

Indhold

Figurer	ii
Tabeller	ii
1 METODE OM NETVÆRKSANALYSE OG MONECA	1
1.1 Grundlæggende begreber og datastruktur i social netværksanalyse	2
1.2 Relativ risiko og styrken af forbindelser	3
1.3 Segmentering og klyngeanalyse	6
1.4 Mål for vurdering af segmentkvalitet i Moneca	8
2 ANALYSE	9
2.1 Hovedkort: ledighed baseret på SOCIO/Socstil	10
2.2 Bikort: DST's definition af arbejdsløse som nettoledige	20
2.3 Bikort: vores definition af arbejdsløshed, minimum seks uger	21
2.4 Bikort: Alle beskæftigede på arbejdsmarkedet (ingen arbejdsløse)	22
3 Litteraturliste	23
Primær litteraturliste	23
Bilag A FIGURER, STØRRE UDGAVER	24

Figurer

1.1	Eksempel på netværk	7
2.1	Hovedkort, SOCSTIL & SOCIO, 1996-2009. Se større version i appendiks A.5. .	10
2.2	Hovedkort, farvelagt efter andel af kvinder. Se større version i appendiks A.2.	12
2.3	Klynge 3.1	14
2.4	Klynge 2.29	14
2.5	Klynge 3.13	15
2.6	Klynge 3.5	15
2.7	Hovedkort, SOCSTIL & SOCIO, 1996-2009	16
2.8	Hovedkort, SOCSTIL & SOCIO, 1996-2009	17
2.9	Hovedkort, SOCSTIL & SOCIO, 1996-2009	18
2.10	Hovedkort, SOCSTIL & SOCIO, 1996-2009	19
A.1	Hovedkort: Vores definition af arbejdsløse, minimum et halvt år	25
A.2	Hovedkort: farvelagt efter andel af kvinder	26
A.3	Bikort: DST's definition af arbejdsløse som nettoledige	27
A.4	Bikort: vores definition af arbejdsløshed, minimum seks uger	28
A.5	Bikort: Alle beskæftigede på arbejdsmarkedet (ingen arbejdsløse)	29

Tabeller

1.1	Eksempel på en adjacency matrice	2
1.2	t	7

1 METODE OM NETVÆRKSANALYSE OG MONECA

I dette speciale tilgår vi arbejdsmarkedet som et netværk af forskellige arbejdsstillinger. I netværket bevæger individer sig mellem forskellige typer af arbejdsstillinger. Det sker når en person går fra at være beskæftiget i en arbejdsstilling til at være beskæftiget i en anden arbejdsstilling efter en mellemliggende periode med ledighed eller uden beskæftigelse. Arbejdsstillingerne tager form som noder i netværket, og personernes bevægelser mellem forskellige arbejdsstillinger er det som frembringer i netværket. Formålet med at anskue arbejdsmarkedet som et netværk er at kortlægge beskæftigelsesmønstre på arbejdsmarkedet for at se, hvilke arbejdsstillinger de ledige bevæger sig imellem, og hvilke arbejdsstillinger de ikke bevæger sig imellem. Vi har tidligere beskrevet arbejdssegmenteringsteorier kort, og vi vil her gå i dybden med vores centrale metodiske værktøj, social netværksanalyse, i en specifik udgave, som Toubøl og Larsen har udviklet specifikt til analyser af mobilitet. Denne metode har de kaldet Moneca - *“Mobility Network Clustering Algorithm”*.

Dette afsnit uddyber og forklarer de centrale antagelser og beregninger i Toubøl og Larsens to artikler om deres nyopfundne metode, samt opridser fundamentale begreber i netværksanalysen efterhånden som de introduceres i forbindelse med gennemgangen af Moneca. Da der er tale om en ny anvendelse af social netværksanalyse, er der ikke meget andet litteratur at benytte sig af end de to forfatters to artikler, da vi er de første udover omtalte skabere af metoden, der benytter sig af den. Et delmål for dette speciale er derfor at undersøge, hvad Moneca-algoritmen er i stand til, samt være opmærksom på dens eventuelle begrænsninger og mulige fejlbehæftninger. Det er derfor et delmål for dette speciale at evaluere hvorledes Moneca fungerer i relation til videnskabelige mål for reliabilitet og validitet. Dette vil løbende blive diskuteret i specialet, men først skal algoritmens grundlæggende funktioner beskrives.

1.1 Grundlæggende begreber og datastruktur i social netværksanalyse

Moneca er en overordnet set en kvantitativ, deskriptiv metode, hvis formål er:

1. at vise tilstedeværelsen, fraværet og styrken af relationer mellem forskellige grundkategorier af interesse.
2. at benytte et sæt kriterier baseret på centrale mål indenfor social netværksanalyse til at slå disse grundkategorier sammen i større grupper, herfra kaldet *segmenter*. Deraf navnet Moneca: *Mobility Network Clustering Algorithm*.

Grundkategorierne bestemmes ud fra det empiriske formål. Toubøl og Larsen har benyttet Moneca til at se på den sociale mobilitet mellem forskellige beskæftigelseskategorier indenfor hele det danske arbejdsmarked. Relationerne defineres som skift fra et arbejde indenfor én beskæftigelseskategori til arbejde indenfor en anden beskæftigelseskategori. Herefter benyttes Moneca-algoritmens 2. skridt til at undersøge, hvorledes denne datadrevne inddeling af arbejdsmarkedet i segmenter kan give ny indsigt i klasseinddelingen.

Det centrale er derfor hvilke grundkategorier, der benyttes, samt hvad der tæller som en relation mellem to grundkategorier. Toubøl og Larsen kommer selv med andre forslag til mulige anvendelser af Moneca, eksempelvis kortlægning af klassemobilitet gennem ægteskaber (Toubøl, Larsen og Strøby 2013, s. 27). Her ville grundkategorierne blive bestemt ud fra en given klasseinddeling af interesse, og relationerne mellem disse klasser defineres som ægteskaber. I netværksterminologi betegnes kategorierne som *noder*, mens relationerne mellem noderne betegnes som *edges*, eller på dansk: forbindelser.¹ I vores speciale ligger vi derfor konceptuelt helt lig Toubøl & Larsen, når vi definerer vores noder som beskæftigelseskategorier, og edges som skift mellem disse kategorier². Vi har det til fælles, at vi ser på en social mobilitetstabel, der viser skift fra beskæftigelseskategorierne i rækkerne til beskæftigelseskategorierne i kolonnerne. Denne er illustreret i tabel 1.1 ved hjælp af empirisk data fra vores mobilitetstabel.

Tabel 1.1: Eksempel på en adjacency matrice

	Til: Tandlæge	Til: Folkeskolelærer	Til: Pædagog	Til: Automekaniker
Fra: Tandlæge	264	0	0	0
Fra: Folkeskolelærer	0	6148	288	6
Fra: Pædagog	0	454	9308	0
Fra: Automekaniker	0	13	13	1861

I netværkstermer kaldes sådan en tabel for en *adjacency matrice*, da den har samme udfald i både rækker og kolonner, og datamatricen derfor er kvadratisk (Scott 2000,

¹Det er ofte en udfordring at oversætte de tekniske termer fra netværksanalyse til dansk. Hvor det ikke har været muligt at benytte et tilstrækkeligt unikt oversat dansk begreb, har vi derfor beholdt de engelske termer.

²Omend forskellen i populationsdefinition - ledige fremfor alle beskæftigede - har stor betydning for den konkrete operationalisering af begreberne, hvilket vil blive behandlet i de efterfølgende kapitler.

s. 55). I vores tilfælde skal det tolkes sådan, at rækkeudfaldene er de beskæftigelseskategorier, man kom *fra*, og kolonneudfaldene er de beskæftigelseskategorier man er gået *til*, efter en ledighedsperiode. Det ses at 264 personer på et tidspunkt har skiftet fra tandlæge til tandlæge i løbet af de 14 år, og ikke har skiftet til nogle af de andre tre erhverv. Skift fra samme beskæftigelse til samme beskæftigelse som fra tandlæge til tandlæge, noteres grundet den kvadratiske form langs diagonalen i en adjacency matrice, og denne har derfor en særlig status. Vi har at gøre med en *retningsbestemt, vægtet* adjacency matrice, hvilket er den mest komplicerede form for data i social netværksanalyse (Scott 2000, s. 61). Med vægtet skal forstås, at de enkelte celler ikke bare tilkendegiver en binær opdeling i tilstedeværelse eller fravær af en forbindelse, eksempelvis mellem folkeskolelærer og automekaniker, men at den også angiver en værdi for styrken af denne forbindelse. At den er retningsbestemt betyder at matricen ikke er symmetrisk langs diagonalen, da bestemmelsen af retningen sker ud fra hvorvidt man ser på et udfald rækkevis eller kolonnevis. Samt - i en vægtet matrice - kan styrken naturligvis også variere alt efter om man befinder sig over eller under diagonalen. I tilfældet automekaniker ses det, at hvis man i matricens nederste del kigger på Automekaniker, er relationen til tandlæge 0, det vil sige fraværende, mens den til folkeskolelærer og pædagog har en styrke på 13 i begge tilfælde. I selve diagonalen, det vil sige *den interne mobilitet*, er styrken forventeligt langt højere. Hvis man ser på automekaniker i matricens øverste del, det vil sige kolonnevis, er der et fravær af forbindelse til tandlæge, en styrke på 6 til folkeskolelærere, og et fravær af forbindelse til pædagoger. Det viser hvad der menes med en retningsbestemt, vægtet adjacency matrice: Styrken i mobiliteten fra automekaniker til folkeskolelærer er på 13, mens den fra pædagog til automekaniker er på 6, altså halvt så kraftig.

1.2 Relativ risiko og styrken af forbindelser

Indtil videre har vi kun talt om styrken af forbindelser som antallet af skift, altså en absolut enhed. Det er imidlertid ikke særlig retvisende, da styrken af forbindelsen dermed ikke står relativt til størrelsen af den kategori, den udspringer af. I 1996 var der eksempelvis 5.207 tandlæger, mens der var 53.676 beskæftigede pædagoger, altså omtrent ti gange så mange. Et skift fra tandlæge til en anden profession bør derfor vægtes højere end et skift fra pædagog til en anden profession. Her kommer konceptet relativ risiko ind i billedet, som en ratio mellem to proportioner (Agresti og Finlay 1997, s. 244, 271):

$$\frac{\pi_A}{\pi_B} \quad (1.1)$$

Den relative risiko (RR) fortæller os hvad chancen er for at begivenhed B sker, relativt til begivenhed A. Begivenhed A kender vi. Det er den simultane sandsynlighed for udfald x og udfald y i de to stokastiske variable I og J () (Malchow-Møller og Würtz 2003, s. 41), givet ved sandsynlighedsfunktionen:

$$f(i, j) = P(I = i, J = j) \quad (1.2)$$

I eksemplet fra tabel 1.1 er den simultane sandsynlighed for at have skiftet fra folkeskolelærer til pædagog: $\frac{45}{45} = 0,45 = \pi_A$. Begivenhed B er til vores formål³ karakteriseret som *den forventede værdi*. Det er derfor en helt igennem teoretisk størrelse - dens værdi er udelukkende bestemt ud fra teoretiske forventninger til hvad chancen *bør* være for at begivenheden indtræffer.

