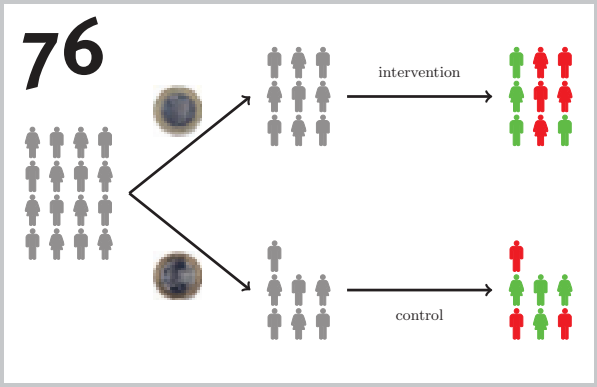
Inhoud | Contents

Een interview met Luc Illusie over de gouden periode van de algebraïsche meetkunde in de jaren zestig.



In zijn oratie aan de UvA sprak Joris Mooij over de relaties tussen oorzaken en hun gevolgen.

Colofon

Het Nieuw Archief voor Wiskunde is een uitgave van het Koninklijk Wiskundig Genootschap en verschijnt vier maal per jaar. Het tijdschrift richt zich op eenieder die zich beroepsmatig met wiskunde bezighoudt, als academisch of industrieel onderzoeker, student, leraar, journalist of beleidsmaker. Het stelt zich ten doel te berichten over ontwikkelingen in de wiskunde in het algemeen en in de Nederlandse wiskunde in het bijzonder.

ISSN 0028-9825

Adres
Nieuw Archief voor Wiskunde
Centrum Wiskunde & Informatica
Postbus 94079, 1090 GB Amsterdam
tel. 020-5924288
redactie@nieuwarchief.nl
www.nieuwarchief.nl

Hoofdredactie
Raf Bocklandt (r.r.j.bocklandt@uva.nl)
Robbert Fokkink (r.j.fokkink@tudelft.nl)

Eindredactie/Bladmanager
Fred Drissen (fred@nieuwarchief.nl)

Redactie
Francesca Arici (UL), Viktor Blåsjö (UU), Hans Cuypers (TU/e), Teun Koetsier (VU), Nikki Levering (UvA), Daan Mulder, Nicolaos Starreveld (UvA), Clara Stegehuis (UT), Hans Sterk (TU/e), Mark Timmer (UT)

Problemenrubriek (problems@nieuwarchief.nl)
Onno Berrevoets en Daan van Gent

Bureauredactie
Anne van Grinsven

Illustraties
Ryu Tajiri

Vormgeving
Susanne Laws (omslag, begeleiding binnenwerk)
Kitty Molenaar (ontwerp)

Druk
Veldhuis Media

Advertenties
advertenties@nieuwarchief.nl

Koninklijk Wiskundig Genootschap
www.wiskgenoot.nl

Platform Wiskunde Nederland
www.platformwiskunde.nl

European Mathematical Society
www.euro-math-soc.eu

International Mathematical Union
www.mathunion.org

Koninklijk Wiskundig Genootschap
Het Koninklijk Wiskundig Genootschap (KWG) is een landelijke vereniging van beoefenaren van de wiskunde. In 1778 opgericht onder de zinspreuk: 'Een onvermoeide arbeid komt alles te boven' is het 's werelds oudste nationale wiskundegenootschap. Leden ontvangen het Nieuw Archief voor Wiskunde als onderdeel van hun lidmaatschap.

Bestuur
Mathisca de Gunst (VU, voorzitter), Alef Sterk (RUG, penningmeester), Jop Briët (CWI, secretaris), Barry Koren (TU/e, vice-voorzitter), Danny Beckers (VU), Rogier Bos (UU), Charlene Kalle (UL), Michael Müger (RU).

Secretariaat
Koninklijk Wiskundig Genootschap
Centrum Wiskunde & Informatica
Postbus 94079, 1090 GB Amsterdam
wiskgenoot@wiskgenoot.nl

Ledenadministratie KWG / Abonnementen
admin@wiskgenoot.nl, tel. 020-5924193
Wijzigingen of opzegging kunt u ook doorgeven door in te loggen op www.wiskgenoot.nl.

Contributie
Tarieven per jaar (bij betaling per automatische incasso krijgt u € 2,50 korting):
– gewoon lid: € 97,50
– reciprociteitslid: € 77,50
– gepensioneerden en werklozen: € 52,50
– aio's, oio's, studenten: € 42,50 (max. 4 jaar)
– lid zonder Nieuw Archief: € 57,50
Studenten en pas afgestudeerden kunnen een jaar lang gratis lid worden.
Voor verzending van het Nieuw Archief naar het buitenland wordt een portotoeslag van € 10 in rekening gebracht.
Instituten in Nederland en buitenland kunnen zich abonneren op het Nieuw Archief voor Wiskunde voor € 105 (Nederland), € 115 (Europa) of € 117,50 (buiten Europa).

Members of foreign societies
Members of *SIAM*, the *Société Mathématique de France*, the *Gesellschaft für Angewandte Mathematik und Mechanik*, the *Deutsche Mathematiker-Vereinigung*, and of the *American*, *Australian*, *Belgian*, *Indian*, *London*, *Spanish*, and *South-African Mathematical Societies* living outside of the Netherlands pay € 77,50 instead of the full membership fee of € 97,50 a year. Conversely, these foreign societies, as well as the VvS+OR and the NVORWO, have reduced membership fees for KWG members. A postage charge of € 10 is added for members living abroad but in Europe, and a charge of € 12,50 for members living outside of Europe.

Exchange subscriptions
Koninklijk Wiskundig Genootschap Library
Centrum Wiskunde & Informatica
P.O. Box 94079, NL-1090 GB Amsterdam
KWG-exchange@cwi.nl

Informatie voor auteurs

Kopij
Kopij dient per e-mail te worden aangeboden aan kopij@nieuwarchief.nl. Voor meer informatie en auteursinstructies zie www.nieuwarchief.nl.

Volgende nummers

nummer	*maand verschijnen*	*deadline kopij*
24(3)	september 2023	verstreken
24(4)	december 2023	01-08-2023
25(1)	maart 2024	01-11-2023
25(2)	juni 2024	01-02-2024

Instituuts- en bedrijfsleden

De publicaties van het KWG worden mede mogelijk gemaakt door de bijdragen van de volgende instituten en bedrijven.

Centrum Wiskunde & Informatica

Universiteit van Amsterdam

Vrije Universiteit Amsterdam

Rijksuniversiteit Groningen

Universiteit Leiden

Radboud Universiteit Nijmegen

Universiteit Utrecht

Technische Universiteit Delft

Technische Universiteit Eindhoven

Universiteit Twente

Elsevier

ORTEC

Op het omslag

Als je veel chocolade eet, nemen dan je cognitieve vaardigheden toe? Veroorzaakt roken longkanker? Heb je met het nieuwe covid-19-vaccin minder kans om in het ziekenhuis opgenomen te worden? In zijn oratie geeft Joris Mooij duiding aan dit soort causale verbanden, zie pagina 76.

Krasse knarren

Wiskunde is voor de meesten onder ons niet enkel een beroep maar ook een passie, en vele wiskundigen blijven vaak actief na hun pensioen. Het hoeft dan ook niet te verbazen dat er in de wiskundewereld heel wat beroemde gezichten zijn die actief bleven tot ver na hun tachtigste, negentigste en soms zelfs honderdste verjaardag.

De recordhouder onder deze mathematische senioren is veruit Leopold Vietoris, die met zijn 110 levensjaren niet enkel de oudste wiskundige was, maar tevens de oudste inwoner van zijn geboorteland Oostenrijk. Meer dan de helft van zijn publicaties schreef hij na zijn zestigste en er verschenen nog artikelen van hem toen hij de honderd al voorbij was. Hoewel hij reeds twintig jaar geleden gestorven is, is zijn wiskundige erfenis nog springlevend, want al wie algebraïsche topologie bestudeert, is bekend met zijn fameuze lange exacte rij om homologiegroepen te berekenen.

In Frankrijk is er Henri Cartan, die 104 werd, maar ook heel wat van de jongere garde van de Bourbaki-groep zoals André Weil, Roger Godement, Jean-Louis Koszul en Jean-Pierre Serre rondden met gemak de kaap van 90. Driekwart eeuw later, is het misschien moeilijk om deze gevestigde waarden voor te stellen als enthousiaste jongelingen, maar in zijn interview met Nieuw Archief gunt Luc Illusie, inmiddels zelf ook een actieve tachtiger, een blik op deze gouden periode uit de Franse wiskunde.

In de Verenigde Staten zijn er eveneens enkele beroemde wiskundigen die actief bleven tot op hoge leeftijd. Saunders Mac Lane, mede-uitvinder van de categorietheorie, overleed op 95-jarige leeftijd, de bekende John Milnor, die onder andere het bestaan van exotische sferen bewees, is nog steeds actief op zijn 91ste en onlangs werd Eugenio Calabi, emeritus aan de Universiteit van Pennsylvania en beroemd om de naar hem vernoemde Calabi–Yau-manifolds, honderd jaar.

Ook in Nederland zijn we gezegend met een paar wiskundige eeuwelingen. Zo bereikte Dirk Jan Struik, nog een differentiaalmeetkundige en expert in de geschiedenis van de wiskunde, de begenadigde leeftijd van 106. Afgelopen januari vierden we de honderdste verjaardag van Jaap Korevaar, mijn beminnelijke collega aan de UvA. Tot voor enkele jaren zag ik hem nog vaak aan de koffieautomaat in het instituut, en voor de gelegenheid van zijn eeuwwisseling trokken Jan Wiegerinck en ik samen naar Bussum voor een interview dat je in dit nummer kunt lezen.

Wiskunde doen is misschien niet een noodzakelijke of voldoende voorwaarde voor een lang en gezond leven, maar het houdt je wel fris en monter. Of, om een Engels spreekwoord te parafraseren:

A theorem a day keeps the doctor away.

Raf Bocklandt, hoofdredacteur
Korteweg-de Vries Instituut, Universiteit van Amsterdam

Agenda

| Upcoming Events

juni 2023

15 juni 2023

❑ **Basisvaardigheden rekenen-wiskunde**

Conferentie van NVvW en Platform Rekenbewust Vakonderwijs voor docenten van onderbouw vmbo en havo/vwo over de basisvaardigheden rekenen-wiskunde in alle vakken.

plaats *Domstad, Utrecht*

info *nvvw.nl*

19–23 juni 2023

❑ **Singularities and Curvature in General Relativity**

Workshop over algemene relativiteit, waarin nieuwe technieken en open problemen worden besproken.

plaats *Radboud Universiteit Nijmegen*

info *www.math.ru.nl/SCGR2023*

23 juni 2023

❑ **Wil's Mathematical Odyssey, so far**

Feestelijk symposium ter ere van Wil Schilders, die met pensioen zal gaan, en tijdens dit symposium ook zijn afscheidsrede zal geven.

plaats *TU Eindhoven*

info *casa.win.tue.nl*

26–27 juni 2023

❑ **NDNS+ Workshop Twente**

Tweedaagse workshop van het cluster NDNS+ (Nonlinear Dynamics of Natural Systems).

plaats *Universiteit Twente*

info *utwente.nl*

28 juni–14 juli 2023

❑ **IMAGINARY**

Reizende tentoonstelling over zichtbare en onzichtbare wiskunde. Georganiseerd door PWN in samenwerking met Nederlandse universiteiten en ondersteund door partnerorganisaties.

plaats *Universiteit Maastricht*

info *imaginarymaths.nl*

juli 2023

3–7 juli 2023

❑ **Analysis and Numerics of Nonlinear PDEs: Degeneracies and Free Boundaries**

Lorentz workshop, mede georganiseerd door Manual Gnann (TU Delft), Joost Hulshof (VU) en Stefanie Sonner (RU).

plaats *Lorentz Center@Snellius, Leiden*

info *lorentzcenter.nl*

3–5 juli 2023

❑ **Motivic and Non-Communitative Aspects of Enumerative Geometry**

Conferentie ter ere van de zestigste verjaardag van Bjørn Ian Dundas.

plaats *Radboud Universiteit Nijmegen*

info *www.math.ru.nl/mnc-conference*

5–7 juli 2023

❑ **Homotopy Theory, K-theory, and Trace Methods**

Workshop over het gebruik van invarianten en algebraïsche variëteiten voor het bestuderen van enumeratieve vraagstukken.

plaats *Radboud Universiteit Nijmegen*

info *www.math.ru.nl/hkt-conference*

6–13 juli 2023

❑ **International Mathematical Olympiad (IMO)**

Internationale wiskundecompetitie voor middelbare scholieren, dit jaar in Japan.

plaats *Chiba, Japan*

info *imo2023.org*

10–14 juli 2023

❑ **Machine-Checked Mathematics**

Lorentz workshop, mede georganiseerd door Sander Dahmen (VU).

plaats *Lorentz Center@Oort, Leiden*

info *lorentzcenter.nl*

17–21 juli 2023

❑ **A Multidisciplinary Approach to Epistasis Detection**

Lorentz workshop, mede georganiseerd door Marleen Balvert (Tilburg)

plaats *Lorentz Center@Oort, Leiden*

info *lorentzcenter.nl*

31 juli–6 augustus 2023

❑ **International Mathematics Competition for University Students**

Internationale wiskundewedstrijd voor studenten.

plaats *Blagoevgrad, Bulgarije*

info *imc-math.org.uk*

Suggesties voor deze agenda zijn welkom.
Redacteur: Nikki Levering
agenda@nieuwarchief.nl

augustus 2023

25–26 augustus 2023, 1–2 september 2023
❑ **Vakantiecursus 2023**
Vakantiecursus van PWN voor leraren en studenten van lerarenopleidingen en voor andere belangstellenden. Dit jaar is het thema 'Priemgetallen'.
plaats *Universiteit Antwerpen, CWI Amsterdam*
info *platformwiskunde.nl/vakantiecursus*

28 augustus–1 september 2023
❑ **Multiscale Modeling and its Applications**
Vakantiecursus van de Universiteit Groningen voor masterstudenten en promovendi.
plaats *Rijksuniversiteit Groningen*
info *rug.nl*

september 2023

4–7 september 2023
❑ **Stochastic Models in Life Science**
Conferentie over de ontwikkelingen in stochastische modellen voor biologische processen.
plaats *Eurandom, Eindhoven*
info *eurandom.nl*

4–8 september 2023
❑ **ALGO 2023**
Jaarlijks evenement waarin het vooraanstaande 'European Symposium on Algorithms' wordt gecombineerd met gespecialiseerde conferenties en workshops over algoritmes.
plaats *CWI, Amsterdam*
info *algo-conference.org*

4–8 september 2023
❑ **Mathematics of FinTech**
Vijftiende editie van de European Summer School in Financial Mathematics.
plaats *TU Delft*
info *tudelft.nl*

6–7 september 2023
❑ **John Einmahl Workshop**
Tweedaagse workshop ter gelegenheid van de afscheidslezing van John Einmahl.
plaats *Hotel Boschoord, Oisterwijk*
info *einmahlworkshop.nl*

15 september 2023
❑ **Finale Nederlandse Wiskunde Olympiade**
Nationale finale van de wiskundeolympiade voor middelbare scholieren.
plaats *Technische Universiteit Eindhoven*
info *wiskundeolympiade.nl*

23 september 2023
❑ **Een bonte verzameling**
Symposium van de Werkgroep Geschiedenis van de NVvW.
plaats *Utrecht*
info *nvvw.nl/werkgroepen/werkgroep-geschiedenis*

25–29 september 2023
❑ **Social Influence Analysis**
Lorentz workshop, mede georganiseerd door Doina Bucur (UT) en Mike Preuss (UL).
plaats *Lorentz Center@Snellius, Leiden*
info *lorentzcenter.nl*

27–29 september 2023
❑ **Woudschoten Conference 2023**
Een jaarlijkse conferentie van de Nederlands-Vlaamse Numerieke Analyse-groepen, georganiseerd door de Werkgemeenschap Scientific Computing (WSC).
plaats *Woudschoten, Zeist*
info *wsc.project.cwi.nl/events*

29 september 2023
❑ **Onderwijs meets onderzoek**
Het thema van deze editie van het symposium is 'Curriculumvernieuwing'.
plaats *Domstad, Utrecht*
info *nvvw.nl*

oktober 2023

9–13 oktober 2023
❑ **Autumn School – Scientific Machine Learning and Dynamical Systems**
Onderdeel van het CWI Research Semester Programma, voor promovendi, postdocs en andere junior onderzoekers.
plaats *Amsterdam Science Park*
info *cwi.nl*

12–13 oktober 2023
❑ **Annual Meeting Dutch Inverse Problems Community**
Derde editie, met onder andere een plenaire lezing van Bas van der Linden (TU/e).
plaats *Rijksuniversiteit Groningen*
info *cwi.nl*

23–27 oktober 2023
❑ **Computing Combinatorial Dynamics in High Dimensional Biological Networks**
Lorentz workshop, mede georganiseerd door Shane Kepley en Bob Rink (VU).
plaats *Lorentz Center@Snellius, Leiden*
info *lorentzcenter.nl*

november 2023

4 november 2023
❑ **Jaarvergadering NVvW**
Jaarlijkse vergadering en studiedag van de Nederlandse Vereniging van Wiskundeleraren.
plaats *Ichthus College, Veenendaal*
info *nvvw.nl/jaarvergadering*

6–10 november 2023
❑ **Integrating Acquisition and AI in Tomography**
Lorentz workshop, mede georganiseerd door Tristan van Leeuwen (CWI).
plaats *Lorentz Center@Oort, Leiden*
info *lorentzcenter.nl*

10 november 2023
❑ **Prijsuitreiking Nederlandse Wiskunde Olympiade**
De prijsuitreiking van de finale van de Nederlandse Wiskunde Olympiade.
plaats *Technische Universiteit Eindhoven*
info *wiskundeolympiade.nl*

Nieuws

| News

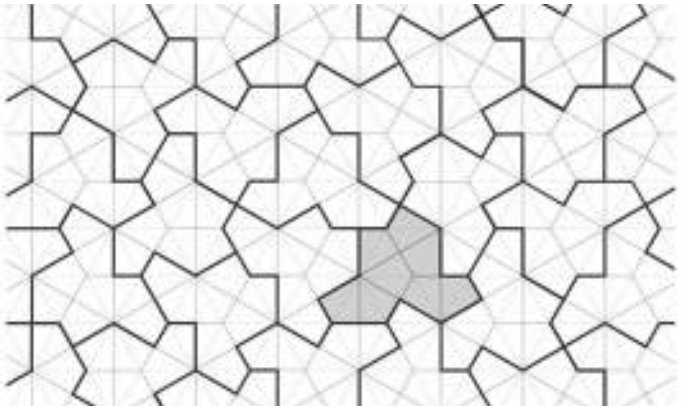
Deze rubriek is een kroniek van wiskundige activiteiten in Nederland. Toekomstige activiteiten worden aangekondigd en van voorbije activiteiten wordt verslag gedaan. Wilt u uw aankondiging of verslag in deze rubriek geplaatst zien? Stuur ons dan uw bijdrage, zo mogelijk met illustratie. De redactie behoudt zich het recht voor berichten te weigeren of in te korten.

Redacteur: Nikki Levering

nieuws@nieuwarchief.nl

Niet-herhalend tegelpatroon

Een klassiek open probleem dat wiskundigen al jaren fascineert, is de zogenaamde aperiodieke betegeling. De vraag is of er een tegelvorm bestaat waarmee een oppervlak van willekeurige grootte volledig bedekt kan worden, zonder dat ergens een herhalend patroon ontstaat. Naar deze tegelvorm wordt al sinds de jaren zestig gezocht, toen Robert Berger een set van meer dan twintigduizend vormen construeerde die, gecombineerd, aperiodiek kunnen betegelen. In de jaren zeventig werd een grote stap gezet: Roger Penrose vond een aperiodieke betegeling met slechts twee tegelvormen. Sindsdien zijn er vele niet-succesvolle pogingen gedaan een *einstein* (van het Duitse 'ein stein', ofwel 'één steen') te vinden. Vorig jaar vond er echter een grote doorbraak plaats: David Smith, Joseph Samuel Myers, Craig S. Kaplan (University of Waterloo) en Chaim Goodman-Strauss (University of Arkansas) hebben met 'de hoed' een enkele tegelvorm gevonden die een bedekkende aperiodieke betegeling kan geven. Verrassend is de relatief simpele vorm van de hoed: hij heeft slechts 13 zijden, en behoudt zijn aperiodieke eigenschappen wanneer de lengte van de zijden gevarieerd wordt. *newscientist.nl, quantamagazine.org*



Illustratie: arXiv:2303.10798

Luis A. Caffarelli ontvangt Abelprijs, Ingrid Daubechies Wolfprijs

De Abelprijs, toegekend door de Noorse Academie voor Literatuur en Wetenschappen, wordt dit jaar uitgereikt aan de Argentijns-Amerikaanse wiskundige Luis A. Caffarelli (University of Texas). Hij ontvangt de prijs voor zijn vele bijdragen aan het begrip van partiële differentiaalvergelijkingen (PDEs). Hij begon met het bestuderen van non-lineaire PDEs tijdens zijn postdoc, na het bijwonen van lezingen van Hans Lewy, wie hij naar open problemen vroeg om aan te werken. Op suggestie van Lewy werkte hij aan het *obstacle problem*, en het algemenere gebied van *free boundary problems*. Later zou hij focussen op de Monge–Ampère-vergelijking, en *Caffarelli's regularity theory* ontwikkelen, welke belangrijke toepassingen kent in de theorie van optimaal transport. In 1982 schreef hij samen met Robert Kohn en Louis Nirenberg een baanbrekend artikel over de Navier–Stokes-vergelijkingen, waarvoor ze in 2014 de AMS Steele Prize ontvingen. Verder ontving Caffarelli onder andere de Bôcherprijs (1984), de Rolf Schock-prijs (2005), de Wolfprijs (2012) en de Shawprijs (2018).

De Wolfprijs, uitgereikt door de Wolf Foundation, wordt dit jaar toegekend aan de Belgisch wis- en natuurkundige Ingrid Daubechies (Duke University, Durham). De prijs is een erkenning voor de vele ontwikkelingen in de wavelet-theorie, die door haar resultaten

tot stand zijn gekomen. Haar onderzoek heeft geleid tot vele vernieuwende algoritmes voor datacompressie, die een revolutie hebben veroorzaakt in de manier waarop signalen en afbeeldingen numeriek verwerkt worden. Zo werd haar werk onder andere gebruikt voor het reconstrueren van afbeeldingen uit de Hubble-telescoop, voor het opsporen van vervalste documenten en vingerafdrukken, en om geluidssequenties om te zetten in MP3-bestanden. Daubechies was van 2011 tot 2014 de eerste vrouwelijke president van de International Mathematical Union, en ontving eerder al de Satterprijs (1997, 2013), de Pioneer-prijs (ICIAM, 2013) en de Nemmersprijs (2013). *abelprize.no, wolffund.org.il, ece.duke.edu*

Foto: abelprize.no / Nolan Zunk, University of Texas at Austin

Luis A. Caffarelli

Foto: wolffund.org.il

Ingrid Daubechies

Wiskunde-examens makkelijker?

Loek Zonnenberg (docent wiskunde en oud-partner van McKinsey) en Paul Rutten (partner McKinsey) hebben onderzoek gedaan naar het niveau van de wis- en natuurkunde-examens van de afgelopen decennia. Ze brachten hun bevindingen uit in het rapport *Toetsen getoetst*, dat eerder dit jaar onder de vlag van McKinsey & Company werd uitgebracht. In hun onderzoek proberen ze een verklaring te vinden voor het stijgen van de vwo CE-cijfers wis- en natuurkunde enerzijds, en het dalen van de wis- en natuurkunde PISA-score anderzijds. Het McKinsey-rapport wijst niet de ontoereikendheid van de PISA, maar de niveaudaling van de CEs wis- en natuurkunde als oorzaak voor de geobserveerde discrepantie aan. Zo zouden de CEs de afgelopen jaren wel 40–50% minder stof bevatten dan dertig jaar geleden. Onderwerpen als driedimensionale ruimtemeetkunde zijn bijvoorbeeld uit de examenstof voor wiskunde verdwenen. Verder concluderen de onderzoekers dat de CEs tegenwoordig makkelijker zijn, omdat er vaak minder denkstappen nodig zijn om een vraag te beantwoorden. Een derde grondoorzaak wordt onder andere gezocht in het feit dat het aantal vwo CE-kandidaten de afgelopen jaren met 30 procent is gestegen. *platformwiskunde.nl*

Scholieren vinden mogelijk nieuw bewijs stelling van Pythagoras

Ne'Kiya Jackson en Calcea Rujean Johnson hebben mogelijk een nieuw bewijs voor de stelling van Pythagoras gevonden. De twee scholieren van de St. Mary's Academy in New Orleans raakten geïnspireerd door de bonusopgave van een wiskundewedstrijd in de kerstvakantie, welke ze als enigen wisten te beantwoorden. Ze mochten hun ontdekking op een regionale bijeenkomst van de American Mathematical Society presenteren. Ze werden hier aangespoord hun resultaten te publiceren, zodat het mogelijk is hun bewijs te controleren.

Het nieuwe bewijs van de 2000 jaar oude stelling zou erg bijzonder zijn, omdat het zou gaan om een trigonometrisch bewijs dat de sinusregel gebruikt. Het bewijs zou daarmee onafhankelijk zijn van de pythagorische identiteiten, welke juist gebaseerd zijn op de stelling van Pythagoras. Dit laatste maakt het moeilijk trigonometrische argumenten te gebruiken voor het bewijs van de stelling, omdat deze vaak leiden tot cirkelredeneringen. Door het gebruik van de sinusregel, oneindige reeksen en fractals van soortgelijke driehoeken weten de twee studenten een cirkelredenering te voorkomen. De scholieren uit New Orleans zouden niet de eersten zijn met een trigonometrisch bewijs die de pythagorische identiteiten niet gebruikt: Jason Zimba, destijds verbonden aan het Bennington College in Vermont, toonde in 2009 aan dat de pythagorische identiteiten zonder de stelling van Pythagoras bewezen kunnen worden. Zijn bewijs gebruikt echter heel andere technieken. *The Guardian*

Vier medailles op EGMO

Op de European Girls' Mathematical Olympiad (EGMO) in Portorož, Slovenië, hebben alle vier de Nederlandse deelneemsters een bronzen medaille veroverd. Hiermee zijn zij het eerste Nederlandse team dat vier medailles weet te behalen op de EGMO. Met hun scores eindigden ze op de 26ste plaats van de 55 deelnemende teams. Het Nederlandse EGMO-team werd begeleid door Gabriëlle Zwaneveld en Wietze Koops (RUG, RU), en werd gevormd door Allie Zong (24 punten), Lieke van Dam (18 punten), Damaris ter Haar (16 punten) en Katya Nikitchenko (16 punten). *wiskundeolympiade.nl*

Foto: wiskundeolympiade.nl

Van Zwet Award en Jan Hemelrijk Award uitgereikt

Tijdens de jaarlijkse bijeenkomst van de Nederlandse Vereniging voor de Statistiek en Operations Research (VVSOR) heeft Collin Drenth (TU/e) de Willem R. van Zwet Award ontvangen voor zijn proefschrift *Structured Learning and Decision Making for Maintenance*. Zijn onderzoek richt zich op systemen die onderhoud nodig hebben om te kunnen blijven functioneren.

Op dezelfde bijeenkomst werd ook de Jan Hemelrijk Award uitgereikt aan Fleur Theulen (Tilburg University en het Analytics for a Better World Institute), voor haar masterscriptie *Solving Large Maximum Covering Location Problems with a GRASP Heuristic*. Het doel van de scriptie was het verbeteren van de toegang tot gezondheidsfaciliteiten in Vietnam.

