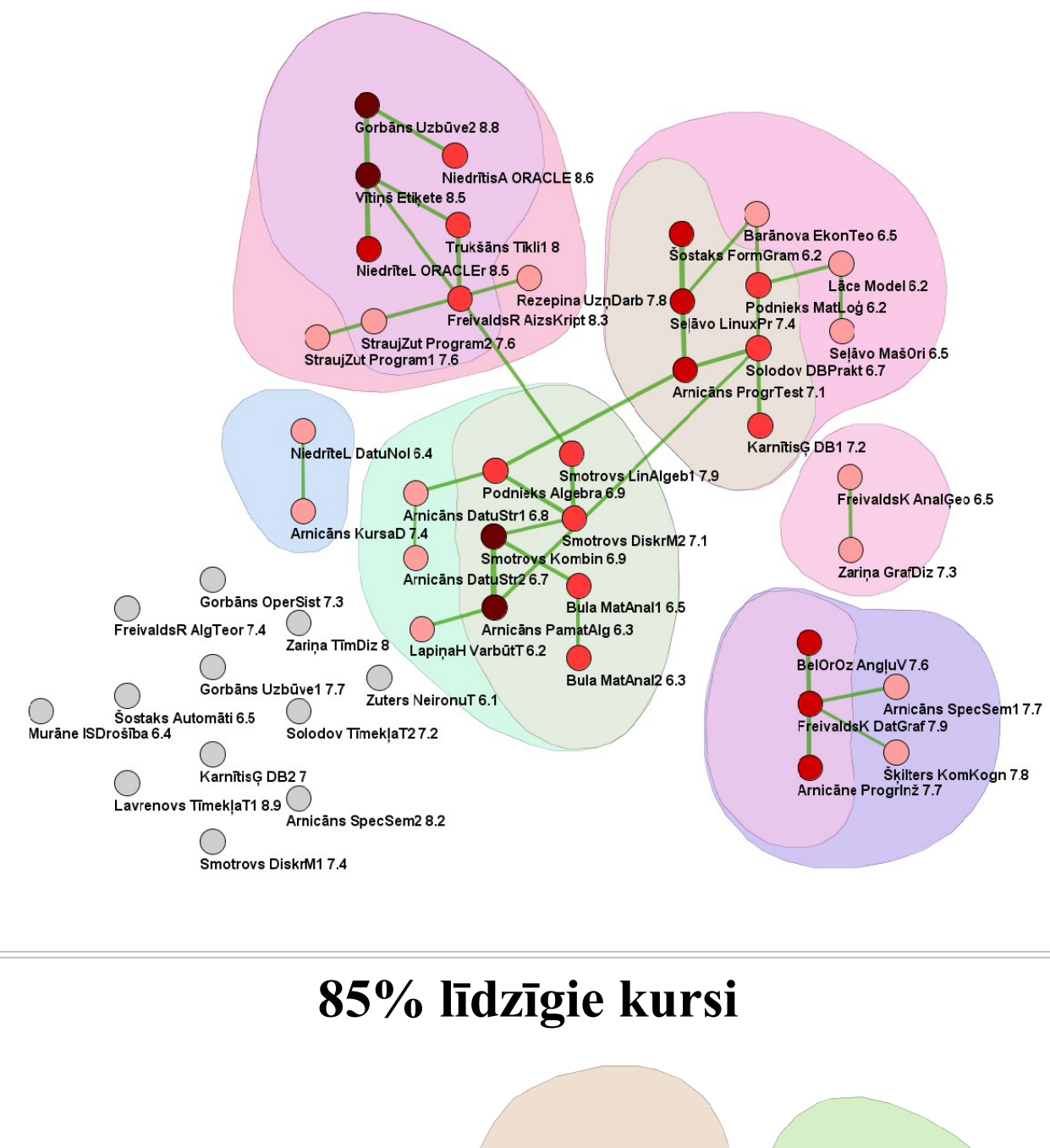
DZ bakalaura programma

Izmantojam datus par jau minētajiem 7.semestra 89 studentu atzīmēm dažus gadus atpakaļ.

Un izmēģinām dažādas kursu līdzības pakāpes.