Toubøl & Larsen foreslår at det forventede udfald π_B bør være uafhængighed mellem $f(I)$ og $f(J)$, hvilket betyder at de marginale sandsynligheder for I og J giver den simultane sandsynlighed (Malchow-Møller og Würtz 2003, s. 44):

$$f(i, j) = f_I(i) * f_J(j) \quad (1.3)$$

Hvis vi ser arbejdsmarkedet som et frit marked, er der således tale om den perfekte markedstilstand; alle søger til og fra jobs som man bør forvente ud fra fordelingen af de givne jobs, uden at være betinget af hvilket job man kommer fra eller går til (ibid., s. 43). Lad π_{ijB} være udtrykket for den simultane sandsynlighed under uafhængighed for den i de række og den j de kolonne, det vi i starten benævnte som begivenhed B. π_{ijA} er den observerede simultane sandsynlighed vi rent faktisk observerer, det vil sige $f(i, j)$. Den relative risiko, beregnet ud fra de observerede marginale sandsynligheder, er derved givet ved udtrykket

$$\frac{\pi_{ijA}}{\pi_{ijB}} = \frac{f(i, j)}{f_I(i) * f_J(j)} = RR_{(i, j)_{observeret}} \quad (1.4)$$

Det betyder at hvis den relative risiko for celle(i, j) er lig 1, er formel 1.3 sand, og dermed eksisterer der uafhængighed mellem job i og job j . Hvis RR er under 1, er sandsynligheden for jobskifte mellem de to jobtyper mindre end forventet givet uafhængighed, hvorimod en RR på over 1 betyder at det er mere sandsynligt end under uafhængighed. Toubøl & Larsen definerer derfor en forbindelse mellem to typer job som uafhængighed eller derover, altså $RR \geq 1$ (Toubøl, Larsen og Strøby 2013, s. 9). Man kan tolke dette som udtryk for, at hvis $RR < 1$, eksisterer der barrierer i tilgangen fra det ene job til det andet, hvorimod $RR \geq 1$ betyder at en sådan barriere ikke eksisterer.

Indtil videre har vi beregnet den simultane sandsynlighed under uafhængighed ud fra de marginale fordelinger af de simultane sandsynligheder observeret i krydstabellens celler. I tilfældet med mobilitet blandt ledige afspejler det dermed hvor mange mobile, der skulle være i hver celle, ud fra hvor mange mobile der i det hele taget er. Det kan af sociologiske årsager være en problematisk antagelse. Hvis et bestemt job ikke fordrer en særlig stor mobilitet i det hele taget, bør det tages med i beregningen af den forventede værdi, hvilket ikke sker ved at benytte de marginale sandsynligheder for de mobile - fordi det netop kun beregnes ud fra dem, der rent faktisk *er* mobile. Hvis der blandt

³Indenfor medicinsk forskning vil relativ risiko typisk blive brugt sammen med oddsratio-værdier til at bestemme risici mellem en gruppe patienter tildelt en ny medicin samt en kontrolgruppe. Relativ risiko bruges derfor typisk til at vurdere forskelle mellem to konkrete grupper. Toubøl og Larsen benytter det i stedet som det bruges indenfor hypotesetest som det kommer til udtryk i eksempelvis χ^2 -testen (Malchow-Møller og Würtz 2003, s. 205).

tandlæger eksempelvis er 5 % mobile, mens der blandt tjenere er 20 %, er det ikke hensigtsmæssigt at at tandlægenes 5 % tæller for 100 %, ligesom tjenernes 20 % mobile nu også tæller som 100 %. Det er ikke desto mindre konsekvensen af at benytte de observerede marginale sandsynligheder fra formel 1.5. Vi vil i stedet have at tjenernes mobilitet på 20 % skal tælle fire gange så meget som tandlægenes 5 % i beregningen af deres respektive π_{ij_B} . Toubøl & Larsen benytter derfor de marginale sandsynligheder, ikke fra den observerede fordeling, men fra den oprindelige fordeling blandt alle beskæftigede, hvorfra de mobile “er trukket fra”, for at blive i jargonen. Dermed bliver forventningen til indholdet i den enkelte celle, π_{ij_B} , udtrykt i form af af de marginale sandsynligheder for det samlede antal beskæftigede. Dermed forhindres det at at tandlægenes 5 % og tjenernes 20 % begge tæller for 100 %, og i stedet tages der udgangspunkt i forventninger, der er informeret af vores viden om den samlede fordeling blandt alle tandlæger og tjenere. Vi laver derfor følgende omskriving af formel 1.5, hvor variablene K og L står for de marginale sandsynligheder for henholdsvis række- og kolonnefordelingerne for alle beskæftigede. Det vil sige de fordelinger, som fordelingerne af variablene I og J er trukket fra K og L , og vi kan derfor opskrive følgende formel:

$$\frac{\pi_{ij_A}}{\pi_{ij_B}} = \frac{f(i, j)}{f_K(k) * f_L(l)} = RR_{(i, j)_{population}} \quad (1.5)$$

Der er yderligere en fordel ved at benytte de marginale sandsynligheder for populationen, og det er at man kan tage højde for strukturelle ændringer i beskæftigelse over tid. Det vil sige, hvis givne fagkategorier skrumper eller vokser, vil π_{ij_B} i den modificerede form, baseret på $f_K(k) * f_L(l)$, tage højde for dette. For at forstå det skal man huske på, at selvom K og L er samme variabel, er der forskel i fordelingen hos de to, da de er forskudt med et enkelt år. I vores tilfælde er række-variablen udtryk for tidsperioden 1996 til 2008, da perioden, vi beskriver, strækker sig fra 1996 til 2009. Man kan derfor ikke inkluderes med et “start-job” i 2009, og det omvendte gælder for kolonnevariablen: Man kan ikke ende i et job i 1996, da vi ikke ved hvilket job man kom fra, i 1995. Det betyder at de to marginale sandsynligheder for henholdsvis K og L er forskudt med et enkelt år, og ændringer mellem disse to år medregnes i forventningerne til den enkelte celle π_{ij_B} .

Vi har hermed redegjort for principperne bag styrken af forbindelser i den retningsbestemte, vægtede mobilitetstabel, som Moneca er baseret på. Det centrale er, at styrken af forbindelser er et teoretisk funderet valg. π_{ij_B} bestemmes ud fra en betragtning om hvilket mønster, de observerede frekvenser i cellerne bør følge. Det vil sige, 1) fra hvilken population mener man at éns egen population er en subpopulation af, og 2) hvilket mønster i denne populations fordeling er det interessant at vurdere som udgangspunkt for om forbindelser eksisterer eller ej. Vi har her valgt, ligesom Toubøl & Larsen, at benytte de marginale sandsynligheder i populationen af alle beskæftigede til at vurdere hvilket mønster, der burde eksistere under uafhængighed. Det gør vi, da vi, jævnfør ovenstående diskussion af mobilitet blandt de hypotetiske tandlæger og tjenere, mener at den bedst viser det mønster der burde eksistere såfremt ingen mobilitetsbarrierer eksisterede, samt

tager højde for erhvervsstrukturelle forhold, der bestemmer mønstrene også hos vores subpopulation af ledige,⁴.

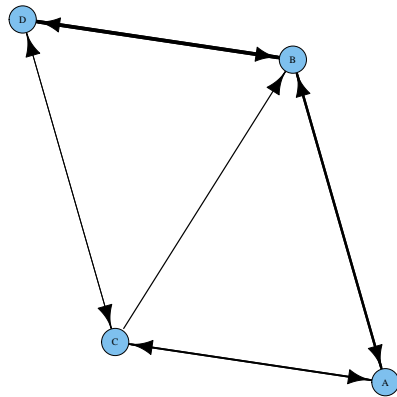
1.3 Segmentering og klyngeanalyse

At kortlægge forbindelserne mellem beskæftigelseskategorier, hvori mobiliteten er så stor at der ikke kan siges at eksisterer barrierer, baseret på definitionen af barrierer fra foregående afsnit, er i sig selv interessant. Moneca har, som beskrevet på side 1.1, også til formål at benytte centrale mål fra social netværksanalyse til at vurdere hvilke beskæftigelseskategorier, hvori styrken af forbindelserne antager en sådan karakter, at de kan siges at tilhøre et segment, eller klynge. Vi vil komme nærmere ind på de sociologiske argumenter for at lave denne opdeling i afsnit ??, mens dette afsnit primært omhandler metoden der benyttes til denne segmentering.

At lede efter sammenhængende subgrupper i et netværk, er en af de helt fundamentale opgaver i social netværksanalyse (Toubøl og Larsen 2015, s. 6), og en række mål indenfor eksisterer for at afgrænse og identificerer sådanne subgrupper, såsom *k-cores*, *m-cores*, *klike* og *klan*. Problemet med disse mål er at de beregnede klynger ikke er diskrete, men overlapper. Det er naturligvis ikke et problem i sig selv. Men for at reducere kompleksitet i komplicerede og og tæt forbundne netværk, så forskelle og ligheder fremtræder tydeligt, kan det være ønskeligt. En anden fordel ved Moneca er muligheden for at håndtere vægtede netværk, hvilket ovenstående koncepter traditionelt er dårlige til. Fordelene ved at benytte social netværksanalyse til at bestemme klynger, er at denne tager udgangspunkt i klyngedannelsen ud fra et princip om *forbindelse*, fremfor *enshed* (ibid., s. 6). Det betyder kort sagt om man ser på en gruppe relationelt i forhold til andre grupper, eller om nodernes tilhørsforhold bestemmes ud fra karakteristika, de har til fælles. Det er en central forskel i videnskabsteoretisk grundlag, hvilket vi kommer nærmere ind på i afsnit ?. Monecas algoritme baserer sig i det store og hele på netværksteoretiske mål klike. En klike er i netværkstermer defineret som en subgraf, hvor alle noder er forbundet med alle noder i subgrafen, og som samtidig ikke er indeholdt i andre klikker (Scott 2000, s. 112). Moneca kan ses som en algoritme, hvis formål er at afgøre de tvivlsspørgsmål, hvor noder er befinder sig i positioner, hvor de kunne tilhøre flere forskellige klikker, hvilket ofte er tilfældet i komplekse netværk med mange, men stærkt differentierede, forbindelser. Figur og tabel 2.1 er en illustration af et simpelt tilfælde af

⁴Vi kunne også have baseret vores forventninger til cellens indhold på eksempelvis mønstrene hos alle mobile, fremfor alle beskæftigede, ud fra en betragtning om at de forventede mønstre blandt ledige burde følge denne fordeling under uafhængighed. Det kunne vi dog ikke se nogle gode argumenter for. Ved at benytte denne alternative fordeling ville vi kun introducere en række usikkerheder omkring hvilke sociale strukturer man prøvede at beskrive. Man kan se det sådan at vi dermed ville benytte mønstrene hos mobile, der i forvejen er trukket fra populationen af alle beskæftigede. Det ville dermed introducere et tredje led af strukturelle fordelingsmekanismer, hvilket kunne være hensigtsmæssigt, såfremt vi mente at vores population af ledige rent faktisk var en subpopulation (af subpopulationen) alle mobile. Det har vi dog intet belæg for at mene, da populationen af alle mobile i langt de fleste tilfælde er kendetegnet ved at gå mere eller mindre direkte fra et job til et andet. Derfor virkede mønstret under uafhængighed blandt alle beskæftigede mere hensigtsmæssigt.

Figur 1.1: Eksempel på netværk



Tabel 1.2: Adjacency matrice til netværk

	A	B	C	D
A	-	-	-	-
B	4	-	-	-
C	3	2	-	-
D	0	6	1	-

et sådant tvivlsspørgsmål om tilhørsforhold. Moneca ville i dette tilfælde komme frem til at netværket i figur 2.1 består af klyngerne $[A|C]$ og $[B|D]$. Det sker ud fra følgende procedure:

1. Først lægges de to kraftigst forbundne noder sammen. Det vil her sige $[B|D]$, hvor styrken af forbindelsen er seks.
2. Derefter gør det den samme med de to næstmest forbundne noder, $[A|B]$, der har en styrke på 4. Eftersom $[B]$ allerede er en del af den foreløbige klynge $[B|D]$, spørger Moneca om det er muligt at indlemme $[A]$ i den allerede etablerede foreløbige klynge. Det kan ikke lade sig gøre, da $[A|B]$ ikke er forbundne.
3. Moneca går derfor videre til den tredje stærkeste forbindelse, $[A|C]$. Hverken $[A]$ eller $[C]$ er en del af en foreløbig klynge, og de lægges derfor sammen.
4. Den fjerde stærkeste forbindelse er $[B|C]$. $[B]$ og $[C]$ er allerede parret med henholdsvis $[D]$ og $[A]$, og Moneca beregner derfor om $[A|B|C|D]$ udgør en klike, altså alle er forbundne med hinanden. Da det ikke er tilfældet, stopper Moneca og har dermed etableret klyngerne $[A|C]$ og $[B|D]$.

