



Discussion paper

Toepassingsmogelijkheden van Complexiteitstheorie bij het CBS

Gert Buiten, Edwin de
Jonge en Frank P. Pijpers

Mei 2018

Inhoudsopgave

1. Wat is Complexiteitstheorie?	4
1.1 Eigenschappen van complexe systemen	4
1.2 Belangrijke concepten bij complexe systemen	6
1.3 Soorten complexe systemen	6
1.4 Relevante methoden voor het beschrijven en analyseren van complexe systemen	7
2. Mogelijke toepassings-gebieden bij het CBS	8
2.1 Vormen van toepassingen	9
2.2 Fenomenen waar het CBS zich op zou kunnen richten	9
3. Hoofdpijnen van een aanpak	10
3.1 Complexiteitsonderzoek op bestaande datasets	10
3.2 Ontwikkelen aparte datasets voor complexiteit	11
3.3 Samenwerking met andere partijen	12
3.4 Voorwaarden voor de langere termijn	13
3.5 Lopende en geplande projecten op de korte termijn	13
4. Eerste resultaten	14
Bijlage	18

Samenvatting

De Complexiteitstheorie is een relatief jong wetenschapsgebied dat de belofte in zich draagt om belangrijke tot nog toe onvoldoende begrepen verschijnselen in bijvoorbeeld het milieu, de samenleving en de economie beter beschrijfbaar en hanteerbaar te maken. Voor het CBS is deze ontwikkeling interessant vanuit de eigen rol van het statistisch beschrijven van maatschappelijke fenomenen ten behoeve van praktijk, beleid en wetenschap. Vandaar dat het onderwerp Complexiteit sinds 2016 onderdeel vormt van het CBS-Onderzoeksprogramma. In dit position paper worden de eerste ervaringen en inzichten over mogelijke toepassingen van Complexiteitstheorie in de 'officiële statistiek' beschreven, evenals de aanpak die het CBS wil volgen bij het realiseren daarvan. Met dit position paper wil het CBS niet alleen belanghebbenden informeren over de plannen, maar ook de dialoog met hen aangaan om de ideeën aan te scherpen en te verbeteren. Bovendien kan het de basis vormen voor het aangaan van samenwerkingen bij het onderzoek.

Keywords

complexiteit, methodologie

1. Wat is Complexiteitstheorie?

Complexe systemen zijn er in allerlei soorten en maten. Denk aan ecosystemen en voedselketens, met ingewikkelde jager/prooirelaties die voortdurend schommelende hoeveelheden dieren en planten laten zien – maar toch een bepaalde logica lijken te hebben. Ook de evolutie van soorten is een voorbeeld van een complex verschijnsel. Op economisch gebied zijn conjunctuurrampen tekenen van complexiteit, op sociaal gebied onder andere sociale uitsluiting en crises. In het verkeer kunnen files genoemd worden, die vaak zonder enige ‘externe’ aanleiding ontstaan.

Deze voorbeelden laten de twee meest essentiële kenmerken van complexe systemen zien. Om te beginnen ontstaan deze systemen uit interacties tussen ‘actoren’, die op het niveau van het systeem als geheel leiden tot ‘emergente eigenschappen’ die niet tot de individuele actoren te herleiden zijn. Zo is een file een eigenschap van ‘het verkeer’ en niet van één auto. Een tweede kenmerk is dat die emergente eigenschappen niet-lineair gedrag vertonen: bij precies hetzelfde aantal auto’s op een weg is de ene keer wel en de andere keer geen sprake van een file.

1.1 Eigenschappen van complexe systemen

Emergente eigenschappen en niet-lineair gedrag zijn kenmerken van alle complexe systemen. Van daaruit ontstaan andere eigenschappen die in veel, maar niet alle complexe systemen voorkomen. Een aantal relevante eigenschappen staat hieronder.

Omslagpunten: complexe systemen hebben vaak verschillende toestanden, waarbij de overgang van de ene toestand (bijvoorbeeld rijdend verkeer) naar de andere (bijvoorbeeld file) plaatsvindt bij een omslagpunt. Voor praktijk, beleid en wetenschap is het erg relevant om te weten waar de omslagpunten zitten en om daar ‘early warning signals’ voor af te leiden die kunnen waarschuwen als bijvoorbeeld de economie in de buurt van een omslagpunt komt.

Multicausaal: belangrijke verschijnselen in complexe systemen hebben meestal meerdere oorzaken, die bovendien op niet-lineaire manieren op elkaar inwerken. Pogingen om zo’n verschijnsel te beïnvloeden door uit te gaan van slechts één van die oorzaken kunnen averechts werken. Het geven van inzicht in het complex van oorzaken is daarom maatschappelijk erg relevant.

Orde c.q. regime: gegeven de initiële structuur en de interactieregels van een complex systeem kan zo’n systeem toestanden aannemen (het geheel van mogelijke uitkomsten gegeven de initiële structuur, interactieregels e.d.). Veranderingen in de structuur en interacties kunnen leiden tot regimeshifts, iets dat voor beleidsmakers belangrijk is.

Padafhankelijkheid en onomkeerbaarheid: bij veel complexe systemen hangt de toekomst mede af van het heden en verleden. Dit geldt zeer zeker voor veel ‘complex adaptive systems’ waarin de actoren hun eigen gedrag kunnen aanpassen op basis van de opgedane ervaringen. Dergelijke systemen zijn ‘historisch’: de tijd kan niet worden teruggedraaid, effecten kunnen niet ongedaan worden gemaakt. Dit

betekent ook dat de context waarbinnen iets gebeurt bepalend is voor het effect ervan.

Evolutionaire dynamica: in veel complexe systemen is sprake van variatie, selectie en versterking. Dit leidt tot een soort 'evolutionaire motor', waarin variaties ontstaan in de interacties tussen actoren en succesvolle variaties versterkt worden en zich verspreiden door het systeem – waardoor ook emergente eigenschappen weer veranderen. In combinatie met padafhankelijkheid kan het betekenen dat voor een bepaalde variatie op een bepaald moment 'de tijd nog niet rijp is' en op een later moment ineens een doorslaand succes is. Ook hier zitten voor praktijk, beleid en wetenschap belangrijke aanknopingspunten.

