



Universiteit  
Leiden  
The Netherlands

## Probabilistic graph inspections through forests

Koperberg, V.T.

### Citation

Koperberg, V. T. (2026, June 25). *Probabilistic graph inspections through forests*. Retrieved from <https://hdl.handle.net/1887/4307047>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4307047>

**Note:** To cite this publication please use the final published version (if applicable).

*I must acknowledge myself to be indeed a very backward scholar; since I cannot now discover an argument which, it seems, was perfectly familiar to me long before I was out of my cradle.*

- Hume; An Enquiry Concerning Human Understanding

# Samenvatting

*Grafen* zijn wiskundige representaties van netwerken. Denk als voorbeelden van netwerken aan het wegnnet bestaande uit splitsingen en kruisingen die met wegen zijn verbonden, sociale netwerken van mensen verbonden door vriendschapsrelaties of aan het netwerk van neuronverbindingen dat zich in de hersenen bevindt. Grafen modelleren deze netwerken als een verzameling punten, *knopen* genaamd, met een bijbehorende verzameling *lijnen*, waarvan elke lijn twee knopen met elkaar verbindt. De lijnen duiden de verbindingen aan die in het netwerk voorkomen. Het analyseren van de wiskundige eigenschappen van grafen kan inzichten verschaffen in vraagstukken waarin een netwerkstructuur voorkomt. Voorbeelden van dergelijke vraagstukken zijn of het sluiten van een specifieke weg verkeersopstoppingen zal veroorzaken of juist zal verhelpen, of hoe snel een gerucht of virus zich door een groep mensen zal verspreiden. Echter, voor grafen met een groot aantal knopen en lijnen kan het oplossen van vraagstukken dusdanig veel berekeningen vergen, dat dit zelfs met behulp van moderne computers niet haalbaar is.

*Bossen* zijn een speciaal type graaf met een simpele structuur: voor alle mogelijke tweetallen knopen bestaat er hoogstens één pad, een aaneenschakeling van lijnen, tussen deze twee knopen, oftewel een bos bevat geen *kringen*. De knopen van een bosgraaf kunnen worden onderverdeeld in één of meerdere groepen genaamd *samenhangscomponenten*, waarbij knopen tot verschillende componenten behoren wanneer ze niet verbonden zijn via een pad van lijnen. Een samenhangscomponent van een bos wordt ook wel een *boom* genoemd. De terminologie 'bossen' en 'bomen' is afgeleid van de grafische representatie van dergelijke grafen. Een afbeelding van een bosgraaf lijkt enigszins op een verzameling bomen met een stam en met takken. Wegens hun relatief simpele structuur zijn bossen gemakkelijker te analyseren dan gecompliceerdere grafen. Een complexe graaf bevat meerdere deelgrafen met een bosstructuur. Deze deelgrafen kunnen worden verkregen uit de oorspronkelijke graaf door alle knopen te behouden en sommige van de lijnen te verwijderen. Een dergelijke bosdeelgraaf bevat informatie over de oorspronkelijke graaf, en zou hierdoor kunnen helpen bij het ofwel exact berekenen ofwel benaderen van de relevante eigenschappen van de graaf. Echter, in veel gevallen gaat in de reductie van complexe graaf tot bosgraaf te veel informatie verloren. Een mogelijke oplossing is om random bosdeelgrafen te gebruiken. De complexiteit van de graaf kan dan deels worden gevat in de kansverdeling van het random bos, en niet alleen in de structuur van het bos zelf.

Het bovengenoemde idee om ingewikkelde grafen te benaderen met random bosdeelgrafen vormt een belangrijke motivatie voor het onderzoek in dit werk. In het eerste deel van dit proefschrift bestuderen we een specifieke kansverdeling op bosdeelgrafen, de zogenaamde *Kirchhoff bosgraafkansverdeling*. Deze kansverdeling leent zich bij uitstek voor mogelijke toepassingen in de analyse van complexe netwerken, door

het bestaan van een efficiënte methode, genaamd *Wilson's algoritme*, waarmee een steekproef met deze kansverdeling genomen kan worden. In dit huidige werk worden Kirchhoff bosgrafen niet direct gebruikt voor het analyseren van complexe netwerken. In plaats daarvan wordt gepoogd een beter theoretisch begrip te krijgen van Kirchhoff bosgrafen, waardoor mogelijk toegepaste technieken in de netwerkanalyse zouden kunnen worden ontwikkeld op basis van dit fundamentele onderzoek.