73% līdzīgie kursi

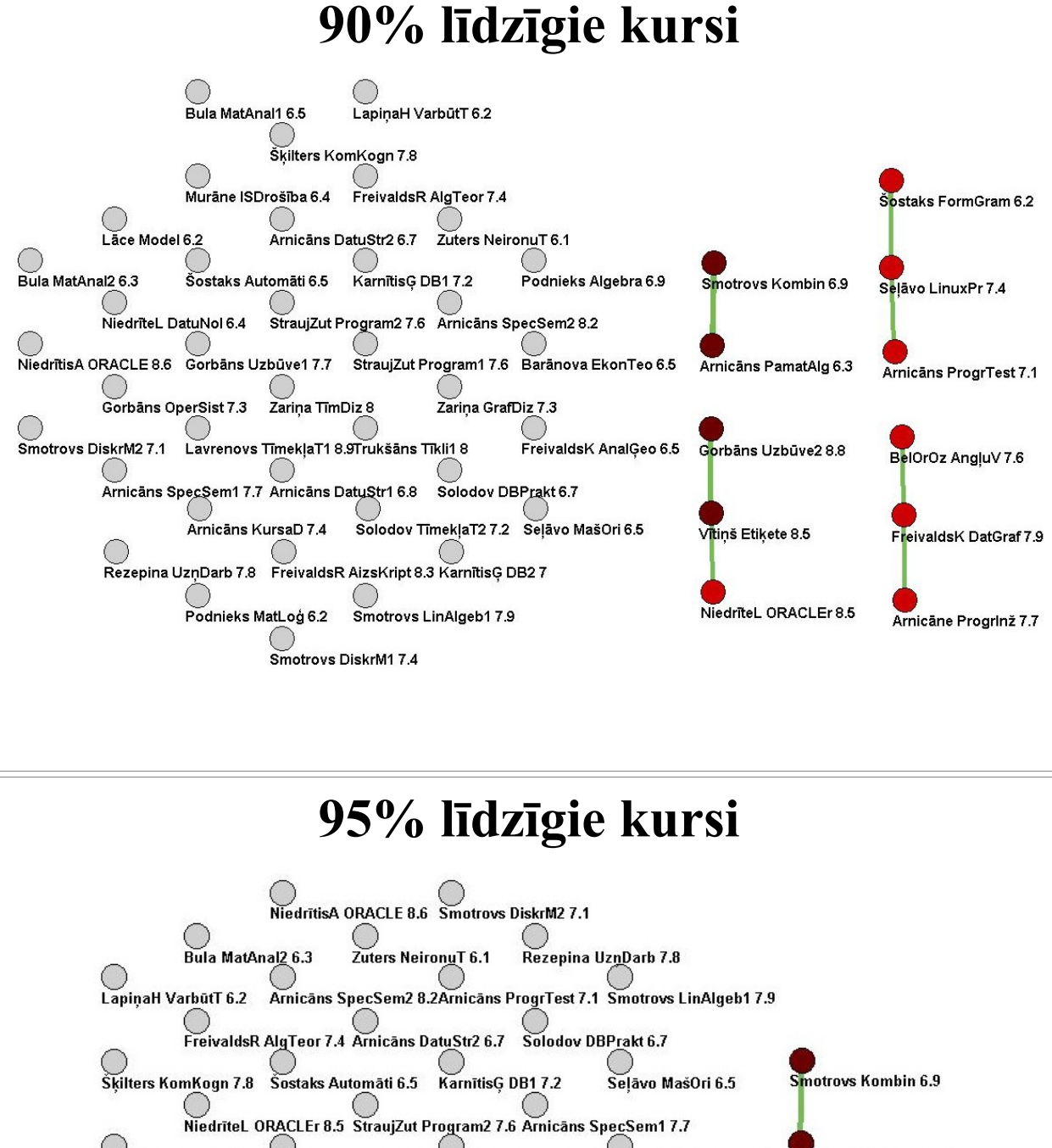
Kā redzam, pat pie 73% līdzības gandrīz visi kursi ir vienā lielā klasterī. Tikai daži kursi no kopējā klastera atdalās (varam padomāt, kāpēc...)



Galvenais secinājums: nekādas īpašas "atzīmju likšanas problēmas" DF nav. Gandrīz katrs kurss ir 73% līdzīgs ar vairākiem citiem kursiem.