Niet normaal verdeeld: bij veel complexe systemen zijn belangrijke eigenschappen van de actoren niet op een statistisch gezien 'normale' manier verdeeld over de populatie. Bij een normaalverdeling heeft de helft van de actoren een lagere waarde dan het gemiddelde en de andere helft een hogere, waarbij waarden minder vaak voorkomen naarmate ze verder van het gemiddelde afwijken. Dit soort verdelingen komen in de wereld om ons heen vaak voor en zijn ook een belangrijke basis voor bijvoorbeeld de kansrekening. Bij complexe systemen is vaak sprake van niet-normale of 'scheve' verdelingen: een klein aantal eenheden heeft hele grote waarden en de rest veel kleiner. Een concreet voorbeeld is de verdeling van inkomens en vermogens over individuen, waarbij bijvoorbeeld 1% van de bevolking van een land evenveel vermogen kan hebben als alle anderen bij elkaar. Onnodig te zeggen dat dit voor makers en gebruikers van statistieken belangrijk is.

Positieve en negatieve feedbackloops: vanuit de interacties tussen de onderdelen van een complex systeem ontstaan meestal terugkoppelingen en vooruitkoppelingen die leiden tot zichzelf versterkende effecten dan wel moeilijk te doorbreken evenwichtstoestanden. Zichzelf versterkende effecten treden bijvoorbeeld op bij economische perioden van hoogconjunctuur en recessie. Een extreem voorbeeld zijn 'bubbels' in financiële of vastgoedmarkten. Een voorbeeld van een moeilijk te doorbreken evenwichtstoestand is de armoedeval waarin huishoudens terecht kunnen komen. Door het ingewikkelde samenspel van allerlei regelingen kan het krijgen van betaald werk dan toch leiden tot een lager inkomen.

Grenzen aan voorspelbaarheid: een aantal complexe systemen is maar in beperkte mate te voorspellen, zeker op de wat langere termijn. Dit geldt onder andere voor systemen waarin evolutiemechanismen met variatie, selectie en versterking een belangrijke rol spelen.

Onvolledig meetbaar: veel complexe systemen kunnen nooit volledig worden gemeten, bijvoorbeeld als het aantal actoren enorm is, de informatieverwerking waardoor actoren bepaald gedrag vertonen niet bekend is, er regelmatig nieuwe interactiesoorten kunnen ontstaan en actoren zelf kunnen verschijnen of verdwijnen. Hierdoor zal iedere beschrijving van dergelijke complexe systemen altijd deels gebaseerd zijn op schattingsmethoden. Dit is opnieuw iets dat zowel voor makers als gebruikers van statistieken relevant is.

1.2 Belangrijke concepten bij complexe systemen

Tegen de achtergrond van deze eigenschappen kunnen enkele belangrijke concepten worden onderscheiden die nuttig zijn bij het beschrijven van en denken over complexe systemen. De belangrijkste worden hieronder genoemd.

Zelforganisatie: in veel complexe systemen ontstaan ‘vanzelf’ structuren, dat wil zeggen zonder een centrale leiding of iets dergelijks. Voorbeelden zijn het ontstaan van vriendengroepen of andere clusters in netwerken. Deze spontane processen kunnen erg krachtig zijn en dan invloed hebben op bijvoorbeeld beleid om structuren te beïnvloeden.

Resilience c.q. veerkracht: door de interacties tussen actoren hebben veel complexe systemen vaak mogelijkheden om externe schokken op te vangen door zich aan te passen. Daarmee hebben ze een bepaalde veerkracht. Sterker nog: sommige complexe systemen hebben zelfs schokken en weerstand nodig om zelf te leren en sterker te worden (waarvoor Taleb (2012) de term ‘anti-fragile’ heeft bedacht). Dit aspect is erg belangrijk om het gedrag van een complex systeem te kunnen begrijpen en te beïnvloeden.

Aanpassingvermogen van actoren: een belangrijke bron voor de veerkracht van een systeem is het vermogen van de onderdelen van een complex systeem om hun gedrag aan te passen aan gewijzigde omstandigheden. Inzicht in de mate waarin onderdelen hun gedrag aan (kunnen) passen is erg relevant, evenals veranderingen in de mate waarin dat kan.

Diversiteit / heterogeniteit: in veel complexe systemen zijn niet alle onderdelen hetzelfde, maar is sprake van heterogeniteit. Dit is een bepalende factor bij de soorten interacties die kunnen optreden in een systeem en beïnvloedt ook de veerkracht. In een multiculturele samenleving en een op samenwerking en specialisatie gebaseerde economie is het cruciaal om dit aspect te kunnen analyseren.

Interconnectivity c.q. verbondenheid: de onderdelen binnen een complex systeem zijn in meerdere of mindere mate met elkaar verbonden en beïnvloeden elkaar. De mate waarin de onderdelen verbonden zijn kan verschillen en heeft ook invloed op het functioneren van een complex systeem. Het inzichtelijk maken van die verbondenheid is belangrijk bij het begrijpen en beïnvloeden maken van complexe systemen.

1.3 Soorten complexe systemen

Complexe systemen zijn er in allerlei soorten en maten. Het is goed om daarbinnen in ieder geval onderscheid te maken naar de onderstaande twee groepen (zie John H. Holland).

Complex physical systems: hier zijn de actoren vooral fysieke elementen die altijd op dezelfde manier reageren, zoals vloeistoffen of elektronische systemen.

Complex adaptive systems: bij deze systemen kunnen de actoren hun gedrag aanpassen aan de omstandigheden door het maken van ALS... DAN... afwegingen. Vaak kunnen ze ook in de loop van de tijd geheel nieuwe afwegingen en zelf geheel

nieuw gedrag vertonen. Denk aan het klimaat, een grote stad, sociale netwerken of de economische kringloop.