SIAM Hackathon 2023

Onderdeel van de SIAM Conference on Computational Science and Engineering, afgelopen februari in Den Haag, was de SIAM Hackathon. Twee dagen focusten meer dan honderd deelnemers zich op nieuwe wiskunde-oplossingen voor zes industriële problemen, aangeleverd door ASML, AMD, Siemens, KUKA, Roche Institute en Amazon. De winnaar was team Go'round, bestaande uit de PhD-studenten Ikrom Akramov (TU Hamburg), Wouter van Harten (RU) en Tjeerd Jan Heeringa (UT). Zij presenteerden een oplossing voor het multi-agent pakketbezorgingsprobleem van KUKA. *4tu.nl*

Foto: Platform Wiskunde Nederland

Leonardo García Heveling wint KWG PhD-prijs

Op het Nederlands Mathematisch Congres (NMC) kregen vier promovendi de kans hun werk te presenteren aan een breed wiskundig publiek. De beste voordracht werd beloond met de KWG PhD-prijs, dit jaar voor de 18de keer uitgereikt. De vier geselecteerde kandidaten waren Anina Gruica, Perfect Y. Gidisu (TU/e), Leonardo García Heveling (RU) en Luis Felipe Vargas (CWI). De jury, onder leiding van Martin van Gijzen (TU Delft), wees Leonardo als winnaar aan. Zijn promotie-onderzoek betreft de metriek van *spacetimes*, welke door algemene relativiteit worden gebruikt voor het beschrijven van zwaartekracht.

Tijdens het NMC werd ook de Pythagorasprijs uitgereikt, voor het best geschreven profielwerkstuk over een wiskundig onderwerp. Er waren drie finalisten, die allen hun werkstuk op het NMC mochten presenteren. De eerste prijs was voor Iris Penninga (Odulphuslyceum, Tilburg), die liet zien hoe men origami kan inzetten om tweede- en derdegraadsvergelijkingen op te lossen. De tweede prijs was voor het profielwerkstuk van Stan de Haas (Rijnlands Lyceum, Sassenheim), over wiskundige modellen voor fileproblemen. De derde finalist, Mai Thy Nguyen (Linecollege, Tiel), schreef een werkstuk over het gebruik van wiskunde in de populatiegenetica en -dynamica. *cwi.nl, pyth.eu*

EMS Young Academy

De European Mathematical Society (EMS) heeft haar Young Academy (EMYA) gelanceerd, een platform waar de nieuwe generatie wiskundigen kan samenwerken aan de toekomst van wiskunde in Europa, en van waaruit zij activiteiten voor Europese wiskundigen kan organiseren. Ieder jaar zullen dertig jonge wiskundigen zich voor een periode van vier jaar inzetten voor EMYA, waardoor EMYA straks 120 leden zal tellen. Onder het eerste cohort van dertig leden bevinden zich onder anderen Simon Telen (CWI, MPI MiS Leipzig) en Raffaella Mulas (VU). *euromathsoc.org*

Sandjai Bhulai in Adviescommissie Analytics Financiën

VU-hoogleraar Sandjai Bhulai heeft zitting genomen in de Adviescommissie Analytics Financiën. Deze onafhankelijke commissie adviseert het ministerie van Financiën, de Belastingdienst, de Douane en de Dienst Toeslagen over het ethisch verantwoord omgaan met data analytics. Naast Sandjai Bhulai bestaat de commissie onder anderen uit experts op het gebied van privacyrecht, ethiek, en beveiliging van digitale data. *rijksoverheid.nl*

Jaap Murre overleden

Op 10 april 2023 is op 93-jarige leeftijd overleden Jacob P. Murre, hoogleraar wiskunde te Leiden vanaf 1961 (na een periode als assistent en lector aldaar). Jaap was een invloedrijke, bewogen en zeer gewaardeerde wiskundige die tot op hoge leeftijd verspreid over de wereld inspirerende voordrachten hield over zijn werk in de algebraische meetkunde. Daarbij stond Jaap bij iedereen bekend als een beminnelijk, zachtaardig en hulpvaardig mens. *universiteitleiden.nl*

Koninklijk Wiskundig Genootschap

❑ **NMC**

Het NMC op 11 en 12 april is goed bezocht met circa 270 deelnemers. De Brouwermedaille is uitgereikt aan Éva Tardos (Cornell University), de Indagationes Mathematicae-prijs is gewonnen door Catharina Stroppel en Jens Eberhardt (University of Bonn) en de Stieltjesprijs door Freek Witteveen (UvA/University of Copenhagen) en Sophie Huiberts (UvA/Columbia University).

❑ **KWG-bestuur**

Op de ALV is Mathisca de Gunst (VU) gekozen tot voorzitter. Barry Koren (TU/e) is nu vice-voorzitter en Charlene Kalle (UL) de nieuwe vice-secretaris. Wioletta Ruszel is teruggetreden. Er wordt nog gezocht naar een nieuwe vice-penningmeester.

❑ **Young KWG**

Op de ALV is unaniem ingestemd met de oprichting van de KWG-sectie Young KWG. Op dit moment bestaat de sectie uit: Bharti Bharti (UvA), Mar Curco Iranzo (UU), Pierfrencesco Dionigi (UL), Martijn Gösgens (TU/e), Nikki Levering (UvA), Sven Polak (Tilburg), Nicos Starreveld (UvA) en Lotte Weedage (UT).

Recent verschenen:

❑ **Indagationes Mathematicae** (www.elsevier.com/locate/indag)

Special Issue on the occasion of Jaap Korevaar's 100-th anniversary, Jan Wiegerinck, Han Peters, Arno Kuijlaars (eds.), Volume 34, Issue 2.

❑ **Epsilon Uitgaven** (www.epsilon-uitgaven.nl)

Zebra 68. Peilingen in de Praktijk, Jelke Bethlehem, € 10.

Zebra 67. Modelleren met naalden en een bolletje wol, Gerd Hautekiet en Luc Van den Broeck, € 10.

Zebra 66. Islamitische Meetkundige Patronen – De geheimen van een eeuwenoude ontwerptraditie ontsluierd, Goossen Karssenberg, € 10.

Giraf 4. Figurenfestival – Spelen met vierhoeken, Els Franken, € 7.

Raf Bocklandt
Korteweg-de Vries Instituut voor Wiskunde
Universiteit van Amsterdam
r.r.j.bocklandt@uva.nl

Interview Jaap Korevaar

Een eeuwige passie

Ter ere van Jaap Korevaars honderdste verjaardag trok het Nieuw Archief naar Bussum, waar we op een zonnige februaridag bij een koffie bijpraatten over zijn bijna tachtigjarige carrière als wiskundige. In dit interview deelt hij een paar van zijn meest opmerkelijke ervaringen uit zijn lange en vruchtbare loopbaan, aangevuld met enkele interessante anekdotes uit zijn ongepubliceerde memoires.

Jaap Korevaar groeide op in Henrik-Ido-Ambacht, een dorpje aan de rand van Dordrecht, waar zijn vader schoolhoofd was van de lagere school. De basisschool deed hij in de school van zijn vader en daarna doorliep hij de middelbare school in Dordrecht, maar net voor zijn eindexamens vielen de Duitsers Nederland binnen. Gelukkig werden de examens slechts een paar weken uitgesteld en kon Jaap alsnog in de herfst van 1940 aan zijn universitaire studies beginnen, maar dat was allesbehalve vanzelfsprekend.

"In september 1940 schreef ik me in aan de Universiteit van Leiden om wiskunde, natuurkunde en astronomie te gaan studeren. Helaas werd de universiteit op 26 november 1940 gesloten door de door de Duitsers gecontroleerde regering, als vergelding voor de protesten tegen het ontslag van de Joodse professoren. De studenten moesten thuisblijven en konden worden gearresteerd voor deportatie naar Duitsland om te werken in de oorlogsindustrie.

Een dag nadat ik terugkeerde naar Hendrik-Ido-Ambacht kwam de politie naar ons huis om me op te halen. Gelukkig was mijn vader gewaarschuwd en kon ik net op tijd onderduiken. Ik ging naar mijn oom Arie Korevaar, die boer was in de Alblasserwaard, waar ik bleef totdat het veilig leek om terug naar huis te gaan. Het werd echter al gauw duidelijk dat Leidse studenten voorlopig niet zouden worden toegelaten tot de universiteit. Mijn vrienden en ik zetten onze studie thuis voort zo goed en zo kwaad als we konden, geholpen door uitstekende notities die oudere studenten hadden gemaakt en uitgewerkt. Die notities moest je met de hand overschrijven want er waren nog geen kopieerapparaten, en de examens kon je stiekem bij de professoren thuis afleggen.


Jaap Korevaar

In januari 1942 mochten de Leidse studenten verder studeren aan een andere universiteit. Samen met Fred van der Blij, een medestudent en een heel goede vriend, besloten we ons in te schrijven in Utrecht. Daar behaalde ik mijn kandidaatsdiploma in wis- en natuurkunde en werd ik student-assistent. Helaas werd de Universiteit van Utrecht ook gesloten door de Duitsers en moest ik weer even onderduiken bij mijn oom. Daarna werd ik assistent op de Technische Hogeschool Delft, maar ook die werd gesloten."

Na de oorlog behaalde Jaap zijn doctoraal examen in Leiden en kon hij aan de slag in het pas opgerichte Mathematisch Centrum in Amsterdam, de voorloper van het CWI.

"In september 1947 waren mijn vriend Fred van der Blij en ik de enige wetenschappelijk medewerkers die effectief in Amsterdam werkten bij het Mathematisch Centrum. In theorie was ons belangrijkste werk het doen van onderzoek, maar bij het destijds nog niet zo goed georganiseerde MC moesten we ook dagelijkse klussen doen zoals het bedienen van de primitieve stencilmachine. Daarnaast gaven we ook avondcursussen in analyse, lineaire algebra en statistiek.

Het Mathematisch Centrum onderhield ook levendige contacten met wiskundigen

Tauberiaanse stellingen

Als $f(x)=\sum_i a_i x^i$ convergeert binnen de eenheidscirkel en de limiet binnen de eenheidscirkel $\lim_{x\to 1} f(x)$ bestaat, wil dit niet noodzakelijk zeggen dat de som van de a_i bestaat en

$$\lim_{x\to 1} f(x) = \sum_{i=0}^{\infty} a_i.$$

In 1896 bewees de Oostenrijkse wiskundige Alfred Tauber [10] dat dit wel het geval is als $\lim_{n\to\infty} n a_n = 0$. Later kon Littlewood [8] de voorwaarde van Tauber versoepelen tot $\sup_{n\to\infty} n a_n < \infty$ en ontstonden er nog vele andere veralgemeningen van het werk van Tauber. Tijdens zijn jaren in Wisconsin werkte Jaap Korevaar vaak aan dit onderwerp en later publiceerde hij een boek over de geschiedenis ervan, dat in 2004 werd uitgegeven bij Springer [6].

uit het buitenland. Paul Erdős gaf in 1948 een lezing aan het MC over zijn zogenaamde elementaire bewijs van de priemgetalstelling [3]. Het bezoek van Paul Erdős aan het MC was erg belangrijk voor mij. Na zijn uiteenzetting over het elementaire bewijs van de priemgetalstelling, zocht Erdős me op en sprak uitvoerig over zijn niet-lineaire Tauberiaanse stelling [4] wat me inspireerde voor mijn latere werk.

Ik herinner me ook een curieus incident uit die tijd. Terwijl ik met Paul Erdős langs de Churchilllaan liep, zag hij een vrouw met een kinderwagen voor ons. Hij liep naar haar toe (terwijl ik een beetje achterbleef) en riep uit: 'What a nice epsilon, is it a slave or a boss epsilon?' (Pauls aanduiding voor een jongen of een meisje.) De vrouw was een beetje geschrokken, dus legde ik snel uit dat dr. Erdős dol was op kinderen en zeker geen kwaad bedoelde.[7]"

Tijdens zijn baan aan het MC werkte Jaap ook aan een proefschrift bij Kloosterman, dat hij verdedigde in 1949. Omdat Jaap graag in de academische wereld wou blijven, werd hem aangeraden eerst een paar jaar naar de Verenigde Staten te gaan. Uiteindelijk zou hij er meer dan twintig jaar blijven: eerst twee jaar in Purdue en dan, na een korte terugkeer naar Delft, tien jaar in Madison, Wisconsin, en tien jaar in San Diego, California.

"De UW Madison had een zeer goed Wiskunde-departement. Het personeel was breed georiënteerd en actief betrokken bij onderzoek. Daarnaast werd goed onderwijs als belangrijk beschouwd. Ik gaf een jaarlijkse cursus Mathematical Methods in Physics and Engineering, waarvoor ik uitgebreide nota's heb ontwikkeld. De cursus was voornamelijk bedoeld voor beginnende masterstudenten in natuurkunde en techniek. Soms trok de cursus meer dan honderd studenten, inclusief enkele van mijn toekomstige promovendi.

Tijdens mijn eerste jaar in Madison bezocht de beroemde L.E.J. Brouwer onze universiteit. Brouwer was net met pensioen gegaan en maakte een grote lezingentournee door Canada en de Verenigde Staten. Brouwer gaf twee lezingen in Madison, een over topologie en een over intuïtionisme. Hij begon de tweede lezing door te zeggen dat al zijn resultaten van de vorige lezing natuurlijk niet correct waren. Aan het einde vroeg ik brutaal aan Brouwer of hij echt bedoelde dat zijn topologische werk niet juist was. Zijn cryptische antwoord was: 'Wat denk jij?'

Het is misschien interessant om te vermelden dat mijn vrouw Jopie en ik Brouwer hadden uitgenodigd voor een diner bij ons thuis. Voor de maaltijd praatte Brouwer geanimeerd over zijn recente safari in Afrika. Jopie had haar best gedaan om een goede maaltijd voor de gast voor te bereiden. Toen we aan tafel zaten, zei Brouwer plotseling: 'Iedereen weet natuurlijk dat ik een vegetariër ben.' Gelukkig had Jopie een goede soep gemaakt waarmee ze Brouwer tevreden kon stellen. [7]"

In de jaren zestig breidde de Universiteit van California uit en opende een nieuwe afdeling in San Diego, een kuststad bij de Mexicaanse grens. Jaap was een van de eerste professoren van het wiskundedepartement en later werd hij er ook departementshoofd.

"We waren de strenge winters in Wisconsin moe en door onze trips naar Stanford hadden we wel interesse in een verhuizing naar California. Het goede klimaat maakte San Diego ook populair bij bezoekers. Onder andere mijn voormalig leraar Cornelis Visser en mijn oude vriend Fred van der Blij kwamen uit Nederland op bezoek via een uitwisselingsprogramma voor leraren. Ik had ook contact met Wim Luxemburg [9], toen in Caltech, wat resulteerde in een gezamenlijk artikel. We hadden een mooi huis met zwembad en gaven soms pool parties voor gasten. Toen Paul Erdős op bezoek kwam, vroeg hij mij of hij zijn moeder mee kon nemen. Zij was in de negentig en had onlangs Engels geleerd. Paul had uitgelegd dat hij haar niet graag alleen achterliet in Boedapest als hij reisde. Op het feest vertelde ze me vertrouwelijk, terwijl ze rustig in een hoekje zat: 'Weet je, nu Paul ouder wordt, wil ik niet meer dat hij alleen reist...'!"

In 1974 keerde Jaap terug naar Nederland samen met zijn tweede vrouw Pia.

"Pia was de dochter van de Zwitserse wiskundige Albert Pfluger en was zelf ook wiskundige. Ik leerde haar kennen op een

Het vermoeden van Bieberbach

Dit vermoeden gaat over complexe machtreeksen van de vorm

$$f(z) = z + a_2 z^2 + a_3 z^3 + \cdots + a_n z^n + \cdots$$

die convergeren voor $|z|<1$, en waarvoor $f(z)$ binnen dat gebied nooit tweemaal dezelfde waarde aanneemt. In 1916 bewees Ludwig Bieberbach dat $|a_2|\leq 2$ moet zijn, en suggereerde dat $|a_n|\leq n$ moet zijn voor alle n. De voortgang voor dit probleem was langzaam. In 1923 vond Loewner een belangrijke partiële differentiaalvergelijking die gerelateerd is aan het probleem en bewees hij dat $|a_3|\leq 3$. Littlewood toonde aan dat $|a_n|\leq en$ voor alle n, en dit globale resultaat werd door anderen verbeterd. Uiteindelijk leverde Louis de Branges een bewijs van het vermoeden in 1984 [2].

Foto: Bob Awick, bussumsnieuws.nl

De 100-jarige Jaap Korevaar

Foto: Peter van de Emde Boas

Raymond Brummelhuis sprak op Jaaps eeuwfeest

meeting in Oberwolfach. Nadat we in 1971 getrouwd waren, vond zij een baan aan San Diego State University, maar toen we kinderen kregen wilden we terugkeren naar Europa. Tijdens een reis naar Zwitserland ontdekte Pia dat ze een functie kon krijgen aan de Hogeschool van Lausanne en ik zou een assistent professorship kunnen krijgen aan de Universiteit van Lausanne. Pia vond echter dat ik zo'n laaggeplaatste functie niet moest overwegen. Bij navraag in Nederland bleek dat ik kon kiezen tussen verschillende hoogleraarschappen en dat Amsterdam en Eindhoven mogelijkheden boden voor ons beiden. We kozen voor Amsterdam waar Pia een aanstelling kreeg in de numerieke analysegroep van Dirk Dekker."

Door de keuze voor Amsterdam was Jaap bij toeval ook getuige van een belangrijk stuk wiskundegeschiedenis en het leverde hem bovendien ook nog een prijs op.

"1984 was een zeer opwindend jaar voor complexe analyse: Louis de Branges loste het beroemde Bieberbachvermoeden op! Ik was een vroege getuige van het bewijs: op de terugweg uit Leningrad, gaf de Branges een lezing over zijn werk aan de Vrije Universiteit in Amsterdam. Hij gaf me een kopie van een vroege versie van zijn bewijs in het Russisch. Ik schreef een rapport over de geschiedenis van het vermoeden en het bewijs, dat verscheen in de *American Mathematical Monthly* [5]. Het artikel leverde me de Chauvenet-prijs op van de Mathematical Association of America. Spoedig zouden veel wiskundigen lezingen geven en schrijven over het bewijs en in Amsterdam hebben we er een workshop over gehouden [7]."

Hoewel hij nog steeds de wereld rondreisde voor conferenties en onderzoeksverblijven, bleef Jaap voor zijn verdere carrière verbonden met de UvA en werd er ook directeur van het Wiskunde-instituut. Zelfs na zijn pensioen bleef hij het instituut bezoeken en tot enkele jaren geleden kon je hem nog vaak aantreffen bij de koffieautomaat. Met een lijst van publicaties die acht decennia bestrijkt, kun je wiskunde met recht en rede zijn levenslange passie noemen.

Eeuwfeest

Op 25 januari 2023 werd Jaap Korevaar 100 jaar. Op het Korteweg-de Vries Instituut is dat gevierd met een kleine bijeenkomst waarbij kinderen van Jaap en zo'n vijftig vrienden en bekenden van Jaap aanwezig waren. Jaap werd toegesproken door Eric Opdam, directeur van het KdVI, en kreeg van Jan Wiegerinck een exemplaar aangeboden van het special issue van *Indagationes Mathematicae* dat ter gelegenheid van zijn honderdste verjaardag verscheen. Raymond Brummelhuis (Reims) — de laatste promovendus van Jaap die nog in de academia actief is — sprak over zijn recente werk 'Spectraal theorie van tweedeorde Sturm–Liouville-systemen'. In deze voordracht passeerden toevallig, maar zeer toepasselijk, onderwerpen uit veel van de doctoraalcolleges die Jaap ooit gaf, op natuurlijke manier de revue. Jaap was zeer verguld met de bijeenkomst en heeft iedereen hartelijk bedankt. Daarna vertrok het gezelschap naar de Jaap Edenbaan voor een gezellige lunch in 'Jaap's Chalet', waarmee de bijeenkomst werd afgesloten.

Jan Wiegerinck

Referenties

1 G. Alberts, F. van der Blij en J. Nuis, eds., *Zij mogen uiteraard daarbij de zuivere wiskunde niet verwaarlozen*, CWI, 1987.

2 L. De Branges, A proof of the Bieberbach conjecture, *Acta Mathematica* 154(1–2) (1985), 137–152.

3 P. Erdős, On a new method in elementary number theory which leads to an elementary proof of the prime number theorem, *Proc. Nat. Acad. Sci. U.S.A.* 35 (1949), 374–384.

4 P. Erdős, On a Tauberian theorem connected with the new proof of the prime number theorem, *J. Indian Math. Soc. (N. S.)* 13 (1949), 131–144. Supplementary note, same journal and issue, 145–147.

5 J. Korevaar, Ludwig Bieberbach's conjecture and its proof by Louis de Branges, *The American Mathematical Monthly* 93(7) (1986), 505–514.

6 J. Korevaar, *Tauberian Theory: A Century of Developments*, Springer, 2004.

7 J. Korevaar, Recollections History of a Dutch-American mathematician, unpublished memoires.

8 J.E. Littlewood, The converse of Abel's theorem for power series, *Proc. London Math. Soc. (2)* 9 (1911), 434–448.

9 W.A.J. Luxemburg en J. Korevaar, Entire functions and Müntz–Szász type approximation, *Transactions of the American Mathematical Society* 157 (1971), 23–37.

10 A. Tauber, Ein Satz aus der Theorie der unendlichen Reihen, *Monatsh. Math. u. Phys.* 8 (1897), 273–277.

Joris M. Mooij
Korteweg-de Vries Institute for Mathematics
University of Amsterdam
j.m.mooij@uva.nl

Inaugural Lecture

Causality: from data

On Thursday 13 October 2022, Joris Mooij delivered his inaugural lecture upon acceptance of the position of professor in Mathematical Statistics at the University of Amsterdam. In many scientific and societal issues the relationship between causes and their effects is an important topic. Mooij explains why both causal models and experimental research remain necessary to answer causal questions (despite recent breakthroughs in machine learning).

In this lecture, I will take you on a tour through what I consider one of the most fascinating scientific disciplines: causality. We will consider a diverse range of scientific questions from a variety of fields. For example:

- *Does smoking cause lung cancer?* There is widespread agreement nowadays that it does, but this question was the topic of a huge debate in the sixties of the last century, involving famous statisticians like Ronald Fisher and Jerzy Neyman.
- *Does chocolate consumption increase cognitive abilities?* In other words: do you get smarter, if you eat lots of chocolate? This (perhaps more innocent) question is still the subject of scientific debate as of today, and the available evidence seems inconclusive.
- *Does the new COVID-19 vaccine protect better against hospitalization?* Pharmaceutical companies have updated their vaccines to protect against the new variants of the SARS-CoV-2 virus. To decide which vaccine to use, it is important to know whether someone who got a booster with the new vaccine has a lower probability to end up in hospital because of COVID-19 compared to someone who got a booster with the old vaccine instead.
- *Does knocking out gene X change the activity of gene Y?* A cell of a typical living organism has thousands of genes, which control the bio-chemical 'machinery' of the cell. Not all genes are active at the same time, and the activity of a gene is regulated by the activity of other genes. If a researcher disables one specific gene in a so-called 'knock-out' experiment, this may result in a change in the activity of one or multiple other genes. Understanding these *gene regulatory mechanisms* is one of the quests in biology.

You might wonder: what do all these questions have in common? The answer is that they are all of the form: *what will happen to Y if action X is performed?*

Similarly, many policy decisions and questions of societal relevance are of a causal nature. For example:

- *How will the revenue of a company change if it increases the price of a product?* A higher price typically means that less products will be sold. Does this decrease in volume compensate for the increased pricing?
- *How much will inflation in the Netherlands decrease if the European Central Bank increases the interest rate by 1 percent point?* An important question, but hard to answer reliably, because of the complex nature of our macroeconomy.
- *Do female graduate students applying for college have lower admission chances than male graduate students?* This appeared to be the case at UC Berkeley (California) in 1973. However, it turned out to be a statistical paradox. Upon closer investigation of the available data by statisticians, it appeared that there was no evidence for gender bias.
- *Would changing gender increase the chance of graduating cum laude for female PhD candidates?* This is a similar type of question, but closer to home, and still relevant today. The Dutch newspaper *NRC* discovered in 2018 that at many Dutch universities, the fraction of male PhD candidates that graduate *cum laude* ('with distinction') is about twice as high as that of female PhD candidates. Is this gender bias, or a naïve (and possibly incorrect) interpretation of the data, like in the UC Berkeley case?

Again, note that all these questions are of the same general form. They all concern to what extent a cause X influences its (potential) effect Y.

Joris Mooij

to science

Historical remarks

For many centuries, humans have been trying to understand the universe by thinking in terms of causes and effects. The importance of this way of thinking can hardly be overestimated. Let me take you on a brief excursion through history to show how various philosophers and scientists thought about causality. While Judea Pearl goes all the way back to Adam and Eve in his recent *Book of Why* [16], I will start with one of the ancient Greek philosophers, Democritus (ca. 460–370 b.C.). Democritus is mostly known for his formulation of an atomic theory of the universe. He clearly appreciated the importance of causality when he wrote:

> "I would rather discover one true cause than gain the kingdom of Persia."

But what does it mean to "discover one true cause"? Philosophers, amongst them David Hume (1711–1776), have been struggling with this question. Hume wrote [7]:

> "Thus we remember to have seen that species of object we call *flame*, and to have felt that species of sensation we call *heat*. We likewise call to mind their constant conjunction in all past instances. Without any farther ceremony, we call the one *cause* and the other *effect*, and infer the existence of the one from that of the other."

While this definition of causality contains many appropriate elements, one could criticize it as being overly simplistic. For example, does the rooster's crow really cause the sun to rise? And does the barometer needle really cause rain? Furthermore, is the "constant conjunction in all past instances" really necessary? If we say "smoking causes lung cancer", we do not necessarily mean that *everyone* who smokes will get lung cancer.

Difficulties like these have led some to throw the towel in the ring. One of them was Bertrand Russell (1872–1970), a famous logician and philosopher, and one of the authors of the *Principia Mathematica* [19] (an attempt to derive all of mathematics by pure logic from a small set of axioms). Russell proposed to abandon the concept of causality completely. He wrote [17]:

> "All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word 'cause' never occurs. The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm."

The difficulties of formally defining the notion of causality also led Karl Pearson (1857–1936), one of the founders of statistics — still well-known today from the *correlation coefficient* named after him — to propose to get rid of the notion. Pearson wrote [14]:

> "Beyond such discarded fundamentals as 'matter' and 'force' lies still another fetish amidst the inscrutable arcana of even modern science, namely, the category of cause and effect."

You may have heard the slogan 'correlation is not causation'. Pearson's point of view was that one should *only* consider correlation. It is striking that even today, I teach BSc mathematics students how to calculate Pearson's correlation coefficient in an introductory statistics course, but I do not teach them anything about causality. And this is not just me: this appears to be typical for most statistics courses taught at most universities in most countries. I am convinced that this is a missed opportunity!

In the last decades, though, things have changed quite dramatically. If we fastforward to today, we see that causality is a thriving and growing scientific discipline. It is scattered across fields, and has seen important contributions from epidemiology, econometrics, genetics, machine learning, statistics, artificial intelligence, computer science, and more.[1] For example, the consultancy company Gartner recently included Causal AI in its *Hype Cycle for Emerging Technologies* [5], because they expect that it will deliver "a high degree of competitive advantage over the next 5 to 10 years."

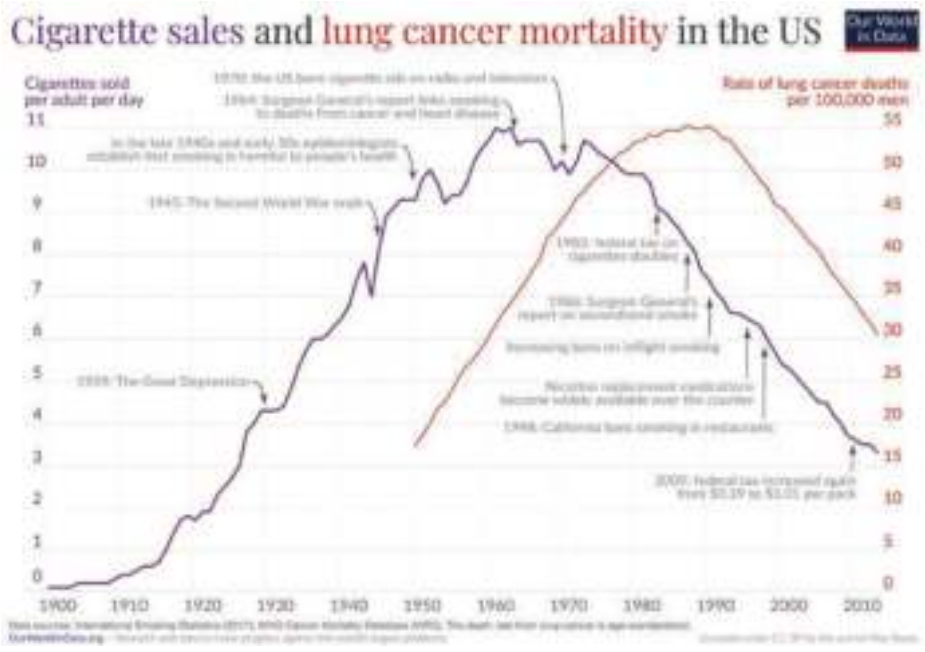


Figure 1 There is a clear (time-lagged) correlation between cigarette sales and lung cancer mortality in the US.[2]

Causation or correlation?