Det vil sige at Moneca ikke etablerer nogen af de maksimale klike, $[A|B|C]$ eller $[B|C|D]$. Den semistrenge stopregel om klike-tilhørsforhold indenfor klynger betyder desuden, at man med en vis sindsro kan sige at en klynge rent faktisk *er* en samlet størrelse - da den kun kan dannes hvis alle noderne har forbindelse til hinanden, hvilket eksemplet illustrerer: De to maksimale klike etableres netop ikke, og hvis de gjorde, ville troværdigheden af grænsedragningen mellem klyngerne være langt mere tvivlsom.

For at opsummere ovenstående i mere generelle termer: Moneca starter med at slå noderne med de to mest intense forbindelser sammen til en foreløbig klynge, og går derefter videre til den næstmest intense forbindelse. Hvis en eller begge af noderne i de

efterfølgende forbindelser allerede har forbindelse til en tredje eller fjerde node, vurderer Moneca, om denne indgår i en klike med den allerede etablerede klynge. Det er her vigtigt at understrege, at denne vurdering ikke er baseret på styrken af forbindelserne, men udelukkende om der eksisterer en forbindelse⁵. Moneca fortsætter med denne procedure i prioriteret rækkefølge fra de mest intense forbindelser til de mindst intense, indtil alle noder er placeret i klike med andre noder, der endnu ikke er "optaget" af en mere intens forbindelse. Kriteriet for, om Moneca tillader at slå de tre noder sammen, er om de tilsammen former en klike, altså alle er forbundet til hinanden. Hvis de ikke er det, går den videre til den næstmest intense forbindelse, og fortsætter med at forbinde noder indtil der ikke længere kan etableres flere klike. Kriteriet om at allerede-etablerede klynger kun kan lægges sammen med nye noder, hvis disse indgår i en klike med alle medlemmer af klyngen, er den stop-regel, der gør at Moneca ikke bare ender med at etablere det redundante stykke information, at hvert enkelt komponent⁶ er en klynge i sig selv (Toubøl og Larsen 2015, s. 8).

Efter denne første segmentering er Moneca beregnet til at gentage proceduren indtil det ikke længere er muligt at skabe større segmenter, fordi den foromtalte klike-regel forhindrer det. Det er vigtigt at fremhæve, at Moneca i de efterfølgende klyngeinddelinger baseret på side 7 beskrevne procedure, *ikke* længere tager de oprindelige, interne forbindelser mellem grundkategorierne i betragtning, hvis disse er blevet lagt sammen med andre grundkategorier. I den efterfølgende klyngeinddeling vil de interne forbindelser mellem noderne på et lavere niveau ikke indgå i beregningerne i forbindelserne mellem de nyskabte noder. Det vil sige at klike-reglen kun tager højde for forbindelser til noder *på det niveau noderne befinder sig på, og ikke forbindelserne på de lavere niveauer*. Når man vurderer kvaliteten af klyngerne på de højere niveauer, er det derfor centralt at se på en række standardmål for klyngens interne forbindelser, for at vurdere rimeligheden af at vurdere klyngen som en samlet størrelse. Det er dette spørgsmål vi nu afslutter gennemgangen af Moneca med.

1.4 Mål for vurdering af segmentkvalitet i Moneca

To mål centrale Moneca: Densitet der viser

Moneca gør det dermed muligt at opdele netværk med mange og komplekse forbindelser mellem noder en på sådan vis, at klynger kan identificeres, hvor man med en vis rimelighed kan sige at noderne hører sammen. Dette

⁵Man kan forestille sig en fremtidig version af Moneca foretage en mere avanceret vurdering i disse tvivlsspørgsmål, hvori styrken af relationen kunne indgå som vurderingsgrundlag.

⁶Et komponent betyder en subgraf, hvor alle noder er forbundne gennem stier. Et netværk kan således bestå af flere komponenter, der per definition ikke er forbundne (ellers ville de være en del af samme komponent), samt noder uden forbindelse til andre noder (*isolates*) (Scott 2000, s. 100).

2 ANALYSE

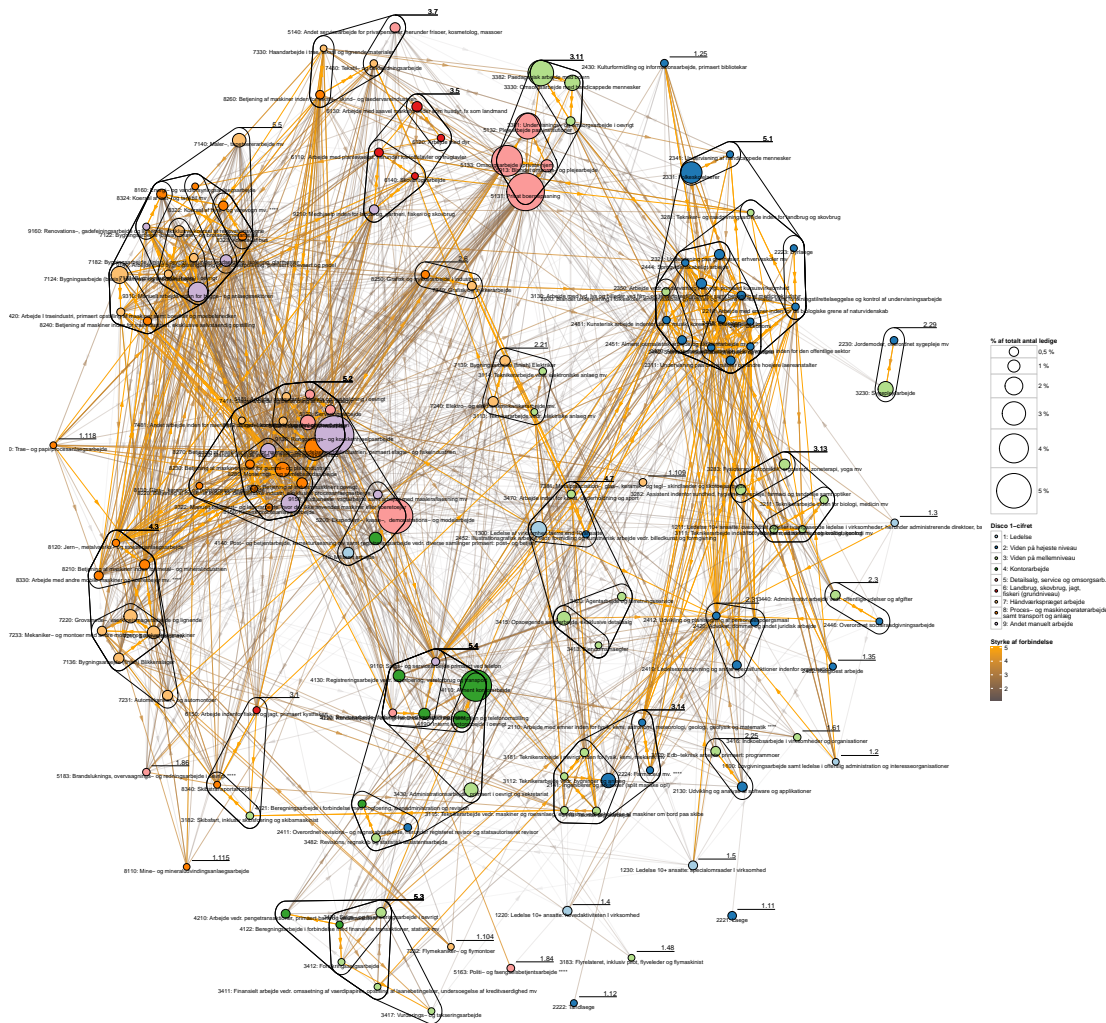
Konklusioner:

- Alle arbejdsløshedskort har en store segmenter sammenlignet med alle beskæftigede. Dette kan forklares med, at arbejdsløse søge arbejde længere væk fra deres fagområde end når beskæftigede skifter arbejde.
- De forskellige analyseudvalg indeholder forskellige grader af marginalisering. Personer som i det indeværende år vil generelt ligge tættere på inkluderet på arbejdsmarkedet end en person som har været arbejdsløs i halvandet år såvel som en nettoledig ligger tættere på at være inkluderet end en kontanthjælpsmodtager. Dette fremgår også af de gennemsnitslængden på arbejdsløshedsperioden se...

2.1 Hovedkort: ledighed baseret på SOCIO/Socstil

Kortlægningen af vores definition af arbejdsløse i minimum et halvt år giver os et kort baseret på 516.812 personer, som er gået fra beskæftigelse til arbejdsløshed til beskæftigelse, altså har foretaget 597.437 skift fra 1996 til 2009. Kortet slår bevægelser for alle årene sammen i én samlet struktur, som viser mobilitetsmønstrene i denne årrække.

Figur 2.1: Hovedkort, SOCSTIL & SOCIO, 1996-2009. Se større version i appendiks A.5.



At aflæse et Moneca-netværkskort

Kortet skal læses således, at nodernes størrelse er udtryk for andelen af personer i hver node. Den største node er 9130: Rengørings- og køkkenhjælpsarbejde, der udgør 5,7 % af alle beskæftigede, svarende til 21.065 mennesker, mens den mindste node er 8110: Mine- og mineraludvindingsanlægsarbejde, der udgør 0,02 % af noderne, svarende

til 87 personer, som beskrevet i DISCO-afsnittet (ikke vedlagt).

Farvelægningen af forbindelserne viser styrken af forbindelserne: desto mere orange, desto stærkere er en forbindelse. Pilene angiver retningen af forbindelsen, hvilken de større udgaver af kortene vedlagt i appendiks A gerne skulle synliggøre.

Farvelægningen af noderne er på dette kort sket ud fra disco-kategoriernes hovedgrupper. Disse kategorier giver et dobbelt indtryk af såvel en vis ensartethed i arbejdsfunktionen og uddannelsesniveau, samt ikke mindst et indtryk af den samfundsmæssigt konstituerede hierkarkisering af jobtyperne.

Det sidste væsentlige element for at kunne tolke kortene, er klyngeinddelingen. Klyngeinddelingen tegner også klyngerne på lavere niveau ind, men er kun navngivet på kortet efter klyngens højeste niveau. Den store klynge, der næsten udleukkende består af blå DISCO-2 kategorier i højre hjørne, er derfor nummeret 5.1, da dens øverste niveau også er det øverste niveau Moneca fandt i mobilitetsmatricen. Overfor dette står en række små "klynger" på niveau 1, hvilket betyder at de slet ikke er slået sammen med andre kategorier.