1.4 Relevante methoden voor het beschrijven en analyseren van complexe systemen

De Complexiteitstheorie kent inmiddels een ruime keuze aan methoden om complexe systemen te beschrijven en te analyseren. In enkele gevallen bouwen die voort op methoden die in de statistische wereld al bekend zijn en toegepast worden. Een belangrijk voorbeeld hiervan is de input/output-analyse uit de economische statistiek, waarin de zichzelf versterkende effecten meetbaar kunnen worden gemaakt door de berekening van multipliereffecten Zie ook Miller & Blair (2009), Samuelson en Nordhaus (2009) en Krugman c.s. (2010).

De nieuwe ontwikkelingen in de complexiteitstheorie bouwen hierop voort en gaan steeds verder – mede door digitalisering, de enorme toename in rekenkracht van computers en het beschikbaar komen van nieuwe digitale bronnen. Voor het CBS relevante methoden worden hieronder genoemd.

Netwerkanalyse: een netwerkbenadering is erg vruchtbaar omdat de interacties tussen actoren goed beschreven kunnen worden als een netwerk – zelfs als de actoren zelf niet fysiek in een netwerk met elkaar verbonden zijn (zie Holland (2017) en Newman, Barabasi en Watts (2006)). De netwerktheorie c.q. grafentheorie kent een aantal belangrijke maatstaven waarmee de eigenschappen van een netwerk en van de onderdelen daarin kunnen worden beschreven, zoals connectiviteit, centraliteit. Daarnaast zijn er methoden om bijvoorbeeld clusters en communities binnen netwerken te onderscheiden van actoren die meer met elkaar interacteren dan met andere actoren. Hierdoor kunnen hiërarchieën in netwerken worden beschreven, maar bijvoorbeeld ook ketens worden onderscheiden.

Dynamische systeemanalyse: hierbij ligt de focus op relaties tussen de emergente eigenschappen op zich en wordt het complexe systeem min of meer als black box beschouwd. Strikt genomen horen dit soort analyses bij complexe fysieke systemen van actoren met een vast reactiepatroon, maar ze kunnen ook zinvol zijn bij systemen waarin actoren een quasi-vast gedragspatroon hebben – al is het maar op de korte termijn. Op economisch gebied sluit dit aan op veel modellen die uitgaan van veronderstellingen over een vast gedrag bij consumenten en producenten, zoals de veronderstelling van rationaliteit en volledige informatie.

Evolutionaire dynamica: hiermee kan bestudeerd worden welke mechanismen op het niveau van de actoren een rol spelen om het emergente gedrag op het niveau van het systeem, waarbij speltheorie en Agent Based Models voorbeelden zijn van concrete benaderingen. Zie onder andere Arthur (2015) voor voorbeelden van de toepassing van deze aanpak op het verklaren van schommelingen in de prijzen van aandelen.

Early Warning Signals: vanuit netwerkanalyse, dynamische systeemtheorie en evolutionaire dynamica kunnen indicatoren worden ontwikkeld die iets zeggen over de mate van weerbaarheid en de nabijheid van omslagpunten van complexe systemen. Vaak gaat het om relatief zwakke signalen op plaatsen in een netwerk die niet direct belangrijk lijken. Denk aan veranderingen in de staart van de verdelingen

van bijvoorbeeld actoren naar grootte. Self-organized criticality kan hierbij ook een rol spelen, een verschijnsel waarbij een systeem gedreven door de interacties tussen de onderdelen vanzelf toegroeit naar een evenwicht dat niet stabiel is.

Het is goed te beseffen dat de Complexiteitstheorie nog relatief jong is en volop in ontwikkeling. Er is veel inzicht verkregen in de mechanismes achter verschijnselen die eerder ongrijpbaar waren, maar tegelijkertijd liggen concrete toepassingen op bijvoorbeeld economische en sociaal gebied nog niet echt binnen handbereik. Soms kan de theorie de achterliggende mechanismes goed repliceren, maar is het nog niet mogelijk bijvoorbeeld een model te maken dat voor een specifiek land en een specifieke periode enigszins redelijke voorspellingen kan doen – wat toch een belangrijk ijkpunt is voor wetenschappelijke theorieën. Sterker nog: het is de vraag of dat soort modellen überhaupt mogelijk zijn, gezien de non-lineariteit van complexe systemen.

Wat dat betreft maakt John H. Holland een zinvol onderscheid in drie soorten modellen:

Datagedreven modellen, die onder andere concrete voorspellingen kunnen doen en die gebruikt kunnen worden om beleidsalternatieven door te rekenen.

‘Existence proof’-modellen, die gebruikt kunnen worden om te bewijzen dat een bepaald verschijnsel of mechanisme überhaupt denkbaar is.

Exploratieve modellen, die gebruikt kunnen worden om te onderzoeken in hoeverre bepaalde interactiemechanismes voldoende verklaring bieden voor het gedrag van complexe systemen. Agent Based Models zijn hier een voorbeeld van.

2. Mogelijke toepassingsgebieden bij het CBS

De bovengenoemde inzichten bieden belangrijke aanknopingspunten voor statistische indicatoren, analyses en dergelijke die meer en beter inzicht bieden in een aantal complexe verschijnselen in de Nederlandse samenleving en economie. Daarnaast kent de complexiteitstheorie overigens ook toepassing op bijvoorbeeld het gebied van bijvoorbeeld organisatie en management, computernetwerken en dergelijke (zie ook Newman (2011)).

Mogelijk zullen lang niet alle hooggespannen verwachtingen rond de Complexiteitstheorie kunnen worden waargemaakt, maar het is de moeite waard om het te proberen. Dit past ook goed in de missie van het CBS om voldoende ‘maatschappelijke toegevoegde waarde’ te creëren.