Let us revisit the causal question "Does smoking cause lung cancer?" One might wonder: what evidence exists that smoking causes lung cancer? I illustrated one piece of evidence in Figure 1: if we plot the number of cigarettes sold per adult per day in the United States (in purple), and the rate of lung cancer deaths for men in the US (in red), both over a period of several decades, we observe a striking similarity in the shapes of these curves (although the lung cancer deaths occur with a delay of about 25 years).

However, following in the footsteps of for example Ronald Fisher, you may argue that this is not a very convincing proof. Indeed, this could also be an example of a so-called *spurious* correlation, that is, a correlation without causation. Two examples of such spurious correlations are shown in Figure 2. If we plot US spending on science, space and technology over the years, and the number of suicides by hanging, strangulation and suffocation, we observe a striking similarity between the two curves. It appears quite implausible, though, that there is any causal relation between the two. The divorce rate in Maine (a state in the US) over the years also shows a strikingly similar pattern as the per capita consumption of margarine. Would one take this as evidence for a causal relation between the two?

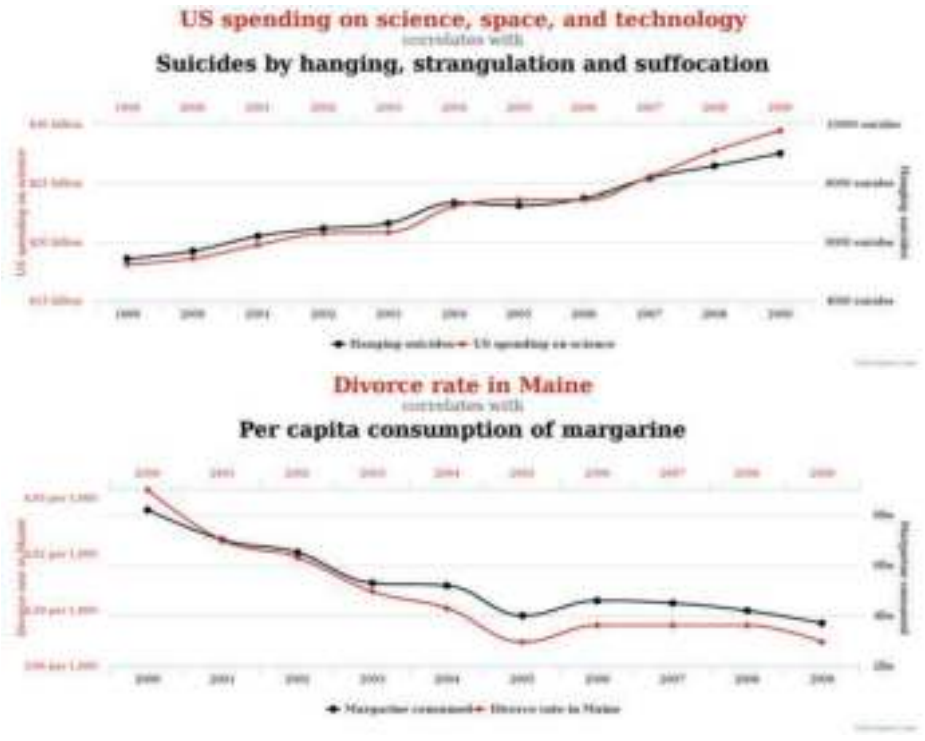


Figure 2 Two examples of correlations between time-series that appear to be spurious.[3]

I took these examples from a website, where you can find many more.[4] The surprising nature of these examples is due to selection bias: the website was created with the help of an algorithm that searches for short pieces of highly correlated time-series in a large database containing many different time-series. Because of the multiple-testing issue, these strong correlations may actually not be statistically significant (indeed, even if you search for such patterns in random data, you will eventually find them).

So, if causation is not correlation, then what is it? Giving a precise definition is not straightforward. It is perhaps as challenging as defining other elementary notions like 'space' and 'time'. I provide here a simplified definition that contains the basic gist (note, though, that it is still so vague that no mathematician or statistician would be satisfied with it!).

First, consider *deterministic* systems, that is, systems in which chance plays no role. Suppose that variables X and Y describe part of the system's state.

Definition 1. We say that X *causes* Y if a minimal external intervention on the system that sets the value of X may change the possible values of Y.

For example, consider a bicycle. If we take for X the position of the break lever, and for Y the rotation angle of the wheels, then X causes Y (since pulling the break lever prevents the wheels from rotating). However, rotating the wheels does not change the possible positions that the break lever can have, so Y does not cause X.

For *stochastic* systems, where chance plays a role, we just need a small change in the definition. As we noted before: not everyone who smokes will get lung cancer.

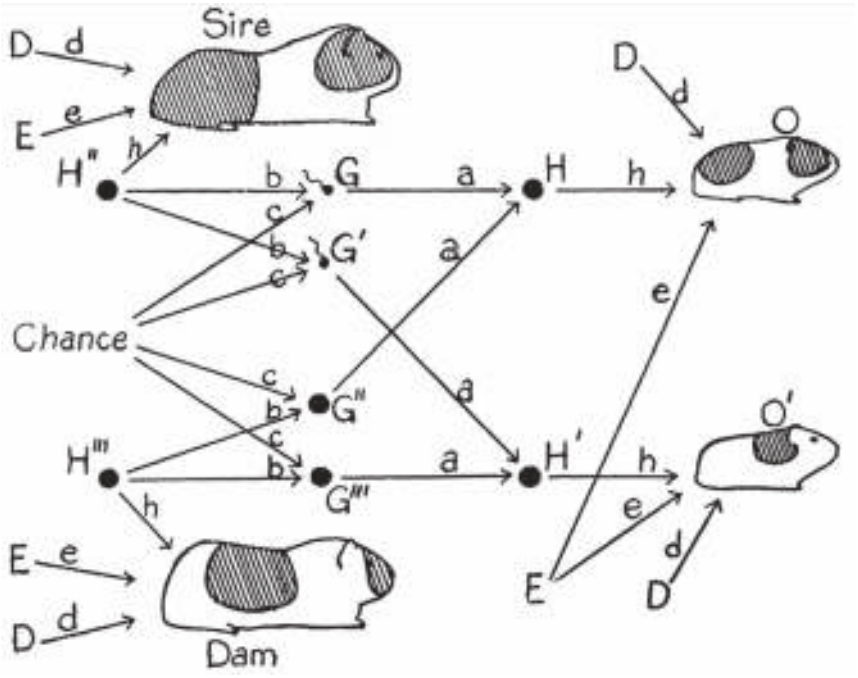


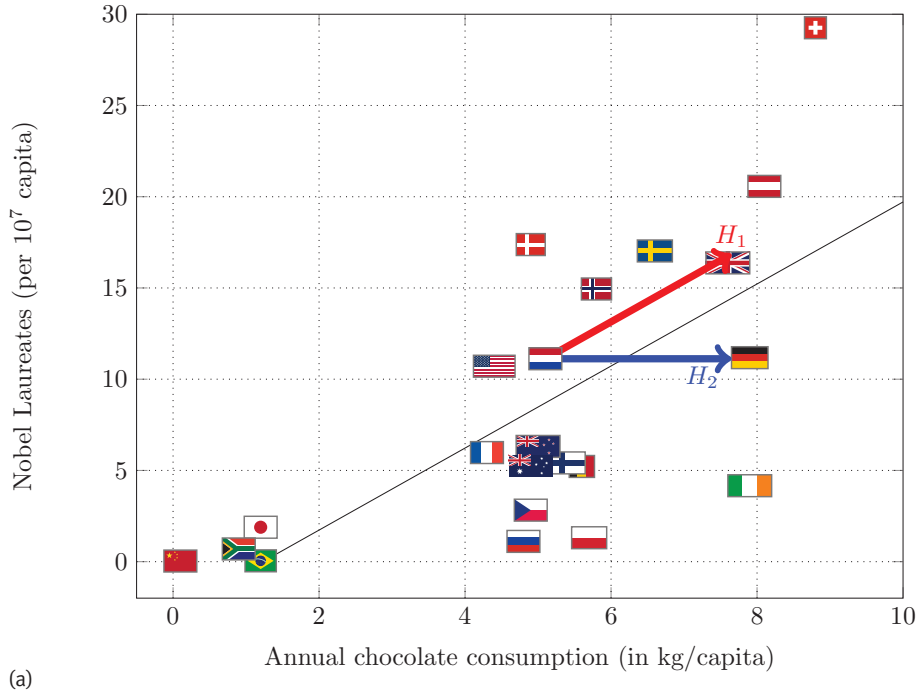
Figure 3 'Path diagram' in [20] expressing causal relations between fur patterns and various genetic and environmental factors. This is perhaps the first instance of a graphical causal model.

Therefore, we replace 'the possible values of' by 'the probability of'. This leads to the following definition for stochastic systems.

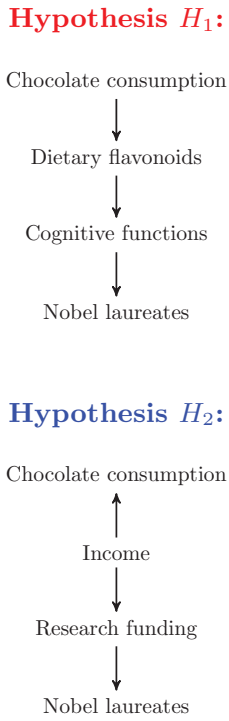
Definition 2. We say that X *causes* Y if a minimal external intervention on the system that sets the value of X may change the probability distribution of Y.

I would like to illustrate this by using the 'path diagram' in Figure 3, used by geneticist Sewall Wright in 1920 to communicate a causal model of how the fur pattern of guinea pigs is determined by various genetic and environmental factors [20]. 'Chance' is explicitly represented here, and plays a role in how the genetic constitution of the offspring depends on that of the parents. If we take for X the genetic constitution of the parent individuals, and for Y the fur patterns of their children, then X causes Y in a non-deterministic way, according to the above definition applied to Wright's model.

As a remark to the statisticians in the audience: note that this definition shows that the standard framework of classical statistics is too narrow and needs to be extended. Indeed, a statistical model describes the joint distribution of X and Y, but not how the distribution of Y changes when we intervene on X. Two such extensions have become popular for modeling stochasticity and causality. The *potential outcome framework* considers jointly defined random variables for each hypothetical intervention (so-called 'potential outcomes'). The other framework models how the probability distribution of the system's variables depends on interventions (thus essentially treating interventions as parameters of the distribution).[5] In both frameworks, graphs can be used to represent causal relations and independence relations between the variables. While there has been a heated debate on which of the two frameworks is superior, the differences are actually minor.[6]



(a)



(b)

Figure 4 (a) There is a significant correlation (with Pearson correlation coefficient $R = 0.69$, and p-value 0.0004) between chocolate consumption and Nobel laureates across countries.[7] (b) Two causal hypotheses that might explain the observed correlation between chocolate consumption and Nobel laureates. For both hypotheses, I indicated with colored arrows in (a) the expected effects of an intervention in which the Dutch government increases the chocolate consumption in the Netherlands by 50%.

Chocolate consumption and Nobel laureates

Now that we have some idea of how causation differs from correlation, let us return to the question "Does chocolate consumption increase cognitive abilities?" In a publication in the prestigious *New England Journal of Medicine*, an interesting observation was reported: the average annual chocolate consumption is significantly correlated with the number of Nobel laureates per capita [10]. The graph in Figure 4(a) visualizes the data. Chocolate consumption is on the horizontal axis, and the fraction of Nobel laureates is on the vertical axis. We indeed observe a correlation, as the data fall roughly on a straight line, with a Pearson correlation coefficient of about 0.7. In particular, we can read off that an average Dutch inhabitant eats 5 kg of chocolate a year. This is about two thirds of the average chocolate consumption of an inhabitant of the United Kingdom. What would happen if the Dutch ate 50% more chocolate? Would the number of Dutch Nobel laureates also increase by 50%? And thus, would it be a good idea for the University of Amsterdam to provide free chocolate for all staff members and students? As we shall see, the answer depends on the causal relations between these two variables.

Messerli provided the following possible explanation of the observed correlation [10], that I will refer to as hypothesis H_1. Chocolate contains certain chemical substances known as 'dietary flavonoids'. There is some evidence in animal studies that dietary flavonoids may have positive effects on brain regions involved with memory and learning. Messerli speculates that the consumption of chocolate may lead to an increased uptake of dietary flavonoids in the brain, which may improve the cognitive functions of the brain, eventually leading to more Nobel laureates. I have illustrated this hypothesis by means of a causal graph in Figure 4(b), where the arrows indicate a direct causal relationship of one variable on another, just like in the 'path diagram' of Wright that we just saw.

But we can also consider an alternative theory, hypothesis H_2: the average income in a country determines how much money the inhabitants have for buying chocolate, and also how much tax payers' money is used to fund scientific research, eventually leading to Nobel laureates. In this hypothesis, the variable 'income' is a common cause of our two variables of interest (chocolate consumption and Nobel laureates), and is called a *confounder* for that reason. This confounder may explain the observed correlation of chocolate consumption and Nobel laureates, even if there is no direct causal relation between the two.

According to these hypotheses, what would happen if the government intervened to increase the chocolate consumption in the Netherlands by 50%? I illustrated this in Figure 4(a). Under hypothesis H_1 (red arrow), chocolate consumption causes Nobel laureates, and therefore, we would expect the number of Nobel laureates to go up as well, perhaps even to the level of that of the UK. In contrast, under hypothesis H_2 (blue arrow), chocolate consumption does *not* cause Nobel laureates, hence we would *not* expect any change, and the Netherlands would probably end up somewhere near Germany.

We may conclude that predictions of the consequences of actions can depend in a very sensitive way on the underlying causal relationships between the variables. This also means that using off-the-shelve machine learning tools for supervised learning (including deep neural networks) to make such predictions may be a bad idea.

Randomized controlled trials

One way to investigate whether chocolate consumption indeed improves cognitive abilities is to setup a so-called *randomized controlled trial*. This approach to estimating causal effects is called a 'gold standard', as it is considered to be the most reliable method. In our case, it would work as follows (see also Figure 5). We first select a sample of individuals (say, a representative sample of inhabitants of the Netherlands). Then, by flipping a coin we divide the individuals into two groups, the intervention group and the control group. On all individuals in the intervention group, we enforce a diet containing large amounts of chocolate. The eating habits of the individuals in the control group are left unchanged. We sustain the diets for several years, and then measure the cognitive abilities of all individuals. Finally, we compare the outcomes in the two groups. If we see that the individuals in the intervention group have become significantly smarter on average than those in the control group, we conclude that there is a causal relationship, and can estimate the size of the causal effect.

The approach of randomized controlled trials was popularized by Ronald Fisher, and nowadays provides the pillar of 'evidence-based' medicine. Also known as 'A/B-testing', it is used extensively by big tech companies to optimize their business algorithms.

The example also points out some of the limitations of randomized controlled trials. First, there are logistic aspects. We need a sufficiently large sample size to arrive at statistically significant conclusions. If the outcome of interest is a rare event (for example, winning a Nobel price), the number of participants in the trial needs to be huge. And how exactly would we enforce this chocolate diet in practice? Another issue is ethics. This is the reason that no randomized controlled trial has yet been performed on humans to investigate whether smoking causes lung cancer: it would simply not be

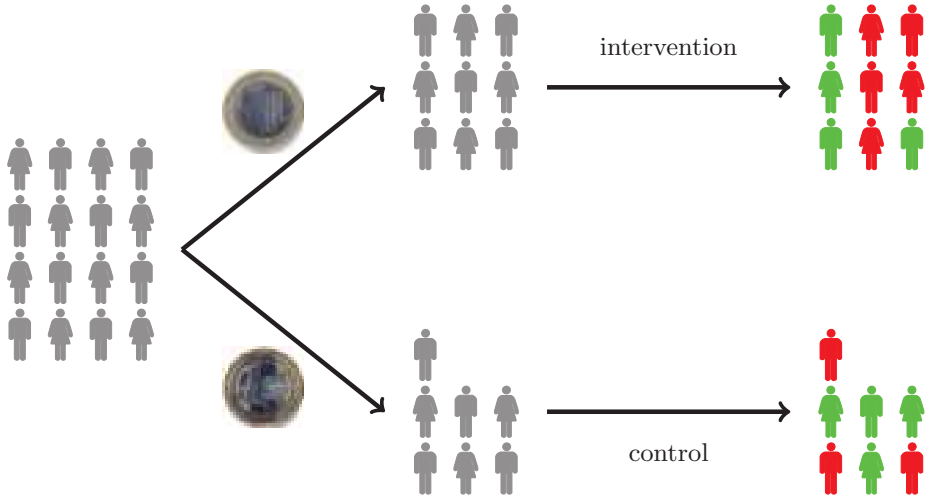


Figure 5. Schematic illustration of a randomized controlled trial.

ethical to force subjects in the treatment group to smoke for many years, given that we expect many of them to develop lung cancer as a result of the experiment. Another limitation lies in the inclusion criteria. In medical studies, subjects are often healthy young males, whereas one might be interested more specifically in the effect of treatment on elderly diseased females. To which extent can the conclusions of a randomized controlled trial be extrapolated to other subpopulations?

So, do we have alternatives? I don't believe that a randomized controlled trial has ever been performed in astronomy, for instance. Yet, astronomers rely on purely *observational* data to arrive at a causal understanding of the universe.[8] In medicine, could we perhaps make use of routinely collected electronic health records to estimate the causal effects of treatments on health outcomes?

Naïve attempts to use observational data (rather than experimental data) to infer causal effects can fail horribly. A striking illustration of this is related to a phenomenon known as *Simpson's paradox*. This paradox can be thought of as a statistical analogue to an optical illusion: we appear to see something that cannot exist.

Simpson's paradox

I will attempt to visualize Simpson's paradox for you. Consider the following hypothetical scenario. Suppose someone has collected data from electronic patient records concerning COVID-19 vaccinations. The question at stake is which of two vaccine types (A or B) is most effective at preventing hospitalization because of COVID-19. I will denote the treatment variable with X, and the outcome variable with Y. Suppose for the moment that there are two possible treatments (A and B, corresponding to the two vaccines), and two possible outcomes: positive and negative (where negative corresponds to hospitalization within six months after vaccination).

In total, the data concerns 10 000 cases of COVID-19. In Table 1 (columns '♂ + ♀') we see for instance that of the 5000 individuals that had vaccine B, 2000 ended up in hospital, but 3000 did not.[9] This means that 60% of the individuals that had vaccine B had a positive outcome. Of the individuals that had vaccine A, only 50% had a positive outcome. I visualized these

	♂ + ♀		♂		♀	
	$Y=+$	$Y=-$	$Y=+$	$Y=-$	$Y=+$	$Y=-$
$X=\mathrm{A}$	2500	2500	1500	2250	1000	250
$X=\mathrm{B}$	3000	2000	375	875	2625	1125

Table 1 Data for the fictitious example of Simpson's paradox. Treatment X (type of COVID-19 vaccine) can take values A, B. Outcome Y (hospitalization after COVID-19 infection) can take values +,−. The aggregated data can be split into gender-specific subgroups.

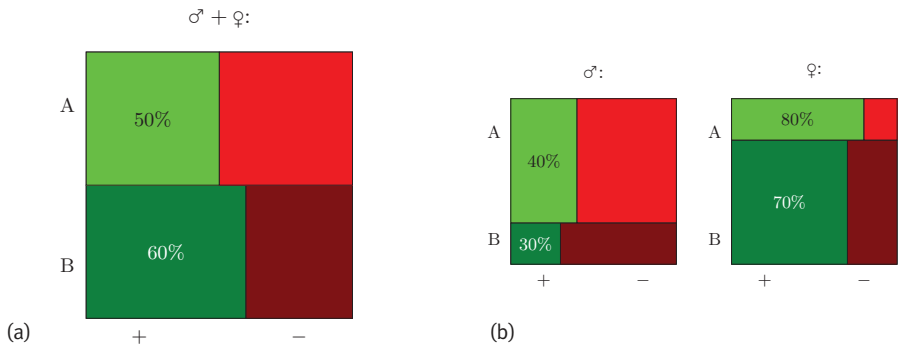


Figure 6 Visualizations of the fractions of positive outcomes corresponding to the data in Table 1; (a) aggregated over genders; (b) gender-specific. Area is proportional to counts.

fractions in Figure 6(a). Which of the two vaccines would you prefer, based on this data?

A closer look at the data reveals something remarkable. The genders of these 10 000 individuals were also recorded. For simplicity of exposition, assume that all of them are either male or female. I provided the counts for both genders separately in Table 1 (columns '♂' and '♀'). For example, of the 3000 individuals that got vaccine B and had a positive outcome, only 375 were male, and 2625 were female. You can check for yourselves that all the numbers add up. Let us look at the fractions, which I visualized in Figure 6(b). For males, those that got vaccine A had 40% probability on a positive outcome, versus only 30% for vaccine B. For females, 80% of those that got vaccine A had a positive outcome, versus only 70% for vaccine B. So, based on this additional information, which of the two vaccines would you now prefer?

Before I proceed with explaining how I would answer this question, let me emphasize the paradoxical nature of these numbers. It is intuitively clear that it cannot be the case that vaccine A is best for men *and* best for women, but worst overall. Indeed, if these data came from a randomized controlled trial, such a paradox could not happen. The paradox stems from an incorrect interpretation of the *correlations* between treatment and outcome as *causal relations*, the very mistake that Karl Pearson warned us against!

To understand this better, we consider different causal hypotheses that might apply in our case. I have illustrated them as causal graphs in Figure 7. Under the hypothesis in (a), gender (Z) causes both vaccine type (X) and hospitalization outcome (Y). One can prove that under this causal hypothesis, the data implies that vaccine A should be preferred. Another hypothesis, the one in (b), assumes an unobserved common cause L of X and Z, which explains the observed correlation between these two variables. Also in this case, one can prove that vaccine A should

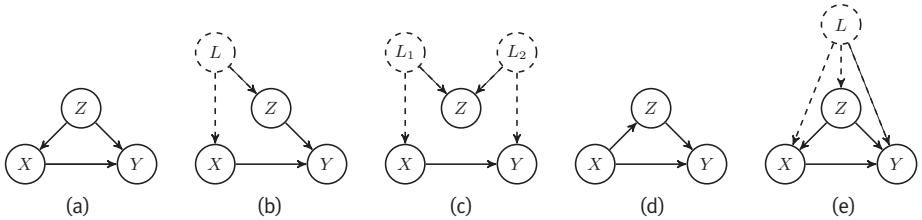


Figure 7 Different possible causal hypotheses for the data in Table 1 (X is treatment, Z is gender, Y is outcome, L is an unobserved confounder). One can show that for (a) and (b), treatment A is better than B, while for (c) and (d), the opposite holds. For (e) it is unclear which treatment is better.

be preferred. But, there also exist causal hypotheses for which one can prove that vaccine B should be preferred! Two such hypotheses are shown in (c) and (d). So we see that the right answer to the question which vaccine to prefer depends not only on the data, but also on our causal assumptions!

However, we can still rule out some of these hypotheses. Indeed, we may assume that treatment does not affect gender (I have heard many rumours and conspiracy theories about what undesired effects COVID-19 vaccines may have, but not yet that they transform males into females or vice versa!). That assumption allows us to rule out hypothesis (d). Furthermore, according to basic biology, gender is determined at conception by the chromosomes in the sperm cell that enters the egg cell and is an independent random event (like flipping a coin). Therefore, it is hard to think of any possible common cause of gender and treatment. That assumption allows us to rule out hypotheses (b) and (c).

Does this then lead us to the final conclusion that one should prefer vaccine A? Not yet! Indeed, we cannot easily rule out the possibility of another variable that causes treatment and outcome, as in hypothesis (e). For example, 'age' could be such a variable. Different vaccines have been used for different age groups, and elderly people generally have a higher risk of ending up in hospital with a virus infection. But including 'age' might lead to another instance of Simpson's paradox![10]

So, is there then nothing we can conclude from the data? One concrete answer to this question is provided by the *natural bounds* [12]. If we cannot rule out unobserved confounders, the best we can do is to calculate lower and upper bounds on the causal effect. For the mathematicians in the audience, I show here the theorem and the proof. Though conceptually advanced, the derivation itself is straightforward.

Theorem 1 [12]. *If treatment X, outcome Y and observed confounder Z are discrete variables, then in the presence of additional unobserved confounding (of X, Y and Z) we can bound:*

$$\begin{aligned} &p(X=x,Y=y \mid Z=z) \\ &\quad \leq p(Y(x)=y \mid Z=z) \\ &\quad \leq p(X=x,Y=y \mid Z=z) \\ &\qquad + 1 - p(X=x \mid Z=z). \end{aligned}$$

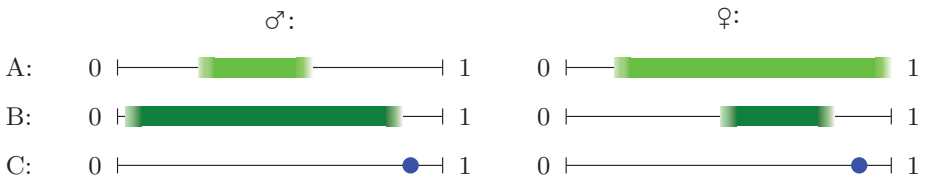


Figure 8 The probability of a positive outcome under enforced treatment x for gender z, $p(Y(x) \mid Z=z)$, must lie within the corresponding intervals (obtained by applying Theorem 1 to the data in Table 1). For a third possible treatment C, the probability of a positive outcome under enforced treatment is assumed to be precisely known (blue dots).

Proof. Using consistency $(Y = Y(X))$ and elementary probability theory:

$$\begin{aligned} &p(X=x,Y=y \mid Z=z) \\ &\quad = p(X=x,Y(x)=y \mid Z=z) \\ &\quad \leq p(Y(x)=y \mid Z=z) \\ &\quad = p(X=x,Y(x)=y \mid Z=z) \\ &\qquad + p(X\neq x,Y(x)=y \mid Z=z) \\ &\quad \leq p(X=x,Y=y \mid Z=z) \\ &\qquad + p(X\neq x \mid Z=z) \\ &\quad = p(X=x,Y=y \mid Z=z) \\ &\qquad + 1 - p(X=x \mid Z=z) \end{aligned}$$

for all z with $p(Z=z)>0$. □

I have visualized these natural bounds for our data in Figure 8. The fact that the lower and upper bounds have to be estimated from a finite sample of individuals yields additional (statistical) uncertainty, which I have visualized here by making the bounds fuzzy. Note, however, that in this case most uncertainty is 'causal' (due to uncertainty about the causal relations) rather than 'statistical' (due to extrapolating from a finite sample to the entire population). In other words, most uncertainty would *not* go away as we collect more and more samples. If we gave all males vaccine A, they would have a probability for a positive outcome between $\sim 30\%$ and $\sim 55\%$ (in the left light-green bar in Figure 8). If, instead, we gave all males vaccine B, that probability must lie between $\sim 7.5\%$ and $\sim 82.5\%$ (in the left dark-green bar). Since we only know that these probabilities must fall within these two ranges, and these ranges overlap, we cannot conclude which vaccine is to be preferred for males (based on the data, and our causal hypothesis). For females, the situation is similar, although the probability ranges differ.

Does this mean that the observational data is useless? Not at all! Suppose there is a third vaccine C, for which a randomized controlled trial has been done, which has a probability of a positive outcome of 90%. Then we know that vaccine C is to be preferred over both vaccines A and B for males, and that it is preferred over vaccine B for females. So, in a randomized controlled trial we only need to compare vaccines A and C for females (which is obviously more efficient than setting up a large randomized controlled trial to compare all three vaccines for both genders).

