Netwerken bekeken vanuit de statistische fysica *

Frank den Hollander,
Mathematisch Instituut, Universiteit Leiden

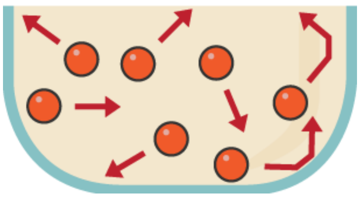
11 september 2016

Samenvatting

In de statistische fysica is de wiskundige beschrijving van veeldeeltjessystemen in evenwicht gebaseerd op Gibbs ensembles. Elk ensemble representeert een *a priori* keuze voor de kansverdeling op de ruimte van configuraties van het systeem, corresponderend met een bepaalde *experimentele situatie*. Een van de hoekstenen van de statistische fysica is de aanname dat, in de thermodynamische limiet wanneer het systeem oneindig groot wordt, alle Gibbs ensembles equivalent zijn. Equivalentie wil zeggen dat de thermodynamische eigenschappen van het systeem hetzelfde zijn ongeacht de keuze van het ensemble. Echter, er zijn voorbeelden van systemen gevonden waarvoor equivalentie wordt gebroken. Een volledige theorie van dit intrigerende verschijnsel is nog niet voorhanden. In dit artikel laten we zien dat breking van equivalentie van ensembles ook kan optreden in *toevallige grafen* waar we *randcondities* aan opleggen. Dit feit heeft belangrijke consequenties voor hoe we grote netwerken dienen te modelleren.

Statistische fysica

In de statistische fysica wordt voor het bestuderen van de eigenschappen van veeldeeltjessystemen in evenwicht (zie Figuur 1) gebruik gemaakt van *Gibbs ensembles*. Welk ensemble wordt gekozen hangt af van de informatie die over het systeem beschikbaar is [3]. Zo correspondeert het *micro-kanonieke ensemble* met de uniforme kansverdeling op de deelverzameling van al die deeltjesconfiguraties waarvoor de energie een voorgeschreven waarde heeft. De keuze voor de uniforme kansverdeling drukt uit dat van het systeem geen andere informatie beschikbaar is dan dat het de voorgeschreven energie heeft, zodat elke configuratie met deze energie een *a priori gelijke kans* heeft. Bij dit ensemble liggen de energie en het aantal deeltjes van het systeem dus vast, en is er sprake van een *gesloten* systeem.



Figuur 1: Deeltjes in een container die met elkaar wisselwerken.

*Dit artikel is een aangepaste versie van [1].

Nu is het rekenen met een harde randconditie in het algemeen lastig. In de praktijk is het handiger om te werken met een zachte randconditie. Het *kanonieke ensemble* correspondeert met de kansverdeling van *maximale entropie* op de verzameling van alle deeltjesconfiguraties, zonder restrictie op de energie, zodanig dat het *gemiddelde* van de energie een voorgeschreven waarde heeft. Bij dit ensemble kan er energie worden uitgewisseld met de omgeving, maar liggen de gemiddelde energie en het aantal deeltjes van het systeem vast. Er is dus sprake van een systeem dat *energetisch open* is, maar verder gesloten. De keuze voor maximale entropie drukt uit dat van het systeem geen andere informatie beschikbaar is dan dat het de voorgeschreven gemiddelde energie heeft. Er wordt een geschikte temperatuur gekozen, die wiskundig gezien de rol speelt van een Lagrange multiplicator die de zachte randconditie realiseert. Het kanonieke ensemble kan gezien worden als de oplossing van een probleem van *statistische inferentie* op basis van de partiële informatie die verpakt zit in de zachte randconditie [5].

Voor systemen waarin de deeltjes over *korte afstanden* met elkaar wisselwerken is aangetoond dat de beide ensembles equivalent zijn in de thermodynamische limiet, d.w.z. wanneer het aantal deeltjes naar oneindig gaat. Het idee hier achter is dat in het kanonieke ensemble de energie weliswaar fluctueert, maar op een schaal die verwaarloosbaar klein is ten opzichte van de gemiddelde energie, zodat het zich effectief gedraagt als het mikro-kanonieke ensemble. Er zijn echter voorbeelden van systemen waarin de deeltjes over *lange afstanden* met elkaar wisselwerken waarvoor de equivalentie wordt gebroken. Dit heeft een belangrijke consequentie: de keuze van ensemble dient met zorg te worden gemaakt omdat het tot verschillend thermodynamisch gedrag leidt. M.a.w. de *experimentele situatie* waarin het systeem verkeert dient nauwgezet te worden vastgesteld en gehandhaafd.