Het CBS staat hierin overigens niet alleen. Ook bijvoorbeeld de OESO, het Ministerie van Economische Zaken en Klimaat en enkele andere statistiekbureaus en organisaties verkennen de mogelijkheden om Complexiteitstheorie toe te passen (zie o.a. OESO (2017), Ministerie van Economische Zaken (2017)). Belangrijke beleidsmatige inzichten daarbij zijn onder andere dat er in complexe systemen

sprake is een circulaire causaliteit, die leidt tot een fundamentele onzekerheid en onvoorspelbaarheid op langere termijn. Het is belangrijk voor beleidsmakers om zicht te krijgen op de range van beleidsopties en de mogelijkheden voor systeeminnovaties. De rol van modellen is dan niet zozeer om de toekomst te voorspellen, maar vooral om achterliggende mechanismes te onderkennen en met verschillende beleidsopties te kunnen experimenteren. Bovendien is het tijdig onderkennen van 'zwakke signalen' als 'early warning indicators' belangrijk (Ministerie van Economische Zaken (2017)).

2.1 Vormen van toepassingen

Er kunnen drie vormen worden onderscheiden van toepassingen van Complexiteitstheorie binnen het CBS.

Betere duiding van bestaande CBS-data, met name voor de niet-lineaire aspecten. Inzichten voor een betere duiding kunnen voortkomen uit empirische onderzoek vanuit een complexiteitsinvalshoek, maar ook uit daarop voortbouwende exploratieve modellen zoals Agent Based Models.

Nieuwe output o.b.v. bestaande CBS-data, waarbij allereerst gedacht kan worden aan indicatoren voor de mate van complexiteit als zodanig, bijvoorbeeld netwerkmaatstaven en indicatoren voor 'scheve' statistische verdelingen. Van hieruit kunnen eventueel ook Early Warning Indicators worden ontwikkeld. Last but not least kunnen er indicatoren worden ontwikkeld voor aspecten die uit exploratieve modellen naar voren gekomen zijn als relevant.

Geheel nieuwe statistische data en output, waarbij in ieder geval moet worden gedacht aan het maken van statistische bestanden over relevante netwerken in de samenleving, zowel op sociaal als economisch en functioneel gebied. Verder kunnen CBS-data op termijn een bron gaan vormen voor data gedreven complexiteitsmodellen en als basis voor Agent Based Models die bijvoorbeeld door universiteiten worden ontwikkeld.

2.2 Fenomenen waar het CBS zich op zou kunnen richten

Er is een aantal workshops gehouden om te zien op welke fenomenen het CBS inzichten uit de Complexiteitstheorie zou kunnen toepassen, zowel intern als in samenwerking met de Universiteit van Amsterdam. Concrete ideeën die daarbij naar voren kwamen zijn de volgende.

- Economie en arbeidsmarkt
- Sociale en maatschappelijke netwerken
- Gezinsvorming en sociale segregatie
- Milieu en energie
- Brede welvaart

Redenen om dit bij het CBS op te pakken zijn onder andere dat het belangrijke maatschappelijke fenomenen zijn waarin urgente problemen spelen die met meer gangbare methoden niet voldoende hanteerbaar zijn gebleken en dat er vaak gekoppelde microgegevens voor nodig zijn die vooral bij het CBS beschikbaar zijn.

Complexiteitstheorie kan bij deze fenomenen belangrijke inzichten toevoegen, met op de eerste plaats inzicht in de aard van de complexiteit en de mate waarin van complexiteit sprake is. Welke achterliggende mechanismen zijn het meest relevant voor de complexiteit van bijvoorbeeld een economisch verschijnsel? En zijn bepaalde bevolkingsgroepen kwetsbaarder voor een bepaald complex verschijnsel op sociaal gebied?

In de tweede plaats kan het CBS de aard en ontwikkeling beschrijven van de emergente eigenschappen van complexe systemen, die per definitie niet meetbaar zijn bij de afzonderlijke bedrijven, huishoudens en andere actoren. Voor een deel beschrijft het CBS dit soort eigenschappen al, zoals bijvoorbeeld het Nationaal Inkomen en inkomensongelijkheid. Toepassing van Complexiteitstheorie maakt het makkelijker om de emergente eigenschappen te koppelen aan gedrag op individueel niveau in het geval er sprake is van niet-lineaire effecten.

In de derde plaats kan Complexiteitstheorie de basis vormen voor indicatoren en analyses over robuustheid en eventuele Early Warning Indicators.

3. Hoofdpijnen van een aanpak

Complexiteitstheorie is dus veelbelovend, er zijn veel ideeën en hoge verwachtingen - maar concrete toepassingen in de officiële statistiek zijn er nog niet of nauwelijks. Voor het CBS ligt hier een rol om ervaring op te doen of, hoe en waar toepassing van de complexiteitstheorie in de officiële statistiek daadwerkelijk leidt tot toegevoegde waarde. Uiteraard kan het CBS dat niet alleen, maar zal dit in nauwe samenwerking met (potentiële) gebruikers en met de academische wereld moeten gebeuren. Doel is uiteindelijk om als CBS 'praktijk, beleid en wetenschap' te helpen beter grip te krijgen op complexe maatschappelijke problemen door die problemen mede vanuit de Complexiteitsinvalshoek statistisch verantwoord te beschrijven.

Hiervoor kan een tweesporenbeleid worden gekozen. Het ene spoor bestaat uit het toepassen van Complexiteitstheorie op bestaande datasets, zowel voor empirisch onderzoek als aanvullend modelmatig onderzoek met bijvoorbeeld Agent Based Models. Het andere spoor bestaat uit het ontwikkelen van nieuwe datasets over complexe systemen en daarbij zonodig nieuwe methodologie ontwikkelen en geheel nieuwe databronnen aanboren. Beide sporen kunnen naast elkaar worden bewandeld, waarbij gaandeweg de nieuwe datasets uit het tweede spoor toegepast gaan worden in het eerste spoor.

3.1 Complexiteitsonderzoek op bestaande datasets

Bij het eerste spoor van onderzoek op concrete datasets kan het onderstaande als een soort standaardaanpak worden gehanteerd:

- Verzamelen en zo nodig koppelen van bestaande data.