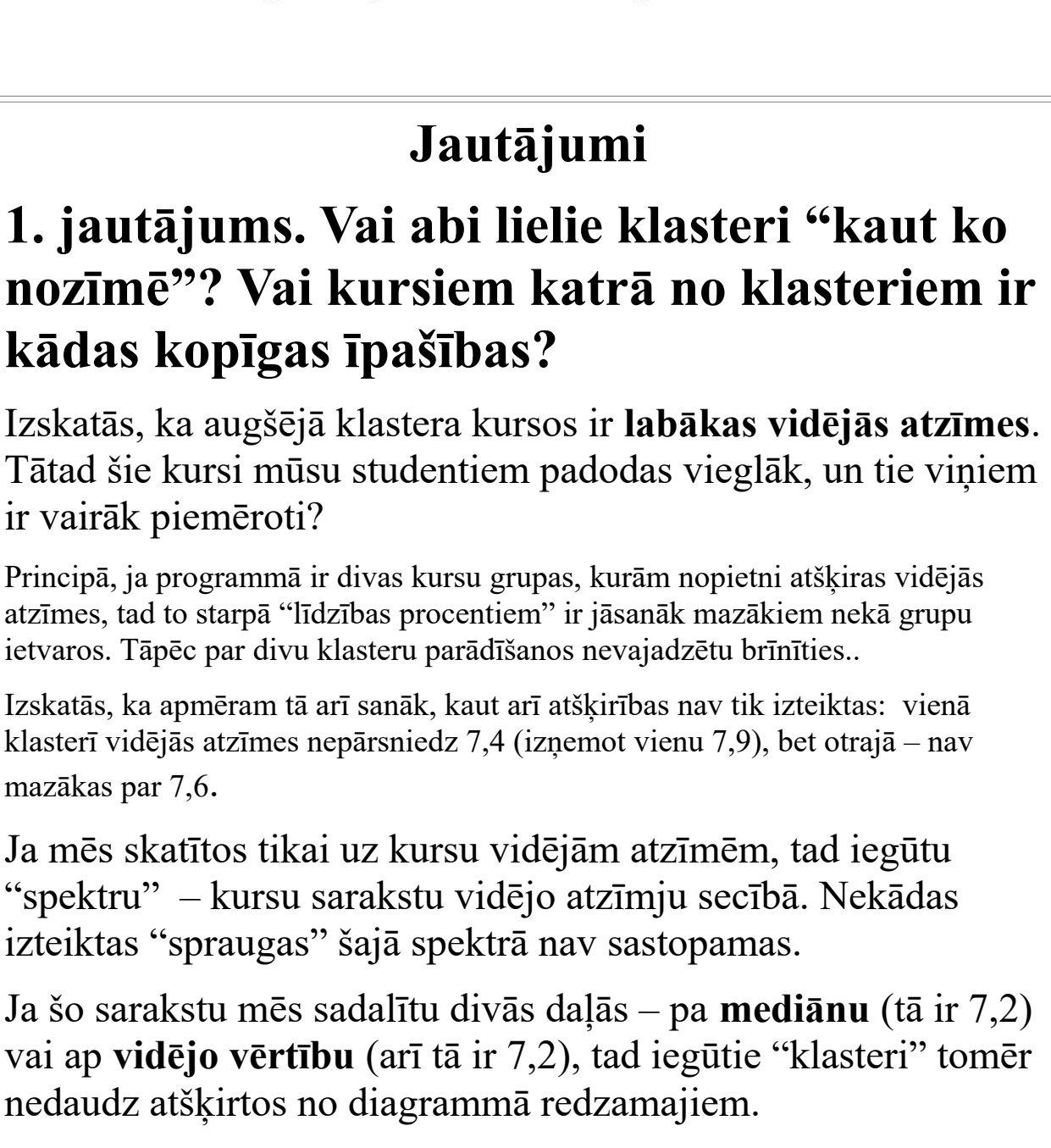
80% līdzīgie kursi

Te izdalījies divi lieli klasteri. Bet ir arī ievērojams skaits kursu ārpus tiem.