De achtergrond van dit verschijnsel is nog niet goed begrepen. Vanuit wiskundig perspectief zijn er diverse manieren om niet-equivalentie te kwantificeren [7]. Een suggestie is om te kijken naar de *relatieve entropie per deeltje* van de twee ensembles en te laten zien dat die niet naar nul convergeert in de thermodynamische limiet. Relatieve entropie is een pseudo-afstand op de ruimte van kansverdelingen en kan derhalve kansverdelingen van elkaar onderscheiden. Onder bepaalde aannamen kan worden aangetoond dat deze vorm van kwantificeren van niet-equivalentie de scherpste is [7].

Kader: relatieve entropie

Zij χ een eindige verzameling. Laten $\vec{p} = (p_i)_{i\in\chi}$ en $\vec{q} = (q_i)_{i\in\chi}$ kansverdelingen zijn op χ. De relatieve entropie van $\vec{p}$ t.o.v. $\vec{q}$ is gedefinieerd als

$$S(\vec{p} \mid \vec{q}) = \sum_{i\in\chi} p_i \log\left(\frac{p_i}{q_i}\right).$$

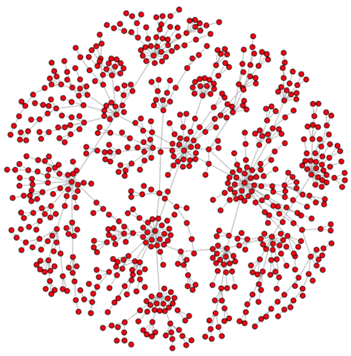
Er geldt dat $S(\vec{p} \mid \vec{q}) \geq 0$, met gelijkheid dan en slechts dan wanneer $\vec{p} = \vec{q}$. Merk op dat $S(\vec{p} \mid \vec{q})$ niet symmetrisch is onder verwisseling van $\vec{p}$ en $\vec{q}$.

Complexe netwerken

We bestuderen ensemble-equivalentie voor *complexe netwerken*: grote toevallige grafen waaraan we randcondities aan opleggen. De rol van de deeltjesconfiguratie wordt daarbij overgenomen door de graaf zelf, en de energie van de deeltjesconfiguratie door een geschikte functie van de graaf.

Kader: toevallige grafen en complexe netwerken

Een graaf bestaat uit een verzameling punten, ook wel knopen genoemd, en lijnen tussen die punten, ook wel verbindingen genoemd. De graad van een punt is het aantal punten waarmee het via lijnen verbonden is. Toeval bepaalt bij een toevallige graaf of er tussen twee punten een lijn is of niet. Grafen worden gebruikt om netwerken te modelleren, bv. van deeltjes en hun interacties, van wetenschappers en hun co-auteur-relaties, van banken en hun financiele transacties, of van personen en hun vriendschappen. Wanneer netwerken groot en complex zijn, dan is het zinvol om toevallige grafen voor de modellering ervan te gebruiken (zie Figuur 2).



Figuur 2: Een voorbeeld van een complex netwerk.

Complexe netwerken vormen een multi-disciplinair onderzoeksgebied, dat zich richt op het begrijpen van de structuur, het gedrag en de algoritmiek van grootschalige netwerken, via modellering, analyse, optimalisatie en control, gedreven door toepassingen. Toeval is daarbij veelal synoniem met complex.

