Gewone grafen kent u waarschijnlijk wel; een netwerk van punten (knopen) en lijnen (kanten) dat van alles voor kan stellen. Wanneer een kant van een graaf er soms wel is en soms niet, kan daar een stochastische graaf van gemaakt worden. **Ionica Smeets** en **Gerard Hooghiemstra** laten u zien hoe dat gaat.

Stochastische grafen in alledaagse modellen

Inleiding

Stochastische grafen zijn grafen waarbij het aantal kanten bepaald wordt door kansverdelingen. Deze grafen zijn belangrijk omdat ze in veel actuele problemen als model gebruikt worden. Bijvoorbeeld om te bepalen hoe lang data erover doen om van de ene naar de andere computer te komen via het internet. Ook veel sociale netwerken kunnen met een stochastische graaf gemodelleerd worden. Voorbeelden zijn de vriendschappen van kinderen op een lagere school; welke acteurs met elkaar in een film gespeeld hebben of de seksuele contacten van volwassenen. Dat laatste is belangrijk om een model te kunnen ontwikkelen voor de verspreiding van HIV. In dit artikel laten we enkele eigenschappen zien van twee belangrijke stochastische grafen: de zogenaamde toevalsgraaf en de zogenaamde schaalvrije graaf.

Deterministische grafen

U kent waarschijnlijk wel voorbeelden van modellen met deterministische grafen. Even ter herinnering: een graaf is een verzameling $G = \{V, E\}$, waarin V de verzameling knopen van de graaf is en E de verzameling van kanten tussen deze knopen. Een bekend voorbeeld van een model met een graaf is het handelsreizigersprobleem. Een vertegenwoordiger moet langs verschillende steden om daar zijn handel te verkopen. Hij wil het liefst de kortste route tussen deze steden kiezen, want tijd is natuurlijk geld. Het standaardmodel voor dit probleem is een gewogen graaf met de steden als de knopen. De kanten in de graaf zijn de wegen tussen de steden. Elke kant heeft een gewicht, namelijk de afstand in kilometers van de weg die de kant voorstelt.

Stochastische grafen

Naast deterministische grafen bestaan er ook stochastische grafen, die een meer bruikbaar model geven voor netwerken waarin nog veel onbekend is. Internet is zo'n netwerk. Met internet bedoelen we het netwerk van fysieke verbindingen tussen computers. In dit grote netwerk kunnen we bijvoorbeeld nooit weten welke computers op een specifiek moment met elkaar verbonden zijn. Ook houden providers de bandbreedten van hun verbindingen geheim, zodat het onmogelijk is deze graaf volledig in kaart te brengen.
Het inbrengen van een stochastische factor in een graaf doen we door een kansverdeling te gebruiken. In een stochastische graaf kiezen we met deze kansverdeling welke kanten wel en welke niet tot de graaf behoren. Dat kan op veel manieren. Zoals gezegd zullen we in dit artikel de toevalsgraaf en de schaalvrije graaf behandelen.
Een belangrijke eigenschap van een graaf is het aantal kanten per knoop, we noemen dit de graad van een knoop. Zo worden in het netwerk van kindervriendschappen de knopen gevormd door kinderen, en is de graad van een kind het aantal vriendjes van dat kind. Voor de toevalsgraaf zullen we de verdeling van de graad van een willekeurige knoop berekenen.
Een tweede belangrijke eigenschap is de afstand tussen twee willekeurige knopen in een stochastische graaf. De afstand tussen twee knopen is het aantal kanten in de kortste verbinding. In het internet is deze afstand het aantal verbindingen tussen routers die door de datastroom wordt gepasseerd. Deze afstand geeft een indicatie voor de tijd die data nodig hebben om van de ene naar de andere computer te komen. We zullen bij de schaalvrije graaf kort ingaan op de gemiddelde afstand tussen twee willekeurig gekozen knopen.
We geven vanaf hier elke kant gewicht 1, dan kunnen we de gewichten verder buiten beschouwing laten.