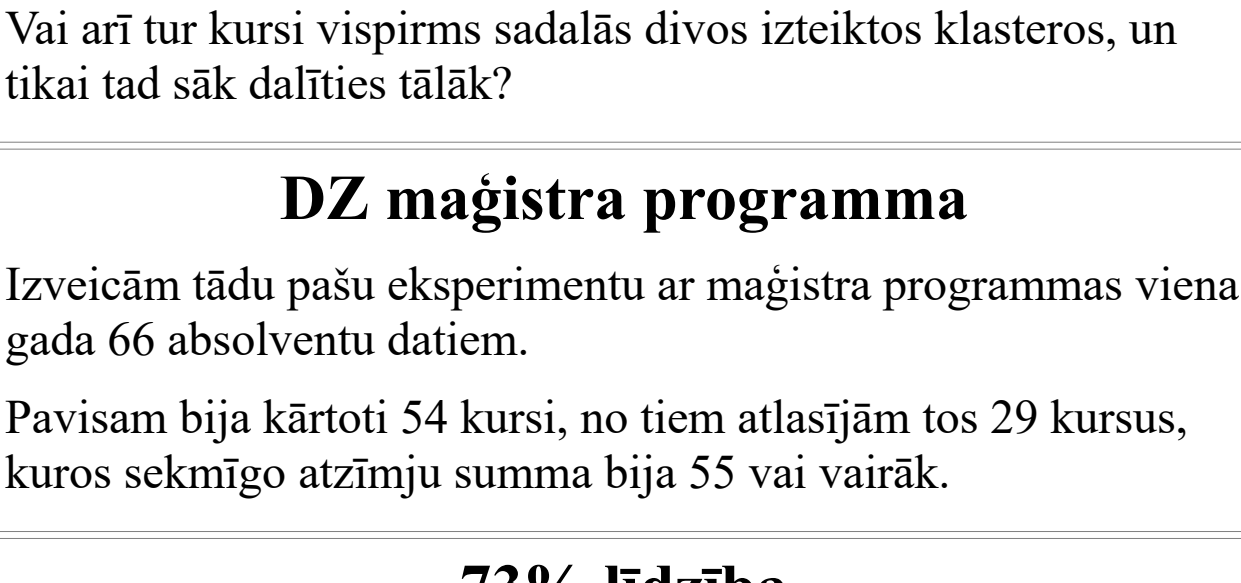


83% līdzīgie kursi

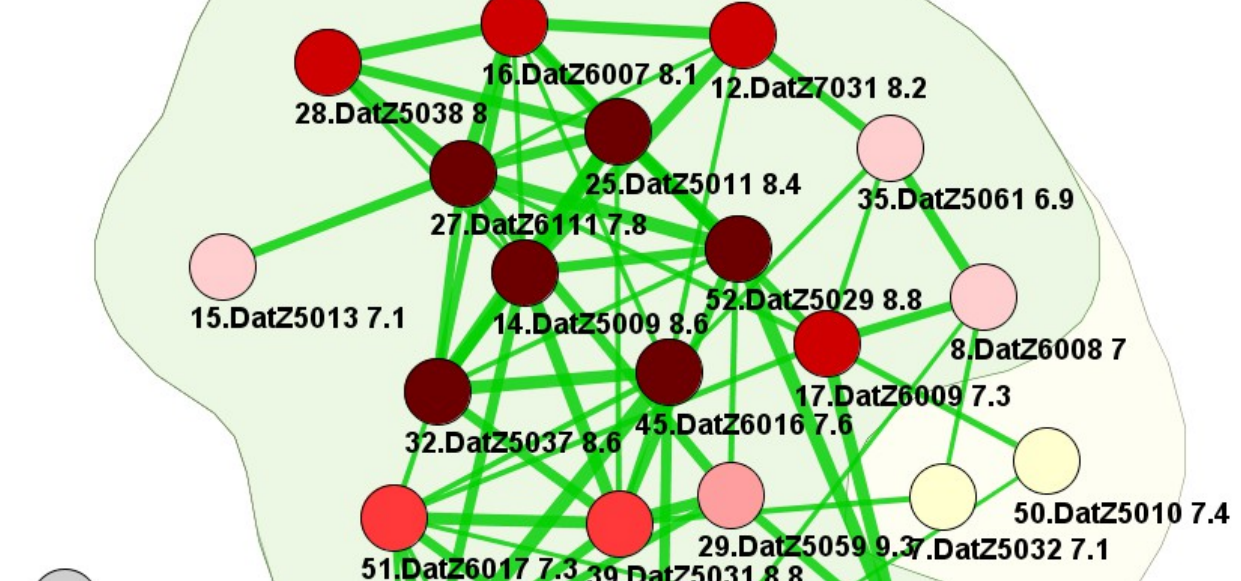
Te "nopietnais" klasteris ir sadalījies divos, un vēl vairāk ir "maļā stāvošo" kursu:



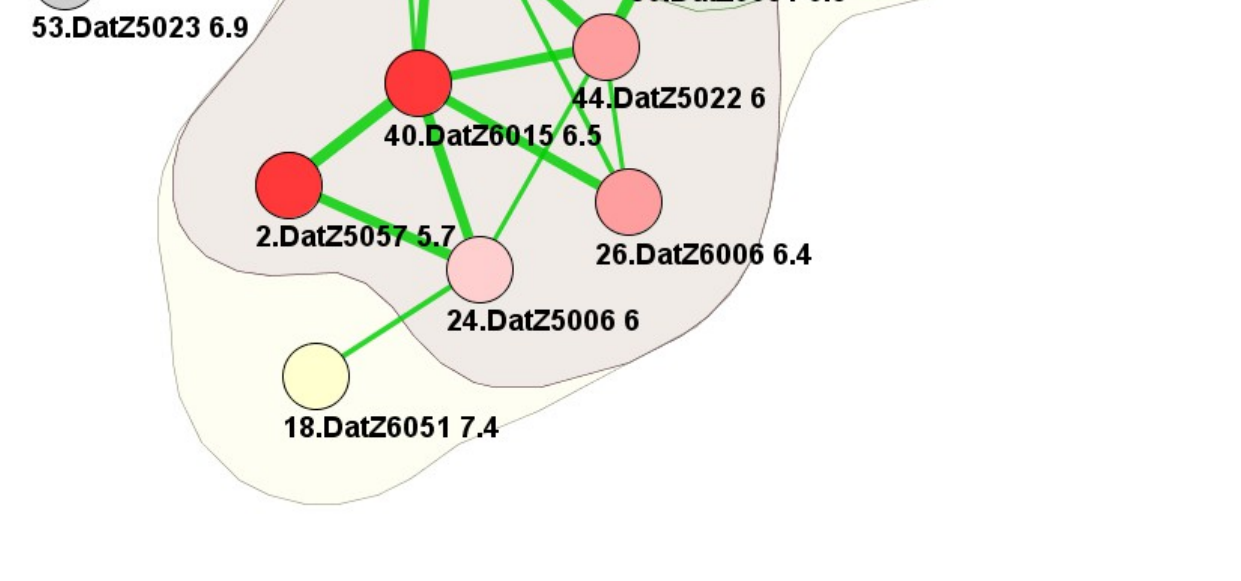
85% līdzīgie kursi



90% līdzīgie kursi



95% līdzīgie kursi



Jautājumi

1. jautājums. Vai abi lieli klasteri "kaut ko nozīmē"? Vai kursiem katrā no klasteriem ir kādas kopīgas īpašības?
Izskatās, ka augšējā klastera kurss ir labākas vidējās atzīmes. Tātad šie kursi mūsu studentiem padodas vieglāk, un tie viņiem ir vairāk piemēroti?
Principā, ja programmā ir divas kursu grupas, kurām nopietni atšķiras vidējās atzīmes, tad starpā "līdzības procentiem" ir jāsanāk mazākiem nekā grupu ietvaros. Tāpēc par divu klasteru parādīšanos nevajadzētu brīnīties..
Izskatās, ka apmēram tā arī sanāk, kaut arī atšķirības nav tik izteiktas: vienā klasterī vidējās atzīmes nepārsniedz 7,4 (izņemot vienu 7,9), bet otrajā – nav mazākas par 7,6.

Ja mēs skatītos tikai uz kursu vidējām atzīmēm, tad iegūtu "spektru" – kursu sarakstu vidējo atzīmju secībā. Nekādas izteiktas "spraugas" šajā spektrā nav sastopamas.

Ja šo sarakstu mēs sadalītu divās daļās – pa **mediānu** (tā ir 7,2) vai ap **vidējo vērtību** (arī tā ir 7,2), tad iegūti "klasteri" tomēr nedaudz atšķirtos no diagrammā redzamajiem.

2.jautājums. Vai citās programmās notiek kaut kas līdzīgs?
Vai arī tur kursi vispirms sadalās divos izteiktos klasteros, un tikai tad sāk dalīties tālāk?