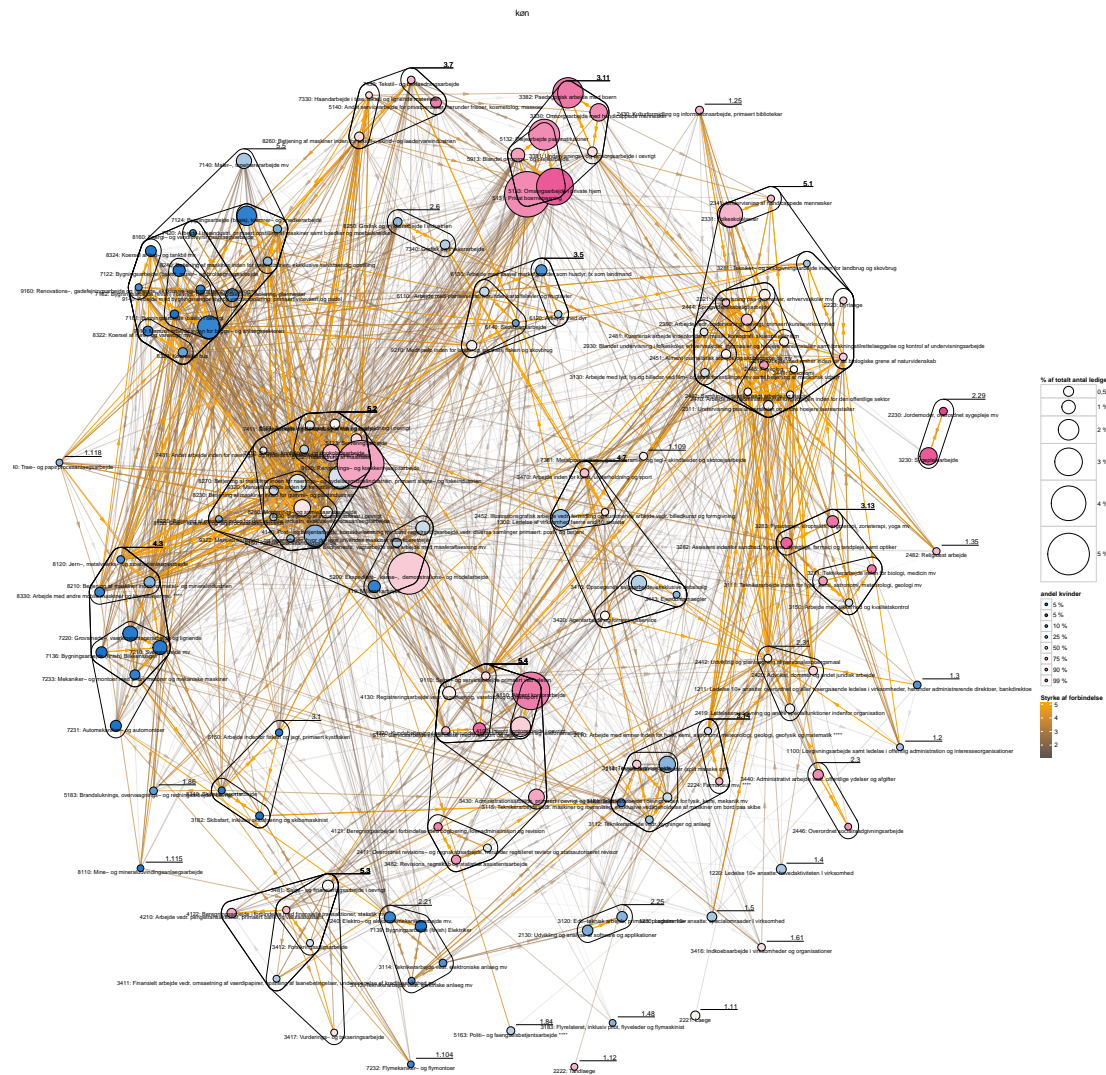
Ved første øjekast

Det første vi lægger mærke til er at et pænt antal klynger indeholder arbejdsfunktioner, der umiddelbart giver god mening er sammenlagt. Således kan vi se at klynge 5.1 indeholder langt de fleste af de lange videregående uddannelser, mens klynge 5.4 indeholder de fleste mennesker, der beskæftiger sig med kontorarbejde. I venstre hjørne kan vi se klynge 5.5 5.2, 4.3 og 3.7, der indeholder mange forskellige DISCOhovedkategorier, men i høj grad karakteriseret ved at være fysisk krævende arbejde indenfor service eller håndværk. Landbruget har sin helt egen klynge kladet 3.5.

Det er ikke overraskende, at der er en genkendelige systematik i den sociale mobilitet på denne måde. Det ville være alarmerende i forhold til metoden samt datakvalitetens pålidelighed, hvis dette ikke var tilfældet.

Før vi går i dybden med hovedkortet, vil vi vise et kort, der er farvelagt efter kønsfordelingen i hver beskæftigelseskategori. Det er centralt i netværksanalyse, at man viser korrelationer med relevante karakteristika, der ikke er anvendt til at definere netværket. Det giver os mulighed for at vurdere, om vores specifikation af netværket stemmer overens med en genkendelige social virkelighed, hvilket øger validiteten af netværksspecifikationen (Laumann, Marsden og Prensky 1983, s. 29). I et netværkskort baseret på Moneca-algoritmen, der jo benytter en klyngealgoritme til at foreslå sandsynlige grupperinger af netværket, virker det ydermere meget relevant at benytte disse eksterne karakteristika til at vurdere klyngernes validitet.

Figur 2.2: Hovedkort, farvelagt efter andel af kvinder. Se større version i appendiks A.2.



På kort 2.2 ser vi andelen af kvinder, således at en næsten lyserød farve angiver en andel af kvinder

gennemgang af segmenter i hovedkort

En nærmere granskning af segmentinddelingen vil give os nogle ideer om sammensætningen af kortet, så vi kan analysere kortet nærmere, såvel som være opmærksom på eventuelle svagheder i data og metoden.

To mål er centrale i gennemgangen af klyngerne: Den interne mobilitet og densiteten. Den interne mobilitet er andelen af skift, der sker indenfor beskæftigelseskategorien. Den giver derfor et mål for hvor uafhængig kategorien er af indsupplering fra andre jobtyper. Vi ser eksempelvis at læger og tandlæger ligger helt i top på den interne mobilitet, med en intern mobilitet på 94 % for begge kategorier. Det er en voldsomt høj grad af in-

tern mobilitet, især taget i betragtning af at alle disse skift sker efter en mellemliggende periode med arbejdsløshed. I en sociologisk tolkning passer det godt ind, at to fag der i så grad er karakteriseret som en profession i klassisk weberiansk forstand, er så ekstremt selvsupplerende. I kontrast hertil står 8140: Træ- og papirprocesanlægsarbejde, der har en intern mobilitet på 18 %, hvilket betyder at lidt under hver 5. skifter tilbage til job af denne type, der primært indeholder job som savskærer og finearbejde, og har omtrent 100 beskæftigede i gennemsnit pr. år i perioden¹

Densiteten er et netværksmål, der siger noget om hvor godt en graf er forbundet med de andre i et komponent eller en subgraf. Desto højere densitet, desto tættere er grafen på at være en "komplet" subgraf, det vil sige at alle noder er forbundne med alle andre noder. I så fald er densiteten lig 1, da alle mulige forbindelser er skabt. Omvendt vil en graf kun bestående af "isolates", det vil sige noder uden forbindelser til nogle andre noder, en sådan graf²- hvis man overhovedet kan kalde det det - vil have en densitet på nul.

I Moneca bruges densiteten som et mål for hvor godt integreret noderne i klyngen er med hinanden. Det er vigtigt, fordi vi husker på at segmentering over 1. sammenlægningsstadiet ikke indebærer at alle noderne fra niveauet under er forbundne med de noder der nu tages i betragtning i klikedannelsen. Derfor vil klynger på 2. niveau altid have en densitet på 1, da det er kravet for at Moneca overhovedet lægger dem sammen. Det gælder ikke efterfølgende, jævnfør metodeafsnittet. Det er ydermere vigtigt, da det, i modsætning til meget andet netværksanalyse, giver meget lidt teoretisk mening at forestille sig forbindelser med en path på længere end 1, til nøds 2. Vi har ikke at gøre med personer, eller bare organisationer, men beskæftigelses kategorier, så at tænke disse kategorier som noget meningsfuldt udover hvad de umiddelbart er forbundet med giver meget lidt teoretisk mening. Medmindre den interne mobilitet i klyngen er meget høj, hvor man med en vis rimelighed så kan forvente at vandringer foregår internt i klyngen. Ikke desto mindre skulle det gerne stå klart at densitet er et helt centralt mål.

gennemgang segmenter - ret ufærdigt

Følgende er en gennemgang, der starter med de klynger der har højest intern mobilitet og bevæger sig nedad.

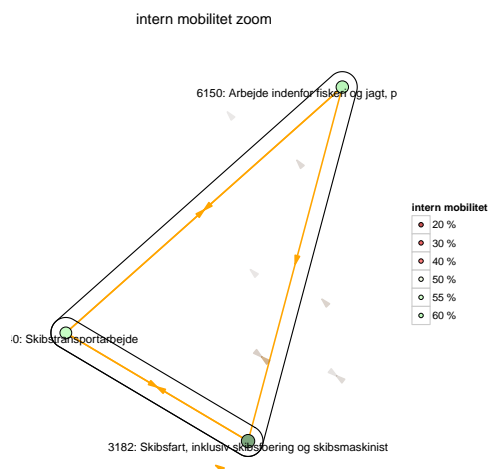
3.1, skibsfart og fiskeri.

Tydelig ensartethed i genstandsfelt, godtgør nok en feltbeskrivelse af en art. Meget stærke forbindelser, dog med den ene forskel, at 6150 går til 3181, men det går ikke den anden vej.

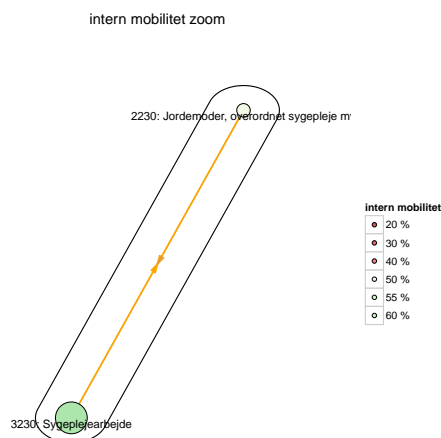
¹Til jens, skal stå et andet sted: eftersom skiftene er registreret over en 14-årig periode, så husk at det er hvor mange der i gennemsnit er beskæftiget med den type arbejde over de 14 år. Dvs antallet af beskæftigede divideret med 14.

²et netværks kaldes ofte en graf indenfor jargonen.

Figur 2.3: Klynge 3.1



Figur 2.4: Klynge 2.29



2.29: jordemoder og overordnet sygepleje mv samt sygeplejerske.

Vi kan se at disse to er sammen i en klynge for sig og ikke i den store omsorgsklynge (LA's kvindefængsel). Vi har at gøre med arbejde på disco-niveau 3 og 2, det vil sige at dette arbejde trods ensheden i genstandsfelt, differentierer sig ud på grund af sit højere færdighedsniveau i form af diplomer, samt muligvis udskilnen hvor dem der har jobbet ligger nærmere grundet den type færdigheder der kræves, i modsætning til kvindefængselklustren. udgør 2,6 % af alle beskæftigede.

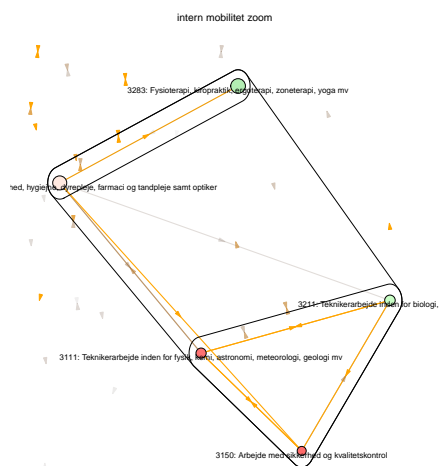
3.13

Interessant klynge. Selvom den interne mobilitet er høj er densiteten noget lavere. Stadig 60 %, hvilket er ok godt, men noget lavere end de allerbedste klynger hvor over 3/4 af noderne er forbundet med hinanden. Ved første øjekast indeholder den også grupper, der umiddelbart må forventes at indeholde meget forskellige genstandsfelter såvel som arbejdsfunktioner.