De toevalsgraaf

We maken de toevalsgraaf uit een volledige graaf met N knopen. In een volledige graaf zijn alle knopen met elkaar verbonden. Het aantal kanten in de volledige graaf op N knopen is dus:

$$K = \binom{N}{2} = \tfrac{1}{2}N(N-1)\,.$$

We gebruiken als voorbeeld een graaf met $N = 5$, met dus $K = \binom{5}{2} = 10$ kanten (zie de linkergraaf in figuur 1). Nummer, in willekeurige volgorde, de kanten van 1 tot K. We laten nu kanten weg door elke kant stochastisch een

0 of 1 toe te kennen. We doen dat door bij kant i (i = 1, 2, ..., K) een Bernoulli verdeelde variabele B_i te kiezen. Dat betekent dat:

$$P(B_i = 1) = p \text{ en } P(B_i = 0) = 1 - p.$$

We kiezen de verschillende $B_1, B_2, ..., B_K$ onafhankelijk van elkaar. We krijgen nu een stochastische graaf door alleen de kanten i te houden waarbij $B_i = 1$. Een voorbeeld van een realisatie van een stochastische graaf met $p = \frac{1}{2}$ en $N = 5$ ziet u rechts in figuur 1.

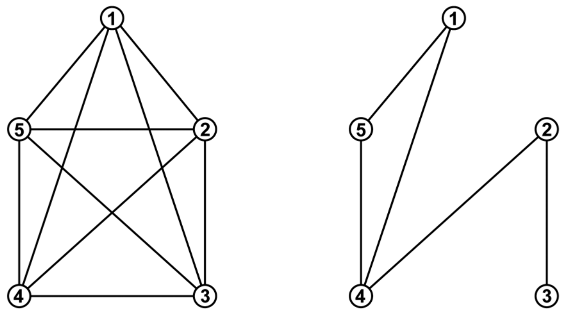

fig. 1 Volledige en toevalsgraaf op 5 knopen

De toevalsgraaf is een eenvoudig model, dat vooral gebruikt wordt om een beter begrip te krijgen van eigenschappen van (grote) netwerken. De toevalsgraaf wordt niet serieus gebruikt als model voor bestaande netwerken, omdat bepaalde eigenschappen van deze netwerken niet overeenstemmen met deze eigenschappen van de toevalsgraaf. We komen hier straks op terug.

Het aantal kanten per knoop

Zoals gezegd is de graad van een knoop in een graaf of netwerk het aantal kanten van deze knoop. In de toevalsgraaf is deze graad een stochast. Voor knoop j noteren we deze stochast als D_j. In de volledige graaf heeft elke knoop graad $N - 1$. Bij elk van deze $N - 1$ kanten hoort een Bernoulli verdeelde variabele. Daarom is in onze toevalsgraaf de graad D van een willekeurige knoop verdeeld als de som van $N - 1$ Bernoulli verdeelde variabelen. Deze verdeling is precies de binomiale verdeling met parameters $N - 1$ en p. De verwachting van het aantal kanten per knoop wordt dan gegeven door $(N - 1)p$. Stel, we kiezen $p = \frac{1}{2}$. Voor de toevalsgraaf op 5 knopen is de verwachting van de graad dan $4 \times \frac{1}{2} = 2$.

Veel realistische netwerken zijn erg groot, met duizenden en soms wel honderdduizenden knopen (denk aan het internet). Bij zulke grote netwerken kijken we vaak naar de limiet voor $N \to \infty$, waarbij N het aantal knopen is. We willen dat de gemiddelde graad van een knoop gelijk blijft; we houden daarom Np constant. We krijgen nu een binomiale verdeling met grote N en kleine p. Zo'n verdeling kunnen we benaderen met de Poissonverdeling:

$$\lim_{\substack{N \to \infty \\ Np \to \lambda}} \mathbf{P}(D = k) = \lim_{\substack{N \to \infty \\ Np \to \lambda}} \binom{N}{k} p (1-p)^{N-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \; k \geq 0 \text{ vast}$$