- Gericht empirisch onderzoek met behulp van methoden uit de Complexiteitstheorie, waarbij voor dynamische systeemanalyse en netwerkanalyse belangrijk zijn. Welke aspecten van de samenleving zijn complex en welke niet? En daar waar sprake is van complexiteit: hoe groot is die? En hoe komt die tot uiting?
 - Aanvullend modelmatig onderzoek dat gebruikt kan worden om hypothesen uit het empirische onderzoek te toetsen en om gaandeweg zicht te krijgen op de achterliggende mechanismen en de relevante interacties tussen actoren. Dit kan goed met exploratieve modellen, zoals Agent Based Models en modellen voor dynamische systeemanalyse (Pijpers (2017)). Deze modellen zijn niet zozeer bedoeld voor het doen van voorspellingen, maar allereerst om inzicht te krijgen in mogelijke achterliggende mechanismes en in de interpretatie van empirische uitkomsten. Bij dat laatste kan bijvoorbeeld worden gedacht aan de interpretatie van de hellingshoek van een variabele als indicator van de marktwerking in een markt. Op basis hiervan kunnen ook aanvullende indicatoren worden afgeleid, zoals Early Warning Indicators. Modellen die door gebruikers voldoende krachtig worden gevonden kunnen door hen ook gebruikt worden voor simulaties en beleidsanalyses. Tot slot kunnen dit soort exploratieve modellen ook gebruikt worden om te kijken wat voor relevante informatie nog ontbreekt, zodat het CBS kan proberen alsnog in dat soort data te voorzien.
 - De resultaten van het empirisch en modelmatig onderzoek kunnen vervolgens allereerst worden gebruikt voor een betere duiding van bestaande cijfers. Daarnaast kunnen ze worden gebruikt voor het ontwikkelen van aanvullende indicatoren op basis van bestaande data.
- Het verdient aanbeveling bij dit onderzoek de samenwerking verder uit te bouwen met onder andere universiteiten, beleidsdepartementen, planbureaus en dergelijke. Projecten zelf kunnen geheel in de geest van de Complexiteitstheorie worden opgezet volgens het principe van ‘think big, start small, fail fast en scale fast’.

3.2 Ontwikkelen aparte datasets voor complexiteit

Het tweede spoor, het ontwikkelen van speciale datasets, draait vooral om het verzamelen en afleiden van gegevens over het interacties tussen actoren (zoals huishoudens, bedrijven, personen, overheid, buitenland). Dit gaat verder dan alleen het beschrijven van het gedrag van actoren, omdat het ook gegevens vraagt over ten opzichte van welke andere actor dat gedrag plaatsvindt. Daarmee gaat het per definitie dus om het samenstellen van netwerkbestanden die de relaties, interacties, transacties en dergelijke tussen actoren weergeven.

Informatie over relaties is binnen de bestaande statistieken vaak niet voorhanden, omdat de aandacht bij het verzamelen van data traditioneel ligt bij waarneming bij individuele actoren. Zo wordt bij bedrijven gevraagd naar inkopen en verkopen en bij huishoudens naar inkomsten en uitgaven, maar vaak niet naar de ‘tegenpartij’ van dergelijke transacties.

Het verzamelen en afleiden van informatie over relaties betekent deels dat geprobeerd moet worden om bronnen aan te boren die wél over relaties gaan. Denk aan informatie van banken en uit bedrijfsboekhoudingen over transacties tussen partijen. Ook kunnen bestaande bronnen op een andere manier worden aangewend

om (potentiële) relaties af te leiden, bijvoorbeeld over relaties met buren, klasgenoten, familieleden en dergelijke.

Hierbij zal voor een deel ook nieuwe methodologie moeten worden ontwikkeld. Denk bijvoorbeeld aan methoden voor het afleiden van relaties uit bronnen die daar oorspronkelijk niet voor ontworpen zijn. Voorbeelden hier zijn het veronderstellen c.q. afleiden van relaties tussen bijvoorbeeld mensen die wonen op hetzelfde adres of in dezelfde straat, werken bij hetzelfde bedrijf, lid zijn van dezelfde sportclub en dergelijke.

Denk ook aan alternatieve steekproefmethoden: de bestaande methoden zijn veelal nog gebaseerd op populaties en steekproeven van individuele actoren. Bij relaties geldt dat ze steeds betrekking hebben op twee actoren en dat er vaak scheve verdelingen zijn: sommige actoren hebben heel veel relaties en andere veel minder. Als je de dataverzameling kan richten op de actoren met veel contacten krijg je vanzelf ook veel van de relaties in beeld van actoren met veel minder relaties. Een ander probleem dat speelt is dat bij zeer grote netwerken er vaak geen populatie bekend is en dat ze zo dynamisch kunnen zijn dat ze veranderen terwijl je ze bijvoorbeeld met 'web crawling' in kaart probeert te brengen. Overigens is het goed om te beseffen dat netwerkdatasets niet alleen bruikbaar zijn in het kader van complexiteitsanalyses, maar ook geschikt zijn voor andere toepassingen en voor analyse met andere methoden.

3.3 Samenwerking met andere partijen

Complexiteit is voor het CBS een nieuw terrein en het verdient aanbeveling om het te verkennen en ontginnen in samenwerking met andere partijen zoals universiteiten, onderzoeksorganisaties als TNO, NWO en Planbureaus, beleidsdepartementen, maatschappelijke organisaties – en specifieke verbanden op het gebied van complexiteit als het Netherlands Platform Complex Systems en de Complex Systems Society.

Daarbij zijn meerdere aspecten van belang. Om te beginnen kan op deze manier sneller kennis worden opgebouwd en ontwikkeld. In de tweede plaats kunnen van daaruit analyses en andere concrete toepassingen worden ontwikkeld die aansluiten op de behoeften vanuit praktijk, beleid en wetenschap. In de derde plaats kan door het bundelen van krachten gezorgd worden voor voldoende mensen en middelen voor dit onderzoek. En last but not least: binnen samenwerkingsconstructies kunnen de data van het CBS op een veilige manier toegankelijk worden gemaakt voor de partners.