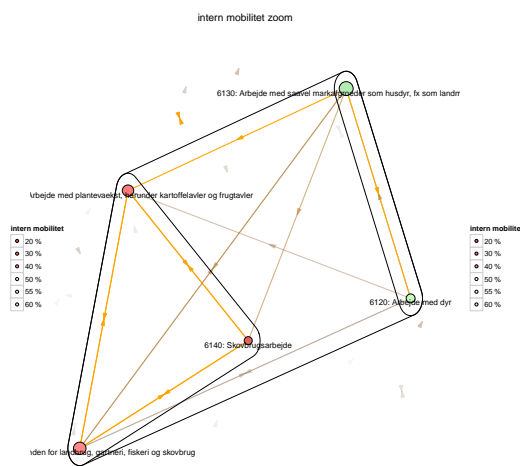
Ved at zoome ind kan vi se denne klynge er slået sammen gennem to klynger skabt på 2. niveau. Her ser vi teknisk betonet arbejde indenfor biologisk produktion, medicin og forskning, samt teknisk arbejde med kemi, fysik, astronomi og geologi. Disse to udgør en klynge på 2. niveau sammen med kvalitetskontrol af fastsatte standarder indenfor diverse produktionsfærdigheder, hvilket giver fin mening. Man må forvente at kompetencer opnået i teknikerarbejde indenfor kemi, fysik, biologi etc kan benyttes netop inden kvalitetskontrol og sikkerhed af eksempelvis kemiske standarder.

Den anden klynge på niveau to består af assistentarbejde indenfor en række sundhedsrelateret arbejde, samt fysioterapi, yoga, ergoterapi og række andre jobtyper, med fokus på behandling af menneskekroppen. Denne sammenlægning er ikke overraskende. Det overraskende sker i sammenlægningen af disse to klynger. Det ses at 3282 rent

Figur 2.5: Klynge 3.13



Figur 2.6: Klynge 3.5



faktisk har stærk forbindelse til 3150, men også en nogenlunde stærk forbindelse til 3111, og en ganske svag forbindelse til 3211, dog kun i ekstern retning. (læseren bør huskes på hvad definitionen er af en forbindelse her - at der ikke eksisterer nogle barrierer i form af den forventede mulighed for at tage job når man kommer fra XXX og går til XXX). 3282 har på samme tid en intern mobilitet på cirka 50 %, og det ses at det er netop denne beskæftigelseskategori, der binder den samlede klynge 3.13 sammen. Kropsbehandlerne, der desuden har en høj intern mobilitet på 67 %, er derfor kun med i kraft af deres forbindelse til assistenterne indenfor kropspleje. Det er desuden sådan, at det primært at assistenterne indenfor kropspleje, der bliver kropsbehandlere, og kun i nogen grad den anden vej rundt.

3.5 Landbrugsklyngen.

Har en god høj densitet, og en intern mobilitet på cirka 50 %. Vi kan igen se at genstandsfeltet er enormt vigtigt, kendskabet til området. I den forbindelse skal man også huske de geografiske dimensioner, den konkrete lokalisering som vi skal se på senere (R Danmarkskort)

ide: lav grafer ud komponenter. Hvilke komponenter indeholder dette netværk? Ret centralt. Hvor er der ihvertfald *slet* ikke forbindelser? Tjek adelsdata igennem.

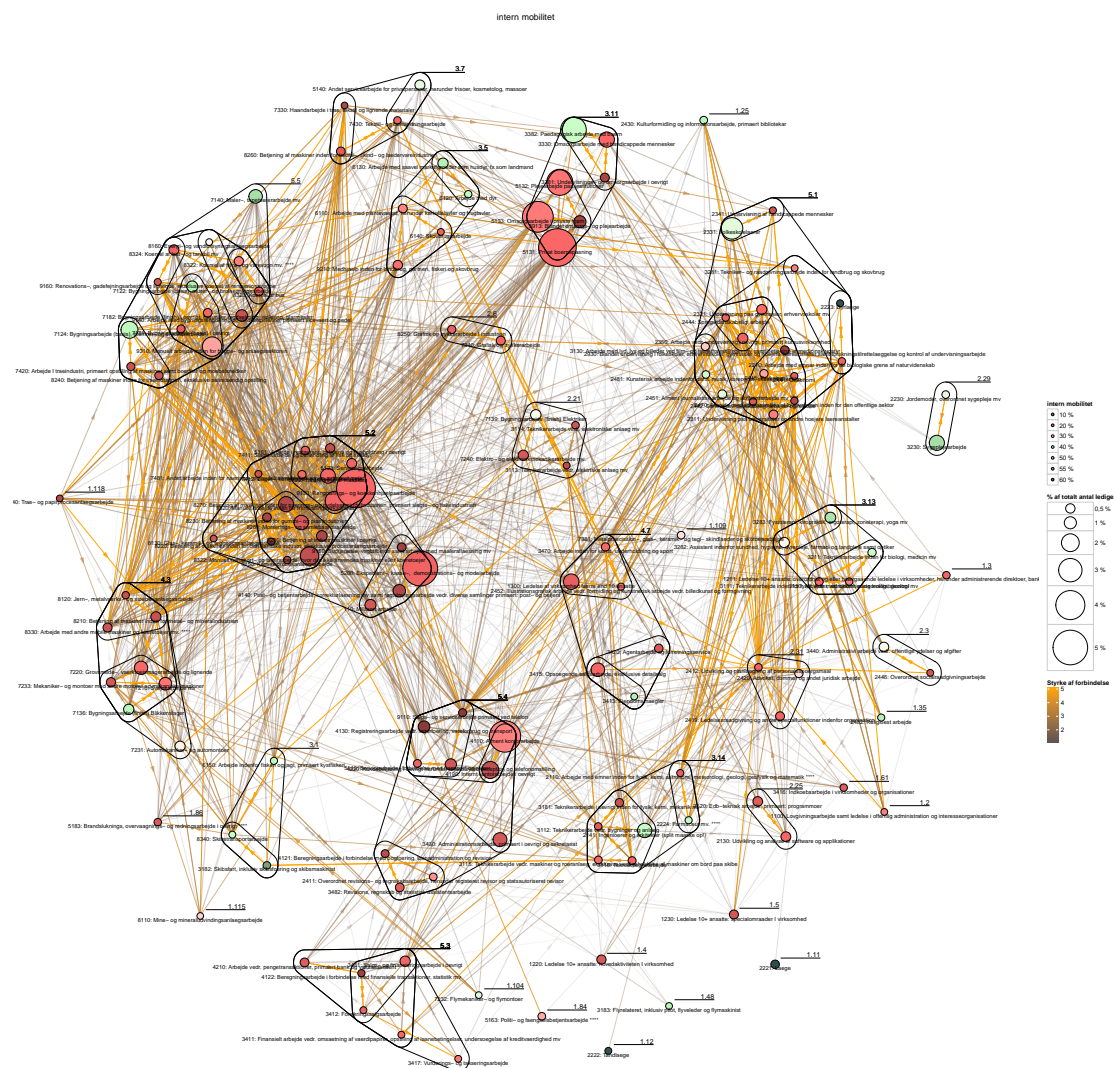
vi kan også sige noget om at hvis man i perioder med arbejdsløshed ikke engang skifter, så er der virkelig tale om stærke barrierer for sammenfald.

1220 har også lav mobilitet, men har 3449 personer og må siges at være vigtig i sig selv.

Hovedkort: intern mobilitet på node

Kortlægningen af vores definition af arbejdsløse i minimum et halvt år giver os **516.812 personer** som er gået fra beskæftigelse til arbejdsløshed til beskæftigelse **597.437 gange** fra 1996 til 2009.

Figur 2.7: Hovedkort, SOCSTIL & SOCIO, 1996-2009



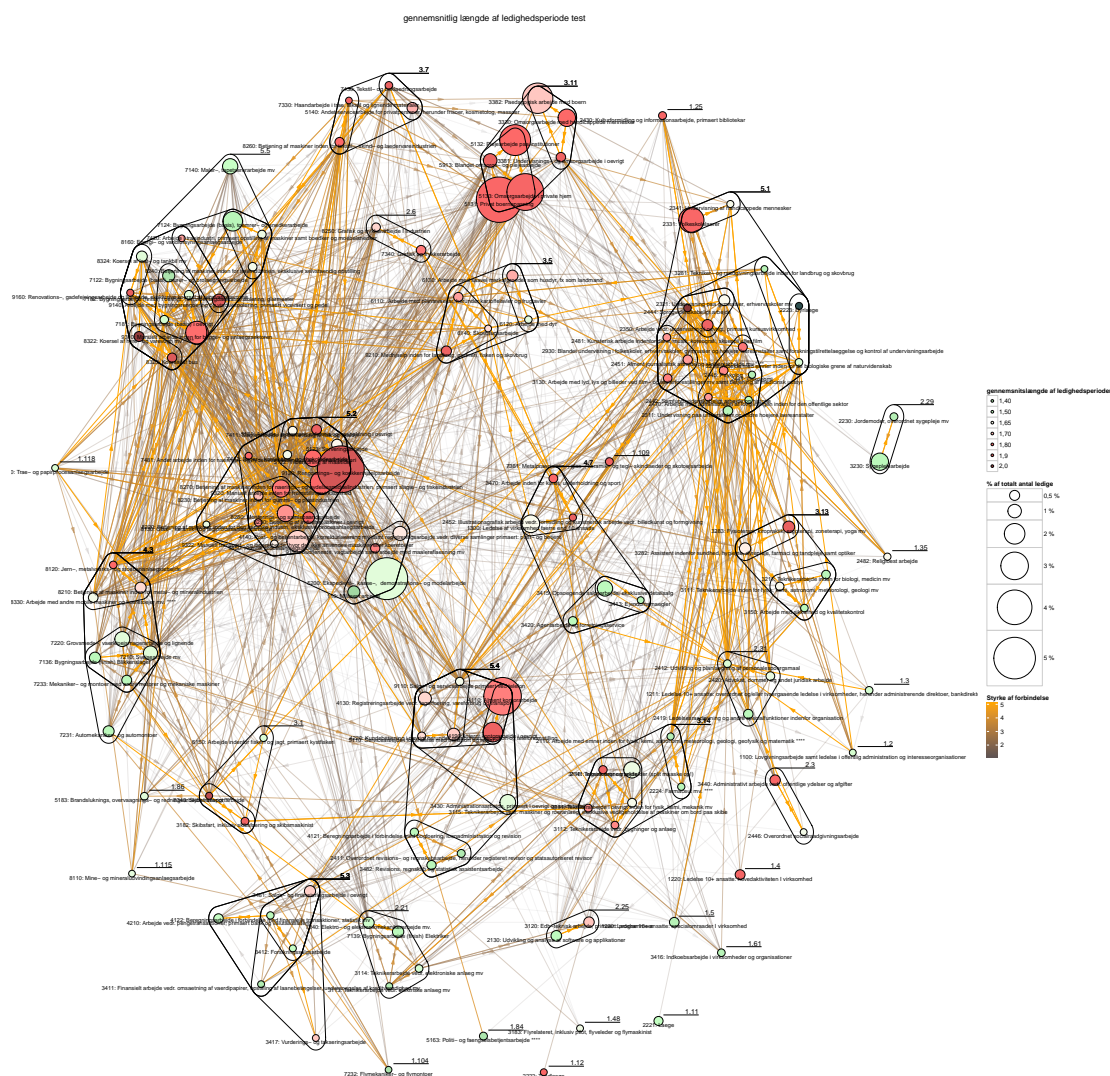
Hovedkort: længden på arbejdsløshedsperioderne

Kortlægningen af vores definition af arbejdsløse i minimum et halvt år giver os **516.812 personer** som er gået fra beskæftigelse til arbejdsløshed til beskæftigelse **597.437 gange** fra 1996 til 2009.