Overigens is bovenstaande limietovergang niet triviaal. (zie Frank Heierman, Rein Nobel en Henk Thijms, *Poisson, de Pruisen en de Lotto*, Epsilon-uitgaven: Zebrareeks 5 (2000)). De betekenis van de bovenstaande limiet is dat voor grote N de graad van een willekeurige knoop in de toevalsgraaf in benadering een Poissonverdeling heeft.
We komen nu terug op onze bewering dat eigenschappen van veel realistische grafen afwijkend zijn van de overeenkomstige eigenschap in de toevalsgraaf. Een van deze eigenschappen is de graad van een willekeurige knoop. We hebben zojuist gezien dat deze voor grote toevalsgrafen in benadering een Poissonverdeling heeft. We nemen als voorbeeld van een groot realistisch netwerk de AS-domeinen van het internet. De afkorting AS staat voor de Engelse woorden 'Autonomous Systems'. Een AS-domein is een groep routers (dit zijn speciale computers die het dataverkeer op internet controleren en regelen) waartussen het verkeer lokaal wordt geregeld. De computers van een groot bedrijf of universiteit vormen samen vaak een AS-domein. De AS-domeinen corresponderen ruwweg met domeinnamen. Er zijn naar schatting bijna 11.000 AS-domeinen op het internet. In de graaf van het model nemen we als knopen deze AS-domeinen. Hun verbindingen met andere AS-domeinen zijn de kanten van de graaf.
In 1999 verscheen een artikel van de drie broers Faloutsos over internetverbindingen (C. Faloutsos, P. Faloutsos en M. Faloutsos, *On power-law relationships of the internet topology*, Proceedings of the ACM SIGCOMM (1999)). Hun tellingen van het aantal verbindingen per AS-domein laten zien dat het aantal verbindingen van een willekeurig gekozen AS-domein helemaal niet verdeeld is volgens een Poissonverdeling. Dit zou wel het geval moeten zijn als de toevalsgraaf een goed model is. De toevalsgraaf kan dus geen goed model zijn voor het internet op het niveau van AS-domeinen. Deze observatie gaf onderzoekers de impuls om stochastische grafen te ontwikkelen, waarbij de verdeling van de graad van een willekeurige knoop wel overeenkomt met de metingen in *On power-law relationships of the internet topology*. Een van deze modellen is de zogeheten schaalvrije graaf, die we hieronder bespreken.

De schaalvrije graaf

De schaalvrije graaf is een graaf waarbij de graad van de knopen uit een gegeven kansverdeling komt. We maken een model voor dit soort grafen op een heel andere manier dan bij de toevalsgraaf. We beginnen met alleen de N knopen. Voor elke knoop j trekken we zijn graad $g(j)$ uit een vaste discrete verdeling met kansen:

$$p_k \geq 0 \text{ , en } \sum_{k=0}^{\infty} p_k = 1 .$$

De bijbehorende stochastische variabele noteren we met D_j. De trekkingen doen we onafhankelijk van elkaar. We tekenen bij knoop j precies D_j zogenaamde spikes; twee spikes zullen we later verbinden, zodat ze samen een kant vormen. Zie figuur 2.

fig. 2 De knopen met hun spikes

Uit de N knopen, met bij elke knoop het aantal spikes, construeren we de schaalvrije graaf als volgt: we beginnen met de eerste van de D_1 spikes van knoop 1. We loten een tweede spike willekeurig uit de overige spikes. In figuur 3 is dit de eerste spike van knoop 3 geworden. De twee verbonden spikes van knoop 1 en knoop 3 vormen nu samen een kant.

