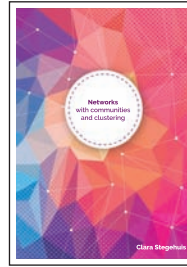


In de verdediging

| In defence

Pas gepromoveerden brengen hun werk onder de aandacht. Heeft u tips voor deze rubriek of bent u zelf pas gepromoveerd? Laat het weten aan onze redacteur.

Redacteur: Nicolaos Starreveld
 FNWI, Universiteit van Amsterdam
 Postbus 94214
 1090 GE Amsterdam
verdediging@nieuwarchief.nl



Networks with communities and clustering

Clara Stegehuis

Op 31 januari 2019 promoveerde Clara Stegehuis cum laude aan de Technische Universiteit Eindhoven bij prof.dr. Remco van der Hofstad en prof.dr. Johan van Leeuwen. De titel van haar proefschrift is *Networks with communities and clustering*. Naast haar onderzoek geeft Clara regelmatig lezingen over de wiskunde voor een algemene publiek en voor middelbare scholieren. Vorig jaar won Clara de KWG-prijs voor promovendi voor de beste onderzoekspresentatie!

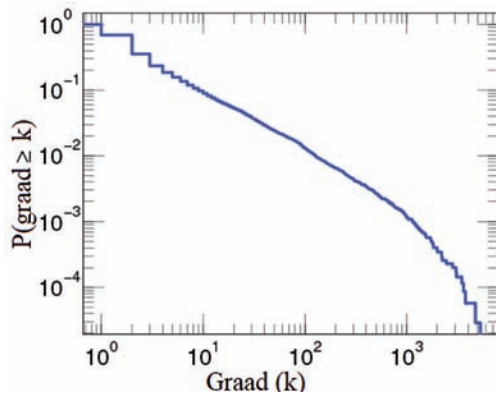
Netwerken zijn overal. Je kunt bijvoorbeeld denken aan sociale netwerken, het internet of je brein. Ondanks dat dit heel verschillende voorbeelden zijn, lijken de verbindingspatronen van deze netwerken vaak veel op elkaar. Veel netwerken hebben bijvoorbeeld een duidelijke groepsstructuur. Groepsleden zijn dan sterk met elkaar verbonden, maar verbindingen tussen verschillende groepen zijn schaars. Een andere eigenschap die vaak wordt waargenomen in netwerken is dat twee bureaus van een knoop waarschijnlijk ook aan elkaar verbonden zijn. Denk bijvoorbeeld aan sociale netwerken, er is een grote kans dat twee van je vrienden ook elkaar goed kennen!

Random grafen

Netwerken worden vaak gemodelleerd met zogenaamde random grafen. Random grafen zijn wiskundige modellen die grote netwerken genereren, vervolgens worden die gebruikt als nulmodellen om echte netwerken te bestuderen. In zo'n model is de structuur van de graaf niet deterministisch, maar stochastisch. Het eenvoudigste voorbeeld is de zogenaamde Erdős-Rényi-graaf, waarin twee knopen zijn verbonden met kans p . Dit model werd geïntroduceerd in 1959 door wiskundigen Paul Erdős en Alfréd Rényi. Het ER-model wordt tegenwoordig niet zo vaak meer gebruikt in de praktijk, voornamelijk omdat de netwerken die het genereert te eenvoudig zijn ter vergelijking met echte complexe netwerken. Dit kunnen we aan de hand van het volgende voorbeeld zien: neem een graaf met een verzameling knopen V en stel dat twee knopen verbonden aan elkaar zijn met kans p . Stel dat $S, T \subset V$. Dan heb je in zo'n graaf tussen S en T gemiddeld $p|S||T|$ verbindingen, en dit is alleen afhankelijk van de grootte van S en T . Dus in een ER-random graaf verwachten we geen lokale structuren te zien, bijvoorbeeld groepen! In tegenstelling tot echte complexe netwerken waar we vaak ingewikkelde lokale structuren tegenkomen!

Schaalvrije netwerkmodellen

Veel complexe netwerken bevatten *hubs*, dit zijn knopen die significant meer bureaus hebben dan de andere knopen in het netwerk. Een typisch voorbeeld zijn sociale netwerken. In een sociaal netwerk zijn er vaak mensen die centraal staan, dat wil zeggen dat ze veel



Figuur 1 Log-loggrafiek van de graadverdeling.

meer mensen kennen dan gemiddeld. In een graaf definiëren we de graad van een knoop als het aantal burens van die knoop. In Figuur 1 zetten we het aantal knopen met graad ten minste k uit tegen k , in een graaf die hubs bevat. De grafiek is getekend op een log-log-schaal. Er is een klein aantal knopen met duizenden burens terwijl de meeste knopen maar weinig burens hebben.

In de complexe netwerktheorie zegt men dat de graadverdeling een zware staart heeft! Dit wordt wiskundig bestudeerd met het gebruik van grafenmodellen met een machtsverband als verdeling. In dit geval, als p_k het aantal knopen met graad k voorstelt, dan geldt

$$p_k \approx Ck^{-\tau}$$

met $C > 0$ constant en τ de graad-exponent. De kans dat een knoop graad minstens k heeft schaalst dan als $k^{1-\tau}$. In een log-log-grafiek krijgen we dan ongeveer een rechte lijn, zoals in Figuur 1!