Wilson's algoritme kan niet alleen worden gebruikt als methode om efficiënt steekproeven mee te nemen. Het algoritme gebruikt *random walks* op de oorspronkelijke graaf om een random bosgraaf volgens de Kirchhoff verdeling te construeren. De kringen die in het traject van een random walk voorkomen worden verwijderd, zodat de overgebleven doorlopen lijnen een bos vormen. Hierdoor kan Wilson's procedure ook gebruikt worden als theoretisch hulpmiddel om vragen over Kirchhoff bosgrafen te vertalen naar vragen over random walks, die eenvoudiger te bestuderen zijn.

De Kirchhoff bosgraafverdeling hangt af van een *intensiteitsparameter*, die kan worden aangepast om het verwachte aantal lijnen in een Kirchhoff bos te bepalen. In hoofdstuk 2 bestuderen we de kans dat twee knopen onderdeel zijn van dezelfde boom in een Kirchhoff bos, en in het bijzonder hoe deze kans afhangt van de intensiteitsparameter.

Hoofdstukken 3 en 4 richten zich op een ander aspect van Wilson's algoritme dan het resulterende bos. In hoofdstuk 3 ligt de nadruk op het *bezettingsveld* van Wilson's procedure, dat voor elke knoop telt hoe vaak deze in totaal door alle random walks is bezocht. Dit bezettingsveld is nauw verwant aan het bezettingsveld van een *random walk loop-soup*, en aan andere modellen in de statistische fysica zoals de *discrete Gaussian free field*. Door niet alleen te kijken naar het bezettingsveld maar naar de verzameling verwijderde kringen, worden in hoofdstuk 4 worden de resultaten uit hoofdstuk 3 uitgebreid. Hier laten we zien dat door Kirchhoff bosverdelingen met verschillende intensiteiten te *koppelen*, deze gekoppelde variant van Wilson's algoritme kan worden gebruikt om een loop-soup te construeren.

Twee kansverdelingen zijn niet noodzakelijkerwijs *onafhankelijk* van elkaar. Als los van elkaar een zeszijdige dobbelsteen en een twaalfzijdige dobbelsteen worden gegooid, dan zijn beide uitkomsten onafhankelijk. Dat wil zeggen, het resultaat van de ene worp geeft geen informatie over het resultaat van de andere. Het is echter mogelijk om de kansverdeling van een zeszijdige dobbelsteen na te bootsen door de uitkomst van de twaalfzijdige dobbelsteen te halveren en naar boven af te ronden. De verdeling van de nagebootste zeszijdige dobbelsteen is dan precies hetzelfde als die van een fysieke zeszijdige dobbelsteen, maar is niet langer onafhankelijk van de uitkomst van de twaalfzijdige. Dit is een voorbeeld van een *koppeling* van twee kansverdelingen, een algemene methode in de kansrekening om meerdere verdelingen gezamenlijk te beschouwen.

De *stelling van Strassen* is een vermaard kanstheoretisch resultaat dat aangeeft of twee verdelingen dusdanig te koppelen zijn dat deze koppeling een aantal gewenste eigenschappen heeft. In hoofdstuk chapter 5 wordt een nieuw eenvoudig bewijs van deze stelling gegeven voor het specifieke geval dat beide verdelingen een eindig aantal uitkomsten hebben. Dit bewijs vertaalt het koppelingsprobleem naar een grafentheo-

retisch probleem, en laat zien dat alle relevante informatie van de resulterende graaf gevat kan worden in één enkele bosdeelgraaf. Verder belicht dit bewijs de connectie tussen Strassens stelling en de *huwelijksstelling van Hall*, een bekend resultaat dat betrekking heeft op *matchings* in bipartiete grafen.