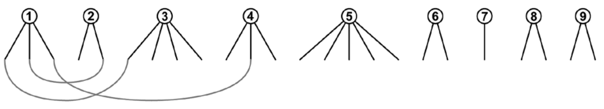

fig. 3 Het verbinden van de spikes

Merk op dat we een trekking doen uit $L_N - 1$ mogelijkheden, waarbij L_N het totale aantal spikes is:
$L_N = D_1 + D_2 + ... + D_N$.
Het is verboden om een spike met zichzelf te verbinden. We gaan nu door met de tweede spike van knoop 1; hiervoor kiezen we willekeurig een verbinding met een van de overgebleven $L_N - 3$ spikes. We mogen immers de eerste twee spikes van knoop 1 en de eerste spike van knoop 3 niet meer kiezen. In figuur 3 wordt de tweede spike van knoop 1 verbonden met de tweede spike van knoop 2. Weer vormen de twee spikes een kant. Zo gaan we door tot alle spikes verbonden zijn, zie figuur 4.

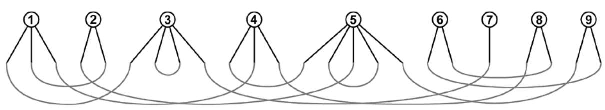

fig. 4 Alle spikes zijn verbonden

De ontstane graaf is getekend in figuur 5. Elke knoop heeft inderdaad de voorgeschreven graad D_j, $j = 1, 2, ..., N$, getrokken uit de verdeling met kansen $\{p_k\}$.

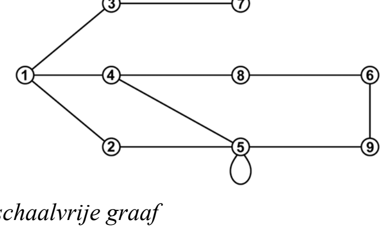

fig. 5 De schaalvrije graaf

We hebben nog een detail weggelaten. In deze constructie moet $L_N = D_1 + ... + D_N$ even zijn. Twee spikes vormen namelijk steeds een kant; we mogen aan het eind geen losse spike overhouden. Als L_N oneven is, lossen we dit probleem op door een extra knoop $N + 1$ met een spike ($D_{N+1} = 1$) toe te voegen. De laatste spike verbinden we met deze ene spike. Het resultaat is dan dat alle knopen op een na de juiste graad hebben. Omdat in de toepassingen N zeer groot is, is de invloed van de ene toegevoegde knoop te verwaarlozen.

Keuze van verdeling $\{p_k\}$ van de graad

We hebben in de vorige paragraaf gezien hoe we een graaf construeren, waarbij de graad van iedere knoop een voorgeschreven kansverdeling heeft. We kiezen nu als verdeling:

$p_k = \text{constante} \cdot k^{-\tau}$, $\tau > 1$

waarbij de constante zo gekozen wordt dat:

$$\sum_{k=1}^{\infty} p_k = 1$$

De graad van elke knoop in de schaalvrije graaf is dan verdeeld volgens $\{p_k\}$. Dus het aantal knopen met graad k zal bij een totaal van N knopen gemiddeld gelijk zijn aan:
$N \cdot p_k = \text{constante} \cdot N \cdot k^{-\tau} = N_0 k^{-\tau}$, $N_0 = \text{constante} \cdot N$.
Volgens de metingen in *On power-law relationships of the internet topology* komt voor $\tau = 2.25$ het aantal AS-domeinen met k verbindingen heel nauwkeurig overeen met $N_0 k^{-\tau}$. Met andere woorden, de schaalvrije graaf met verdeling $\{p_k = \text{constante} \cdot k^{-2.25}\}$ is wat de graad van de knopen betreft een uitstekend model voor het netwerk van het internet op het level van de AS-domeinen.

De afstand tussen twee willekeurige knopen

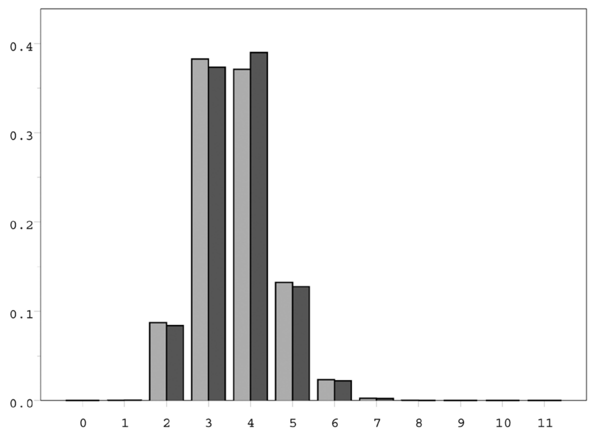

fig. 6 Genormaliseerd histogram van afstanden tussen twee AS-domeinen met $\tau = 2.25$ en $N = 10.900$; links de metingen, rechts de simulatie.