Ensembles voor toevallige grafen

Voor $N \in \mathbb{N}$, zij $\mathcal{G}_N$ de verzameling van alle grafen met N punten. Zij $\vec{C}$ een vector-waardige functie op $\mathcal{G}_N$. De *micro-kanonieke kansverdeling* met *harde randconditie* $\vec{C}^\star$ is gedefinieerd als

$$P_N^{\text{mic}}(\mathbf{G}) = \begin{cases} 1/\Omega_{\vec{C}^\star}, & \vec{C}(\mathbf{G}) = \vec{C}^\star, \\ 0, & \vec{C}(\mathbf{G}) \neq \vec{C}^\star, \end{cases} \qquad \mathbf{G} \in \mathcal{G}_N,$$

waar $\Omega_{\vec{C}^\star} = |\{\mathbf{G} \in \mathcal{G}_N \colon \vec{C}(\mathbf{G}) = \vec{C}^\star\}|$ het aantal grafen is dat $\vec{C}^\star$ realiseert. De *kanonieke kansverdeling* wordt gedefinieerd als de kansverdeling op $\mathcal{G}_N$ die de *Shannon entropy*

$$S(P_N^{\text{can}}) = -\sum_{\mathbf{G} \in \mathcal{G}_N} P_N^{\text{can}}(\mathbf{G}) \log P_N^{\text{can}}(\mathbf{G})$$

maximaliseert onder de *zachte randconditie* $\langle \vec{C} \rangle = \vec{C}^\star$, waarbij $\langle \cdot \rangle$ het gemiddelde is m.b.t. P_N^{can}. Deze kansverdeling wordt gegeven door

$$P_N^{\text{can}}(\mathbf{G}) = \frac{\exp[-H(\mathbf{G}, \vec{\theta}^\star)]}{Z(\vec{\theta}^\star)}, \qquad \mathbf{G} \in \mathcal{G}_N,$$

waar $H(\mathbf{G}, \vec{\theta}) = \vec{\theta} \cdot \vec{C}(\mathbf{G})$ de *Hamilton-functie* op $\mathcal{G}_N$ is en $Z(\vec{\theta})$ de *partitie-functie*, die als normalisatiefactor optreedt. De parameter $\vec{\theta}$ moet gelijk worden gekozen aan de waarde $\vec{\theta}^\star$ waarvoor $\langle \vec{C} \rangle = \vec{C}^\star$. Deze waarde maximaliseert de *statistical likelihood* van $\mathbf{G}$.

Kader: specifieke relatieve entropie van ensembles

We zeggen dat P_N^{mic} en P_N^{can} equivalent zijn dan en slechts dan als hun *specifieke relatieve entropie* nul is [7], d.w.z.

$$s = \lim_{N\to\infty} N^{-1} S(P_N^{\text{mic}} \mid P_N^{\text{can}}) = 0.$$

Omdat zowel de micro-kanonieke als de kanonieke kans hetzelfde is voor alle configuraties met dezelfde waarde van de randconditie, geldt dat

$$S(P_N^{\text{mic}} \mid P_N^{\text{can}}) = \log \frac{P_N^{\text{mic}}(\mathbf{G}^\star)}{P_N^{\text{can}}(\mathbf{G}^\star)},$$

waarbij je voor $\mathbf{G}^\star$ elke willekeurige graaf mag kiezen in $\mathcal{G}_N$ is waarvoor $\vec{C}(\mathbf{G}^\star) = \vec{C}^\star$. Deze observatie is niet alleen van theoretisch belang, hij vereenvoudigt de berekening van s aanzienlijk.

Configuratie Model

Het *Configuratie Model* genereert een toevallige graaf bestaande uit N punten met een van te voren vastgelegde *gradenrij* $\vec{k}^\star = (k_1^\star, \dots, k_N^\star)$, waarbij $k_i^\star \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ het aantal lijnen is dat met knooppunt i is verbonden (zie Figuur 3). [1] In dit voorbeeld is de randconditie dus $\vec{C}^\star = \vec{k}^\star$. Het aantal grafen $\Omega_{\vec{k}^\star}$ dat aan deze randconditie voldoet is niet bekend, maar er zijn wel asymptotische resultaten voor het geval dat $N \to \infty$ en de graden van de punten niet al te groot worden:

$$k_{\max} = \max_{1\le i\le N} k_i^\star = o(\sqrt{N}).$$

M.a.w. de maximale graad $k_{\max}$ mag van N afhangen, maar moet minder hard groeien dan $\sqrt{N}$ als $N \to \infty$. Namelijk, onder deze conditie kan worden aangetoond dat [6]