Er zijn legio manieren om dit soort samenwerkingen vorm te geven, variërend van meedenken met het CBS tot en met het gezamenlijk investeren van mensen en middelen voor een gecombineerd onderzoeksproject. Daar tussenin zitten mogelijkheden als het bij het CBS inzetten van stagiaires of gedetacheerden van andere partijen, het formuleren van onderzoeksvragen, het meedenken over toepassingen, het houden van workshops, het opzetten van een klankbordgroep of community, het financieren van maatwerkonderzoek en dergelijke.

3.4 Voorwaarden voor de langere termijn

Om complexiteitsonderzoek op statistische basis goed van de grond te krijgen moet aan een aantal voorwaarden worden voldaan. Naast algemene voorwaarden als voldoende mensen en middelen zijn er twee die specifiek voor complexiteitsonderzoek van belang zijn.

Allereerst is het belangrijk dat het CBS voldoende toegang heeft tot microdata over transacties en relaties op het niveau van personen en bedrijven. Het gaat daarbij niet om de individuen maar om de relaties die de basis vormen van het empirisch beschrijven van interacties. Voor een deel beschikt het CBS al over dit soort gegevens, bijvoorbeeld over de relaties tussen werkgevers en werknemers uit de Polisadministratie van Belastingdienst en UWV. Tot een aantal andere bronnen heeft het CBS nog geen of maar beperkt toegang, zoals betalingstransacties van banken, klantenbestanden van bedrijven en telefoniedata. Deels hangt dat samen met het beperken van administratieve lastendruk, maar bovenal met de grote maatschappelijke terughoudendheid en juridische beperkingen rond het gebruik van dit soort privacy- en concurrentiegevoelige data. Op de korte termijn probeert het CBS op vrijwillige basis toegang te krijgen tot dit soort gegevens – met wisselend succes. Op langere termijn is een wettelijke basis nodig voor toegang.

Een tweede voorwaarde ligt in het verlengde hiervan: het afdekken van risico's rond onthulling van privacy- en concurrentiegevoelige data bij netwerkanalyses. Hiervoor zullen manieren moeten worden ontwikkeld om over netwerken te publiceren waarbij geen onthulling van gegevens over individuele personen en bedrijven ontstaat.

3.5 Lopende en geplande projecten op de korte termijn

In 2016 is een aantal workshops rond het thema complexiteit georganiseerd, zowel binnen het CBS als met de Universiteit van Amsterdam. Op basis daarvan is een aantal eerste vragen en onderzoeken geformuleerd die passen in de bovengenoemde aanpak. De belangrijkste daarvan zijn opgenomen in het CBS-Onderzoeksplan voor 2017 en voor 2018. Het gaat om de volgende onderzoeksgebieden.

De economie als complex dynamisch systeem, met als concrete lopende onderzoeken:

- Empirisch en modelmatig onderzoek naar de structuur van het Nederlandse bedrijfsleven als resultaat van een complex proces
- De dynamiek in winkelgebieden

Verder zijn er ideeën geformuleerd voor onderzoeken op het gebied van mondiale waardeketens, deeleconomie en de huizenmarkt. De laatste twee zijn geformuleerd in samenwerking met de Universiteit van Amsterdam.

Sociale en economische netwerken, met als concrete lopende acties:

- Het afleiden van een personennetwerk uit een aantal bestaande bronnen
- Het samenstellen van een netwerkmicrobestand van de Nederlandse economie op basis van een mix van bronnen en modelmatige schattingen, met het bedrijvennetwerk als eerste stap
- Onderzoek naar geschikte tools voor de analyse en visualisatie van netwerken.

In samenwerking met de Universiteit van Amsterdam is nog een idee geformuleerd rond 'urban dynamics'.

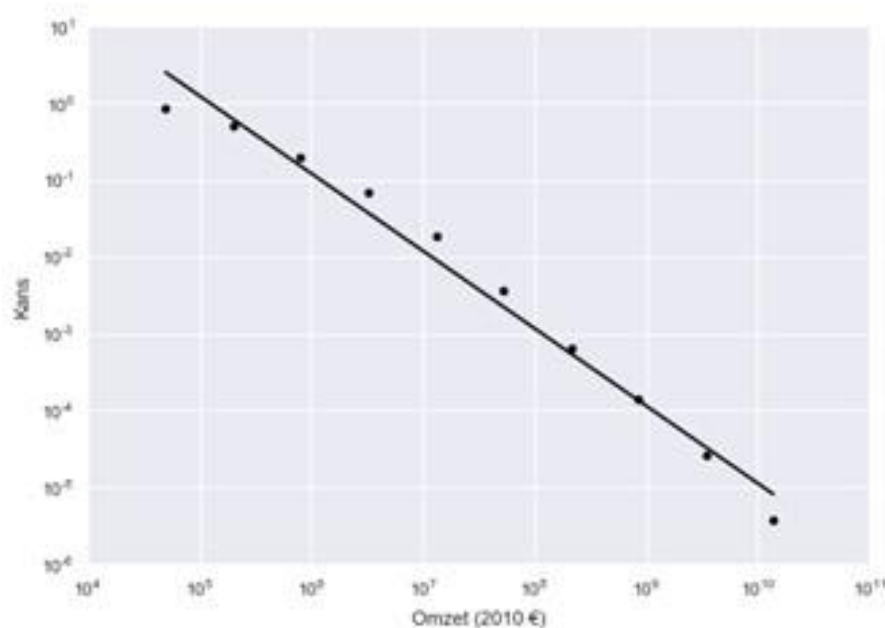
Early Warning Indicators, waarvoor het onderzoek zal starten als de vorige thema's verder zijn uitgewerkt.