Estimating causal effects from observational data is done routinely by medical researchers. What I don't understand, though, is why these bounds are typically not reported. Instead, researchers usually only report confidence intervals for point estimates, making the strong, typically untestable (and likely wrong), assumption that there is no unobserved confounding between treatment and outcome.

Gender bias at UC Berkeley?

The numbers in the previous example of Simpson's paradox were made up. But one also encounters this paradox in real life. A famous case concerned the admission of graduate students at University College Berkeley [1]. In 1973, university officials noticed in pooled data that while $\sim 45\%$ of all male applicants were admitted, only $\sim 30\%$ of all female applicants were admitted.[11] This being a statistically significant difference (with an astronomically small p-value of 10^{-22}), the officials called for a closer investigation as they were anxious the university might get sued for gender discrimination. Statisticians looking into the data noticed that the difference in acceptance probability mostly disappeared when looking at the level of individual departments, while some departments even showed a slight bias *in favor* of females. What would you conclude: does this provide evidence of gender bias, or not?

It is again helpful to consider possible causal hypotheses, three of which are illustrated in Figure 9. The simplest hypothesis is (a): gender (Z) causes department choice (M), which then influences admission (Y) — but without an unfair *direct* effect of gender on admission. Hypothesis (b) and (c) add an unobserved confounder

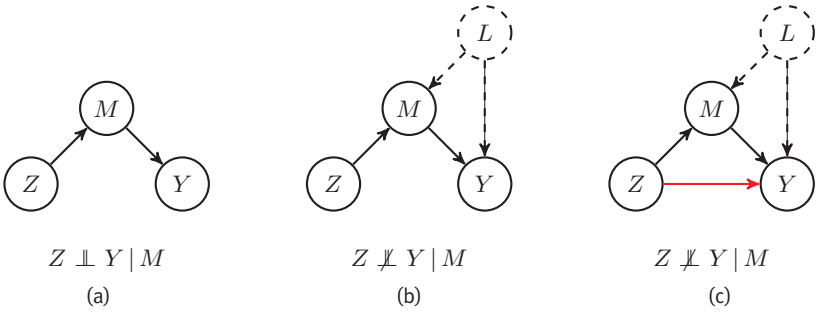


Figure 9. Three possible causal hypotheses for the Berkeley admission data (Z: gender; Y: admission; M: department choice; L: latent confounder). (c) is considered unfair, whereas (a) and (b) are considered fair. (a) implies a conditional independence between Z and Y given M, whereas for (b) and (c) this would generically not be the case.

(L) of department choice and admission. An example of such a confounder could be someone's math skills, which may influence both their department choice (as someone with bad math skills might be more likely to apply at a literature department than at an engineering department) and their admission (if math skills are taken into account in the selection procedure). Hypothesis (c), finally, incorporates the possibility of unfair gender bias as a direct effect of gender on admission (the arrow marked in red).

For (a) one can prove that admission rates are the same for males and females within each department, while for (b) and (c) this is not necessarily the case.[12] Testing for this conditional independence (of Z and Y given M) in the data, we find that hypothesis (a) is almost compatible with the data, but there is still some evidence of a weak conditional dependence of Z and Y given M.[13] If we decide to therefore reject hypothesis (a), we can proceed by performing a statistical test whether the data would favor hypothesis (c) over (b).[14] It turns out that the data is perfectly compatible with hypothesis (b), which describes a fair selection process, and contains no evidence that hypothesis (c) would be more likely. This means that based on the data itself, and our causal hypotheses, we *cannot* conclude that there must have been gender bias in the selection process.[15]

Gender bias at Dutch universities?

In the Netherlands, about 5% of the PhD students that graduate, do so *cum laude* ('with distinction'). Naïvely, one would expect that this percentage should be the same for male and female PhD students. However, in 2018, the Dutch newspaper *NRC* reported that at most Dutch universities, the fraction of male PhD students that graduate *cum laude* is about twice as large as the fraction of female PhD students that do so [3]:

> "Aan alle Nederlandse universiteiten hadden mannen de afgelopen jaren meer kans om *cum laude* te promoveren dan vrouwen. De criteria voor *cum laude* promoveren zijn niet objectief gedefinieerd, dus er is volop ruimte voor genderbias."

The University of Amsterdam is no exception: 6.1% of the male PhD candidates graduated cum laude in 2010–2017, versus 3.7% of the female PhD candidates according to the newspaper article.

This story starts out analogous to that of the UC Berkeley admission case, but I do not know how it unfolds. Could this be an instance of Simpson's paradox, and would the difference disappear if we conditioned on, say, faculty? Or is this perhaps a real case of gender bias? I do not know, but it certainly deserves closer investigation![16]

My research

After this introduction to the field of causality, I would like to say a few words about my own research. I have tried to summarize the research I have been doing over the past 15 years on a single slide. The result is shown in Figure 10.

It all starts with research questions related to *modeling*: in other words, what are appropriate and useful mathematical representations of causality in various types of systems.

Once the modeling framework has been established, the next stage consists of developing mathematical results that aid causal *reasoning*, for example, Markov properties, or the natural bounds that we have seen before. Some challenging aspects in these first two stages are feedback loops (or 'causal cycles', more on those in a minute), modeling and reasoning with data from different domains (for example, combining observational and interventional data), and modeling dynamical systems (where variables become time-dependent). Generality and complexity both increase when not ruling out unobserved confounding and selection bias.

The next stage, 'problem solving', is where most of the hard work has to be done: in order to answer causal questions, we develop statistical algorithms

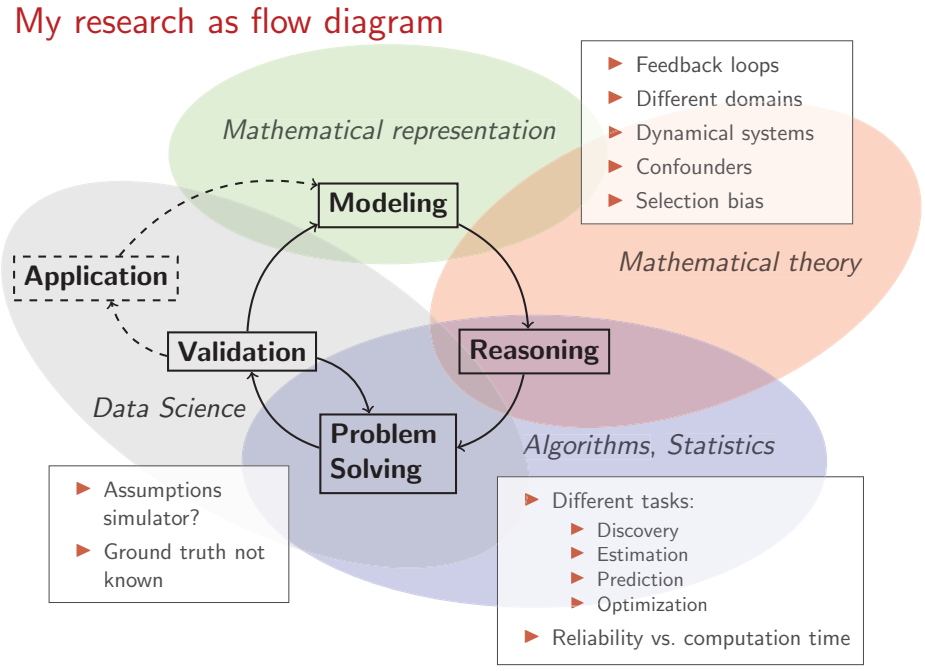


Figure 10 Summary of my research over the past 15 years.

that estimate quantities of interest from data. We improve existing algorithms to become more generally applicable, or more efficient in terms of statistical power or computation time. We attempt to give theoretical guarantees for our algorithms: we try to prove that with a high probability, the algorithms give the correct answer for sufficiently large data sets, under suitable assumptions. Here, one has to distinguish different tasks: discovering causal relations, estimation of model parameters, predicting results of actions, or optimizing a reward. Each problem can often be solved in many different ways, yielding different trade-offs between statistical power, computation time and generality.

As the British say, "the proof of the pudding is in the eating". The next, and perhaps most important, stage is to *validate* how well the algorithms work on real data, by establishing benchmarks. This is the part of the field that is still most in its infancy, especially compared with other parts of AI (like computer vision and natural language processing), where the availability of good benchmarks has led to rapid and impressive developments. It is easy to run computer simulations to get a feeling for how well our algorithms perform on finite data sets, but how realistic are such simulations? When using real data, we often face the problem that the ground truth is not known, and we cannot assess how reliable our algorithms are. In my experience, results on real data often turn out to be disappointing, which then motivates going back to the 'problem solving' stage, and trying harder to design an efficient algorithm, or to go back entirely to the 'modeling' stage.

I would love to tell you more about several of my research projects, but time is limited. Therefore, I just picked two topics that I am especially passionate about.

Feedback loops

The research I am most proud of is our work on *feedback loops*. Many causality researchers have, for a long time, considered feedback loops as exotic, complicated, and some even went as far as to question their very existence. The typical mindset of many causality researchers regarding causal cycles seemed to be: an intriguing possibility, but unlikely to be relevant in the real world.

However, once you start looking for feedback loops, you will see them everywhere. In other fields, no one questions their existence (see also Figure 11). As an example, take climate science. Citing a report of the United Nations Environment Programme [13]:

> "Part of the uncertainty around future climates relates to important feedbacks between different parts of the climate system: air temperatures, ice and snow albedo (reflection of the sun's rays), and clouds."

The word feedback appears 43 times in this 238-page report! Or, in biology. I cite from an editorial of the American Society of Clinical Oncology daily news [9]:

> "Feedback mechanisms may be critical to allow cells to achieve the fine balance between dysregulated signaling and uncontrolled cell proliferation (a hallmark of cancer) as well as the capacity to switch pathways on or off when needed for physiologic purposes."

Or, take econometry. In simple models of an ideal economic market, a feedback loop determines the equilibrium price of a commodity trading good: the price is determined by supply and demand, while supply and demand both depend on price. This is why it is difficult to predict, for example, the gas price.

In a series of papers that my collaborators and I wrote over the past 10 years, we extended the popular causal modeling framework of path diagrams (pioneered by Wright one century ago) to incorporate feedback loops, and we generalized causal reasoning theory and algorithms to allow for feedback loops.

Causal discovery

A research topic that I am also passionate about is *causal discovery*, that is, inferring the presence or absence of causal relations from data. An intriguing alternative to randomized controlled trials is to exploit conditional independences. About three decades ago, causality researchers realized that certain statistical patterns in data can be considered as `fingerprints' of the causal relationships between the variables. These conditional independence patterns can be probed with statistical tests, and under certain assumptions one can draw some conclusions on what the causal relationships between the variables must have been. I illustrated the inference process in Figure 12. Interestingly, this idea works even when allowing for unobserved confounders and selection bias (and, as we have shown more recently, also causal cycles).

Based on this principle, many causal discovery algorithms have been devel-

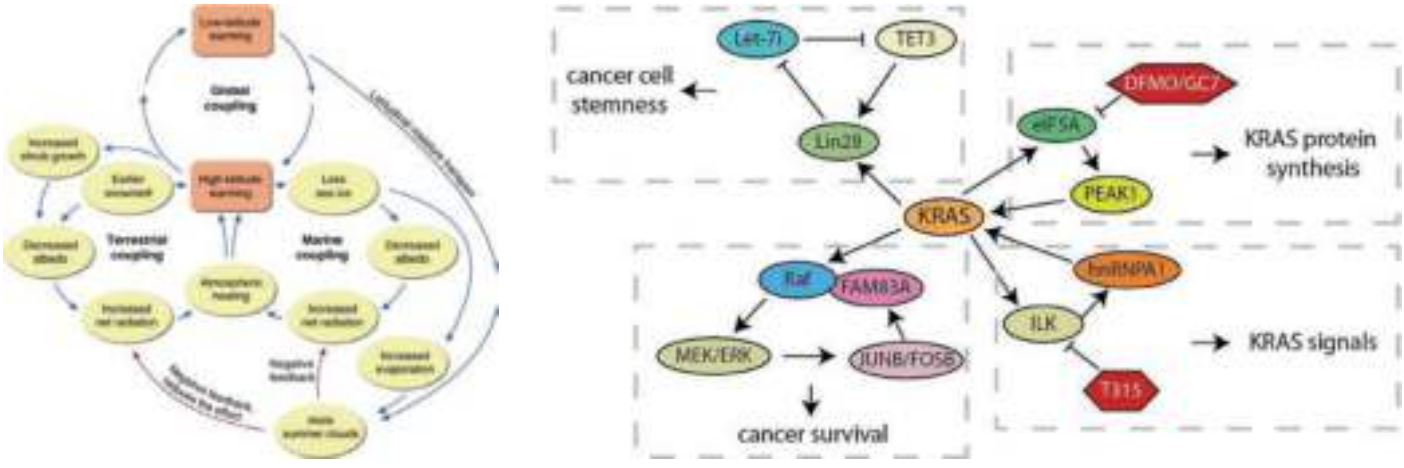
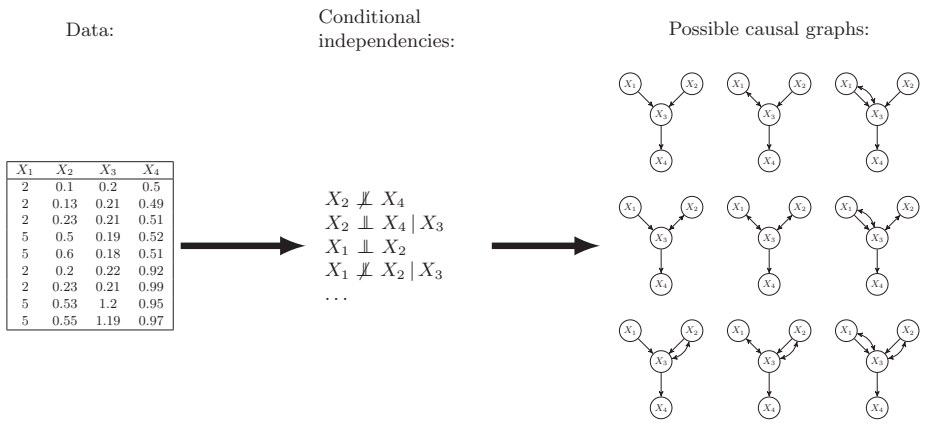


Figure 11 Feedback loops occur in many different systems, for example in climate science[17] (left) and in biology[18] (right).



X_1	X_2	X_3	X_4
2	0.1	0.2	0.5
2	0.13	0.21	0.49
2	0.23	0.21	0.51
5	0.5	0.19	0.52
5	0.6	0.18	0.51
2	0.2	0.22	0.92
2	0.23	0.21	0.99
5	0.53	1.2	0.95
5	0.55	1.19	0.97

Figure 12 Constraint-based causal discovery algorithms infer possible causal structures from conditional independences (certain statistical patterns that can be tested in data).

oped. We can prove that these algorithms work under suitable assumptions. We can demonstrate empirically that they work in computer simulations. But do they work in the real world? Answering this question turned out to be surprisingly hard.

One of our attempts involves yeast. Yeast is a pretty useful organism: you can use it to bake bread, to brew beer, or, as we did, to validate causal discovery algorithms. We made use of a large-scale gene expression data set for yeast, measured in a huge experimental effort by researchers from UMC Utrecht and Utrecht University [8]. The expression levels of about 6000 yeast genes were measured under many different experimental conditions, including almost 1500 single gene knockout interventions. This gigantic randomized controlled trial allows us to estimate the gene regulatory network. The resulting causal graph (with almost 6000 variables!) is depicted in Figure 13, and as you can see, it looks pretty complicated (while it is not even complete!).

The challenge we took on was to predict from this data which gene expression levels change if we knock-out a certain gene *without actually using the data for that knock-out experiment*. This challenge, at first sight, appears similar to reading off from the data in Figure 4 whether chocolate consumption causes Nobel laureates. I have visualized the expression data for two examples of correlated gene pairs in Figure 14. On the horizontal axis we have the expression of one gene, and on the vertical axis, the expression of another gene. The data points correspond with relative gene expression levels in colonies of yeast cells. The blue points are the observational data, and the red points are interventional data corresponding with different single gene knockouts. The causal discovery algorithm is solving a challenging task: it has to decide from the data what will happen with the expression of the gene on the vertical axis if we reduce the expression of the gene on the horizontal axis by knocking it out. There are millions of such gene pairs that we can consider!

A key difference with the data in Figure 4 is that we have an additional variable: the experiment type (observational or interventional). With a conditional independence test, we can use this additional information to decide whether an observed correlation between gene expression levels implies a causal relationship of one on the other. Figure 14(a) shows a case in which the causal discovery algorithm correctly identified a causal relation: the expression of gene YPL273W causes the expression of gene YMR321C.[20] In Figure 14(b) we see a case where a causal relation found by the causal discovery algorithm is incorrect: there, knocking out YPL154C does not significantly change the expression of YDR032C.

With collaborators of ETH Zürich we validated how well causal discovery algorithms perform in this challenge [11]. Here I will focus on LCD, one of the simplest constraint-based causal discovery algorithms [4]. Out of more than 19 million candidate cause-effect gene pairs, I preselected roughly 1000 using an algorithm called L_2Boosting, and of those 88 were selected by the causal discovery algorithm LCD. For 9 of those, the data provides the ground truth causal effect. One can see clearly in

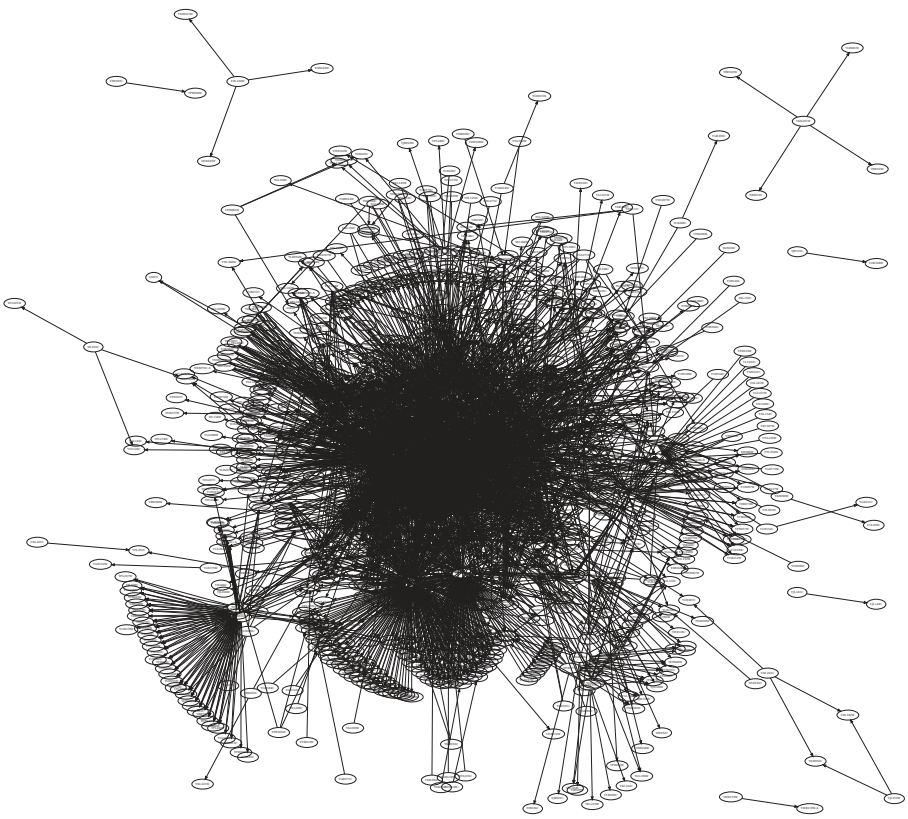


Figure 13 Estimated subgraph of the gene regulatory network for yeast, based on experimental data from [8].[19]

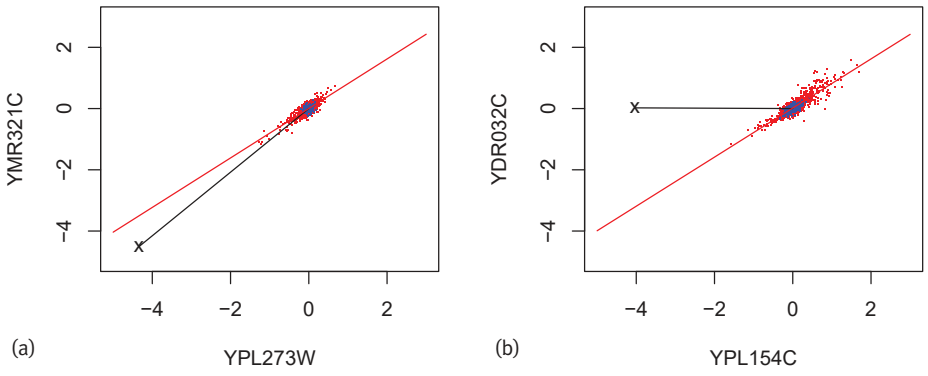


Figure 14 Examples of correlated gene pairs in the yeast gene expression data from [8]. Observational data are in blue, interventional (some gene knockout) in red, and the black cross shows the result of knocking out the gene on the horizontal axis. (a) is causal (YPL273W causes YMR321C), while (b) is not (YPL154C does not cause YDR032C).

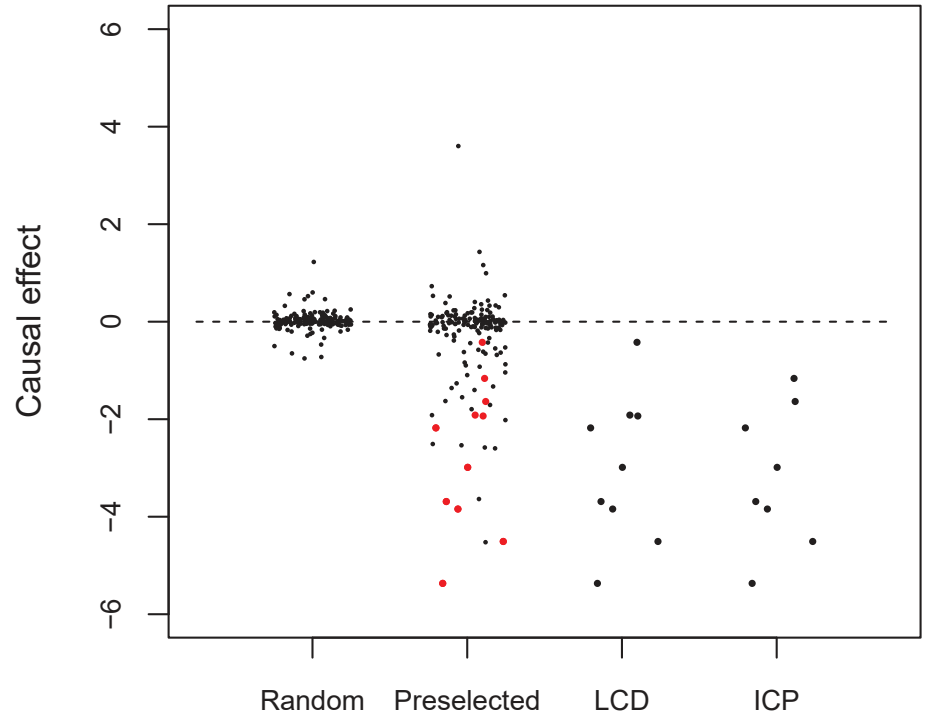


Figure 15 Validation of causal discovery algorithms with the knockout data from [8]. Random: randomly selected gene pairs (baseline); Preselected: gene pairs preselected using L_2Boosting; LCD/ICP: further selected using LCD/ICP.

Figure 15 that these 9 gene pairs indeed correspond with a nonzero causal effect, and hence, a true gene regulatory relation. To my knowledge, this was the first convincing proof that causal discovery based on conditional independences can indeed work in practice.

Applied research

While the core of my research has always been theoretical, I also enjoy more applied research projects where we try to put the theory to use. I will mention a few examples. I am involved in the AI4Science project of the Faculty of Science of the University of Amsterdam, where we collaborate with biologists to apply and further develop causal discovery methods to improve the understanding of gene regulation in a bacterium, *bacillus subtilis*. Microsoft Research sponsors one of my PhD students to study how causality can be used for developing personalised medication and treatment strategies. The Mercury Machine Learning Lab, one of the labs of the Innovation Centre for Artificial Intelligence, is funded by booking.com. I am involved in this lab as co-director and as a supervisor of two PhD students and a postdoc. The aim of the lab is to carry out fundamental scientific research on learning from controlled sources, for example for taking informed business decisions. It should come as no surprise that causality is considered to be of essential importance. These collaborations are great opportunities to see some of the algorithms being deployed in practice.

Ik heb gezegd.

Acknowledgments

I thank my colleagues of the Korteweg-de Vries Institute for Mathematics, and in particular Eric Opdam, Michel Mandjes, and the others who played a role in my transition from the Informatics Institute. I hope that I will successfully follow in the footsteps of my precursors on this chair. I am indebted to those who funded my research, amongst them NWO, ERC, various universities and companies, but also all European tax payers. I have had the privilege of doing research with many different people, and I would like to thank them all. In particular, many thanks to my former and current PhD students Tineke Blom, Stephan Bongers, Philip Versteeg, Noud de Kroon, Teodora Pandeva, Philip Boeken and Leihao Chen, and my former postdocs Thijs van Ommen, Patrick Forré and Sara Magliacane, for the interesting discussions and great collaborations.

Notes

1 While this multi-disciplinarity emphasizes its scientific and practical relevance, it also bears a cost: many ideas are lost in translation. The often isolated, scientific communities have developed different (sometimes contradictory) terminology and conceptual frameworks. Even within statistics, different frameworks are being used. This is unfortunate, as it puts a considerable hurdle on scientific exchange and interaction.

2 Reproduced from https://ourworldindata.org/smoking-big-problem-in-brief, licensed under CC BY 4.0.

3 Reproduced from https://tylervigen.com/spurious-correlations, licensed under CC BY 4.0.

4 https://tylervigen.com/spurious-correlations.

5 A more precise definition is the following. Consider modeling a system with two real-valued variables X and Y, where X is the cause and Y the effect (for example, X could be the dosage of ibuprofen administered to a patient, and Y the patient's body temperature measured 60 minutes later). The potential outcomes are random variables $Y(x) : \Omega \to \mathbb{R}$ for $x \in [0, \infty)$, where $Y(x)$ is distributed as the effect after an intervention that sets X to x, and Ω is a probability space. At most one of the po-

tential outcomes can be measured (indeed, it is not possible to administer two different dosages simultaneously). Alternatively, one can define a statistical causal model that consists of the family of distributions $x \mapsto \mathbb{P}(Y \mid \mathrm{do}(x))$, with $\mathbb{P}(Y \mid \mathrm{do}(x))$ the distribution of $Y(x)$. Here, x is treated as a parameter, and we have avoided to introduce the potential outcomes as random variables jointly defined on the same probability space Ω. While the potential outcomes allow for specification of their joint distribution, the alternative approach only provides access to their marginal distributions.

6 For an account on the history of the various frameworks by some of the pioneering researchers, see *Journal of Observational Studies* 9(2), 2022.

7 I reproduced the visualization in [10], but made use of newer data from https://www.theobroma-cacao.de/wissen/wirtschaft/international/konsum on chocolate consumption in 2017, and from https://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita on scientific Nobel laureates until 2019. This explains deviations from Figure 1 in [10].

8 That is actually a simplification. Instead, astronomers also rely on experiments performed in our solar system, but they extrapolate the results of these experiments to the rest of the universe, assuming that the same physical laws hold everywhere.

9 Keep in mind that these numbers are not supposed to be realistic at all, the numbers were just chosen to make it easier to tell this story. I took them from [18].

10 It could be the case that for each combination of gender and age, vaccine B seems preferable, while vaccine A seems preferable when aggregating the data over all age groups.

11 For my calculations, I used the 4257 samples concerning the six largest departments, available as `UCBAdmissions` in the `R` package `datasets`. This appears to be a subset of the 12763 samples reported in [1], which may explain any discrepancy with their numbers.

12 This amounts to showing that hypothesis (a) implies the conditional independence of Z and Y given M, something that follows for example by using the d-separation criterion [15]. In (b), conditioning on department choice may create a dependence between gender and admission via the 'explaining away' phenomenon (Berkson's paradox). In (c), a direct effect of gender on admission can also lead to a conditional dependence between gender and admission given department choice.