Netwerkstructuren

Een van de kerndoelen van Clara's onderzoek was de structuur van random grafen te bestuderen en te begrijpen. Bovendien wilde Clara ook bekijken of de verschillende netwerkeigenschappen en structuren die men in echte netwerken waarneemt terugkomen in die random grafenmodellen.

Een belangrijk deel van haar onderzoek gaat over netwerken met een groepsstructuur. Traditionele random grafenmodellen creëren echter netwerken die geen groepsstructuren bevatten. Clara heeft daarom een nieuw random grafenmodel geïntroduceerd, het Hiërarchische configuratie model (HCM), die deze groepsstructuren wel meeneemt. Door analyses van dit nieuwe model te vergelijken met traditionele random grafen zonder groepsstructuren onderzocht ze de invloed van groepsstructuren op hoe iets zich over een netwerk verspreidt. Denk aan virussen, nieuws of roddels bijvoorbeeld!

Clara's analyse en haar resultaten blazen nieuwe ideeën in de complexe netwerktheorie! Tot nu toe werd gedacht dat groepsstructuren weliswaar zorgen voor makkelijkere verspreiding binnen de groep, maar leiden tot bottlenecks voor de verspreiding over het gehele netwerk. Uit haar werk blijkt nu dat dit niet altijd het geval is. Het kan twee kanten op: sommige groepsstructuren maken het makkelijker voor een proces om zich over het netwerk te verspreiden, terwijl andere vormen van groepsstructuren het juist moeilijker maken!

Groepen zoeken in Parijs

De invloed van groepsstructuren op hoe iets zich over een netwerk verspreidt is niet het enige dat we willen weten. We zouden voor

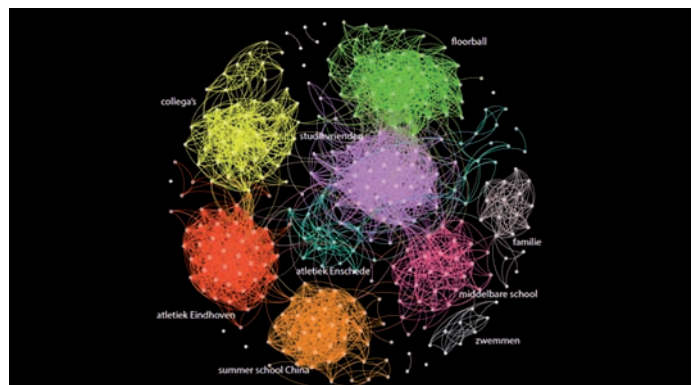
dit soort netwerken graag ook in één keer willen weten waar de groepen zich bevinden! Dit probleem wordt in de wiskunde *community detection* genoemd. Dit blijkt een erg moeilijk probleem te zijn. Er zijn namelijk ontzettend veel verschillende manieren om een netwerk op te delen in groepen. Een netwerk van twintig punten kun je bijvoorbeeld al op honderd miljard verschillende manieren in groepen opdelen! Tijdens haar promotietraject is Clara voor vier maanden naar Parijs gegaan om te werken aan dit probleem. Ze onderzocht hoe nauwkeurig een specifiek *community detection*-algoritme is in het stochastische blockmodel met twee groepen.

Kleine subgrafen

Naast groepsstructuren onderzocht ze ook kleine subgrafen van netwerken. De driehoek komt in echte netwerken vaak voor, bijvoorbeeld doordat op social media twee van je vrienden vaak ook met elkaar bevriend zijn. Ze onderzocht de aanwezigheid van driehoeken en andere subgrafen in verschillende random grafenmodellen. Bovendien was het van belang in haar onderzoek om te weten waar in het netwerk deze subgrafen het meest voorkomen. Clara vond ook dat subgrafen vaak een 'optimale structuur' hebben. Driehoeken zijn in veel random grafenmodellen van grootte n vaak te vinden op punten met graad ongeveer \sqrt{n} . Andere subgrafen komen juist voor op punten met andere graden. Voor iedere subgraaf kunnen we classificeren wat de meest voorkomende structuur van deze subgraaf in het grotere netwerk is.

Faces of Science

Faces of Science is een online-plek waar bezoekers aan de hand van filmpjes, blogs en artikelen over onderzoek een kijkje in de wereld van de onderzoeker krijgen. Doel is promovendi uit verschillende vakgebieden in de schijnwerpers te zetten en vooral aan jongeren die voor een studiekeuze staan, te laten zien wat het leven als wetenschapper inhoudt. Faces of Science is een project van de KNAW, De Jonge Akademie en NEMO Kennislink. Clara is ook een van de Faces of Science. Hierbij hoort ook dat ze regelmatig gastlessen geeft voor middelbare scholieren over leuke toepassingen van wiskunde, of lezingen of theatervoorstellingen voor een algemeen publiek hield. Met bloggen voor Faces of Science probeert Clara aan een algemeen publiek te laten zien dat er "superveel leuke toepassingen zijn van wiskunde". Verspreidingsprocessen over netwerken hebben bijvoorbeeld belangrijke toepassingen, zoals het viral gaan van memes of de verspreiding van een griepvirus. Maar op haar blog zijn nog veel meer leuke toepassingen van netwerken te vinden, van wapenhandel tot muesli tot Sinterklaas. ☺...



Figuur 2 Sociaal netwerk, we zien de verschillende groepen.