DZ maģistra programma

Izveicām tādu pašu eksperimentu ar maģistra programmas viena gada 66 absolventu datiem.

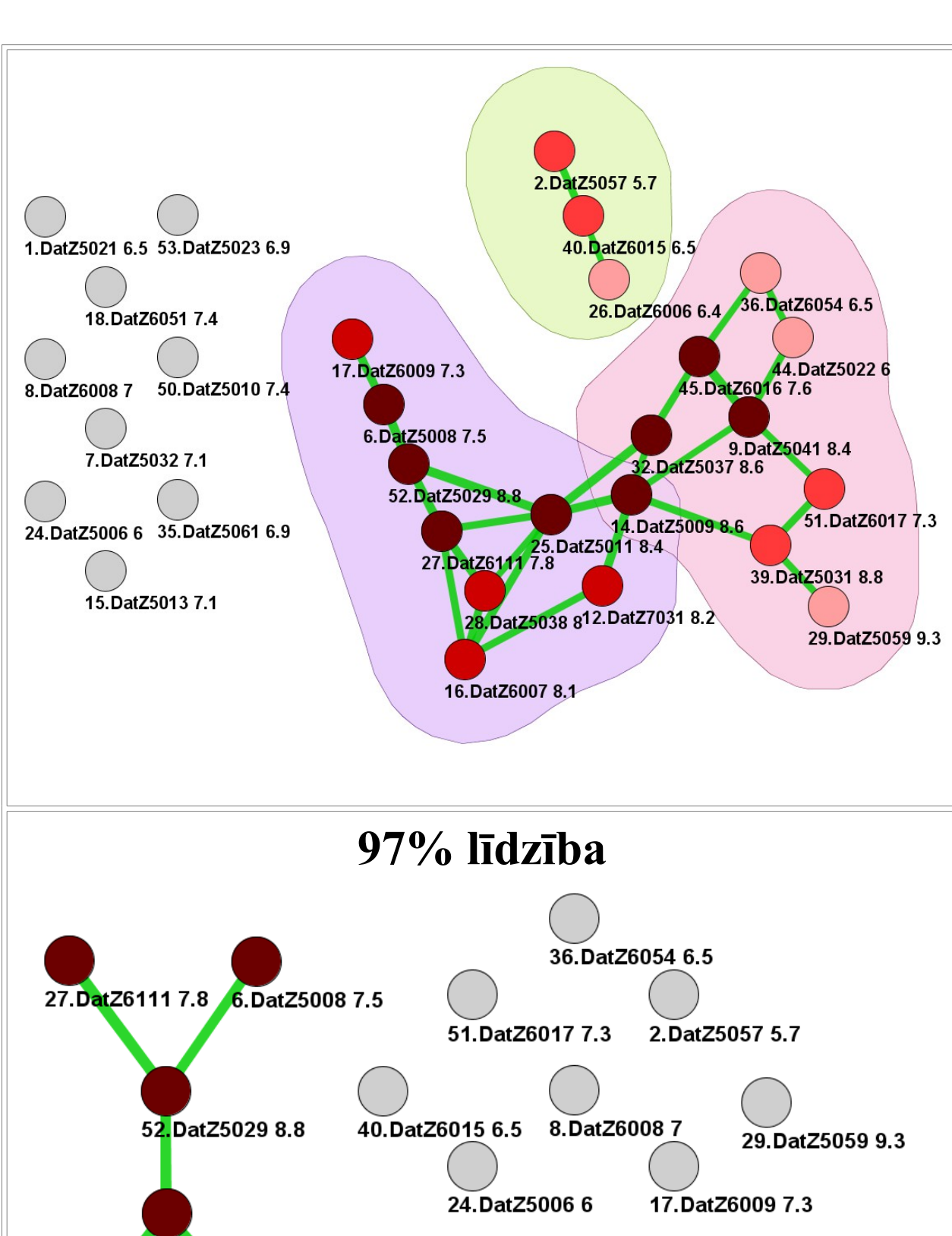
Pavisam bija kārtoti 54 kursi, no tiem atlasījām tos 29 kursus, kuros sekmīgo atzīmju summa bija 55 vai vairāk.

