



UvA-DARE (Digital Academic Repository)

Information processing in complex networks

Quax, R.

Publication date
2013

[Link to publication](#)

Citation for published version (APA):

Quax, R. (2013). *Information processing in complex networks*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Samenvatting

Stel, in een groep identieke deeltjes bevindt elk afzonderlijk deeltje zich in één van N toestanden, en op een eenvoudige wijze verandert ieder deeltje de eigen toestand over de tijd. Dit vormt een ‘systeem’. Het is eenvoudig om de toestand van het gehele systeem te begrijpen en te voorspellen; we kunnen immers de toestand van ieder deeltje analyseren onafhankelijk van alle andere deeltjes, en dit herhalen voor het hele systeem.

Het begrijpen en voorspellen van de systeemtoestand wordt aanzienlijk moeilijker als we interacties toevoegen tussen deeltjes. Als een deeltje een interactie heeft met een ander deeltje, dan is zijn toestand op één of andere manier afhankelijk van de toestand van het andere deeltje. De onafhankelijkheid verdwijnt, en collecties van deeltjes moeten nu tegelijkertijd worden geanalyseerd, in plaats van afzonderlijk. De interacties vormen nu een ‘netwerk’. De afgelopen zestig jaar is er veel onderzoek gedaan naar bepaalde ‘eenvoudige’ netwerkstructuren, zoals het verbinden van alle mogelijke deeltjesparen, of het plaatsen van een bepaald aantal interacties op volstrekt willekeurige plaatsen. Hoewel dergelijke systemen aanzienlijk moeilijker zijn te begrijpen dan de voorgaande, is er reeds een aanzienlijke hoeveelheid theorie en algoritmieken ontwikkeld. We zeggen dat deze systemen een ‘reguliere’ netwerktopologie hebben.

Maar ongeveer 15 jaar geleden bleek dat veel systemen in de natuur bestaan uit ‘deeltjes’ met niet-reguliere netwerken van interacties. Nieuwe topologieën werden geïntroduceerd, waaronder ‘schaalvrij’, ‘kleine-wereld’, en modulair. De bestaande theorie en algoritmieken waren niet toepasbaar, dus tot op de dag van vandaag heten dergelijke topologieën ‘complexe netwerken’. Voorbeelden uit de natuur die worden beschreven door complexe netwerken zijn de hersenen (neuronen verbonden door synapsen en axonen), de interne huishouding van cellen (genen die andere genen

reguleren), en allerlei sociale netwerken zoals het World Wide Web en vriendennetwerken (zoals Facebook). Een kenmerk van dergelijke ‘complexe systemen’ is als volgt: zelfs als het gedrag van ieder deeltje afzonderlijk volledig bekend zou zijn, nog is de wetenschap niet in staat om het systeem als geheel te begrijpen en de systeemtoestand te voorspellen.

Wat ontbreekt is een generieke theorie die in staat is om de invloed van ieder deeltje te volgen door het netwerk van interacties. Eén deeltje beïnvloedt de toestand van ieder direct verbonden ‘buurdeeltje’, die elk op hun beurt weer andere deeltjes beïnvloeden, enzovoort. Een dergelijke theorie moet universeel zijn, dat wil zeggen, het moet verschillende systemen kunnen beschrijven met één en dezelfde taal.

In deze dissertatie worden de beginstappen beschreven van een kandidaat-theorie genaamd *informatieverwerking*, geïnspireerd door de informatietheorie van Claude Shannon uit 1948. Het idee is als volgt. Stel dat een deeltje zich in N toestanden kan bevinden. Dan zijn er minstens $\log_2 10$ ja/nee vragen nodig om te bepalen in welke toestand het deeltje zich bevindt. (Meerdere vragen zijn toegestaan, maar deze zullen ofwel irrelevant zijn ofwel overlappen met eerdere vragen.) Dit leidt tot een kwantitatieve definitie van ‘invloed’: als één deeltje 3 van de 5 ja/nee vragen beantwoordt over de toestand van een ander deeltje, dan heeft het op dat moment 60% van de toestand van het andere deeltje bepaald. In de informatietheorie is één ja/nee vraag gelijkgesteld aan 1 bit.

Nu is het mogelijk om een mechanistische beschrijving te geven hoe informatie wordt verwerkt in een complex netwerk, gedurende het natuurlijke verloop van de systeemtoestand. De toestand van ieder deeltje bestaat impliciet uit een vast aantal bits. Stel dat in de begintoestand de bits van een deeltje alleen nog betrekking hebben op de toestand van het deeltje zelf. Zodra de systeemtoestand begint te ontwikkelen in de tijd, en ieder deeltje steeds interactie heeft met naburige deeltjes, worden de bits (informatie) op een lokale manier ‘verwerkt’. Sommige bits worden verplaatst van het ene deeltje naar een naburig deeltje door middel van

interacties; andere bits worden gekopieerd, en weer andere bits worden vervangen en raken verloren. Eén bepaalde bit, dat één ja/nee vraag beantwoordde van één bepaald deeltje in de begintoestand, kan op een later tijdstip ook meerdere vragen beantwoorden van andere deeltjes. Onze hypothese is dat als we kunnen beschrijven hoe de bits in een complex netwerk worden verwerkt, we het gedrag van het systeem weer kunnen herleiden uit de individuele gedragingen van de deeltjes. Dit zou leiden tot een beter begrip van het gedrag van complexe systemen, zoals netwerken van neuronen of de interacties van genexpressies.