13 Using the G-test to test for dependence between gender and admission, one finds that gender and admission are strongly correlated when aggregating over departments (p-value 4×10^{-22} for testing $Z \perp\!\!\!\perp Y$), whereas after conditioning on department only a slight correlation is left (p-value 0.0014 for testing $Z \perp\!\!\!\perp Y \mid M$).

14 This is more complicated than testing for a conditional independence, but can be done by defining 'response variables' and using linear programming to characterize the probability distributions compatible with hypothesis (b) as the convex hull of a finite number of extreme points (see also [2]), which can be translated into a finite set of inequality constraints on $p(Z, Y \mid M)$. Since the empirical distribution turns out to fall inside this convex hull, hypothesis (b) need not be rejected in favor of hypothesis (c).

15 This conclusion differs from that of [1], who conclude that the data suggest evidence of gender bias *against* males.

16 Unfortunately, the data is not publicly available.

17 Reproduced from https://www.grida.no/resources/5261 with permission of the creator (Hugo Ahlenius, UNEP/GRID-Arendal).

18 Reproduced with permission from [6], licensed under CC BY 4.0 (Creative Commens).

19 In this graph, directed edges represent an estimated subset of the direct causal effects of gene knockouts on gene expressions. I defined a (possibly indirect) causal effect as a $\log_2$ fold change of at least ± 2, according to the data from [8], and then took the transitive reduction of the corresponding directed graph to obtain a minimal set of direct causal effects.

20 Indeed, the black cross in Figure 14(a) denotes the data point that corresponds with the knockout of YPL273W. In this knockout experiment, we see that the expression of YPL273W is substantially reduced (so normally this gene is active), but also the expression of YMR321C; hence, YPL273W causes YMR321C. Interestingly, YPL273W and YMR321C turn out to be *paralogs*: these two genes contain an almost ($> 98\%$) identical subsequence of more than 300 base pairs. Therefore, I do not know whether this finding that YPL273W causes the expression of YMR321C is biologically interesting, or is just an artefact of the experimental knockout procedure.

References

1 Peter J. Bickel, Eugene A. Hammel and J. William O'Connell, Sex bias in graduate admissions: Data from Berkeley, *Science* 187 (1975), 398–403.

2 Blai Bonet, Instrumentality tests revisited, in *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence (UAI 2001)*, 2001, pp. 48–55.

3 Ellen de Bruin, Helft van de promovendi is vrouw, maar cum laude krijgen ze zelden, *NRC Handelsblad*, 20 October 2018.

4 Gregory F. Cooper, A simple constraint-based algorithm for efficiently mining observational databases for causal relationships, *Data Mining and Knowledge Discovery* 1(2) (1997), 203–224.

5 Gartner, Gartner identifies key emerging technologies expanding immersive experiences, accelerating AI automation and optimizing technologist delivery, gartner.com, 10 August 2022.

6 Weigang Gu, HongZhang Shen, Lu Xie, Xiaofeng Zhang and Jianfeng Yang, The role of feedback loops in targeted therapy for pancreatic cancer, *Frontiers in Oncology* 12 (2022), 800140.

7 David Hume, *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*, John Noon, 1740.

8 Patrick Kemmeren, Katrin Sameith, Loes A. L. van de Pasch, Joris J. Benschop, Tineke L. Lenstra, Thanasis Margaritis, Eoghan O'Duibhir, Eva Apweiler, Sake van Wageningen, Cheuk W. Ko, Sebastiaan van Heesch, Mehdi M. Kashani, Giannis Ampatziadis-Michailidis, Mariel O. Brok, Nathalie A. C. H. Brabers, Anthony J. Miles, Diane Bouwmeester, Sander R. van Hooff, Harm van Bakel, Erik Sluiters, Linda V. Bakker, Berend Snel, Philip Lijnzaad, Dik van Leenen, Marian J.A. Groot Koerkamp and Frank C. P. Holstege, Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors, *Cell* 157(3) (2014), 740–752.

9 Grant A. McArthur, The RAS/RAF/MEK/ERK pathway in cancer: Combination therapies and overcoming feedback, Editorial of the *ASCO Daily News*, June 2014

10 Franz H. Messerli, Chocolate consumption, cognitive function, and Nobel laureates, *New England Journal of Medicine* 367 (2012), 1562–1564.

11 Nicolai Meinshausen, Alain Hauser, Joris M. Mooij, Jonas Peters, Philip Versteeg and Peter Bühlmann, Methods for causal inference from gene perturbation experiments and validation, *Proceedings of the National Academy of Sciences of the United States of America* 113(27) (2016), 7361–7368.

12 Charles F. Manski and Daniel S. Nagin, Bounding disagreements about treatment effects, *Sociological Methodology* 28(1) (1998), 99–137.

13 James E. Overland, John E. Walsh and Muyin Wang, Why are ice and snow changing?, in UNEP/GRID-Arendal, ed., *Global Outlook for Ice & Snow*. UNEP, 2007.

14 Karl Pearson, *The Grammar Of Science*, Adam & Charles Black, 1892.

15 Judea Pearl, *Causality: Models, Reasoning and Inference*, Cambridge University Press, 2009.

16 Judea Pearl, *The Book of Why: The New Science of Cause and Effect*, Basic Books, 2018.

17 Bertrand Russell, On the notion of cause. *Proceedings of the Aristotelian Society* 13 (1913), 1–26.

18 Larry Wasserman, *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics, Springer, 2004.

19 Alfred N. Whitehead and Bertrand Russell, *Principia Mathematica*, Cambridge University Press, 1925–1927.

20 Sewall Wright, The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs, *Proceedings of the National Academy of Sciences* 6 (1920), 320–332.

Piet Groeneboom
Delft Institute of Applied Mathematics
TU Delft
p.groeneboom@tudelft.nl

Column Piet takes his chance

Buffon's needles and noodles

Piet Groeneboom regularly writes a column on statistical topics in this magazine.

The dual problem and Oberwolfach

What mathematical subject could be of interest to secondary school pupils? My colleague Chris Klaassen and I had (completely independently) the same idea: Buffon's needle problem. Chris talked about it in a lesson for pupils of the Leiden city gymnasium during his study and I wrote about it in a flyer for secondary schools of the Mathematical Institute of the University of Amsterdam in 1987. The flyer was intended to seduce the pupils of these schools to study mathematics (at the University of Amsterdam).

I recently understood from Chris that his main conclusion after the lesson at the Leiden school was that any desire to be a teacher at a secondary school had left him. And my own conclusion after writing about Buffon in the flyer of the Mathematics Institute of the University of Amsterdam was that I should perhaps write about something else if I would have to do it again.

After I had written the draft of my contribution on Buffon's needle for the flyer I received visits from nearly all my colleagues at the mathematics institute who told me that they thought it was too difficult for secondary school children. Only my colleague Ietje Paalman said that she thought it was an enjoyable piece to read, but also thought it was too difficult. When I moved from Amsterdam to Delft University in 1988 my contribution to the flyer was immediately replaced by an interview with Annoesjka Cabo (who wrote a dissertation on stochastic geometry, though, of which the Buffon needle problem is a kind of beginning).

So let's try Buffon's needle problem again, this time aimed at readers who may be somewhat beyond high school age. And conveniently in this magazine where Casper Albers, my predecessor as columnist, has already mentioned this problem and laid out the standard solution using trigonometry. But for this article I'm going to look at some altogether more interesting pathways to the proof.

My preferred proof (the proof I also tried to explain in the flyer for the secondary schools in Holland) is sketched in the introduction of the remarkable book on combinatorial integral geometry [2] by the Armenian mathematician Rouben Ambartzumian. Rouben is a recipient of the Rollo Davidson Prize (1982). I met him in Oberwolfach. He had a lot of jokes about mathematicians in Moscow, which resembled the Dutch jokes about Belgians (and possibly the Belgian jokes about the Dutch).

Now that I am mentioning Oberwolfach again, I realize that some readers might not know what I am talking about. The 'Mathematisches Forschungsinstitut Oberwolfach' (MFO) originally consisted (until 1969) of one building, 'the old castle', on a hill outside the village Oberwolfach in the Schwarzwald, Germany, where mathematicians could be invited to stay for a week to discuss a particular subject in mathematics. Then a new guest house and bungalows were constructed (in total, 3 buildings). In 1975 the old castle was replaced by a modern conference and library building.

It was founded during the Second World War (1944!). Nowadays there is one building where the bedrooms are and meals are served, and the library building is for the lectures. In the latter building there is also a billiard table, a ping pong room and a music room. The music room has a grand piano and several string instruments (not enough to play string quartet, though, but people often bring their own instrument) and also a pretty good choice of music scores. There is also a choice of wines and other drinks in this building.

Originally, one could only stay there by being the organizer of a workshop or by being invited by the organizers. Presently there are more buildings for guests and also the rules for staying there have somewhat changed. For example, one can try to get a grant for stay-

ing there with a colleague to work on a special project ('Oberwolfach Research fellows', previously called 'Research In Pairs', RIP). The door to one's bedroom had no key, "mathematicians trust one another" (this has changed too, since 2008 it is possible to lock the door from the inside, apartments for longer stays also have different rules).

Depending on the type of mathematicians gathering for such a week in Oberwolfach, there might be a lot of wine drinking and going to bed in the middle of the night/early morning or early to bed, early to rise. The week on stochastic geometry where I met Rouben Ambartzumian was a week of the first type. I had guessed this correctly beforehand (weeks of this type are very exhausting) and did not stay for the next week, for which I was also invited (this was a week on Lie algebras, if I remember correctly, where Tom Koornwinder was one of the organizers). So I do not know what the habits of this particular group were.

Anyway, coming back to Ambartzumian's sketch of proof of the probability in Buffon's needle problem, I now quote the first two paragraphs of the introduction of his book.

> "The idea of introducing measures into the space of lines in the plane was already implicit in Buffon's classical *needle problem*. Let us recall its formulation. The plane is ruled by a fixed lattice of parallel lines unit distance apart. A needle ν of length $|\nu|<1$ is 'thrown' at random onto the plane. What is the probability of the event that in its final position the needle will be intersected by a line of the lattice?
>
> In an equivalent formulation, needle and lattice exchange roles and one assumes that it is now the needle which is fixed in the plane with the lattice being thrown down at random."

When I first read this, "picking up the lines, throwing them on the non-moving needle", I liked this idea very much. And I thought this might also appeal to the secondary school pupils. Whether it did I do not know. In any case it did not appeal to my colleagues at the Mathematics Institute in Amsterdam.

The first bottleneck is the introduction of a measure on the space of lines. The invariant measure for lines in the plane is what Lebesgue measure ('area') is for points in the plane. The measure is invariant under translation and rotation. I tried to explain this concept in the flyer and this was something that my colleagues found particularly offensive. I said something like "invariant measure for lines is what area is for points". Unfortunately, I do not have the flyer anymore, so I do not know what I said exactly.

One would think that this flyer, which presumably was sent to all Dutch schools, should be traceable. I indeed contacted the Mathematics Institute of the University of Amsterdam in view of the present column, but the director of the institute told me that they could not look thoroughly for it because they were understaffed. This is the more regrettable, since Tobias Baanders, then at the Mathematical Centre (CWI), drew a very good picture for my contribution, where one could see a soldier (or officer?) of the American civil war with one arm in a sling, determining π with his non-wounded arm by throwing a stick on a rotating wooden disc (see below for the occurrence of π in this context). This described a historical event (which I had picked up from a book on 250 years of Buffon's needle).

Returning to Ambartzumian's book: he continues his exposition in the Introduction by saying:

> "Without loss of generality one may assume that the needle lies within some fixed open disc K of unit diameter. Then in all possible outcomes of the lattice-throwing experiment the disc K is intersected by exactly one line of the lattice, if we assume that the case of tangency is 'impossible' (i.e., of probability zero). Since other lines of the lattice now play no role, we may fix attention to this single line, g_K say, intersecting K, and Buffon's original problem is now replaced by the following one: what is the probability p that the random line g_K intersecting K should also intersect the needle? We may refer to this as the *dual* problem to the classical Buffon's needle problem."

And then, skipping a paragraph which is not directly relevant:

> "In the classical solution of Buffon's problem it is assumed that the centre of the needle and its orientation are independently and uniformly distributed, that is that the projection of the centre onto a line perpendicular to the lines of the lattice is uniformly distributed on some segment of unit length, and the angle between the line containing the needle and the lines of the lattice is independently and uniformly distributed on $(0,\pi)$. With these assumptions
>
> $$p=\frac{2}{\pi}|\nu|.$$
>
> (This example is the earliest instance of the calculation of a 'geometrical probability'.)
>
> To this result corresponds the following solution of the dual problem: there is a unique distribution P of the random lines g_K such that
>
> $$P\{g_K \text{ intersects } \nu\}=\frac{2}{\pi}|\nu|, \tag{1}$$
>
> for every needle ν within K. This distribution P is proportional to the restriction, to the set of lines intersecting K, of the so-called *invariant measure* on the lines of the plane."

Ambartzumian does not explain how the $(2/\pi)|\nu|$ arises in the dual problem, but I will explain this now. There is the following 'easy to believe' theorem in stochastic geometry.

Theorem 1. *Let $A\subset B$ be two bounded convex sets in the plane. Then the probability that a random line meets A, given that it meets B is given by*

$$p=\frac{\ell(\partial A)}{\ell(\partial B)}, \tag{2}$$

where ∂A (∂B) is the boundary of A (B) and ℓ denotes length.

The length of the boundary of the needle ν, as a degenerate convex set in the plane, is $2|\nu|$ in this interpretation. In measuring the length of the boundary we consider a lower side and an upper side. If we apply this to the computation of the probability on the left in (1) we get: $2|\nu|$ (length of boundary of needle) divided by π (length of circumference of circle with diameter $d=1$).

This argument also immediately gives that if the distance between the lines of the lattices is d instead of 1, the probability is given by

$$\frac{2|\nu|}{\pi d},$$

if $|\nu|<d$.

Buffon's noodles

In [1] the 'Buffon's noodle' proof is given. This proof starts with the trivial observation that the probability we are looking for can also be interpreted as an expectation of number of crossings. This phenomenon is due to the 'hit or miss' character of the event of hitting a line of the lattice, the expectation of 'the number of crossings' is $p \cdot 1 + 0 \cdot (1-p) = p$, if p is the probability of crossing.

We can then consider the expectation of the number of crossings of lines of the lattice by a 'polygonal needle' P (a *noodle*), consisting of several needles. By the linearity of the expectation operator, the expected number of crossings of lines in the lattice by this polygonal needle will be proportional to its length. In expectation notation we get, for some constant $c > 0$,

$$\mathbb{E}\{\#\text{crossings of } P \text{ with lines of lattice}\} = c\ell(P) \tag{3}$$

if $\ell(P)$ is the length of the polygonal needle P.

In particular, we can consider polygonal needles approaching a circle with diameter d if the distance between the lines of the lattice is d. A circle with diameter d, thrown down on this lattice, will always have *two* crossings. Considering polygonal needles P_n of this type, approaching the circle, we get from (3), disregarding circles which hit two lines and have tangents there (this has probability zero)

$$\lim_{n\to\infty} c\ell(P_n) = c\pi d = 2.$$

implying $c = 2/(\pi d)$. This in turn implies, again by (3), that for our simple (degenerate polygonal) needle ν with length $|\nu|$ the expected number of crossings is $2|\nu|/(\pi d)$. And hence, by the trivial observation at the beginning of this argument, this is also the probability sought for.

This is Barbier's solution [4], which also avoids any calculation with sines or cosines. Still I prefer Ambartzumian's proof. It would be great if someone could still locate the flyer of the Mathematics Institute of the University of Amsterdam from (around) 1987 with Ambartzumian's proof and the drawing of Tobias Baanders.

Postscriptum

After having written this column and having sent it to Tobias Baanders, he kindly made a new drawing, based on a silly sketch I made in the train from my memory of his drawing in the flyer. So here is a Tobias Baanders drawing after all with this column. He depicts the situation where the needle is longer than the distance between the lines, and one counts the number of crossings, which in principle leads to more correct digits in the determination of π in this way (which is a rather pointless exercise anyway).

Illustration: Tobias Baanders

Determining π

Acknowledgements

I want to thank Tobias Baanders for making again a nice drawing and Professor Stephan Klaus (Scientific Administrator of the MFO) for providing additional information on Oberwolfach.

References

1 Martin Aigner and Günter M. Ziegler, *Proofs from The Book*, Springer, 2018, sixth edition.

2 R.V. Ambartzumian, *Combinatorial Integral Geometry: With Applications to Mathematical Stereology*, Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics, Wiley, 1982.

3 E. Barbier, Note sur le problème de l' aiguille et le jeu du joint couvert, *J. Mathematiques Pures et Appliquées* 5(4) (1860), 273–286.

Jan Hogendijk
Mathematisch Instituut
Universiteit Utrecht
j.p.hogendijk@uu.nl

Jan Hogendijk

Afscheidsrede

Friese wiskunde en het digitale tijdperk: de strijd tussen Pibo Gualtheri en Jan Beerntsen (1612–1613)

Op 7 september 2022 hield Jan Hogendijk, hoogleraar aan het Mathematisch Instituut van de Universiteit Utrecht, zijn afscheidscollege. Dit artikel is hierop gebaseerd.

Het onderzoek in de geschiedenis van de Nederlandse wiskunde is door digitalisering veel gemakkelijker geworden. Tussen 1580 en 1850 zijn honderden Nederlandstalige boeken over wiskunde en aanverwante onderwerpen verschenen. Tot voor kort konden deze boeken alleen worden bekeken in gespecialiseerde bibliotheken. Uitlenen was onmogelijk vanwege de hoge ouderdom en de zeldzaamheid. Inmiddels is meer dan 80 procent van deze boeken gedigitaliseerd en gratis toegankelijk. Dat betekent dat iedereen deze oude Nederlandse wiskundeboeken op de eigen laptop kan lezen en downloaden. Alle wiskundeboeken van voor 1700 worden beschreven in de bibliografie van Klaas Hoogendoorn.[1] Digitale versies kunnen gevonden worden door zelf te zoeken op het internet, en in de lijsten op mijn website www.jphogendijk.nl.

Achter veel van deze boeken zit een leuk verhaal. Voor mijn afscheidscollege heb ik een boekje uitgekozen dat te maken heeft met de plaatsen waar ik ben geboren (Leeuwarden) en opgegroeid (Franeker). Het boekje is anderhalve eeuw ouder dan het beroemde planetarium van Eise Eisinga (1774–1781). Het heeft de intrigerende titel:

> *ANTWOORT op de Spits-feninighe selfs niet ghemaeckte voor-reden en Tael Quael Solutie (pag. 16. 17. en 18. gestelt) by eenen IAN BEERNTSEN wtgegeven, over t'Geometrische Vraeghstuck inden Iare 1612. binnen* Leeuwarden *openbaer aengheslaghen.*

Dit betekent in modern Nederlands:

> "Analyse van het venijnige en geplagieerde voorwoord en de oplossing in de huidige (slechte) staat (op p. 16, 17, 18 hieronder uitgelegd) door ene Jan Beerntsen, van het meetkundeprobleem dat in het jaar 1612 in Leeuwarden in het openbaar is opgehangen."

De culturele achtergronden van het boekje zijn onderzocht door Arjen Dijkstra in zijn proefschrift.[2] Wij gaan de wiskundige inhoud bekijken.

De auteur van het boekje was Pibe[3] Wouters, die de Latijnse naam Pibo Gualtheri aannam. Deze kleurrijke figuur werd rond 1580 in Franeker geboren en studeerde wis- en sterrenkunde bij Adriaan Metius (1571–1635)[4] aan de Franeker universiteit. Na het afronden van zijn studie werkte hij in een hoge functie bij de Friese overheid in Leeuwarden. Hij liet in 1611 een nieuw huis aan de Eewal in Leeuwarden bouwen en leidde daar een luxueus leven. In zijn vrije tijd bleef hij aan wiskunde doen. Zijn vrouw Trijntje overleed in 1618 en doordat hij kort daarna opnieuw wilde trouwen moest er een opsomming worden gemaakt van zijn inboedel.[5] In deze opsomming staan honderden titels van boeken over wiskunde en sterrenkunde die hij had verzameld. Hij had ook sterrenkundige instrumenten[6] en muziekinstrumenten, en hij hield een papegaai, die we verderop in het verhaal zullen tegenkomen.

Er is weinig bekend over het leven van de tweede hoofdpersoon van ons verhaal, Jan Beerntsen.[7] Hij was voor 1615 officieel als landmeter toegelaten in Friesland, maar verdiende de kost als schipper. Hij had niet aan een universiteit gestudeerd en was dus geen academicus.[8] Jan volgde lessen bij Nicolaus Mulerius (1564–1630).[9] Mulerius had in Leiden gestudeerd en werd daarna stadsarts in de Friese havenstad Harlingen (1590–1603). Van 1608–1614 was Mulerius rector van de Latijnse school in

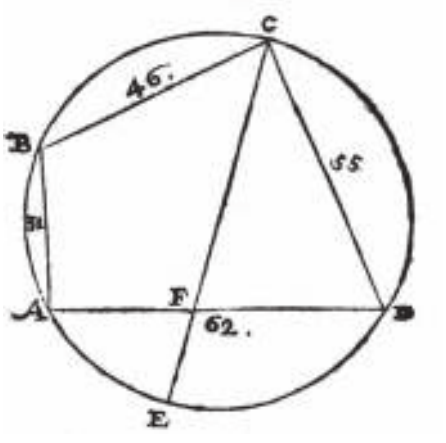
Leeuwarden, daarna werd hij hoogleraar in Groningen. Hij is vooral bekend geworden door zijn uitgave van Copernicus' *De Revolutionibus*.

Pibo en Jan waren in eerste instantie wiskundevrienden. In die tijd was het de gewoonte dat rekenmeesters, landmeters en liefhebbers van wiskunde, al dan niet academisch geschoold, elkaar uitdaagden met problemen, die op een pamflet in het openbaar werden opgehangen. Een eerder meetkundeprobleem van Pibo had al tot ongenoegen tussen de twee geleid. Hier is het probleem dat Pibo in 1612 in sierlijke letters aan de deur van zijn huis in Leeuwarden had opgehangen, samen met Figuur 1:

> D'Linien DF.AF. *van* Diameter *des Circkels uyt C. doorsneden,, snel*
> *Met d'lanckte CF.EF. door d'bekende alhier te produceren,, staen:*
> *Areâ Quadrilateri gedeelt met* $\sqrt{11-\sqrt{82\frac{3}{4}}}$. *ghy zult fourneren,, gaen* T'getal *des* Iaers. *Laet Konst blijcken door goede Bewijsreden,, wel.*

In modern Nederlands betekent dit: Gegeven een vierhoek $ABCD$ die concyclisch is, dat wil zeggen: ingeschreven in een cirkel. De zijden van de vierhoek zijn bekend, namelijk $31, 46, 55, 62$ als in Figuur 1. De middellijn CE van de cirkel snijdt zijde AD in F. Eerste deel: bereken de lengtes van DF, AF, CF, EF. Tweede deel: bereken de oppervlakte van de vierhoek, deel dit getal door $\sqrt{11-\sqrt{82\frac{3}{4}}}$, en zet het quotiënt om in een datum: het gehele deel is een jaartal, het breukdeel wordt omgezet in maand, dag, uur, enz. Het jaartal moet 1612 zijn. Pibo wilde een beredeneerde oplossing.

Op 16 januari 1613 ontving Pibo Jans oplossing van het probleem uit handen van een gemeenschappelijke vriend. Pibo citeert deze oplossing in zijn boekje, en probeert daarna aan te tonen dat de oplossing waardeloos is en Jan Beerntsen een "driedobbelen konst-verachter ende hater derselvighen".



Figuur 1

Jan begon zijn oplossing met een sarcastisch gedicht van vijftig regels, dat door Pibo in zijn geheel wordt geciteerd. Het gedicht bevat verwijzingen naar de Griekse en Romeinse mythologie, kennelijk om de academische stijl van Pibo te imiteren. Pibo verklaarde dat Jan het gedicht niet zelf heeft geschreven, maar hij kan niet aangeven wie de auteur dan wel geweest was.

In het gedicht zegt Jan dat Pibo's probleem weinig diepgang heeft, en te vergelijken is met een angstaanjagende leeuw op een schilderij:

> *Maer als ick met aandacht t'zelfd' wel ginck overwegen,*
> *Vond' ick t'* Chierliick *geprónckt maer weynich angelegen;*
> *Gelijck* Geschilderd' Leeuw *door gramschap toornich siet,*
> *Seer wijt te gapen schijndt maer ganslick bijttet niet.*

Pibo citeert daarna Jans oplossing van het eerste deel van het probleem. De getallen zijn zo lang dat ze niet horizontaal op de pagina's van het boekje pasten, daarom zijn ze verticaal gedrukt:

$$\mathrm{CF} = \sqrt{\frac{5436301662402227469352 5}{43866653126040244500}} + \sqrt{\frac{76784396560888433459224387329088427548859193820419961016088816712565906250 0}{25609864704432475328418265737271609357456029741756721755531727872562500 00}},$$

$$\mathrm{FE} = \sqrt{\frac{5436301662402227469352 5}{43866653126040244500}} - \sqrt{\frac{76784396560888433459224387329088427548859193820419961016088816712565906250 0}{25609864704432475328418265737271609357456029741756721755531727872562500 00}},$$

$$\mathrm{AF} = 31 - \sqrt{\frac{1257801146340365655152109646235636054259963274690612500}{5838116856293710789773543598536468209316402647561250 0}},$$

$$\mathrm{FD} = 31 + \sqrt{\frac{1257801146340365655152109646235636054259963274690612500}{5838116856293710789773543598536468209316402647561250 0}}.$$

Het is meteen te zien dat de breuken kunnen worden vereenvoudigd door nullen weg te strepen, en door het wegdelen van extra gemeenschappelijke factoren 5. We zullen straks ingaan op de vraag waarom Jan zijn oplossing zo formuleerde. We doen nu eerst een poging om de feitelijke gebeurtenissen na ontvangst van de oplossing te destilleren uit Pibo's woedende tirades.[10]

Op 20 januari 1613 werd een ontmoeting tussen Pibo en Jan georganiseerd in het huis van Mulerius, de leraar van Jan. Er waren ook andere, niet met name genoemde, liefhebbers van wiskunde bij aanwezig. In Pibo's geschrift vinden we een paar citaten uit de verhitte discussie die plaatsvond.

Om te beginnen ergerde Pibo zich aan de manier waarop hij door Jan bejegend werd met *Dat heb ick u noch te segghen Maet*, en *du biste my te rap in die beck*. Jan veronderstelde dat Pibo het jammer vond dat een niet-academicus het probleem had opgelost: *T'is dy leet dat een Schipper ende geen Doctor d'Questie ontbonden heeft*.

Wat de oplossing betreft, verklaarde Jan:

> *met d'bevestinge* lae *ter presentie D[oktor] Muleri[us] als andere Konstlievende persoonen … , sulcx alles zo wel Theoricè … als practicè syne ende niemants anders werck te zijn*

Dat kon volgens Pibo niet waar zijn omdat Jan geen bewijzen had gegeven:

> *Alwaar contrarie, een kindt zoude t'mercken, ut syne sotte hiernae ghesettede antwoorden ghebleecke, niet het alderminste bewijs sijns wercx gegeven te hebben, dan seggende metten Pape-gay. Dat heb ick u noch te segghen.*

We gaan nu aan de hand van twee van deze 'sotte antwoorden' het probleem bekijken. Figuur 2 toont de concyclische vierhoek met diagonalen AC en BD die elkaar snijden in G. Jan werkte algebraïsch en stelde $AG = r$, waarbij r de afkorting was van de onbekende wortel (radix) in de cossische algebra, die voorafging aan de algebra van Descartes (1637). Jan gebruikte nu $AG : BG = AD : BC$. Pibo klaagde dat Jan hier geen bewijs voor gaf, maar de verhouding is correct: in de concyclische vierhoek staan gelijke hoeken op gelijke bogen, en daarom zijn de driehoeken AGD en BGC gelijkvormig.