73% līdzība

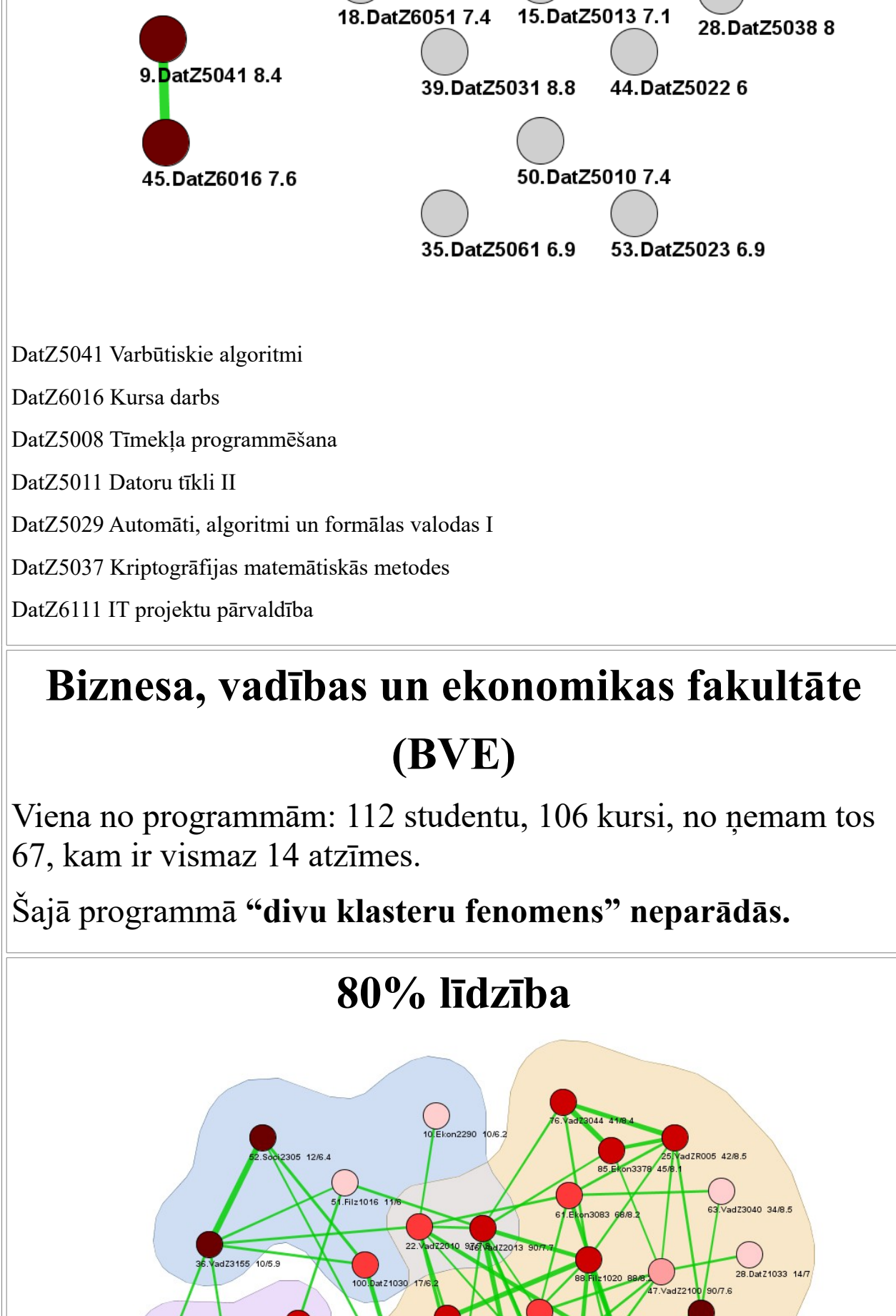
80% līdzība

85% līdzība

DatZ5022 Zināšanu inženierija
DatZ6006 UML
DatZ6015 Lietiskā kriptogrāfija
DatZ5057 Datu apstrādes sistēmas
DatZ5006 Ātrie algoritmi



97% līdzība



DatZ5041 Vārbutiskie algoritmi
 DatZ6016 Kursa darbs
 DatZ5008 Tīmekļa programmēšana
 DatZ5011 Datoru tīkli II
 DatZ5029 Automāti, algoritmi un formālas valodas I
 DatZ5037 Kriptogrāfijas matemātiskās metodes
 DatZ6111 IT projektu pārvaldība