Når vi kigger på selve arbejdsløshedsperioderne kan vi dele dem op i længde og antal og opdele noderne og segmenterne i 1) lange arbejdsløshedsperiode og flere arbejdsløshedsperioder 2) 3) 4)

Når vi kigger på den gennemsnitlige længde af arbejdsløshedsperioderne inden for de forskellige noder og segmenter fremgår det af figuren, at den røde farve markerer lange arbejdsløshedsperioder, og den grønne farver markerer korte arbejdsløshedsperioder.

Figur 2.9: Hovedkort, SOCSTIL & SOCIO, 1996-2009

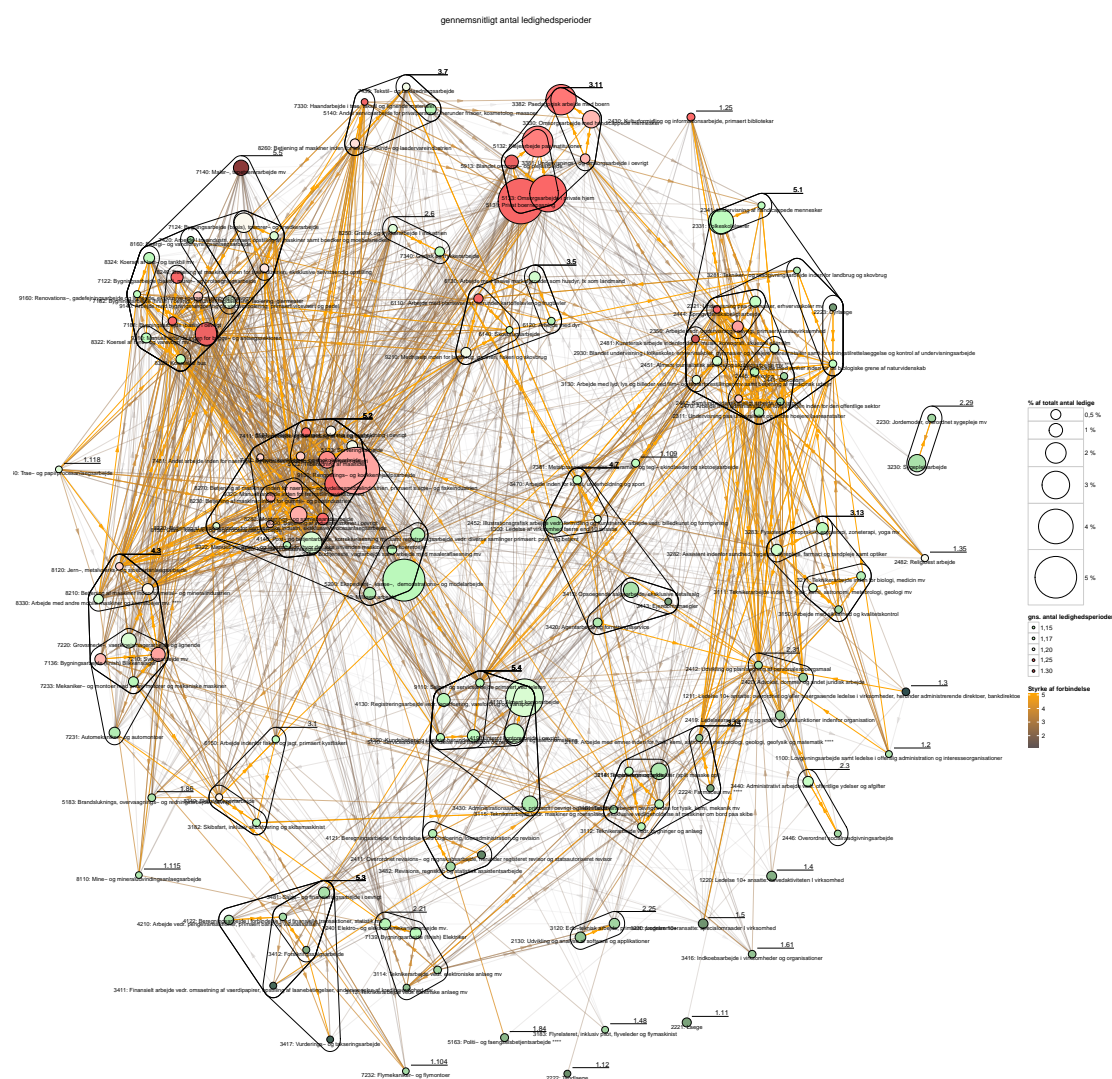


Hovedkort: antal arbejdsløshedsperioder

Kortlægningen af vores definition af arbejdsløse i minimum et halvt år giver os **516.812 personer** som er gået fra beskæftigelse til arbejdsløshed til beskæftigelse **597.437 gange** fra 1996 til 2009.

Når vi kigger på antallet af arbejdsløshedsperioderne inden for de forskellige noder og segmenter fremgår det af figuren, at den røde farve markerer flere arbejdsløshedsperioder, og den grønne farver markerer færre arbejdsløshedsperioder.

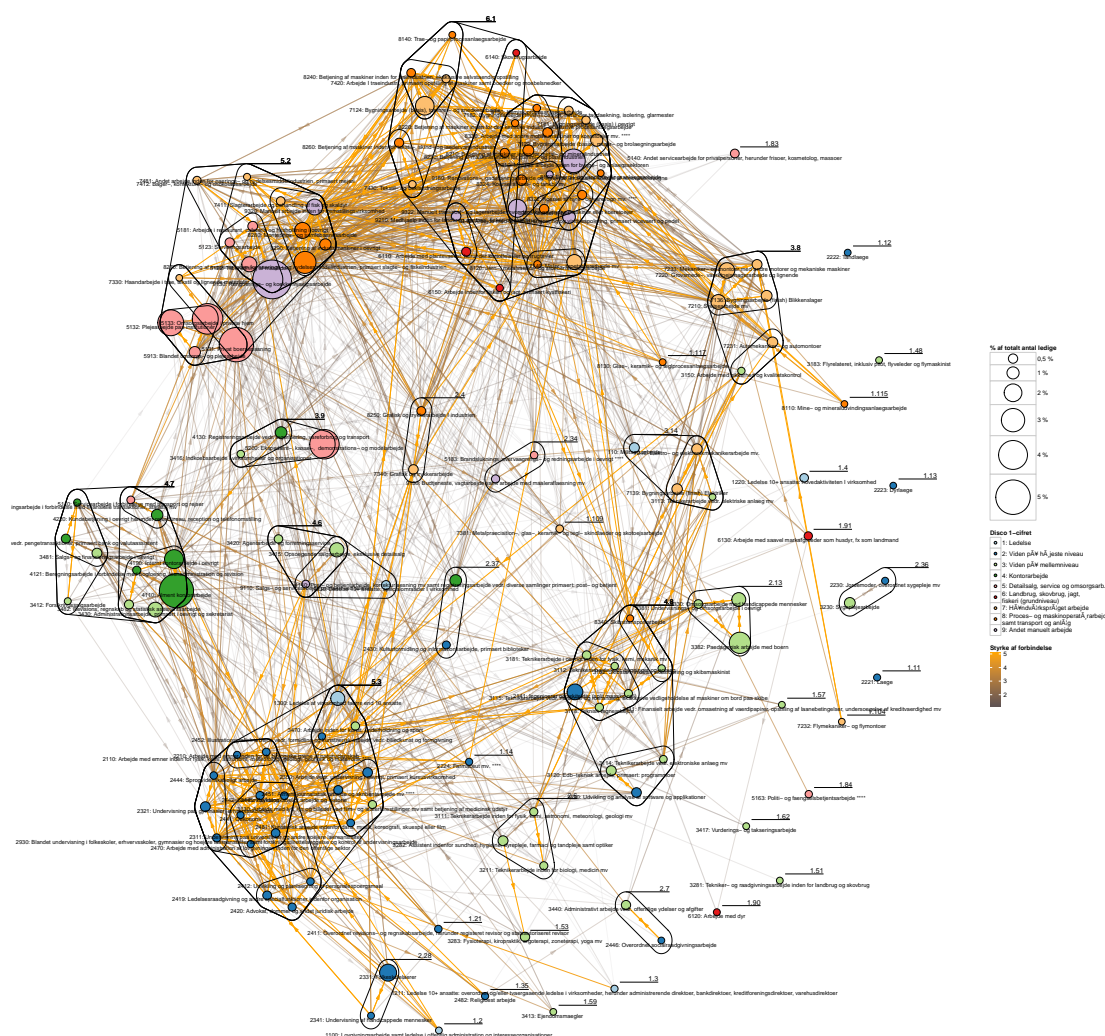
Figur 2.10: Hovedkort, SOCSTIL & SOCIO, 1996-2009



2.2 Bikort: DST's definition af arbejdsløse som nettoledige

Kortlægningen af DST's definition af arbejdsløse som nettoledige giver os **292.926 personer** som er gået fra beskæftigelse til arbejdsløshed til beskæftigelse **317.046 gange** fra 1996 til 2009.

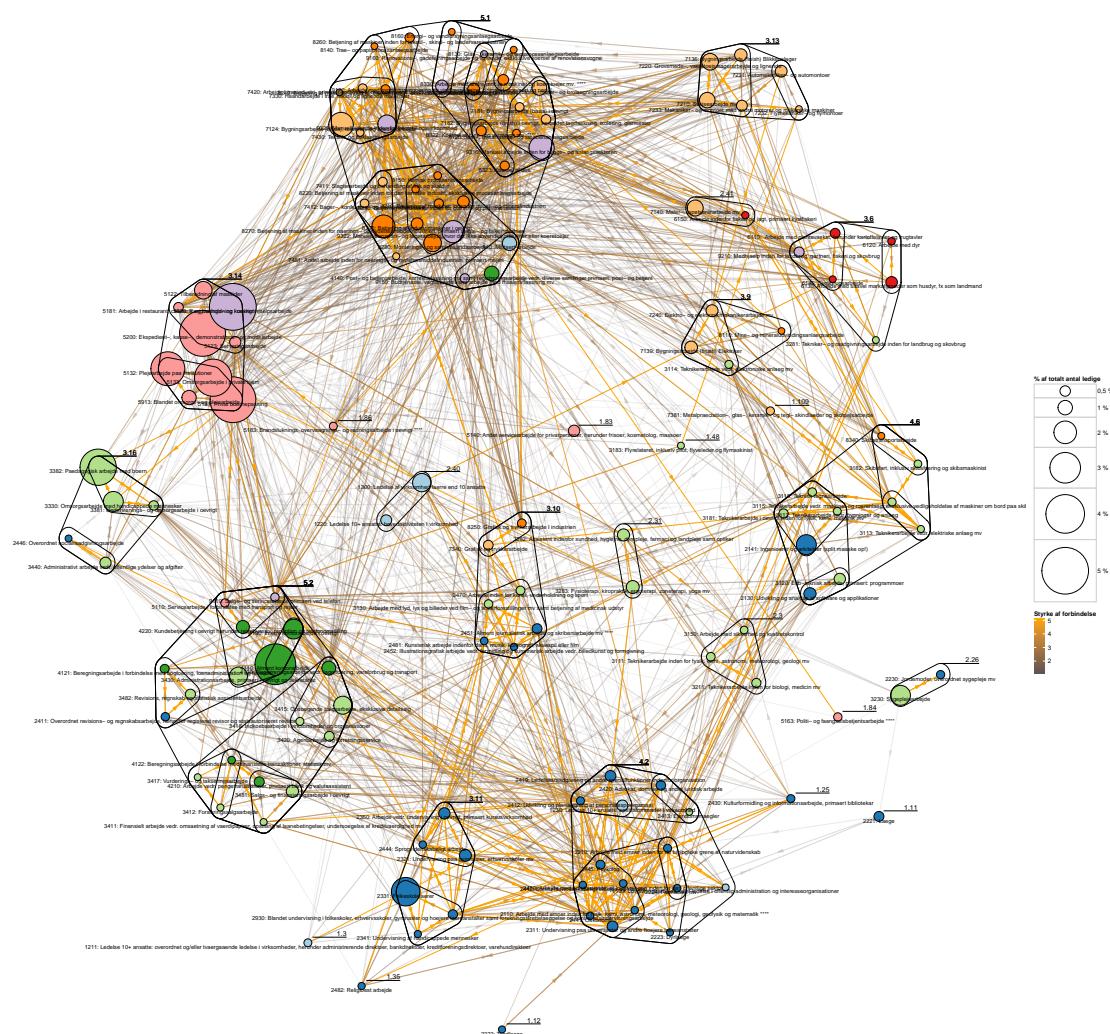
De fem største segmenter er: 1) Det største segment (5.2) består af pleje-, omsorgs, rengørings-, køkkenarbejde mv. (16 noder, 20% af samlet). 2) Det anden største segment (6.1) består af en bred blanding af forskellige faglærte og ufaglærte (28 noder, 14% af samlet). 3) Det tredje største segment (4.7) består blandet kontorarbejde (11 noder, 13% af samlet). Det fjerde største segment (5.3) består primært af forskellige akademikere samt ledelse og diverse kunstnere (20 noder, 8% af samlet). Det femte største segment (3.8) består primært af faglærte håndværkere (6 noder, 5% af samlet).