Is de schaalvrije graaf nu ook een goed model voor de AS-domeinen van het internet wat betreft de tweede be-

langrijke eigenschap van stochastische grafen: de afstand tussen twee willekeurige knopen? Dat blijkt ook zo te zijn. Simulaties van afstanden tussen twee willekeurige knopen in de schaalvrije graaf (met $N \approx 10.000$ en $\{p_k\}$ als boven met $\tau = 2.25$) laten zien dat de verdeling van deze afstanden ook perfect overeenkomt met gemeten afstanden, zie figuur 6.
We zien in figuur 6 dat de afstand tussen twee knopen in een groot netwerk verbazingwekkend klein is. In een netwerk met 11000 domeinen beslaat een pad tussen twee willekeurige adressen gemiddeld slechts drie of vier domeinen. Deze relatief kleine afstand tussen twee willekeurige knopen staat bekend als het *kleine wereld effect*. Ook in veel sociale netwerken zijn de verbindingen kort. De kans is bijvoorbeeld groot dat u een hand heeft gegeven aan iemand die een hand heeft gegeven aan iemand die een hand heeft gegeven aan Koningin Beatrix. In het handenschudnetwerk van Nederland bent u waarschijnlijk via hooguit zes tussenpersonen verbonden met onze koningin.

Wilt u meer lezen over dit onderwerp, dan kunnen wij het overzichtsartikel M.E.J. Newman, *The structure and function of complex networks*, SIAM Review 45, 167-256 (2003) aanbevelen.

Met dank aan Geert Donker voor de illustraties, Hongsuda Tangmunarunkit voor het beschikbaar stellen van de internetdata in figuur 6, Dmitri Znamenski voor de simulatie van afstanden in de schaalvrije graaf van eveneens figuur 6, Pim Hooghiemstra, Swier Garst en Remco van der Hofstad voor het proeflezen en bruikbare aanwijzingen.

Ionica Smeets en Gerard Hooghiemstra
Faculteit EWI, Technische Universiteit, Delft

Na twintig jaar op papier: Uitwiskeling live!

Al twintig jaar confronteert het tijdschrift *Uitwiskeling* de wiskundeleerkracht in Vlaanderen met nieuwe ideeën uit de wiskundedidactiek. Om de afsluiting van dit tweede en de aanhef van het derde decennium te vieren, organiseert *Uitwiskeling* op zaterdag 20 november een jubileumviering in de gebouwen van de Ehsal, op wandelafstand van het centraal station in Brussel.

Feesten en werken gaan hand in hand. Daarom wordt aan de meevierende leerkrachten gevraagd tekenmateriaal en een grafische zakrekenmachine mee te brengen en zich op voorhand in te schrijven voor drie workshops. Op de website www.uitwiskeling.be kunnen belangstellenden een elektronische inschrijfstrook vinden met tien keuzeonderwerpen voor deze workshops, bv. 'rekenen met tandwielen', 'is de Belg een bedreigde diersoort?', 'wiskunde en aardrijkskunde', 'meetkundig denken rond een cirkel'…
Bij de selectie van de onderwerpen voor deze seminaries is er rekening gehouden met de onmiddellijke bruikbaarheid in de klas, met de nieuwe leerplannen en met de evenwichtige spreiding over de drie graden.

Als afsluiter van deze (inter-)nationale feestdag werd de Schotse vuurspuwer-éénwieler-jongleur-en-doctor-in-de-zuivere-wiskunde (Cambridge University) Collin Wright uitgenodigd. Of zijn activiteiten direct toepasbaar zullen zijn in onze klassen kan niet waterdicht en vuurbestendig door de organisatoren verzekerd worden.