$$S_N(P_{\text{mic}} \mid P_{\text{can}}) \sim \sum_{i=1}^{N} g(k_i^\star), \qquad N \to \infty,$$

met

$$g(k) = \ln\left(\frac{k!}{k^k e^{-k}}\right), \qquad k \in \mathbb{N}_0.$$

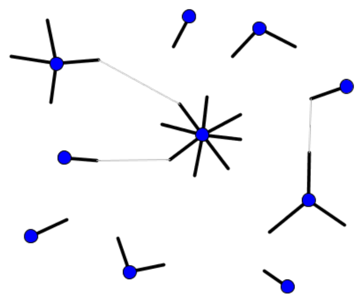
Om s uit te rekenen kijken we naar de *empirische kansverdeling van de graden*, die gegeven wordt door de uitdrukking

$$f_N^\star = N^{-1} \sum_{i=1}^{N} \delta_{k_i^\star}.$$

Hier is δ_k de Dirac-kansverdeling, gedefinieerd door $\delta_k = 1_{\{l=k\}}$, $l \in \mathbb{N}_0$. De betekenis van $f_N^\star$ is dat $N f_N^\star(k)$ het aantal punten telt dat precies graad k heeft. Een andere interpretatie van $f_N^\star$ is de kansverdeling van de graad van een punt waarvan we het label i willekeurig en uniform uit $\{1,\dots,n\}$ trekken. Onder de veronderstelling dat f_N voor $N \to \infty$ convergeert naar een kansverdeling f op $\mathbb{N}_0$, en wel zodanig dat $\lim_{N\to\infty} \sum_{k\in\mathbb{N}_0} |f_N^\star(k) - f(k)| g(k) = 0$, vinden we dat

$$s = \sum_{k\in\mathbb{N}_0} f(k) g(k).$$

[1] De gradenrij $\vec{k}^\star$ dient “grafisch” te zijn, d.w.z. er dient minstens één graaf te bestaan die gradenrij $\vec{k}^\star$ heeft. Dit legt enige beperkingen op aan de keuzen die we voor $\vec{k}^\star$ kunnen maken. Zie [4] voor meer details.



Figuur 3: Een voorbeeld van een toevallige graaf gegenereerd volgens het Configuratie Model met aantal punten $N = 10$ en gradenrij $\vec{k}^\star = (4,1,2,1,3,1,2,1,1,8)$. Aan elk knooppunt i worden $k_i^\star$ *halflijnen* verbonden, waarvan de uiteinden achtereenvolgens op willekeurige wijze met elkaar verbonden worden tot volledige lijnen. Op deze wijze ontstaat een graaf met gradenrij $\vec{k}^\star$, en elke graaf met gradenrij $\vec{k}^\star$ heeft dezelfde kans om aldus gegenereerd te worden. De kansverdeling van deze toevallige graaf is dus het mikro-kanonieke ensemble.

Omdat $g(k) > 0$ voor alle $k \in \mathbb{N}_0 \backslash \{0\}$, volgt dat $s > 0$ zodra $f \neq \delta_0$, en dus zijn de ensembles niet-equivalent. M.a.w. in de limiet $N \to \infty$ geldt dat onder de kanonieke kansverdeling de meeste grafen *niet* een gradenrij hebben die in de buurt ligt van de gemiddelde gradenrij. De keuze voor micro-kanoniek of kanoniek maakt dus een *fors verschil* voor het gedrag van $\mathbf{G}$. In de praktijk betekent dit dat we bij de modellering van een complexe netwerken m.b.v. het Configuatie Model goed moeten oppassen welk ensemble we gebruiken: mikro-kanoniek en kanoniek leiden tot verschillende toevallige grafen.

De laatste formule heeft een interessante interpretatie. Namelijk,

$$g(k) = s\big(\delta_k \mid \mathrm{POI[k]}\big)$$

is de relatieve entropie van de Dirac-kansverdeling δ_k t.o.v. de Poisson-kansverdeling POI[k] gedefinieerd door $\mathrm{POI}[k](l) = e^{-k}\,\frac{k^l}{l!}$, $l \in \mathbb{N}_0$. Dit laat zien dat in het kanonieke ensemble de graden niet Dirac verdeeld zijn, zoals in het micro-kanonieke ensemble, maar Poisson verdeeld.