2.3 Bikort: vores definition af arbejdsløshed, minimum seks uger

Kortlægningen af vores definition af arbejdsløse i minimum seks uger giver os **687.708 personer** som er gået fra beskæftigelse til arbejdsløshed til beskæftigelse **812.533 gange** fra 1996 til 2009.

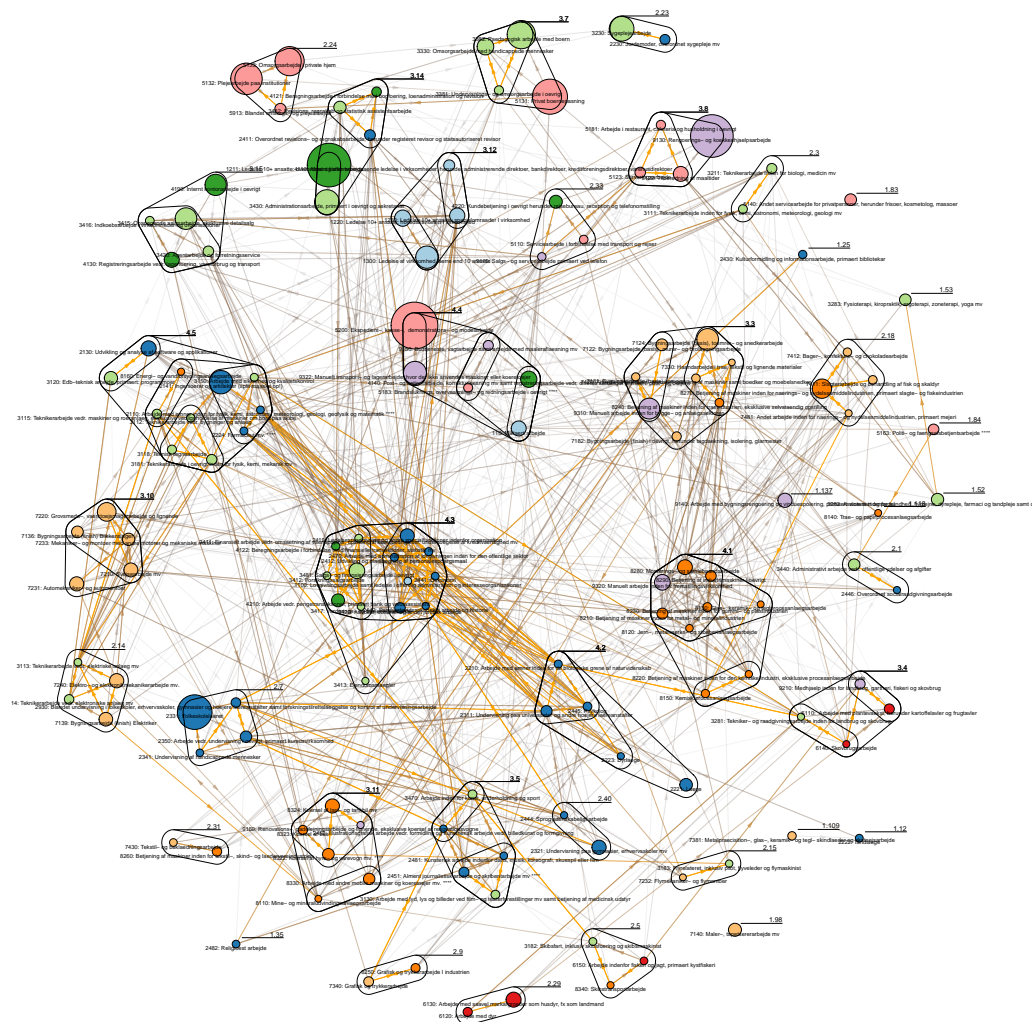
De fem største segmenter er: 1) Det største segment (3.14) består af pleje-, omsorgs, rengørings-, køkken, ekspedientarbejde, mv. (9 noder, 20% af samlet). 2) Det anden største segment (5.1) består af en bred blanding af forskellige faglærte og ufaglærte (35 noder, 20% af samlet). Det tredje største segment (5.2) består af blandet kontorarbejde (19 noder, 17% af samlet). Det fjerde største segment (3.16) består pædagogisk og offentligt administrationsarbejde (5 noder, 5% af samlet). Det femte segment (3.11) består af undervisere (6 noder, 5% af samlet)-



2.4 Bikort: Alle beskæftigede på arbejdsmarkedet (ingen arbejdsløse)

Kortlægningen af alle beskæftigede giver os **2.778.841 personer** som er gået fra beskæftigelse til arbejdsløshed til beskæftigelse **3.298.468 gange** fra 1996 til 2009.

De fem største segmenter i kortlægningen af nettoledige er: 1) Det største segment (4.4) består af ekspedient-, lager, transport, post, brandslukning- og militært arbejde. 2) Det anden største segment (3.14) består af kontorarbejde. 3) Det tredje største segment (3.7) består af pædagogisk, omsorgs- og pasningsarbejde. 4) Det fjerde største segment (2.24) består af pleje og omsorgsarbejde. 5) Det femte segment (3.8) består af rengørings- og køkkenarbejde.