Pibo gaf de rest van de redenering niet weer maar we kunnen nu als volgt verder gaan in moderne notatie. We noteren de gegeven zijden $AB = a$, $BC = b$, $CD = c$, $DA = d$. Uit $AG = r$ en $AG : BG = AD : BC = d : b$ volgt $BG = \frac{b}{d}r$. Met precies hetzelfde argument vinden we $BG : CG = BA : CD$ en daardoor $CG = \frac{bc}{ad}r$, dus $AC = \frac{ad+bc}{ad}r$. Ten slotte $DG : AG = DC : AB = c : a$, dus $DG = \frac{c}{a}r$ en $BD = \frac{ab+cd}{ad}r$. Hieruit $AC/BD = \frac{ad+bc}{ab+cd}$.

Pibo geeft ook de informatie dat Jan uit het rekenboek van Nicolaus Petri van Deventer (gestorven 1602)[11] de stelling van Ptolemaeus voor concyclische vierhoeken gebruikte: $AC \times BD = ac + bd$. We vermenigvuldigen de twee formules en krijgen $AC^2 = \frac{(ac+bd)(ad+bc)}{ab+cd}$, dus is AC bekend. We kunnen nu verder werken in driehoek ACD. Het probleem voor driehoek ACD en de middellijn door C van de omgeschreven cirkel was al opgelost in het hierboven genoemde rekenboek van Nicolaus Petri van Deventer,[12] weliswaar met andere getalswaarden maar met een algemene methode.

We weten niet of Jan op deze manier heeft geredeneerd. In elk geval was Pibo niet overtuigd, en hij vond ook dat men niet zomaar de Stelling van Ptolemaeus mag gebruiken zonder bewijs. Pibo merkte verder op dat Nicolaus Petri de diagonaal van een concyclische vierhoek berekent in een getallenvoorbeeld waarin twee overstaande zijden loodrecht op elkaar staan, en hiervan wordt in de berekening gebruik maakt.[13] Pibo beschuldigde Jan ervan, deze foute aanname gebruikt te hebben. Mulerius en de andere aanwezige wiskundigen waren blijkbaar niet onder de indruk. De discussie keerde zich tegen Pibo en een van de aanwezigen antwoordde op zijn argumenten met: *Iae. is't anders niet?* Ten slotte wilde Pibo Jan een nieuw wiskundeprobleem opgeven dat te maken had met een muziekinstrument en een boek over muziek dat Pibo in zijn verzameling had. Jan en ook Mulerius weigerden dit te accepteren.

Na de mislukte bijeenkomst in huize Mulerius had Pibo volgens eigen zeggen Jan nog een keer uitgenodigd om tegen vergoeding van (reis)kosten de oplossing te komen uitleggen. Jan kwam niet opdagen, maar ging wel door met tegen anderen te zeggen dat hij niets te vrezen had van Pibo. Daarop besloot Pibo dat hij gedwongen was om aan "alle ware Arithmetische ende Geometrische onpartydighe Liefhebberen", en in het bijzonder aan zijn eigen leraar Adriaan Metius, uit te leggen hoe de vork in de steel zat.

Pibo eindigt zijn boekje met de volgende redenering om aan te tonen dat de getallen in Jans oplossing niet correct kunnen zijn. Figuur 3 toont weer de concyclische vierhoek met I het middelpunt van de omgeschreven cirkel, en loodlijnen CH en IK op AD. Lijn $LINM$ evenwijdig aan AD snijdt CH in N. Segmenten AC, AE en ED zijn niet getekend maar worden wel uitgerekend.

Pibo berekende eerst

$$AC = \sqrt{4195\frac{2}{13}},\quad CN = \sqrt{\frac{698659055667}{607409660}},\quad CI = \sqrt{\frac{4634903}{3740}},$$
$$CH = \sqrt{2560\frac{4085}{162409}},\quad ED = \sqrt{1932\frac{108}{935}}.$$

Deze lengtes zijn correct.

Omdat $CN : CI = CH : CF$ en CN, CI, CH vierkantswortels zijn, volgt $CF = \sqrt{}$ (Pibo gebruikt een liggend streepje onder een wortelteken om een rationaal getal aan te geven dat hij niet uitrekent.) Omdat $AC : CF = ED : DF$ geldt ook $DF = \sqrt{}$, en op soortgelijke manier leidt Pibo af $AF = \sqrt{}$, $EF = \sqrt{}$.

Pibo concludeert dat niet kan gelden wat Jan zegt in zijn oplossingen, namelijk

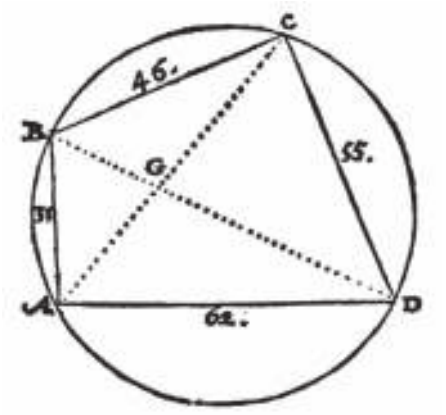
$$AF = 31 - \sqrt{},\quad DF = 31 + \sqrt{},\quad CF = \sqrt{} + \sqrt{},\quad EF = \sqrt{} - \sqrt{}.$$

Tot zover Pibo.

We zullen nu Jans oplossing gaan analyseren met moderne digitale methoden, die geen wiskundige voorkennis vereisen. Het is voldoende om de breuken zonder typefouten in te voeren in het vakje op het beginscherm van wolframalpha.com, en op de enter-toets te drukken: het programma vereenvoudigt dan direct. Bij de lengtes van AF en DF zien we zo:

$$\frac{125780114634036565515210964623563605425996327 4690612500}{583811685629371078977354359853646820931640264756 12500} = \frac{5017452161089}{232886351889}.$$

Als we het woord `factorize` intypen en daarachter de gemeenschappelijke factor van teller en noemer 250685229466616267700091938835792693012500 vinden we $2^2 \times 3^3 \times 5^5 \times 7^4 \times 11^5 \times 13^{10} \times 17 \times 31^{10}$, dus een product van kleine priemfactoren.



Figuur 2

Door de teller en de noemer van de vereenvoudigde breuk ook te factoriseren vinden we dat beide kwadraten zijn:

$$\sqrt{5017452161089} = 2239967$$

en

$$\sqrt{232886351889} = 482583.$$

En zo vinden we dat

$$AF = 31 - \tfrac{2239967}{482583} \quad \text{en} \quad FD = 31 + \tfrac{2239967}{482583},$$

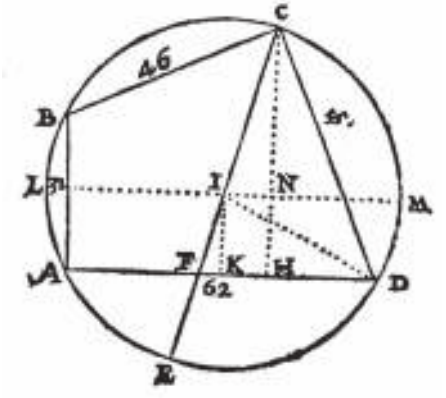
beide rationale getallen.

Op dezelfde manier kunnen de lengtes CF en FE in Jans oplossing vereenvoudigd worden tot

$$\sqrt{\frac{4634903}{3740}} \pm \sqrt{\frac{261144769333524167}{870994956064860}}.$$

De gemeenschappelijke factoren in de oorspronkelijke breuken onder het wortelteken zijn $3 \times 5^2 \times 7^2 \times 11^2 \times 13^4 \times 31^4$ en $2^2 \times 3^4 \times 5^7 \times 7^6 \times 11^7 \times 13^{14} \times 17 \times 31^{14}$, alweer producten van kleine priemgetallen.

Met behulp van moderne notatie (ingevoerd door Descartes in 1637) is nu gemakkelijk te controleren dat Jans oplossing correct is. We noteren $p = AC$, $x = AF$, $y = FD$, $R = \frac{1}{2}CE$ de straal van de omgeschreven



Figuur 3

cirkel, en K de oppervlakte van de concyclische vierhoek. Dan is

$$p^2 = \frac{(ac+bd)(ad+bc)}{ab+cd}$$

zoals we hierboven hebben gezien,

$$K^2 = (s-a)(s-b)(s-c)(s-d)$$

met

$$s = \tfrac{1}{2}(a+b+c+d)$$

(Brahmagupta's formule) en

$$R^2 = \frac{(ab+cd)(ac+bd)(ad+bc)}{16K^2}.$$

Door geschikte gelijkvormige driehoeken te bekijken vinden we

$$\frac{x}{y} = \frac{p^2}{c^2} \cdot \frac{c^2+d^2-p^2}{c^2+p^2-d^2.}$$

Hieruit blijkt dat als a, b, c en d rationale getallen zijn, de lengtes AF en FD ook rationaal zijn.

Als $a = 31$, $b = 46$, $c = 55$, $d = 62$, zoals in Pibo's probleem, dan vinden we

$$s = 97, \quad K^2 = 4948020, \quad R^2 = \tfrac{4634903}{3740},$$

$$p^2 = \tfrac{54537}{13}, \quad \frac{x}{y} = \tfrac{205163}{277420}, \quad x+y = 62,$$

$$x = AF = 31 - \tfrac{2239967}{482583}$$

en

$$y = FD = 31 + \tfrac{2239967}{482583}.$$

Stel $CF = R+z$, voor zekere nog uit te rekenen z, dan is $FE = R-z$. Er geldt dan

$$R^2 - z^2 = CF \cdot FE = AF \cdot FD$$
$$= 31^2 - (\tfrac{2239967}{482583})^2$$

en hieruit

$$z^2 = \frac{261144769333524167}{870994956064860}.$$

We concluderen dat Jans oplossing correct is. Dit kan onmogelijk het resultaat zijn van toeval. Het betekent dat Jan een correcte methode bezat voor het berekenen van de diagonalen van de concyclische vierhoek, vermoedelijk in de trant van wat wij hierboven uit zijn 'sotte antwoorden' hebben gereconstrueerd. Jan heeft Pibo op het verkeerde been gezet door de breuk in AF en FD te kwadrateren en er vervolgens de wortel uit te trekken. Daarna heeft hij de tellers en noemers van alle breuken herhaaldelijk met kleine priemfactoren vermenigvuldigd. Pibo heeft de enorme getallen correct weergegeven, en de drukker heeft ze zonder drukfouten gezet. Jan had niet de methode van Nicolaus Petri gebruikt voor de berekening van de diagonaal van een concyclische vierhoek, maar Pibo kon dit niet weten doordat hij Jans oplossing niet had kunnen controleren.

De competitie tussen Pibo en Jan heeft te maken met verschillende stijlen van wiskundebeoefening in het begin van de zeventiende eeuw. Enerzijds waren er academisch geschoolde wiskundigen die zich sterk op de klassieke oudheid richtten, en wiskunde wilden baseren op de *Elementen* van Euclides. Pibo behoorde tot deze groep.

Anderzijds waren er rekenmeesters die algebra gebruikten, en ook meetkundige figuren en redeneringen, maar geen bewijzen in de stijl van Euclides. Tot deze groep behoorden Nicolaus Petri en ook Jan Beerntsen. Deze wiskundigen waren niet academisch geschoold en kenden meestal geen Latijn.

We kunnen het verschil uitleggen aan de hand van een concyclische vierhoek. Het is intuïtief duidelijk dat er voor elk viertal gegeven lijnsegmenten, zodat het grootste lijnsegment kleiner is dan de som van de drie andere, een concyclische vierhoek bestaat met vier zijden gelijk aan die vier lijnsegmenten. Voor wiskundigen uit de tweede groep was dit genoeg. Wiskundigen uit de eerste groep zouden een euclidische constructie met passer en liniaal willen zien, zoals die gegeven werd door F. Vieta in 1595.[14] Pibo had gevraagd waarom Jan er zeker van was dat de concyclische vierhoek met zijden met lengtes 31, 46, 55 en 62 in de probleemstelling wel bestond, waarop Jan alleen antwoordde *Daeromme*.

De oplossing van Jan is een uitdaging door een niet academisch gevormde rekenmeester aan een academische volgeling van Euclides. De getallen in de oplossing waren te groot om door Pibo en zijn academische collega's te kunnen worden gefactoriseerd, en daarom kon met de standaard methoden van de *Elementen* van Euclides niet worden vastgesteld of de oplossing correct was. Hierdoor werd het twijfelachtig of de methoden van Euclides wel de basis kunnen zijn van alle wiskunde.

Met moderne digitale methoden heeft het boekje van Pibo nu een onbedoeld effect gehad dat Pibo nooit had kunnen voorzien. Vier eeuwen later is de oplossing van Jan Beerntsen in ere hersteld.

Noten en referenties

1 Klaas Hoogendoorn, *Bibliography of the Exact Sciences in the Low Countries from ca. 1470 to the Golden Age (ca. 1700)*, Leiden, 2018, voor Pibo Gualtheri zie p. 446.

2 Arjen Dijkstra, *Between Academics and Idiots: A Cultural History of Mathematics in the Dutch Province of Friesland (1600–1700)*, Leeuwarden, 2012, pp. 49–51, 101–106.

3 Andere schrijfwijzen van zijn voornaam zijn Pybe, Pijbe.

4 Zie Arjen Dijkstra, Goffe Jensma, Djoeke van Netten en Piter van Tuinen, *Wiskunde als Familiebedrijf: Menelaus Winsemius' lijkrede op Adriaan Metius (1571–1635)*, Groningen, 2012.

5 De lijst is bewaard in het Historisch Centrum Leeuwarden (inv. nr. 3370), online beschikbaar via www.allefriezen.nl, door te zoeken met Inventarisatieboek 1618–1620. De papegaai staat in het origineel op pagina 41. Pibo's boekenlijst is getranscribeerd door M. H. H. Engels https://www.mpaginae.nl/At/PiboG.htm (geraadpleegd 7 februari 2023).

6 Pibo maakte zelf ook instrumenten: een prachtig astrolabium uit 1606 wordt in Leeuwarden bewaard.

7 Andere spellingen van zijn naam zijn Beernts, Barents.

8 Behalve de wiskunde in Pibo's boekje is één meetkundig probleem van Jan Beerntsen bekend: no. 25 uit de lijst van meetkundige problemen aan het eind van Theodorus Hoen, *Natuerlycke Astrologie*, p. 216, 'van Jan Barents Lantmeeter tot Harlingen'. In hetzelfde boek staat ook een probleem van Pibo, zie p. 206.

9 Over Mulerius zie Djoeke van Netten, *Nicolaus Mulerius (1564–1630), Een geleerde uit Groningen in de discussies van zijn tijd*, Groningen, 2010.

10 Zie voor deze tirades ook de getypte versie van het boekje, beschikbaar op http://www.jphogendijk.nl/sources/pibo.html.

11 Claes Pietersz van Deventer, *Practique om te Leeren Rekenen cijpheren ende Boeckhouwen / met die Regel Coss ende Geometrie seer profijtelijcken voor allen Coopluyden. Van nieus Ghecorrigeert ende vermeerdert*, vele drukken, bijvoorbeeld Haarlem, 1596. De stelling van Ptolemaus wordt in deze druk genoemd in probleem no. 94, fol. 227b.

12 Probleem no. 85, fol. 222a.

13 De zijden AB en CD staan loodrecht op elkaar wanneer $(ab+cd)^2 + (ad+bc)^2 = (d^2-b^2)^2$. Dit is het geval in het voorbeeld van Nicolaus Petri, $a = 6\frac{2}{5}$, $b = 10$, $c = 13\frac{1}{5}$, $d = 24$. In het voorbeeld van Pibo is er geen loodrechte stand van tegenoverstaande zijden.

14 F. Vieta, *Pseudo-Mesolabum et alia quaedam adiuncta capitula*, Parijs, 1595, herdruk in F. van Schooten, ed., Francisci Vietae Opera Mathematica, Leiden, 1646, pp. 275–283, reprint Olms, Hildesheim–New York, 1970. Ludolph van Ceulen gaat hierop verder in het begin van Boek 5 van zijn *Arithmetische en Geometrische Fondamenten, Leiden 1615*.

Olga Lukina
Mathematical Institute
Leiden University
o.lukina@math.leidenuniv.nl

Column New Impulse

Back to the Netherlands after time abroad

In this column, researchers who have been recently appointed to one of the Dutch mathematics institutes, introduce themselves.

My name is Olga Lukina. In July 2022 I started as a tenure-track Assistant Professor at the Mathematical Institute of Leiden University.

For me, coming to the Netherlands was returning back to the place where I started my scientific career (at a postgraduate level), as I previously spent four years here, doing my PhD at the University of Groningen. While abroad, working in the UK, USA and Austria, rather than losing the connection with the Netherlands, I met new researchers from the Dutch mathematical community. So it was very natural for me to return to the country I already knew, and join the mathematical community to which, I felt, I already belonged.

I grew up in Russia, in the Perm region in the Ural mountains. Since I can remember myself, I was curious about exact sciences, languages and other countries. I remember being fascinated around the age of 8 by a book with physics experiments which a child could do at home. I learned my first words of English around that age from my mother. My favorite book around that time was about a group of sailors traveling on a ship around the world, visiting different countries. While in Japan, they were invited to visit a Japanese family who cooked for them sukiyaki, the marble beef hotpot. I tried sukiyaki when visiting Ritsumeikan University near Kyoto in 2019, and it was as delicious as the book said!

At the time when I was growing up, Perm State Technical University, which I later attended, was running programs for talented children, helping them learn mathematics and physics outside of the school curriculum. I participated in those programs. A few times my classmates and I went to the Math Camp, organized by the university. One of the activities we had in the Math Camps was Math Fights: two teams were given math olympiad type problems, which they had to solve. Then one team would present a solution to the other one (the opponent), whose task was to criticize the solution; then they would switch the roles. I also just read science books on my own. My father had an engineering degree in physics, and my mother in chemistry, and they still had a few science books at home. I remember learning logic from my father's book on electronics, and trying to learn probability from *The Feynman Lectures on Physics*.

After getting my Master's degree in Mathematical Modeling at Perm State University, I worked for a few years in IT. I enjoyed that at the beginning; but after an initial period of learning, a job at

a company settles into a routine, and I was missing the joy and challenge of learning new mathematics. So I decided to pursue a PhD, and see a little bit of the world at the same time. That was the reason I looked for a PhD position abroad, and that was how I came to the Netherlands, to the University of Groningen.

After finishing my PhD in integrable Hamiltonian systems, I worked at the University of Leicester in the UK, University of Illinois at Chicago in the USA, and University of Vienna in Austria. My research interests gradually shifted to minimal sets of foliations, and from there to the study of group actions on Cantor sets, and their applications. I still work in dynamical systems, but with a very different type of objects. I am never bored, since I get to learn and create new mathematics every day. I was lucky to experience and learn the culture of four very interesting countries. I now can speak four foreign languages (but only one well).

Educational systems in different countries are of course influenced by the way their society is organized, and the problems universities in different countries face and solve are different. Most recently, I taught in the USA, at a public university, University of Illinois at Chicago (UIC). In the USA (and especially in very large cities like Chicago), there is a huge divide between the opportunities in early education that the rich and the poor have, and public universities play the role of social lifts, sometimes giving a second chance to those who did not have access to a good high school. At UIC, we payed a lot of attention to helping students to stay on track and finish the course, especially in the large first-second year service courses. Professors and teaching assistants were required to hold two or three office hours per week, where students could come for individual help with the course material. There was also a walk-in Math and Computer Science Learning Center open five days per week from 8 am to 6 pm, where students could get help with their homework. I found that in general students had very good attitude towards learning, they studied hard and wanted to succeed. A rewarding part of teaching at UIC was to feel that I am a part of a social change, helping people to change their life to the better.

The problems I described above are not applicable to the Netherlands, where the level of preparation of students when they enter the university is more uniform. Students here are considered to be independent, and it is their responsibility to keep up with the workload. So far, I found students very well-prepared and motivated to learn. Questions they ask are about further applications of theories they learn, rather than about solutions to the homework.

Comparing the research environment in the countries where I worked, there seem to be a lot more joint activities happening in the Netherlands at the national level than in other countries. For instance, there are monthly national seminars like the Mark Kac seminar on Stochastics and Physics, and regular joint events with participation of research groups from a few universities. This is possibly due to different cities in the Netherlands being very close and easily accessible by train. For instance, in Chicago it took me forty minutes to get from my apartment to the campus of the university. It takes me approximately the same time to get from Leiden to Utrecht. Overall, I feel that the Netherlands has a very vibrant research atmosphere, with a lot of interactions between groups of researchers and universities. I look forward to being a part of this environment!

The country has changed during the years I was away. It has moved firmly into a digital age; it seems to be impossible to get by without a smartphone. On the other hand, the variety and quality of food available when one wants to eat out has improved dramatically; although it seems to be impossible to get a table on a Friday evening without a reservation. When a PhD student in Groningen, I enjoyed shopping at the Vismarkt on Saturdays. Now I often buy fish, vegetables and flowers at the Saturday market on Nieuwe Rijn in Leiden; I am glad this Dutch tradition hasn't changed!

Raf Bocklandt
Korteweg-de Vries Institute for Mathematics
University of Amsterdam
r.r.j.bocklandt@uva.nl

Interview Luc Illusie

Tales from the golden age of algebraic geometry

Last December the University of Utrecht organized a new edition of the Kan memorial lectures. Founded in 2015 by Ieke Moerdijk in honour of the famous Dutch topologist Daniel Kan, these lecture series are intended to highlight various aspects of algebraic topology. In each edition an acclaimed expert is invited to give one public lecture and two more specialized lectures about a topic of his choice. This year the French algebraic geometer Luc Illusie gave three lectures on the de Rham complex. Nieuw Archief seized the opportunity to talk with him about his life, mathematics and the golden age of algebraic geometry in the sixties.

From Nantes to Paris

Luc Illusie was born on 2 May 1940, just days before the start of World War II. By the middle of June his hometown of Savenay, a village in the West of France near the estuary of the Loire, had been conquered by the Germans.

"I was born in Nantes, but I spent my first five years in Savenay, a small village 30 kilometers north of Nantes. We were occupied by the Germans at the time, but by January 1945 some parts of France had been liberated. Together with my parents and my brother we escaped German controlled Savenay and we moved to Nantes. There I attended primary school and secondary school up to 1957.

Both my parents were teachers. My father taught history and literature, a combination that was still possible at the time. I owe him a lot because he knew so many things. He had an extraordinary memory. He could listen to a speech and then recite it afterwards. He knew so many pieces of the great poets.

My mother was teaching mathematics. She helped me discover the joys of algebra. When I was nine years old, she taught me how to put concrete problems into equations. She was my mathematics teacher in school when I was between 10 and 12 years old. She taught me geometry and I was fascinated by the beautiful properties of triangles. I think that today many boys and girls find mathematics a little boring in secondary school, not just because the foundations are messy. The point is that they are not presented with very beautiful results. They see Thales' and Pythagoras' theorems but these are not so exciting. However when you look at triangles, they have a lot more fascinating properties, such as the Euler line or the Euler circle or many other mysterious configurations. You start with something very irregular and you find a certain harmony hidden in plain sight."

Photo: www.imo.universite-paris-saclay.fr

Luc Illusie

Although he was already fascinated by mathematics at an early age, Illusie had diverse interests and a career in mathematics was at that time not on his mind.

"I thought mathematics was fun, it was really a wonderful thing, it was challenging but it did not take me too much time. What took the most of my time was French, Latin and Greek, and the humanities. My teachers were fantastic and I was very involved in those subjects. I remember when I was 15, I could spend hours at night writing some French texts on literature, such as poems by Victor Hugo and other people. I enjoyed that very much, so in view of that I should have continued in literature or humanities.

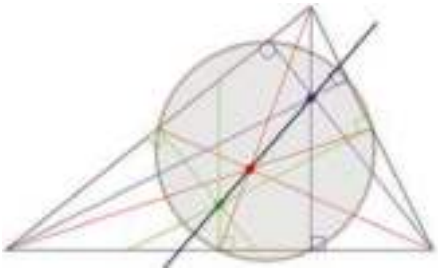


Illustration: eulercircle.com

The Euler line and Euler circle of a triangle

At that time, when you were 16 or 17 you had to pass the Baccalauréat. This was a serious examination in two parts spread over two years and in the second part you had to choose between literature or science. I chose science because I thought there were more opportunities for jobs in mathematics and physics. The idea was that I would compete for the École Normale or École Polytechnique afterwards. Therefore my parents decided it was better for me to go to a good school in Paris. So we all moved to Paris and I enrolled in the Lycée Louis-le-Grand. Getting into that lycée was not so easy, but I had excellent marks in Latin and Greek. Today I have forgotten all my Latin and all my Greek.

At the lycée I had an extraordinary teacher of mathematics in the second year. His name was André Magnier and in fact he knew Grothendieck very well. He was from Montpellier, where Grothendieck had studied, and he helped him come to Paris. He discovered that he got a special talent and then had him contact Cartan and the Bourbaki group. The rest is history… Magnier was a very good teacher and it is certainly thanks to him that I also got drawn into mathematics."

Cartan and the caimans

Despite his teacher's connections, Illusie's personal contacts with Cartan, Grothendieck and the rest of the mathematics scene in Paris would only start later, during his student years at the École Normale.

"At the École Normale there were students called caimans (crocodiles) who would take care of the younger students. My first caiman was Adrien Douady. He was a wonderful character, he knew a lot of things, especially a lot of examples and counterexamples. He had a way of explaining very abstract things in a very concrete way. It was a marvelous person and I discovered afterwards that he was at Bourbaki. From time to time he would give me some secret papers of Bourbaki. The Boubaki group was not so secret, actually these secret papers were circulating quite a lot. Roger Godement, who was teaching me analysis at the time, was also giving me some Bourbaki papers about commutative algebra.

The École Normale offers a broad degree, which not only contains mathematics but also physics and chemistry. In the first year I was still hesitating between physics and mathematics. My physics teachers were quite interesting. One was Alfred Kastler, the Nobel Prize winner, and the other Yves Rocard, the father of the former prime minister Michel Rocard. Both were very brilliant, Kastler taught physics a little bit like mathematics. Rocard on the other hand was not so neat. I said to myself: 'This is not the type of thing I want to do.'

Then I discovered the teachings of Cartan and this was a revelation for me. In high school you learn to do many calculations, but the theory and the conceptual aspect of mathematics is lacking. In the first course by Cartan, he introduced abstract concepts like the characteristic of a field. Cartan was marvelous. Sometimes he had not prepared so much and he would get stuck, but I really liked his approach to the matter."

The fifties and sixties were the haydays of the French mathematical seminars. They were joint efforts of students and researchers who came together under the lead of one of the university professors to study a contemporary topic in mathematics. The most influential one was the seminar of Henri Cartan and it was there that Illusie made his first marks as a mathematical researcher.

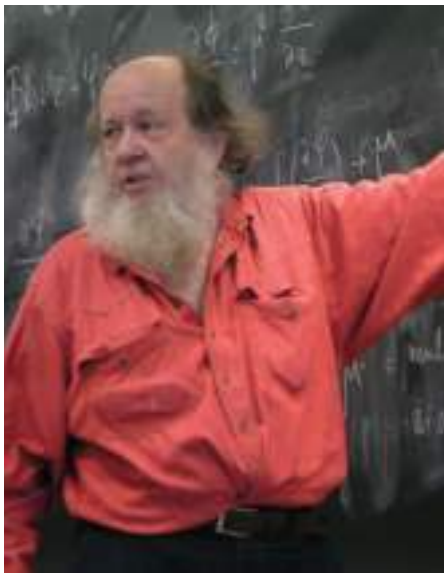


Photo: Wikipedia.org / François Tisseyre

Adrien Douady

"At the end of the fourth year, Cartan had arranged for me to have a temporary position at CNRS and he enrolled me in his seminar on the Atiyah–Singer index formula [1, 9]. It was here that I first felt that maybe I could become a mathematician. Sometimes I made a new observation of which Atiyah, Singer and other very impressive people had not thought. Although it was maybe an epsilon, it gave me great confidence that I could continue.

The idea of the Cartan seminar was a continuation of Hadamard's seminar. We took on a big subject, a new theory, and we started all from the beginning. Black boxes were not allowed, everything had to be proven. For the Atiyah–Singer formula we had to discuss characteristic classes and how they are defined. I remember that the first talk Cartan asked me to give was on the Chern character and the Todd class. I told him: 'Well Mr. Cartan, I do not know any of this. I cannot do it.' He said: 'It is not that difficult. Just come to my office and I will explain it to you.' For one hour he explained to me classifying spaces, their cohomology and characteristic classes. It was so clear and simple.