In deze dissertatie formuleren we een tweetal maten waarvan we vermoeden dat ze een belangrijke rol innemen in een theorie van informatieverwerking. Een volledige formele beschrijving van een dergelijke theorie ligt nog buiten ons bereik. Niettemin, tonen we aan dat deze nieuwe maten reeds nieuwe inzichten en analyse opleveren. Dit versterkt ons vermoeden dat een volledige theorie van informatieverwerking inderdaad een grote potentie heeft in de studie van complexe systemen.

Als eerst bestuderen we de *informatie-dissipatie-tijd*, kortweg IDT. Dit is een maat van hoe lang het duurt voordat de bits in de toestand van een bepaald deeltje volledig zijn verdwenen uit de systeemtoestand. Met andere woorden, het geeft aan hoe lang de systeemtoestand beïnvloedt wordt door de toestand van één deeltje op één tijdstip. Het is dus een maat voor hoe belangrijk het deeltje is voor het verloop van de gehele systeemtoestand. We formuleren deze maat analytisch voor grote netwerken met een brede verdeling van connectiviteit, onder de aanname dat de netwerkstructuur willekeurig is uitgezonderd de verdeling van connectiviteit. Na enige algebra vinden we een paradoxaal resultaat: des de meer interacties een deeltje heeft, des de minder invloed uitoefent het uit op de korte-termijn ontwikkeling van de systeemtoestand. Dit resultaat wordt bevestigd door berekeningen van een computermodel van 6000 Ising-spins in verschillende schaalvrije netwerken. Verder tonen we drie empirische observaties die consistent zijn met dit resultaat, namelijk de propagatie van actiepotentialen in netwerken van neuronen, het evolutionair behoud van proteïnen, en het

succes van mond-op-mond promotie van producten, als functie van de connectiviteit van de neuron, proteïne, en persoon, respectievelijk. Al met al vermoeden we dat deze inverse relatie tussen connectiviteit en korte-termijn invloed een wijdverbreid fenomeen zou kunnen zijn in de natuur.

Als tweede bestuderen we het gerelateerde *informatie-dissipatie-lengte* (IDL), hetgeen een maat is voor de afstand dat de informatie over de toestand van een deeltje kan afleggen voordat het verloren gaat. Om de potentie van deze maat aan te tonen passen we het toe op een gedetailleerde dataset van financiële derivaten, namelijk de tijdreeksen van prijzen van renteswaps van verschillende looptijden. Het voortdurend afsluiten van renteswapcontracten tussen banken is de grootste derivatenmarkt waar 504 duizend-miljard dollar in omgaat. We interpreteren het verloop van deze markt als een dynamisch netwerk, en beredeneren hoe een verandering van IDL in dit (niet-observeerbare) netwerk zich vertaalt in een verandering van IDL in de (observeerbare) tijdreeksen. We tonen aan dat deze IDL de groei van instabiliteit detecteert in de markt gedurende een aantal maanden voorafgaand aan het faillissement van Lehman Brothers, dat het een scherpe piek vertoont vlak voor het faillissement, en dat het na de piek weer sterk terugvalt naar het normale niveau. De val van Lehman Brothers wordt alom erkend als het ‘startsein’ van de recente kredietcrisis. Alternatieve indicatoren zijn niet in staat om de groeiende instabiliteit te detecteren, en de IDL is zover wij weten voor het eerst in staat om een onmiskenbaar waarschuwingssignaal af te geven een aantal maanden vóór de crisis.

Het voorgaande betreft de vertaling van lokale (microscopische) informatie naar globale (macroscopische) informatie. Als de globale informatie inderdaad is opgebouwd uit allerlei stukken lokale informatie, dan moet het zo zijn dat gegevens en metingen van een globaal niveau informatie bevatten over lokale interacties. In het laatste stuk passen we deze redenering toe op globale gegevens van de HIV-epidemie in Engeland, namelijk de verdeling van cluster-groottes van 14560 genotypen van het HIV-virus van anonieme patiënten. De vraag is: kunnen we hieruit parameters afleiden van lokale interacties, met name de kans op infectie van

persoon tot persoon, de verandering van gedrag na een diagnose, en de verdeling van connectiviteit van interacties (infectiewegen)? We gebruiken informatietheorie en gedetailleerde (geïndividualiseerde) computermodellen om aan te tonen dat er inderdaad informatie is opgeslagen in de globale gegevens over deze parameters van lokale interacties. Zover wij weten is dit de eerste studie dat fylogenetische gegevens koppelt aan geïndividualiseerde parameters op een dergelijke schaal, en we vermoeden dat deze methode een significante stap is in de studie en preventie van infectieziekten.