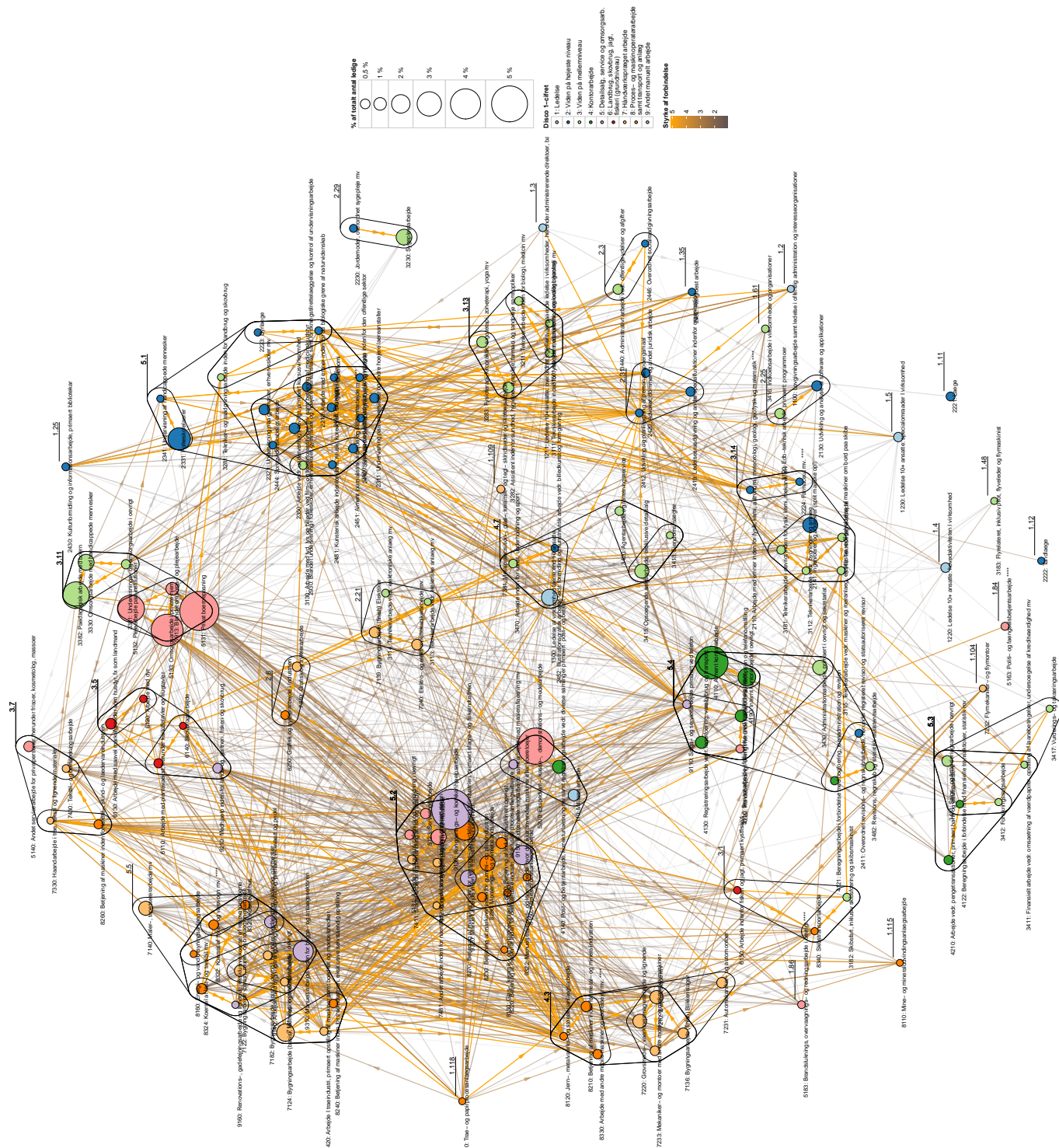
3 Litteraturliste

Primær litteraturliste

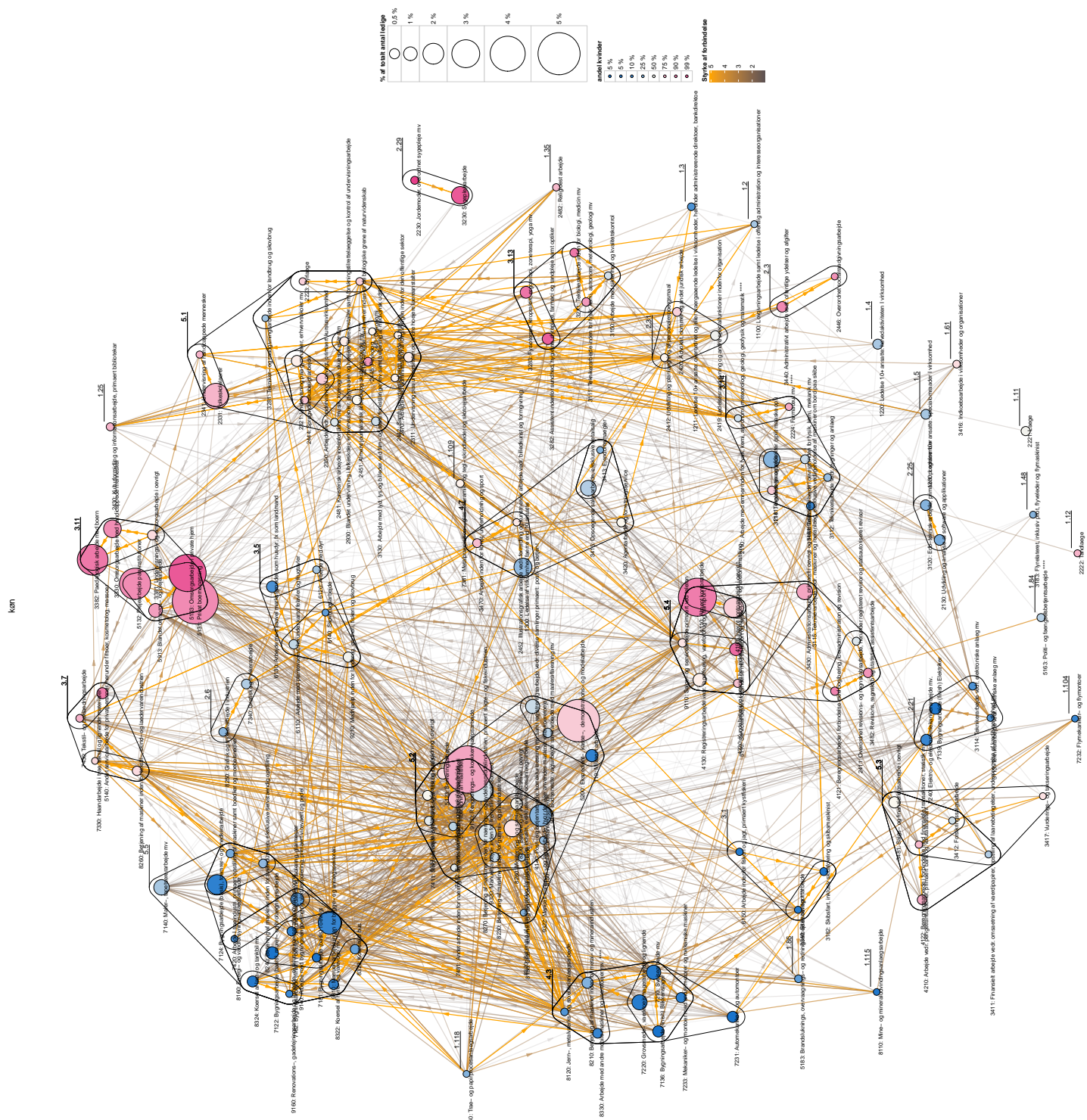
- [1] Alan Agresti og Barbara Finlay. *Statistical Methods for the Social Sciences*. New Jersey: Prentice Hall, 1997.
- [2] Edward Laumann, Peter V. Marsden og David Prensky. „Applied Network Analysis: A Methodological Introduction“. I: udg. af Ronald S. Burt og Michael J. Minor. Sage, 1983. Kap. The Boundary Specification Problem in Network Analysis, s. 18–34.
- [3] Nikolaj Malchow-Møller og Allan Würtz. *Indblik i Statistik*. København: Gyldendal, 2003.
- [4] John Scott. *Social Network Analysis: A Handbook*. Sage, 2000.
- [5] J. Toubøl og A. G. Larsen. „Mapping the class-structure using network analysis: From occupational mobility to social class categories“. I: (2015).
- [6] J. Toubøl, A. G. Larsen og C. Strøby. „A network analytical approach to the study of labour market mobility“. I: *Paper presented at XXXIII Sunbelt Social Networks Conference of the International Network for Social Network Analysis (INSNA), Hamborg, Germany*. (2013).

A FIGURER, STØRRE UDGAVER

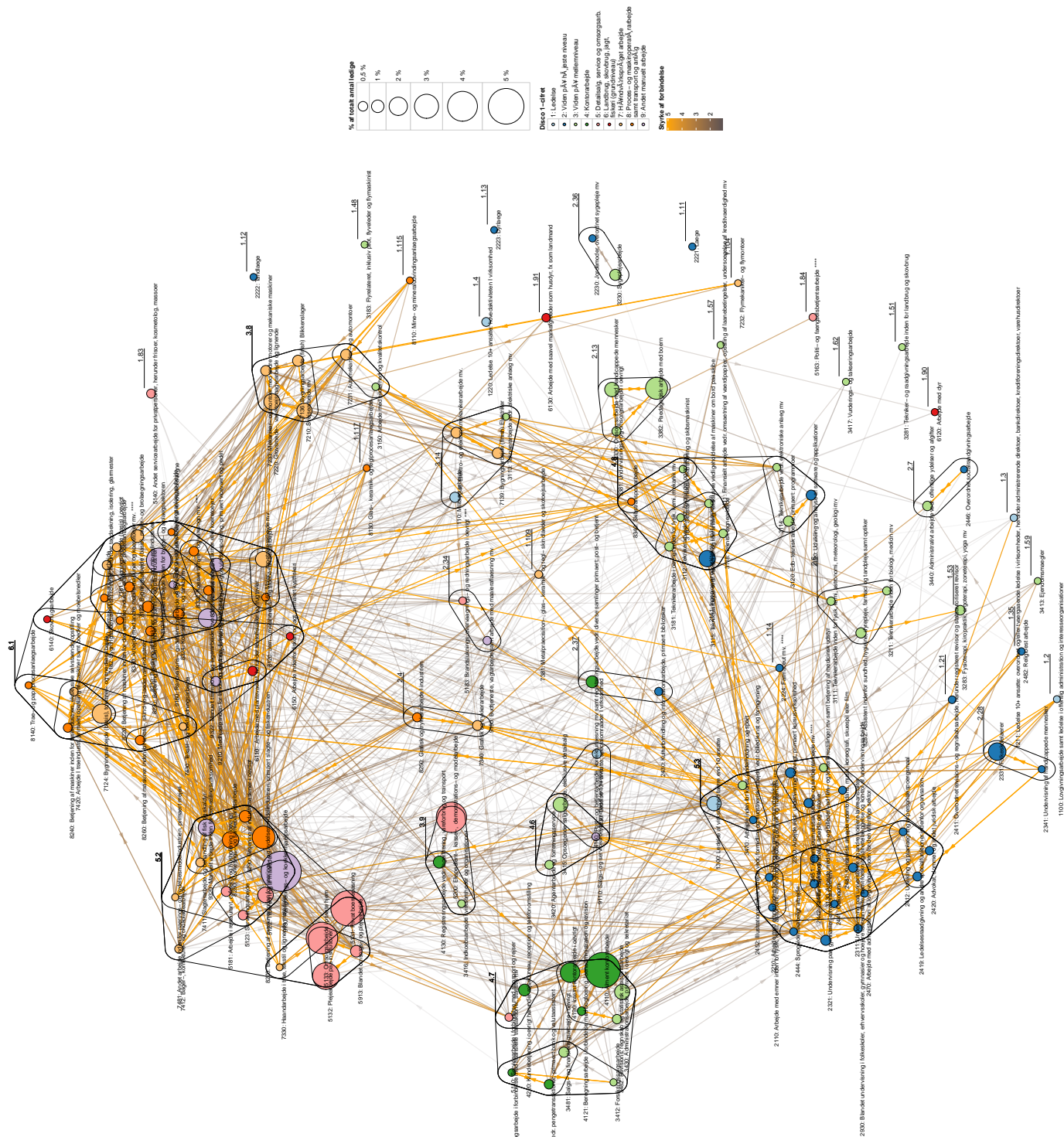
Figur A.1: Hovedkort: Vores definition af arbejdsløse, minimum et halvt år



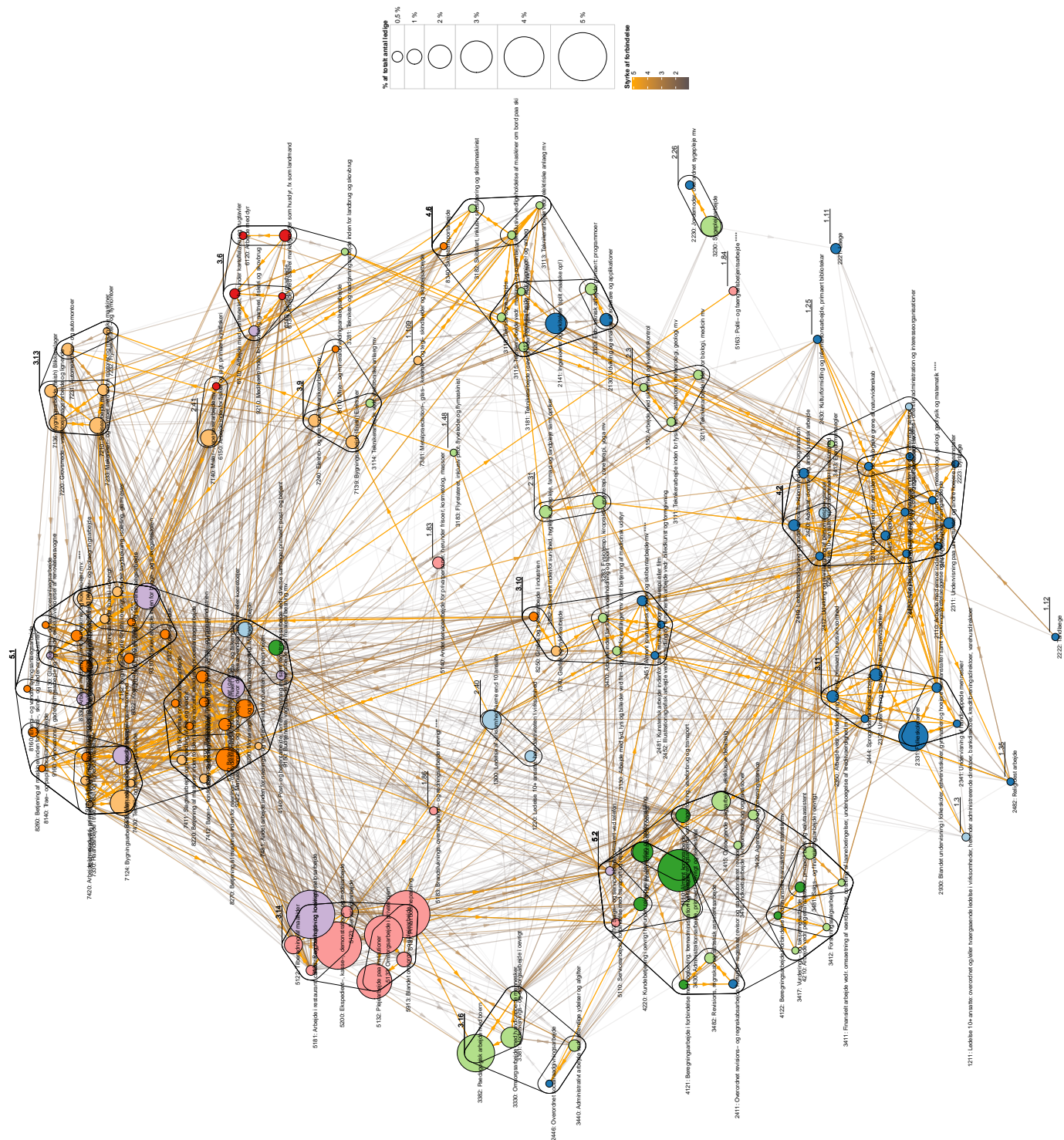
Figur A.2: Hovedkort: farvelagt efter andel af kvinder



Figur A.3: Biktort: DST's definition af arbejdsløse som nettoledige



Figur A.4: Bikort: vores definition af arbejdsløshed, minimum seks uger



Figur A.5: Bikort: Alle beskæftigede på arbejdsmarkedet (ingen arbejdsløse)