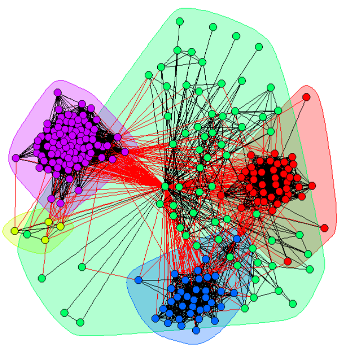
Kader: twee voorbeelden

Reguliere netwerken. Kies $k_i^\star \equiv k^\star$ met $k^\star = o(\sqrt{N})$. Dan geldt $s = g(k^\star)$. Merk op dat wanneer $k^\star$ divergeert als $N \to \infty$, er een extreme vorm van niet-equivalentie optreedt omdat $g(k^\star) \sim \log(\sqrt{2\pi k^\star})$, $k^\star \to \infty$. M.a.w. de relatieve entropie is niet alleen strikt positief, maar divergeert zelfs als $k^\star \to \infty$, zodat de niet-equivalentie van de beide ensembles hand over hand toeneemt.

Schaalvrije netwerken. Kies $\vec{k}^\star$ zodanig dat $f_N^\star(k) = A_{k_c} k^{-\gamma}\, 1\{1 \leq k < k_c(N)\}$, met $\gamma \in (1,\infty)$ en met $k_c(N)$ zodanig gekozen dat $\lim_{N\to\infty} k_c(N) = \infty$ en $k_c(N) = o(\sqrt{N})$. Wanneer $f_N^\star$ benaderd wordt door een continue kansverdeling, dan volgt dat $A_{k_c} \approx \gamma - 1$. Dit leidt tot de approximatieve formule $s \approx \frac{1}{2(\gamma-1)} + \ln\sqrt{2\pi}$. Merk op dat wanneer γ afneemt de mate van niet-equivalentie toeneemt. Hoe vaker punten met grote graad voorkomen, hoe sterker de niet-equivalentie is.

Grafen met een modulaire structuur

Vele variaties op het bovenstaande thema zijn mogelijk. Complexe netwerken hebben doorgaans een *modulaire structuur*, d.w.z. ze vallen uiteen in subnetwerken waarbinnen er relatief veel verbindingen zijn en waartussen er relatief weinig verbindingen zijn (zei Figuur 4). In dat geval is het natuurlijk om voor elk paar subnetwerken randcondities op te leggen aan het aantal verbindingen tussen de subnetwerken. Opnieuw blijkt dat equivalentie wordt gebroken. Zie [2] voor nadere details.



Figuur 4: Een voorbeeld van een graaf met een modulaire structuur.

Referenties

[1] D. Garlaschelli, F. den Hollander, A. Roccaverde, Complexe netwerken vanuit fysisch perspectief, Nieuw Archief voor Wiskunde 5/16 (2015) 207.

[2] D. Garlaschelli, F. den Hollander, A. Roccaverde, Ensemble nonequivalence in random graphs with modular structure, to appear in J. Phys. A: Math. Gen.

[3] J.W. Gibbs, *Elementary Principles of Statistical Mechanics*, Yale University Press, New Haven, Connecticut, 1902.

[4] R.W. van der Hofstad, *Random Graphs and Complex Networks*, Cambridge University Press, 2016.

[5] E.T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (1957) 620.

[6] T. Squartini, J. de Mol, F. den Hollander, D. Garlaschelli, Breaking of ensemble equivalence in networks, Phys. Rev. Lett. 115 (2015) 268701.

[7] H. Touchette, General equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels, J. Stat. Phys. 159 (2015) 987.

Bionote

Frank den Hollander is hoogleraar wiskunde in Leiden. Na zijn PhD in 1985 was hij werkzaam aan de universiteiten van Delft, Utrecht, Nijmegen en Eindhoven. Zijn onderzoeksinteresse gaat uit naar de stochastiek, de mathematische fysica, de mathematische biologie en complexe netwerken. Hij is auteur van 150 wetenschappelijke artikelen en 3 monografieën. Zijn onderzoek wordt ondersteund door een ERC Advanced Grant 2011-2016 en een Zwaartekracht Grant 2014-2023.