4. Eerste resultaten

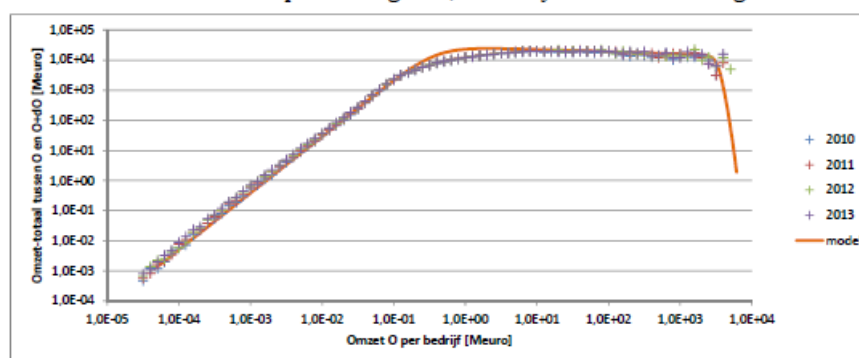
Vanuit de lopende onderzoeken beginnen inmiddels eerste resultaten beschikbaar te komen. In vogelvlucht hieronder enkele uitkomsten.

Complexiteit in het Nederlandse bedrijfsleven: zoals in 1.1 is genoemd zijn verdelingen die statistisch gezien niet-normaal zijn een aanwijzing dat er sprake is van complexe systemen. In het bedrijfsleven is dit een bekend fenomeen: vaak zijn er enkele heel grote bedrijven en veel kleine (zie bijvoorbeeld Axtell (2001) voor de Amerikaanse economie). Bij het CBS is op basis hiervan een onderzoek gestart om te zien hoe dit bij het Nederlandse bedrijfsleven zit - waarbij gebruik gemaakt is van een microbestand van 640.000 bedrijven met onder andere omzetgegevens en een aantal achtergrondkenmerken (zie Roest, (2017)). Figuur 1 laat zien dat er inderdaad ook in Nederland sprake is van scheve verdelingen, wat erop wijst dat de ontwikkeling van bedrijven plaatsvindt binnen een complex proces.

Figuur 1: De staart van de cumulatieve verdelingsfunctie van de omzet van Nederlandse bedrijven in 2010 in een log log plot. (Bron: Roest (2017))



Figuur 2: de verdelingsfunctie van de omzet: het model (doorgetrokken lijn) en de data van het bedrijvenregister voor de jaren 2010 tot en met 2013. (Bron: Pijpers (2017))



Verder verdiepend empirisch onderzoek heeft vervolgens laten zien dat in alle bedrijfstakken hetzelfde grondpatroon te zien is - ook in de minder commerciële. Tussen bedrijven van zelfstandigen enerzijds en BV's, NV's en dergelijke anderzijds zijn wel verschillen te zien, maar vooral als variaties op hetzelfde grondpatroon. In 2018 zal dit empirisch onderzoek worden voortgezet in de vorm van een stage. Aanvullend op het empirisch onderzoek vindt modelmatig onderzoek plaats. Enerzijds gaat het om dynamische systeemanalyse, waarbij een eerste model is ontwikkeld dat op het niveau van emergente eigenschappen een mechanisme beschrijft dat de gevonden verdelingen reproduceert (zie Pijpers (2017) en figuur 2). Een volgende stap is om dit model te vertalen naar economische termen en naar concrete indicatoren.

Anderzijds wordt gewerkt aan een eerste versie van een Agent Based Model, waarmee van hypothesen uit het empirisch onderzoek kan worden getest of ze de gevonden verschijnselen zouden kunnen verklaren. De uitkomsten moeten gebruikt worden voor duiding van uitkomsten en het ontwikkelen van indicatoren. De eerste resultaten laten zien dat hiermee ook relaties kunnen worden gelegd met geaggregeerde indicatoren uit de empirische datasets.

Samenstellen bestand personennetwerk: er is een personennetwerk afgeleid met verschillende relaties en bronnen. Om te beginnen een familienetwerk met derdegraadsfamilie-verbanden uit de ouder-kindrelaties uit de GBA en het SSB. In de tweede plaats een huishoudennetwerk op basis van het SSB. In de derde plaats een burennetwerk op basis van adresinformatie. In de vierde plaats klasgenoten van schoolgaande kinderen op basis van het Opleidingsregister. En last but not least een 'colleganetwerk' op basis van de Polisadministratie.

De afzonderlijke netwerken hebben al veel toegevoegde waarde die in een later stadium verder onderzocht zullen worden. Ze zullen echter ook in onderlinge samenhang worden geanalyseerd. Voor een subset (Rotterdam) is gekeken of communities konden worden gedetecteerd en hoe homogeen die communities zijn

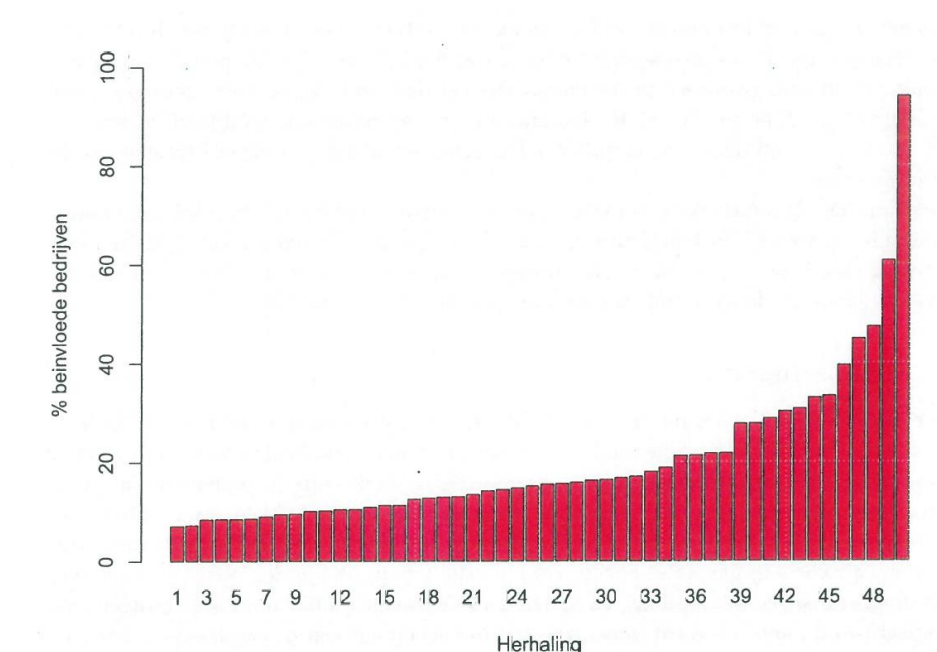
wat betreft allerlei achtergrondkenmerken (herkomst, inkomen en opleidingsniveau). Dit zegt iets over de mate van segregatie.