Eventually I could give my talk, but then there was a constraint. You had to write it up in one month. My handwriting was not so good and Cartan said that I should get a typewriter and type it. I bought a German typewriter, an enormous machine, and I learned how to type with ten fingers. I typed my piece on the Chern character and then I handed it in. To my surprise, he said it was okay apart from a few technical remarks here and there.

At the seminar, there was one student who was very impressive. After my talk he came to me and we discussed passionately for maybe half an hour. At the end I asked 'who is this young guy' and it turned out to be Jean-Louis Verdier. There was also Demazure, who was of course two three years in advance and knew much more than I did. I talked with all these people and I learned how to do mathematics and how to write it down."

Meeting the mentor

The seminar brought Illusie into contact with many of the great mathematicians of

Photo: mathshistory.st-andrews.ac.uk

Henri Cartan

his time, but the meeting that influenced his career the most was the one with Alexandre Grothendieck, the father of modern algebraic geometry.

"The big turning point came when I met Grothendieck. Under Cartan I had started working on a relative version of the Atiyah–Singer index formula. I had done some work on Hilbert bundles in the wake of what Atiyah had done. Unfortunately, I got stuck and Cartan said: 'Perhaps you should ask Grothendieck. Maybe he has some idea about that.' I contacted Grothendieck and he suggested that I use sheaves because they are very powerful and allow to treat singularities. This was somehow revolutionary: sheaves had been used for ten years in complex analytic geometry but not in the C-infinity context. Although his suggestion worked I did not continue with it because by then Atiyah and Singer had already solved the relative formula by other means.

Instead I kept working with Grothendieck on other projects. Cartan was really generous because he did not object to my going to Grothendieck and working with him. He could have said: 'I am losing a student and this is not good', but he was not like that at all."

Parallel to Cartan, Grothendieck ran his own seminar: the Séminaire de Géométrie Algébrique du Bois Marie (SGA). The seminar took place at the Institut des Hautes Études Scientifiques (IHÉS), a newly founded mathematical research institute south of Paris in Bures sur Yvette. The institute was situated in a forested domain, hence the name of the seminar. For ten years Grothendieck and his collaborators rewrote the foundations of algebraic geometry and incorporated many of the recent developments in the fast evolving subject.

In the fall of 1964 Grothendieck started a seminar on ℓ-adic cohomology and asked me to write down notes for some exposés [8]. I objected 'This is impossible! I know almost nothing in algebraic geometry, even the concept of an affine scheme is not so clear to me', but he took my 'no' for an agreement and replied: 'You will learn quickly!'

In the seminar he took the pain of writing all the definitions and explaining how they worked. It was crystal clear and gradually I learned the necessary basic material. For example, at the time I did not know anything about Zariski's main theorem. At the same time I was reading the preprints that came from for SGA4 [11], the basic theorems in étale cohomology and the general formalism of sites and topoi, which was abstract and quite hard. But Grothendieck had a way of making that somehow easy.

Grothendieck was a new spirit, a new flame, because of the functorial language. This functorial language is a whole new world. Grothendieck topologies are amazing, you see, so I was fascinated by that. The other aspect was derived categories. I had studied Cartan and Eilenberg [4] at the École Normale, but I found this new approach so wonderfully simple. For example the spectral sequence of the composition of two functors is complicated but in derived categories it just becomes $R(G \circ F) = RG \circ RF$.

Writing down the notes turned out to be not difficult at all, because his lectures were so clear. I would write down a first draft and present it to him. Then he would invite me to his place and we would go line by line through it. He had many comments on details or spelling and general comments on presentations and proofs. It usually took the whole afternoon and evening.

Grothendieck would not accept a pack of notes of less than 15 pages. Once one of my former caimans at the École Normale handed him maybe 10 or 15 pages and he said: 'When you have elaborated on it, come back again.' It had to be at least 40 or 50 pages.

That was how I learned how to work and how to write. Even though Cartan and Douady had already been very strict and taught me how to write in the Bourbaki style, Grothendieck was at another level. I think I can write in French reasonably well, but he was also making comments on that and although French was not his mother tongue he was always right."

While the first interactions between Illusie and Grothendieck concentrated on the redactions of the exposés of SGA5 and SGA6, Illusie also had to find a good topic for a PhD thesis. This turned out to be more tricky than expected.

"You could say that Grothendieck was quite unsuccessful with me. I was a bad student. He asked me several problems

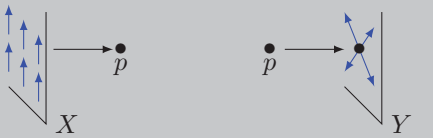
Photo: www.thetimes.co.uk

Grothendieck lecturing at IHES

What is the cotangent complex?

In his thesis, which was later published in *Springer Lecture Notes in Mathematics* [7], Illusie constructed a broad generalization of the cotangent bundle. For each morphism $X \to Y$ of ringed topoi he defined a complex $L_{X/Y}$ that captures the deformation theory of that morphism. More precisely $L_{X/Y}$ is a complex of sheaves over X such that $\mathrm{Hom}(L_{X/Y}, \mathcal{O}_X)$ describes the infintesimal automorphisms of $X \to Y$, $\mathrm{Ext}^1(L_{X/Y}, \mathcal{O}_X)$ the first order deformations, and the higher exts the obstructions.

Let us look at some examples to get a feeling of what is going on. There are two extreme cases: the map of a smooth variety to a point, $X \to p$, and the embedding of a point in a smooth variety, $p \to Y$.



$L_{X/p} = \Omega^1_X$ $\qquad$ $L_{p/Y} = T^*_p Y[1]$

In the former case, the infinitesimal automorphisms are given by vector fields on X and there are no deformations. This shows that (as an object in the derived category of X) $L_{X/p}$ is isomorphic to the cotangent bundle of X concentrated in degree 0. In the latter case, the morphism $p \to Y$ has no infinitesimal automorphisms but its infinitesimal deformations are unobstructed and given by tangent vectors at p, so $L_{p/Y}$ is the cotangent space to Y at p concentrated in degree -1.

In a similar way $L_{X/Y}$ will be the conormal sheaf concentrated in degree -1 for an embedding of smooth varieties, and the sheaf of relative differentials concentrated in degree 0 for a smooth morphism (e.g a flat family of smooth varieties).

Things become trickier if X, Y are singular schemes or even more complicated objects such as algebraic sets, stacks or ringed topoi.

which were all interesting, but I could not get anywhere. For example, he knew that I loved derived categories so he asked me to develop a formalism that could do away with the restrictions on the degrees. Think of the choice between D^-, D^+ or D^b. At the time pro- and ind-objects were very popular, so I tried something along these lines but it did not work. In fact we had to wait for Spaltenstein in the 1980s to get a real new approach.

Another very important problem concerned Kunneth's formula. I loved this formula. In its treatment in EGA3 there are maybe a dozen spectral sequences. This is ridiculous, so he asked me to clean that and write it down neatly in the form of derived categories. But again I was stuck. In Kunneth's formula the ring structure is important, but there is no good ring structure in the derived setting. A few years later it was solved by Quillen with homotopical algebras.

A third problem concerning the notion of properness in algebraic geometry also led to nowhere and only very late in 1968 he proposed to me the questions on deformation theory that would lead to my work on the cotangent complex. But then in a few months I had essentially all the basics theorems of the first part of my thesis."

Carpets

The circle around Grothendieck included mathematicians from all over the world and Illusie has fond memories of interacting with many of them. Two of them stand out: Pierre Deligne, who was his mathematical brother and Daniel Quillen, whose work formed the basis for his PhD thesis and who invited Illusie to visit MIT in 1970.

"Starting in 1965, there was a new timid looking young man attending the seminar: Pierre Deligne. From then onwards, anytime I was stuck in some place I did not understand I asked him to explain it to me. He knew everything already and it was just marvelous. He would come to my place, we would discuss mathematics the whole afternoon and then we would move to Parc Montsouris for a walk and further discussions. I had a carpet in my place in Paris and he liked to lie down on the carpet and do the pear tree. You put your head on the carpet and your legs in the air to make the blood run to your brain.

Quillen was also a wonderful person. In 1968 when I discovered a very simple way to extend his construction to topoi, I wrote him a letter. He replied immediately and invited me to visit MIT, but before I went there I already had the opportunity to meet him in person at IHES. Grothendieck was very impressed actually by Quillen. He wrote some notes entitled 'Tapis de Quillen' (Quillen's carpet). Marchand de Tapis was the Bourbaki term for someone who had a new theory and who wanted to sell it to everybody.

Anyway, I went to MIT and Quillen invited me to give a course on my thesis and we discussed it a lot. It was fabulous, he was extremely clear and precise. He knew a lot of algebraic topology, a lot of number theory, finite groups and algebraic geometry. All that was coming together in his mind wonderfully. I arrived in September 1970 but I could not stay as long as I wanted because my mother had had a stroke. I had to come back in the spring of 1971 and then I defended my thesis at that time. My visit to MIT was really quite a personal discovery."

Writing and redacting

In 1976 Illusie became a professor at the Université Paris-Sud in Orsay and remained there until his retirement in 2005.

"Around 1976 people at Orsay invited me to their university to become a professor there. At first I was a bit reluctant because at CNRS I had quite a lot of freedom to do research but eventually Raynaud and other colleagues convinced me to accept the position that they had for me. I had to do quite some teaching, sometimes to groups of 200 students. These were the beginning students who had not yet decided between math, physics or engineering. Some were very motivated and some were less motivated. I typed notes for them on a weekly basis and then at one point, in 1981 or 1982, I had a small group of very interested students and I said next time you will write the notes and I will teach you how to write properly in Bourbaki style. We would meet at the end of the day and go through the annotations I made, just like Grothendieck did when I was a student. After the corrections their notes were dis-

Photo: mathshistory.st-andrews.ac.uk

Pierre Deligne

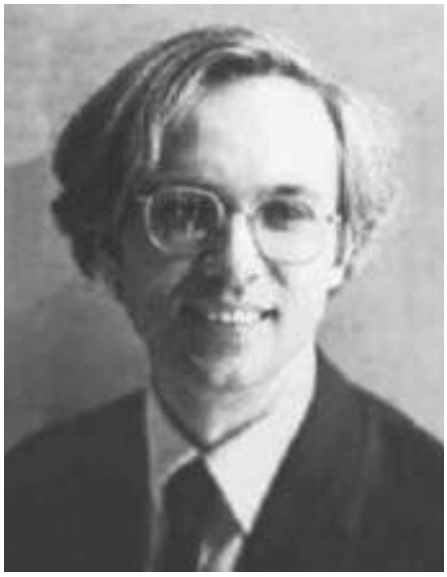


Photo: mathshistory.st-andrews.ac.uk

Daniel Quillen

tributed to the students with their names on and they were very proud. That was a very nice experience."

The Bourbaki style of writing mathematics has a mixed reputation. Some praise it for its precise statements and clear expositions, while others complain about the lack of motivation and examples. Illusie acknowledges this but adds that many good texts combine elements of Bourbaki with a more concrete approach.

"Well, let me give an example. Mumford and Atiyah's way of writing is based on Bourbaki but also very much illuminated with examples and concrete applications. So it looks concrete though at the bottom it is still Bourbaki. So it is a question of style, where you put the emphasis. If you look at the text by Giraud for example, *Méthode de la Descente*, it is very abstract and very terse. That is not so nice, on the substance it is certainly a great text, but it lacks that balance. Even Grothendieck was able to see the difference if you take something very abstract like a topos. I remember that Verdier had written Exposé IV in SGA4 on topoi and there were categories and functors everywhere. Grothendieck said: 'I cannot make sense of all that, you need to write in a more geometric language.' Starting from Verdier's initial text, Grothendieck wrote a completely new version with many examples. From this very abstract substance he made something which was still Bourbaki style but also motivated and concrete. I use the term motivated because I think this is what is often lacking in Bourbaki. Never complain, never explain; the reader will understand by himself why we are doing this. But Grothendieck was very careful to explain why he was doing things.

Today people like Bhargav Bhatt and Peter Scholze write in some kind of Bourbaki style illuminated with examples. Their theories are very abstract but very concrete at the same time with key examples and striking observations."

The role of examples in theory building

Illusie also thinks that Grothendieck's reputation of eschewing examples is misplaced.

"It is not correct to say that Grothendieck did not know any examples. He knew simple examples, maybe he did not know much about $\mathrm{E}8$ or other special things, but he knew basic examples. For example he had a lot of knowledge about abelian varieties and algebraic groups and he was an expert in functional analysis and topology. This helped him a lot and he had a broad view of topology and differential geometry.

He started with simple examples. I think already with curves, abelian varieties and hypersurfaces you have enough. When you want to develop a big theory, the idea is to look for extreme examples where some abstract principle still works and you want to join those extremes together. Often it then turns out that the existing theory is not stable and you need to modify it. You can try to find a new definition which is much more amenable.

For example, in 1966 Grothendieck dreamed of a theory of the determinant of a perfect complex. He proposed this problem to Daniel Ferrand. When you have a short exact sequence, the determinant in the middle should be the product determinant of the two ends, so when you have a triangle then the determinant should be the product in a suitable functorial sense. Ferrand tried to work this out, but this turned out to be problematic. As you know the derivative of the determinant is the trace, so the trace in the middle should be the sum of the traces of the two sides. Alas, triangles in the derived category are not functorial. Nowadays, you pray that infinity categories will do the job, but at the time it was not the case. In fact Ferrand found that in general the trace in the middle is not always the sum of the traces. Usually when there was this sort of problem, you asked Deligne how to repair that. He came up with some sort of refined triangles, which he called true triangles, and then somehow the formula became correct.

I was looking at that and I thought, well, you cannot take tensor products of true triangles. If you have an exact sequence the left term is a subobject of the middle term. This is a small filtration, so we should take filtrations in many steps. This naturally led me to filtered derived categories, which was of course very successful afterwards and Deligne used it very much in Hodge theory. This is just an example of how you see something and want to make it more amenable to generalization and then you are sort of forced to do these things."

Present and future

In the last decade there has been a renewed interest in Illusie's work on the cotangent complex and the de Rham complex. This originated in the work of Alexander Beilinson who in 2011 gave a totally new proof of the p-adic comparison theorem [2].

"I am quite excited with the new developments, but I feel I am running after some people who are much faster than I. I was excited when in 2012 I received a preprint from Beilinson about his comparison theorem, which used to derive de Rham complex in a very striking way. That was fantastic but at the same time there was something I was very shocked about. In the paper he looked at some kind of a presheaf of categories and then he took

the associated sheaf. To me this did not make sense because we all know that objects in derived categories do not glue. We have known that for sixty years.

I asked him and he told me that there is a general formalism by Lurie that enables you to do that. So then I wrote to Jacob Lurie and asked: 'Please help me. Teach me how this works.' I discovered you could in fact take the associated sheaf using infinity categories. The whole thing was completely rigorous. I also discovered that Bhargav Bhatt had some other approach, so I also exchanged a lot with him. Those guys are young, they work so fast and they do so much.

In the sixties things were also evolving very fast. Think of the revolution of étale cohomology, toposes, crystalline cohomology, semi-stable reduction, stacks and so on. At that time I thought it was all quite natural and it was only in retrospect that I realized I was so privileged to have been part of this exciting adventure. But it seems that the period we are now living through with all these developments in derived algebraic geometry and all the new concepts around Scholze and many others is also a very exciting period. I think it is quite similar in spirit to the so-called golden age of the sixties."

These new exciting developments were also the topic of the Kan lectures. Illusie gave two talks on de Rham complexes in mixed characteristic, in which he talked about new work due to Bhatt-Lurie [9], Drinfeld [6], and Petrov that builds further on his own work with Deligne about this topic [5]. These talks were accompanied by a lecture about the history of the de Rham complex.

From Hodge to de Rham in characteristic p

The de Rham complex $\Omega^\bullet_X$ originates in differential geometry, where it is used to measure the difference between closed ($d\omega = 0$) and exact ($\omega = df$) forms. In algebraic geometry one can define a purely algebraic version of the de Rham complex. Because it is a complex of sheaves, you have to compute the hypercohomology instead of the ordinary cohomology to combine the contributions of the differential d and the cohomology of the sheaves.

Let us illustrate this for the Riemann sphere $\mathbb{P}^1$. There are only two terms in the de Rham complex: the sheaf of functions and the sheaf of one forms. To calculate their sheaf cohomology we use the cover $\mathbb{P}^1 = \mathrm{Spec}\,\mathbb{C}[z] \cup \mathrm{Spec}\,\mathbb{C}\left[\frac{1}{z}\right]$. This results in the following hypercomplex and hypercohomology:

$$\begin{array}{lccc} \Omega^0_{\mathbb{P}^1}: & \mathbb{C}[z]\oplus\mathbb{C}[\frac{1}{z}] & \xrightarrow{d-d} & \mathbb{C}[z,\frac{1}{z}] \\ \downarrow d & & & \\ \Omega^1_{\mathbb{P}^1}: & & \mathbb{C}[z]dz\oplus\mathbb{C}[\frac{1}{z}]\frac{dz}{z^2} & \xrightarrow{d} \mathbb{C}[z,\frac{1}{z}]dz \\ \\ \mathbb{H}(\Omega^\bullet_{\mathbb{P}^1},d): & \mathbb{C}(1,1) & 0 & \mathbb{C}\frac{dz}{z}. \end{array}$$

As expected for a sphere, the cohomology is 1-dimensional in degree 0 and 2 and 0 in degree 1.

Surprisingly, if you omit the d's and only calculate the sheaf cohomology, you will get the same result. This holds in general if you work over $\mathbb{C}$ and your scheme is smooth and complete. This is called the Hodge to de Rham degeneration.

$$H^\bullet_{dR}(X) = \mathbb{H}(\Omega^\bullet_X, d) \cong \mathbb{H}(\Omega^\bullet_X, 0) = H^\bullet_{Hodge}(X).$$

In characteristic p the situation is more complicated but in their seminal work of 1987 [5] Deligne and Illusie found conditions for which the Hodge to de Rham degeneration holds in characteristic p. They showed that for $p > \dim X$, it is sufficient that the scheme X lifts to a scheme $\tilde{X}$ which is defined modulo p^2 over the Witt ring.

"For the Kan Lecture I had prepared some slides but I did not use them. We put them on the web site but I consider in fact making a text from these slides. The problem is that in the coming months I am really too busy. I do not see the possibility of remodeling all that right now, but I hope I have a little more time next fall when my travel schedule is a bit more quiet.

I enjoy traveling. This is a privilege of mathematicians, that's why it's the greatest job of all. We can travel a lot and meet wonderful people. I was quite frustrated during the covid period not to be able to have contact in person. But now that we are back to normal, I would like to travel more again, as long as my health enables me."

With that, we wish Luc Illusie 'Bons voyages' and hope he will find some time to write his essay on the history of the de Rham complex.

References

1 M. Atiyah and I. Singer, The Index of Elliptic Operators on Compact Manifolds, *Bull. Amer. Math. Soc.* 69(3) (1963), 422–433.

2 A. Beilinson, p-adic periods and derived de Rham cohomology, *J. Amer. Math. Soc.* 25(3) (2012), 715–738.

3 B. Bhatt and J. Lurie, Absolute prismatic cohomology (2022), arXiv:2201.06120.

4 H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton University Press, 1956.

5 P. Deligne and L. Illusie, Relèvements modulo p^2 et décomposition du complexe de de Rham, *Invent. Math.* 89 (1987), 247–270.

6 V. Drinfeld, Prismatization (2020), arXiv: 2005.04746.

7 L. Illusie, *Complexe cotangent et déformations I, II*, Lecture Notes in Mathematics 239 and 283, Springer, 1971 and 1972.

8 *Cohomologie l-adique et fonctions L, Séminaire de Géométrie Algébrique du Bois Marie 1965–66 (SGA 5)*, dirigé par A. Grothendieck et Luc Illusie, Lecture Notes in Mathematics 589, Springer, 1977.

9 *Théorème d'Atiyah–Singer sur l'Indice d'un Opérateur Différentiel Elliptique*, Séminaire Cartan–Schwartz 16(1) (1963/64), Secrétariat Mathématique, 1965.

10 *Théorie des intersections et théorème de Riemann-Roch*, Séminaire de Géométrie Algébrique du Bois Marie 1966–67 (SGA 6), dirigé par P. Berthelot, A. Grothendieck et L. Illusie, Lecture Notes in Mathematics 225, Springer, 1971.

11 *Théorie des topos et cohomologie étale des schémas. I, II, III*, Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964 (SGA 4), dirigé par M. Artin, A. Grothendieck et J. L. Verdier, avec la collaboration de N. Bourbaki, P. Deligne et B. Saint-Donat, Lecture Notes in Mathematics 269, 270, 305, Springer, 1972 and 1973.

Nelly Litvak
University of Twente, and
Eindhoven University of Technology
n.litvak@utwente.nl

Research

Randomness and structure in complex networks

How to mathematically describe real-life networks, such as social networks or the World Wide Web, as so-called random graphs? How do particular rules of placing random edges result in graphs that share some of the fundamental empirical properties of real-life networks? This article of Nelly Litvak is based on her lectures of PWN Vakantiecursus 2022.

Complex networks, modeled as random graphs

Many real-life systems are *networks*. A network is a set of objects connected by some relationship. For example, a railroad is a collection of stations connected by rails. In a social network, people are connected by friendships. The Internet is a network of routers connected by wires. In our brain, neurons are connected if they fire together.

A *graph* is a natural mathematical model for a network of any kind. In a graph, each object is represented as a *vertex*, and if there is a relationship between two vertices, then there is an *edge* between them. *Undirected edges* represent symmetric relations, and we draw them as lines. For example, communications between two Internet routers usually go in both directions. If the relation is not symmetric, we model this using *directed edges*, and draw them as arrows. For example, if somebody follows you on Twitter, you might not follow back. In Figure 1, we give some examples of networks, directed and undirected.

This article is about how to mathematically describe real-life networks, such as social networks or the World Wide Web, using so-called *random graphs*. Usually in research, and always in this article, the vertices of a random graph are fixed, while the edges

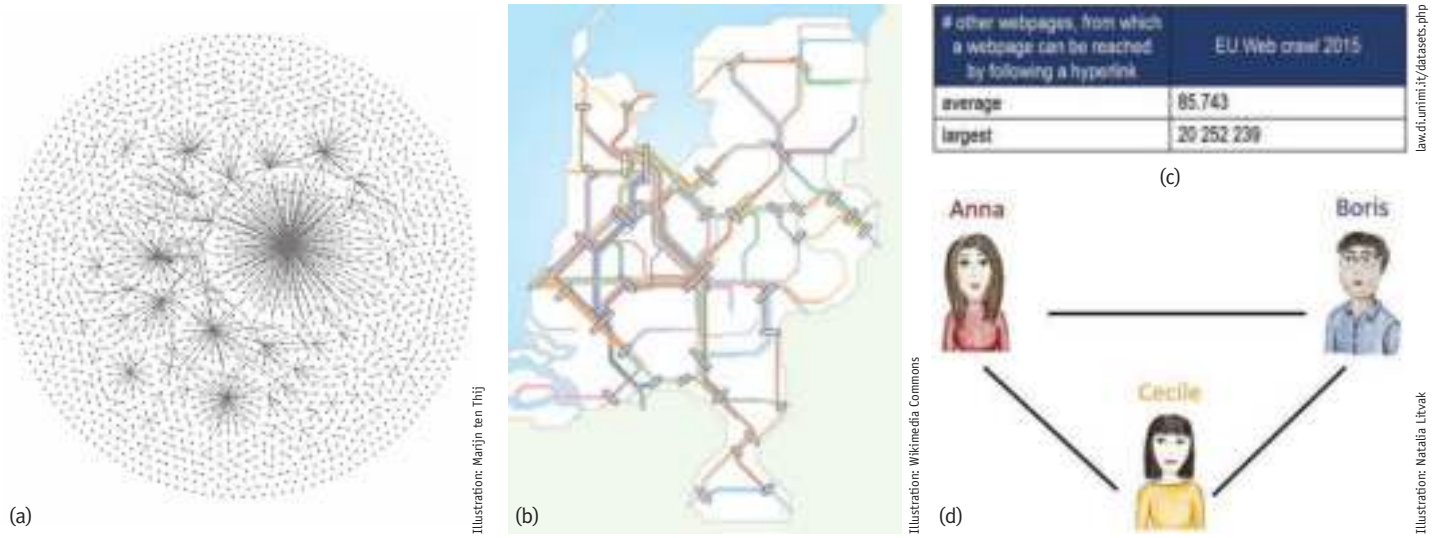


Figure 1 (a) Network of retweets about Project X in Haren, 21-09-2012, 7:00. The average number of retweets per user is small compared to the network size [11]. This network is *sparse*. (b) Any railway station can be reached by train from any other. The railway network is *connected*. (c) The number of other webpages, from which a webpage can be reached, varies greatly per webpage. The Webgraph is *scale-free* [3]. (d) Anna is a friend of Boris and Cecile. Then, Boris and Cecile are likely to be friends, too. Social networks have many *triangles*.

are placed at random. This makes sense because relationships between objects often emerge at random, like friendships in a social network. Also, even if a network is not random, such as the Internet, its structure is so complicated that it is often useful to describe it through statistical summaries and model it as a random object. A *random graph model* is in fact a set of rules, according to which the random edges are chosen. Different rules result in different models with different mathematical properties. For example, in one model, edges between all vertex pairs can be equally likely, while other models assign higher probabilities to some vertex pairs.

In this article, I will tell about several random graph models that are by now well understood, even classical. My goal is to explain how particular rules of placing random edges result in graphs that share some of the fundamental empirical properties of real-life networks. There are many such properties, but I will address four of them, listed in the next section.

Four fundamental properties of real-life networks

Surprisingly, many networks of a completely different nature, such as social networks and the Internet, share common properties. This is why we talk about a *structure* of a network. In this article I will discuss four such common structural properties.

- *Sparse*. Consider a social network. Even if the network is very large, a person can maintain only so many friendships. We say that social networks are *sparse*, meaning that the number of connections per person is limited, and does not increase very much even if the network grows. Figure 1(a) shows an interesting example: the network of retweets about Project X in Haren in 2012. A birthday invitation of a 16-year-old girl went viral in social media and resulted in a destructive riot. Dots are Twitter users and each tiny arrow (a directed edge) represents a retweet from one user to another. The figure shows this network in the morning before the riot. We see that the network is sparse, on average there are only 1.5 retweets per user. Over the night of the riot the network increased in size more than 10 times, but the average number of retweets remained small, it went only a little bit above two.
- *Connected*. Consider the network of railway stations connected by railroads, as the NS network in Figure 1(b). The vertices are the stations, and the (undirected) edges are the railway connections between them. A passenger can travel by train from any station to any other. The railroad network is *connected*. The Internet is another powerful example of a connected network. Internet is extremely complex and completely decentralized, yet, the data can be transferred across the planet from any Internet router to any other!
- *Scale-free*. In the Web graph, vertices are the webpages, and (directed) edges are the hyperlinks. By clicking on a hyperlink, we can go from one webpage to another. From how many other pages a typical webpage can be reached? In Figure 1(c) we show the average and the maximum of this number in the `.eu` domain of the Web graph in 2015. We see that on average, a webpage can be reached from 85.7 other webpages. However, the maximum is over $200\,000$ times larger than average! We say that such network is *scale-free*. This unusual term means that there is no such thing as a 'typical webpage'. The number of hyperlinks pointing to a webpage can have very different scales — from a few, to hundreds, thousands, and millions.
- *Triangles*. How do people in social networks usually meet? Often I know friends of my friends, they can be my friends, too. Groups of friends create clusters with many *triangles*, such as in Figure 1(d): if Anna is a friend of Boris and Cecile, then it is not surprising that Boris and Cecile know each other as well. Having many triangles is another typical structural property of complex networks.

Now that we have discussed the common properties of complex networks, let's see how to model them mathematically using random graphs.

Modeling sparse networks with Erdős–Rényi random graphs

Erdős–Rényi random graph

Recall that in a random graph, vertices are fixed, and edges are placed at random. Suppose we have n vertices. Possibly the simplest rule for placing random edges is to connect any pair of vertices i and j by an edge with the same probability p, independently of all other edges. This model is called an *Erdős–Rényi (E-R) random graph* after the founders of the theory of random graphs Paul Erdős and Alfréd Rényi [7]. We denote this random graph by $\mathrm{ER}_n(p)$. Figure 2 is a small example of a social network as an Erdős–Rényi random graph. The solid lines are existing connections and the dashed lines are possible but not realized connections.