Biznesa, vadības un ekonomikas fakultāte (BVE)

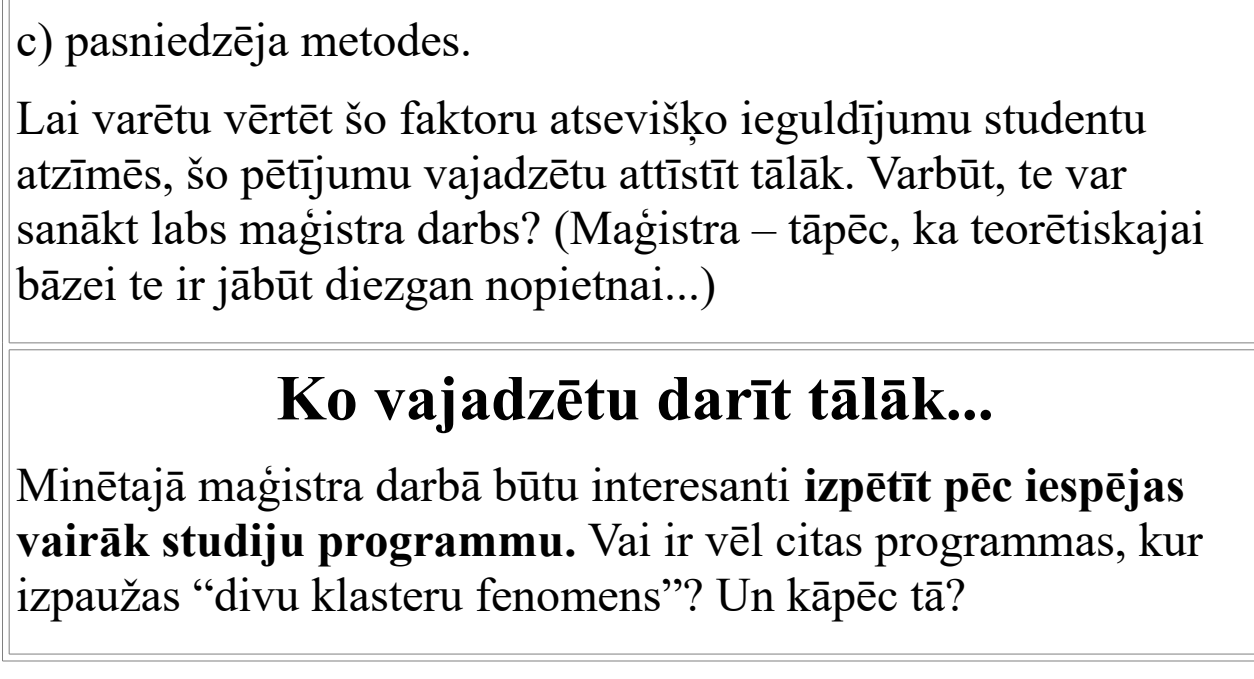
Viena no programmām: 112 studentu, 106 kursi, no ņemam tos 67, kam ir vismaz 14 atzīmes.

Šajā programmā “divu klasteru fenomens” neparādās.

80% līdzība



85% līdzība



Ko nevajadzētu darīt...

Nevajadzētu darīt to, kas anonsēts referēta virsrakstā: nevajadzētu “šķelt kolektīvu”, koncentrējot uzmanību uz DF pasniedzēju it kā sadalīšanos divos klasteros – “liberāļos” un “kanibālos”.

- Katra atzīme rodas **triju faktoru** kopdarbībā:
- kursa saturs (cik tas objektīvi ir grūts vai viegls);
 - studenta priekšzināšanas un spējas (zināms, ka ar to mums “ir problēmas”);
 - pasniedzēja metodes.

Lai varētu vērtēt šo faktoru atsevišķo ieguldījumu studentu atzīmēs, šo pētījumu vajadzētu attīstīt tālāk. Vārbut, te var sanākt labs maģistra darbs? (Maģistra – tāpēc, ka teorētiskajai bāzei te ir jābūt diezgan nopietnai...)

Ko vajadzētu darīt tālāk...

Minētajā maģistra darbā būtu interesanti **izpētīt pēc iespējas vairāk studiju programmu**. Vai ir vēl citas programmas, kur izpaužas “divu klasteru fenomens”? Un kāpēc tā?