Samenstellen microbestand economisch netwerk: er is een project gestart dat als doel heeft een dataset te maken die op microniveau het netwerk van de Nederlandse economie in kaart brengt. Als eerste stap is een eerste versie van het bedrijvennetwerk gemaakt, waarbij de relaties tussen bedrijven grotendeels modelmatig geschat zijn op basis van randtotalen, parameters op basis van onderzoek uit andere landen en microdata over bedrijven en hun omzetten. In 2018 wordt deze methode verbeterd en wordt het netwerk uitgebreid met relaties tussen Nederlandse bedrijven en het buitenland, tussen bedrijven en werknemers en eigenaren en tussen bedrijven en huishoudens.

Parallel aan het maken van de dataset worden ook analyses en toepassingen ontwikkeld, die voor een deel weer feedback geven over gewenste kwaliteitsverbeteringen in de dataset.

Een eerste voorbeeld van een analyse is een simulatie waarin gekeken is hoe een 'besmetting' van een kleine groep bedrijven zich kan verspreiden door de economie - waarbij de Fipronil-crisis bij eieren als vertrekpunt fungeerde (zie Bogaart (2017)). Het blijkt dat in veel gevallen de besmetting relatief beperkt blijft, maar in bijzondere gevallen juist wel een groot deel van de economie raakt. In vervolgonderzoek kan worden uitgezocht welke factoren hierbij een rol spelen, waarna daar eventueel indicatoren uit afgeleid kunnen worden. Dit voorbeeld laat zien dat deze oorzaak in complexe systemen lang niet altijd hetzelfde effect hoeft te hebben.

Figuur 3: uiteindelijk percentage 'besmette' bedrijven voor vijftig uitgevoerde simulaties met 'at random' bepaalde startpunten



Analyse dynamiek winkelgebieden: er is een onderzoek gestart naar de dynamiek in de ontwikkeling van winkelgebieden, om te zien of hierbij sprake is van complexe processen die bijvoorbeeld leiden tot zichzelf versterkende processen. Daarbij zijn hypothesen geformuleerd als: 1) wanneer het aantal winkels in een winkelstraat beneden een bepaalde drempel komt zullen de resterende winkels ook vertrekken of stoppen, 2) een winkelgebied zal harder groeien naarmate het een gevarieerder winkelaanbod heeft en 3) winkelgebieden in de buurt van een winkelgebied dat krimpt zullen juist sterker groeien doordat consumenten uitwijken - en daardoor mogelijk gevarieerder worden en nog harder kunnen groeien. Het onderzoek heeft tot nog toe geen concrete resultaten opgeleverd, onder andere omdat het lastig blijkt een goede dataset te construeren. Enerzijds komt dit doordat beschikbare externe bronnen onderling soms grote verschillen laten zien en anderzijds omdat CBS-data niet altijd goed naar het niveau van winkelstraten te vertalen is (zoals gegevens over winkelketens).

Steekproefmethoden voor netwerken: er is een verkennend onderzoek gestart naar de problemen die naar voren komen bij het 'samplen' van grote netwerken, zoals het internet, Facebook en LinkedIn. Dit soort netwerken zijn dermate groot en dynamisch dat het ondoenlijk is op één moment een compleet beeld te krijgen. Ook is er geen populatie van bekend, traditioneel het vertrekpunt voor het bepalen van representatieve steekproeven. Er is inmiddels een paper verschenen waarin een aantal eerste ideeën zijn uitgewerkt voor het 'samplen' van 'Restricted Access Networks' (Willenborg (2018)).

Bijlage

Referenties:

- Arthur, W. B., Complexity and the economy, 2015
- Axtell, R., *Zipf Distribution of U.S. firm sizes*, Science 293, 1818-1820, 2001
- Beinhocker, E. D., The origin of wealth. Evolution, complexity, and the radical remaking of economics, 2005
- Bogaart, P., Constructie van een micro-netwerk van Nederlandse bedrijfseenheden, 2018 (nog te verschijnen)
- Holland, J. H., Complexiteit, Elementaire deeltjes 53, 2017
- Johnson, N. Johnson, Simply complexity. A clear guide to complexity theory, 2007
- Johnson, S., Emergence, 2001
- Krugman, P., Wells, R. and Graddy K. Essentials of economics, 2010
- Laan, D.J. van der, E. de Jonge, Complex Networks 2017, conference on complex networks and their applications, 2017
- Miller, R.E., and P.D. Blair, Input-output analysis. Foundations and extensions, Second edition, 2009
- Ministerie van Economische Zaken, Sturen in een verweven dynamiek. Perspectieven op complexiteit en oriëntaties voor beleid, 2017
- Newman, M.E.J., Complex systems, a survey, arXiv:1112.1440v1 [cond-mat.stat-mech] 2011
- Newman, M.E.J., A.L. Barabasi, D.J. Watts, The structure and dynamics of networks, 2006
- OESO, Debate the issues: complexity and policy making, OECD Insights 2017
- Pijpers, F., Applications of complexity theory in official statistics, 2017
- Roest, J., Complexiteit en het Nederlandse bedrijfsleven, stagerapport, 2017
- Samuelson, P. en Nordhaus, W. Economics, 2009
- Taleb, N. N., Anti-fragile. Things that gain from disorder, 2012
- Willenborg, L. Sampling restricted access networks, 2018

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress

Centraal Bureau voor de Statistiek, CCN Creatie en visualisatie

Ontwerp

Edenspiekermann

Inlichtingen

Tel. 088 570 70 70, fax 070 337 59 94
Via contactformulier: www.cbsl.nl/infoservice

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen/Bonaire, 2018.
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.