You may see how E-R graphs look like for different n and p on the *Network pages* website [16]. The simulator also includes parameter $\lambda = np$, the significance of which I will explain in the next section. Look at the figures produced by the simulator: do the graphs look as you expected?

Of course, in reality, social connections are formed not at all as in the E-R model. This simple model ignores complex dependencies, heterogeneity between people, polarization, et cetera. Moreover, it is not even supposed to include all these things because Erdős and Rényi invented this model for completely different purposes: to solve difficult problems in combinatorics! This ingenious idea yielded a by now well established approach in Graph Theory, the so-called 'probabilistic method' [1]. However, as it often happens in mathematics, a model invented for solving one problem, becomes very useful for solving completely different problems in

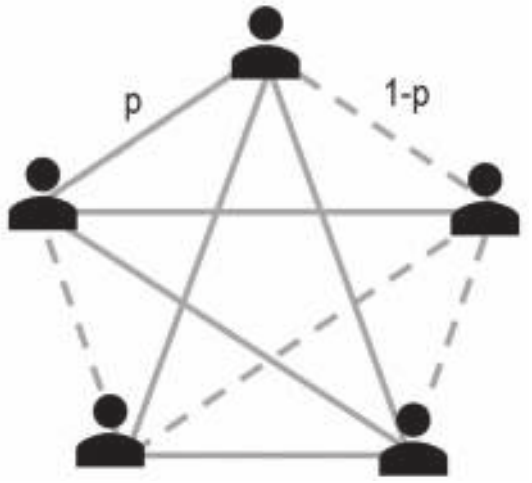


Figure 2 Social network is modeled as an Erdős–Rényi (E-R) random graph: each friendship exists with probability p.

the unknown future. With arrival of abundant network data, and explosive growth of Network Science, the E-R model is a fundamental cornerstone in studying real-life networks, because deep understanding of the simplest possible mathematical models is crucial for explaining real-world phenomena. In this section the phenomenon in question will be sparsity of real-life networks.

Sparse networks

The notion of a *sparse* network has to do with the number of edges and the number of vertices. Loosely speaking, in a sparse network, there are not too many edges per vertex on average. But how many is 'not too many'? Suppose we have a networks of 1000 vertices. If there are on average, say, 2 edges per vertex, then we will probably all agree that this network is sparse. If there are on average 500 edges per vertex, then, we will probably say that this network is not sparse, because 500 is a lot compared to 1000. But what about 20 edges per vertex? Or 50? Or 70? When does a network 'start' being sparse?

We could give a mathematical answer to this by choosing a threshold of some sort. However, the Theory of Random Graphs takes a different approach. We say that a network is *sparse* when the average number of edges per vertex remains bounded when the number of vertices n grows to infinity. In other words, sparsity is a so-called *asymptotic notion*, it is defined only in the limit, when $n \to \infty$.

Sparse Erdős–Rényi random graphs

Let us now see how sparsity works out in the E-R random graph, $\mathrm{ER}_n(p)$.

In any undirected graph of n vertices, there are in total $\frac{n(n-1)}{2}$ possible edges. Since in the $\mathrm{ER}_n(p)$ model each edge exists with probability p, there are on average

$$\frac{n(n-1)}{2} \cdot p$$

edges. To get the average number of edges per vertex, we divide the last expression by n. The result is:

$$\frac{(n-1)}{2} \cdot p \quad \text{edges per vertex on average.} \tag{1}$$

Intuitively, the network is sparse when p is small because then there are less edges per vertex. But how small should p be? Suppose we choose a very small *fixed* p, say, $p = 0.0001$. Then the average number of edges per vertex is

$$\frac{(n-1)}{2} \cdot 0.0001.$$

The most important observation about this expression is that it is *not* bounded, it grows to infinity when n grows to infinity, so our asymptotic definition of sparsity is violated.

Since even a very small fixed p doesn't give us a sparse network, we need a different approach. Here we use a powerful technique called *parametrization*. The idea is to not view p as a given constant in $(0,1)$ but make it a function of n, so $p = p(n)$. Why is it useful? Because, then we can choose the function $p(n)$ in such a way that it will 'compensate' for the growth of n in expression (1) for the average number of edges per vertex.

In order to create a sparse network, we choose

$$p := p(n) = \frac{\lambda}{n}, \quad \lambda > 0.$$

This is, by the way, the same λ that you saw in the simulator before. If we substitute this in (1), the average number of edges per vertex becomes

$$\frac{n-1}{2} \cdot \frac{\lambda}{n} < \frac{\lambda}{2}.$$

Since $\lambda > 0$ is fixed, the average number of edges per vertex is bounded by a fixed number $\frac{\lambda}{2}$, hence the network is indeed sparse.

We denote a sparse E-R random graph by $\mathrm{ER}_n(\lambda/n)$. It is important to notice that the parametrization does not change the number of parameters of the model. The original model $\mathrm{ER}_n(p)$, has two parameters: n and p. The sparse model $\mathrm{ER}_n(\lambda/n)$, too, has two parameters: n and λ. The difference is that in the sparse model, $p = \lambda/n$ decreases with n, and this gives us a sparse network for any fixed λ. This is how a very simple model allows us to describe mathematically a very essential property of real networks — their sparsity, and remains flexible enough to create sparse graphs with less or more edges, by varying λ.

I want to close this section with a small remark from my experience in teaching random graphs. When students learn about the Erdős–Rényi model, they easily remember that all edges have equal probability p, while the independence of edges is often overlooked. The independence of edges indeed doesn't play any role in modeling sparsity (because sparsity concerns only the average number of edges per vertex), but it will take a central stage later in this article, explicitly so when we will discuss the number of triangles.

Almost sure guarantee that a random graph is connected

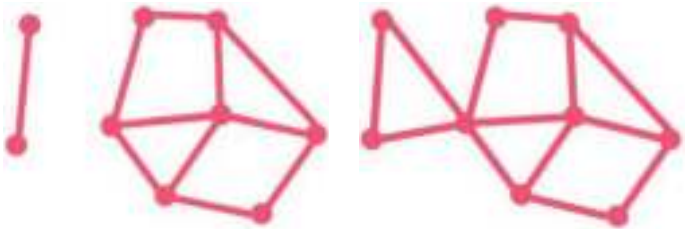
A graph is *connected* if there is a route along the edges from any vertex of the graph to any other. Graphs in Figure 1(b) and Figure 3(b) are examples of such connected graphs.

As we already know, many important real-life networks are connected. For example, in the Internet, information can be delivered from any router to any other. But when we model a network as a random graph, we place edges at random. Will such graph be connected? In this section we will solve this beautiful mathematical puzzle for the Erdős–Rényi random graph.

Isolated vertices in E-R random graph

It is easy to realize that if there are isolated vertices (vertices with no edges attached to them), then the graph is definitely disconnected. Of course, there are other ways to disconnect the graph: there can be isolated islands of $2, 3, \ldots$ vertices. Yet, as we will see, isolated vertices are a good start. Moreover, surprisingly, in the E-R model, isolated vertices define whether the graph is connected or not.

To begin with, let us compute the average number of isolated vertices. We denote this number by N_0; the sub-index zero reflects



(a) Disconnected graph. (b) Connected graph.

Figure 3 In a connected graph, all vertices are connected by a path of edges

that isolated vertices have zero edges attached to them. There are n vertices, each of which can be isolated with some probability, and we need to compute this probability first. There are $n-1$ possible edges, each edge is absent with probability $1-p$, and the edges are independent. Then, the probability that all $n-1$ possible edges of a vertex are absent, is $(1-p)^{n-1}$. The result is:

$$\mathbb{P}(\text{a vertex is isolated}) = (1-p)^{n-1},$$

where $\mathbb{P}(\cdot)$ denotes the probability of an event in the brackets. There are n vertices in total, and the average number of isolated vertices is

$$N_0 = n \cdot (1-p)^{n-1} \tag{2}$$

due to the very useful and somewhat counter-intuitive result in probability, called *linearity of expectations*. The linearity of expectations tells us that when we add random variables, the average of the sum is *always* the sum of averages, even if random variables are dependent. In this case, we can think of counting isolated vertices as adding random zeros and ones. We add 0 if the vertex is *not* isolated, and 1 if the vertex is isolated. These zeros and ones are *dependent*: for example, if we know that $n-1$ vertices are isolated, then for sure the n-th vertex is isolated as well (I leave it to the reader to explain why). Nevertheless, on average, each vertex adds $(1-p)^{n-1}$ to the sum, and the total average is $n \cdot (1-p)^{n-1}$ thanks to the linearity of expectations.

Sparse E-R random graphs are disconnected

Expression (2) is already sufficient to conclude that large sparse E-R graphs are very likely to be disconnected! Indeed, take a sparse E-R random graph $\mathrm{ER}_n(\lambda/n)$, where $p = \lambda/n$. Then (2) becomes

$$n\left(1-\frac{\lambda}{n}\right)^{n-1} \approx ne^{-\lambda}, \tag{3}$$

where the approximation is quite good for large n, and follows from the famous limit:

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}. \tag{4}$$

This approximation (3) tells us that the number of isolated vertices in a sparse E-R random graph grows roughly linearly with n. When there are so many isolated vertices on average, it is very unlikely that there will be none of them in the graph. The exact mathematical result is as follows:

$$\lim_{n\to\infty}\mathbb{P}(ER_n(\lambda/n) \text{ is connected}) = 0.$$

Connectivity threshold

Let us now think of connectivity of the $\mathrm{ER}_n(p)$ model when p changes from 0 to 1. When $p=0$, there are no edges, so the graph is disconnected. When p grows, there will be more edges, and when $p=1$, all edges are present, so the graph is connected and complete. Then, there should be a range of values of p, for which the graph is likely to be connected. We already saw that $p=\lambda/n$, as in sparse E-R graphs, is not large enough to connect all vertices. For connectivity, we need larger p, but it turns out that 'a little bit' larger p is already sufficient. The transition from disconnected graphs to connected ones occurs when we choose a slightly different parametrization:

$$p = \frac{a\ln(n)}{n}, \quad \text{for some } a>0. \tag{5}$$

Indeed, let us substitute this p in (2). Then we get

$$\begin{aligned} N_0 &= n\cdot(1-p)^{n-1} = n\cdot\left(1-\frac{a\ln(n)}{n}\right)^{n-1} \\ &\approx n\cdot e^{-a\ln(n)} = n\cdot n^{-a} = n^{1-a}, \end{aligned} \tag{6}$$

where the approximation holds analogously as in (4). We see that the average number of isolated vertices is approximately n^{1-a}. Whether this number is big or small when n goes to infinity, depends on a. When $a<1$, we have that $1-a$ is positive, so n^{1-a} grows to infinity. In this case, a large network has many isolated vertices, the graph is disconnected. However, if $a>1$, then $1-a$ is negative, and n^{1-a} is vanishing. In this case, on average, there are no isolated vertices, and the probability that a graph is connected, converges to 1 as n goes to infinity. In probability theory we say that in this case the graph is connected *almost surely*. For large finite graphs of size n, this means that the corresponding probability is smaller than 1 but very close to 1, and becomes even closer when n grows. The exact mathematical result is:

$$\text{If } a<1, \text{ then } \lim_{n\to\infty}\mathbb{P}\left(\mathrm{ER}_n\left(\frac{a\ln(n)}{n}\right) \text{ is connected}\right) = 0.$$

$$\text{If } a>1, \text{ then } \lim_{n\to\infty}\mathbb{P}\left(\mathrm{ER}_n\left(\frac{a\ln(n)}{n}\right) \text{ is connected}\right) = 1.$$

The abrupt transition from disconnected to connected graphs is a beautiful example of a phenomenon known as *phase transition*. Like water turns into ice at zero degrees Celsius, isolated vertices vanish when connection probability p exceeds the so-called *connectivity threshold*

$$p_{\text{connectivity}} = \frac{\ln(n)}{n}.$$

We see this in Figure 4(a) and 4(b), a graph here has 100 vertices, and $p_{\text{connectivity}} \approx 0.046$.

Disappearance of disconnected islands

One may wonder, what about other ways of making a graph disconnected? Can there be islands of $2, 3, \ldots$ vertices? To answer this question, we consider *isolated pairs*. An example of a graph with an isolated pair is shown above in Figure 3(a).

Denote the average number of isolated pairs by N_2. Let us now compute this number. There are $\frac{n(n-1)}{2}$ pairs in total. A pair is isolated, if the two vertices in the pair are connected to each other, this happens with probability p. Furthermore, each of the two vertices in the pair must be isolated from the other $n-2$ vertices in the graph. The probability of this for each vertex is $(1-p)^{n-2}$, and for two vertices simultaneously, is $[(1-p)^{n-2}]^2$. Altogether, we get

$$N_2 = \frac{n(n-1)}{2}\cdot p\cdot\left[(1-p)^{n-2}\right]^2.$$

If we now substitute $p = \frac{a\ln(n)}{n}$ as in (5), we get

$$\begin{aligned} N_2 &= \frac{n(n-1)}{2}\cdot\frac{a\ln(n)}{n}\cdot\left[\left(1-\frac{a\ln(n)}{n}\right)^{n-2}\right]^2 \\ &\approx \frac{n^2\left(1-\frac{1}{n}\right)}{2}\cdot\frac{a\ln(n)}{n}\cdot n^{-2a} \\ &\approx \frac{1}{2}\cdot\frac{a\ln(n)}{n}\cdot n^{2(1-a)}. \end{aligned}$$

Compare this to the expression (2) for N_0. If $a>1$ then $n^{2(1-a)}$ goes to zero even faster than n^{1-a}. In addition, $\frac{a\ln(n)}{n}$ goes to zero,

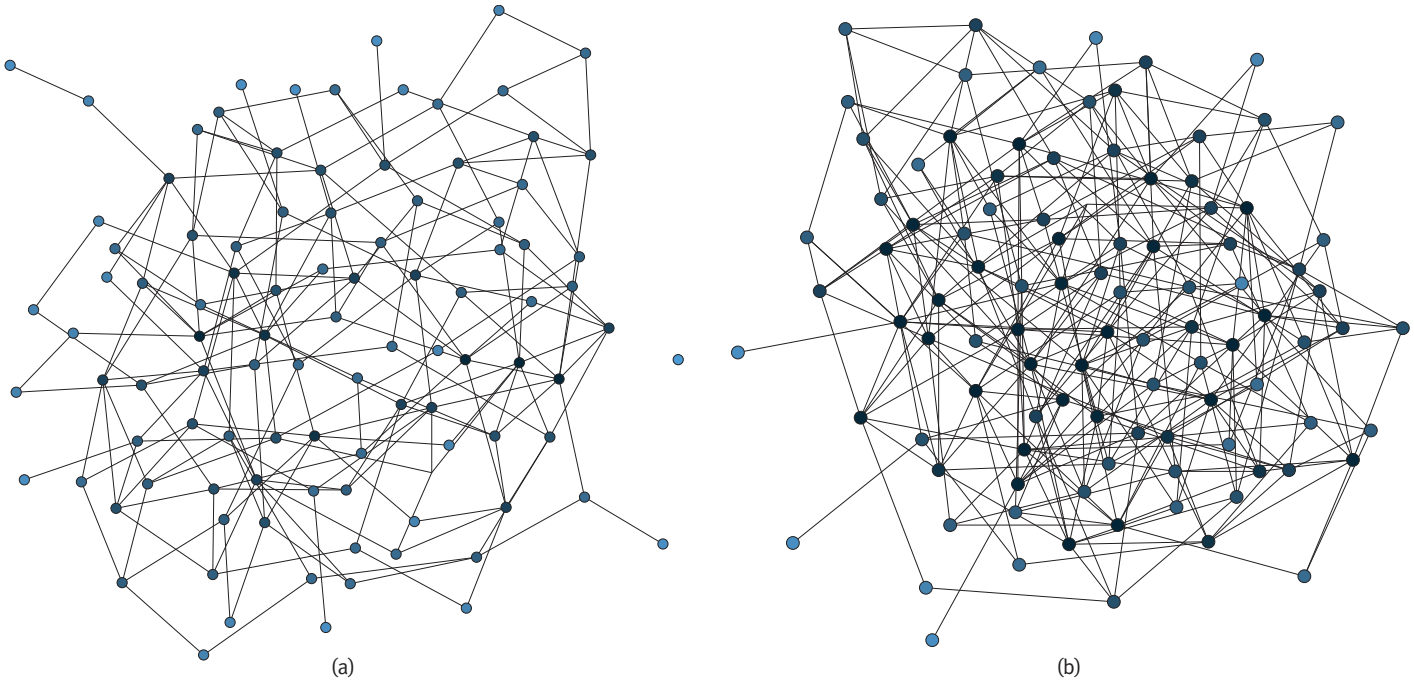


Figure 4 (a) E-R random graph with $n = 100$, $p = 0.04$. The graph is disconnected. (b) E-R random graph with $n = 100$, $p = 0.05$. The graph is connected.

as well. Hence, we conclude that above the connectivity threshold, isolated pairs are even less likely than isolated vertices.

In fact, we can mentally visualize how the graph 'gets frozen' when p increases. You can observe how this happens in the simulator on Network Pages [16]. The simulator adds edges one by one, and this is equivalent to gradually increasing p. The process starts with completely disconnected vertices. Then, p and the number of edges grow, and connected islands appear. As the graph becomes denser, large disconnected islands start merging together. Gradually we do not see anymore islands of 5, 4, 3, 2 vertices. Finally, when p exceeds $p_{\text{connectivity}}$, the last isolated vertices disappear, and the graph becomes connected. And now you also know why the process evolves this way, and you can explain it with just a couple of formulas!

Mathematics of scale-free networks

Real-life networks are often *scale-free*} In this section we want to express this scale-free property in mathematical terms, and get some insight in how this phenomenon emerges.

It will be convenient to introduce a bit of terminology. The number of edges attached to a vertex is called the *degree* of this vertex. For instance, in Figure 5, the degree of vertex 2 is equal to 3.

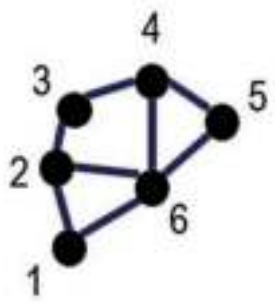


Figure 5 A degree of a vertex is the number of edges attached to it. For example, the degree of vertex 2 is 3.

The scale-free property is nothing else but a large variability of degrees. In real-life networks, there is often a relatively small but notable group of vertices with extremely high degrees, much higher than average. This is the case, for instance, in the World Wide Web, as we saw in Figure 1(c).

Sparse E-R graphs are not scale-free

The title of this subsection already gives away the result, but we are mainly interested in why sparse E-R graphs are not scale-free. An intuitive explanation is that the scale-free property also means large *heterogeneity* of vertices: some vertices attract many more connections than others. In E-R random graph this is clearly not the case: all edges have the same probability, and in that sense, all vertices are symmetric. Yet, since the edges are placed at random, it is also true that degrees of the vertices are random, so there will be some variability among them. The main message of this section is that this variability, solely due to randomness of the edges, is insufficient to create a scale-free network.

Let us see how this looks in formulas. Every vertex has $n-1$ possible edges, each edge exists with probability λ/n, independently of anything else. Then the degree of a vertex follows the well-known *Binomial distribution*, and we can write down the probability that the degree of a vertex is k:

$$P(\text{degree of a vertex is } k) = \frac{(n-1)!}{(n-1-k)!k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-1-k}, \quad k = 0,1,\ldots,n-1. \tag{7}$$

Now, if we take the limit $n \to \infty$, this converges to

$$\lim_{n\to\infty} P(\text{degree of a vertex is } k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0,1,\ldots. \tag{8}$$

I leave it to the reader to derive this limit. A hint: the limit (4) will be useful again. In particular, the limiting probability that a degree is zero (obtained by substituting $k = 0$), is $e^{-\lambda}$. This is of course

the same expression as the probability that a vertex is isolated, that we used in (3). The probability distribution in (8) is called the *Poisson* distribution. The parameter λ is the mean, or, the average of the Poisson distribution. This is consistent with the sparse E-R model because

$$\text{average degree in } \mathrm{ER}_n(\lambda/n) = (n-1)\cdot\frac{\lambda}{n} \approx \lambda.$$

Again, the approximation is accurate when n is large.

We can now use the Poisson formula to check the scale-free property. For instance, assume that the average degree is $\lambda = 100$ (similar to the data in Figure 1(c)), and let us compute the probability that a vertex has degree $k = 1000$, a modest 10-fold of the average, much smaller than typical large degrees in the real data. Substituting the numbers in (8) we get

$$\lim_{n\to\infty} P(\text{degree of a vertex is } 1000 \mid \lambda = 100) = \frac{100^{1000}}{1000!}e^{-100}. \quad (9)$$

Although 100^{1000} is impressive, the $1000!$ in the denominator is a humongous number. Expression (9) has a stunning 610 zeros after the decimal point! Moreover, the probability that degree is *at least* 1000 is not much larger:

$$\begin{aligned}
&\sum_{k=1000}^{\infty}\frac{100^k}{k!}e^{-100}\\
&= \frac{100^{1000}}{1000!}e^{-100}\left(1+\sum_{k=1}^{\infty}\frac{100^k}{(1000+k)\cdot(1000+k-1)\cdots 1001}\right)\\
&< \frac{100^{1000}}{1000!}e^{-100}\left(1+\sum_{k=1}^{\infty}\left(\frac{1}{10}\right)^k\right)\\
&= \frac{100^{1000}}{1000!}e^{-100}\cdot\left(1+\frac{1}{9}\right).
\end{aligned}$$

which is the same expression as in (9), only multiplied by ≈ 1.11, so we basically get the same answer. This is because the Poisson distribution is known to stay very close to its average. It is fundamentally not scale-free!

Power laws

A model that is able to capture the scale-free phenomenon, is the so-called *power law*. In this model the probability that a vertex has degree k is approximately proportional to a negative power of k, thus the name 'power law'. For the time being, we will write this as

$$\begin{aligned}
&P(\text{degree of a vertex is } k) \approx C\cdot k^{-\tau},\\
&\text{where } \tau > 2, C > 0, k > k_{\min} > 0.
\end{aligned} \quad (10)$$

Clearly smaller τ means higher probability of k. In particular, even very large k are quite likely to occur when τ is small. However, we need τ to be greater than 2 because otherwise large values are so likely that the average degree is infinite, and therefore the network is not sparse anymore. I encourage you to check this using the formula for the average, or, mean degree

$$\text{average degree} = \sum_{k=k_{\min}}^{\infty} k\cdot P(\text{degree of a vertex is } k).$$

(*Hint:* Think of harmonic series.)

Let us now verify that the power laws indeed model the scale-free phenomenon. Take $\tau = 2.5$, $k_{\min} = 33$, $C = 288.67$. Then the average degree is approximately again 100, but now we get

$$\begin{aligned}
P(\text{degree of a vertex is } 1000) &= 288.67\cdot 1000^{-2.5}\\
&= 0.0000091287 \approx 10^{-5}.
\end{aligned}$$

This probability, about one in a hundred thousand, is not very small in a network of millions, even billions vertices such as the World Wide Web. And the probability that the degree is *at least* 1000, is even much bigger:

$$\begin{aligned}
\sum_{k=1000}^{\infty} 88.67\cdot k^{-2.5} &\approx 288.67\cdot\int_{1000}^{\infty} x^{-2.5}dx\\
&= 192.45\cdot 1000^{-1.5} \approx 0.006.
\end{aligned}$$

This is a sizable proportion of vertices especially in a very large network. These results are even not very surprising because, definitely, the $k^{-\tau}$ decreases much slower than $\lambda^k/k!$.

When a model correctly captures a phenomenon however, it does not mean that the model is accurate, and there can be other suitable models. After power laws were found for the first time in the paper by three brothers Faloutsos [8] on the connectivity of the Internet, and then, explosively, in most other types of networks [2], there has been a lot of discussion about whether this model really holds. An interesting milestone in this discussion was the 2019 paper 'Scale-free networks are rare' by Broido and Causet [5]. The authors' statistical analysis resulted in the conclusion that (10) holds in but a modest fraction of real-life networks. Almost immediate answer to these results was the paper by Voitalov, van der Hoorn, van der Hofstad and Krioukov [12], called 'Scale-free networks well done', with a playful hint to the steak analogy. This paper has reminded of the fact that the power law is a much broader notion than the rather stringent formula (10). Mathematically, power laws are defined more precisely as

$$P(\text{degree of a vertex is } k) = C(k)\cdot k^{-\tau}, \quad \text{where } \tau > 2,\ k > 0, \quad (11)$$

with function $C(k)$ being a so-called *slowly varying* function. The term 'slowly varying' means that $C(k)$ may grow or decrease not faster than any arbitrarily small power of k. Formally, this can be written as

$$\lim_{k\to\infty}\frac{\log(C(k))}{\log(k)} = 0. \quad (12)$$

Of course, $C = \text{const}$ is a special case of a slowly varying function. The point in [12] is that (12) is an *asymptotic* expression, so for finite k we can never be sure whether (11) holds or not. The authors in [12] suggested their own method how to check whether the power law model is acceptable, and many real-life networks indeed do pass this test.

The debate on power laws in 2019 attracted a lot of attention and even hit the media [15]. I enjoyed this scientific discussion, but personally I do not find it extremely important whether power laws exist in reality or not. Mathematics describes the world with abstract objects, and in this case I find power laws a very suitable object because it does capture the essence of the phenomenon in hand: the high variability of degrees in real-life networks.

Preferential attachment

While we can observe, measure and model the scale-free phenomenon, the question remains, why are the networks scale-free? The *preferential attachment* model is an attempt to answer this question with a dynamic mathematical model of network growth.

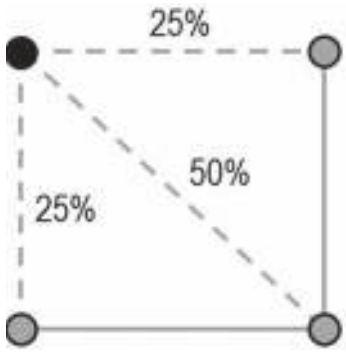


Figure 6 A new (black) vertex arrives in the network, and connects to existing vertices with probabilities proportional to their current degrees. Dash lines denote possible edges, and the number next to a dashed line is the probability of this edge.

This model formalizes the mechanism known as 'rich get richer', or 'popular get ever more popular'. It works roughly as follows. Assume a network starts with three vertices, like the three gray vertices in Figure 6. When the next vertex arrives (the black vertex in the figure), it can make one of three possible connections (dashed lines). For that, the new vertex uses the rich get richer mechanism: the probability to connect to any of the gray vertices is proportional to the number of connections they already have. In Figure 6, one of the gray vertices has two connections, and therefore it has a higher chance to get one more. Clearly, the more connections a vertex gets, the easier it is to attract even more new connections. This is the rich get richer mechanism at work! It is quite natural that with such dynamics, some vertices become extremely well connected.

Barabási and Albert in 1999 [2] put forward the preferential attachment model as a plausible mathematical explanation for the emergence of scale-free networks. This became a very influential work, paper [2] has more than 43K citations on Google Scholar, and counting! That said, it rarely happens that such an influential model and its brilliant application in a rising new area of science, has no history in earlier work. Indeed, the model is not new. In 1965, Derek de Solla Price suggested a very similar model for networks of scientific citations [10]: papers that are already frequently cited, tend to receive many new citations. Even earlier, in 1927, Yule proposed a similar model for the evolution of biological species [13]: large populations of species have large offspring. And even before that, in 1925, great mathematician George Pólya and his PhD student Florian Eggenberger published a paper about a mathematical model that is now called *Pólya's urn scheme*. The scheme works as follows. We have a number of green and red balls in an urn. We choose one ball at random, and then return it to the urn together with another ball of the same color. Figure 7 illustrates how it works.

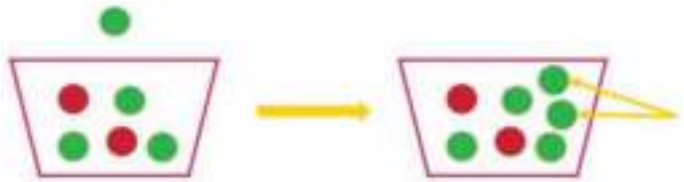


Figure 7 Pólya's urn scheme. Initially we have 4 green balls and 2 red balls. We randomly chose a green ball, and therefore we add another green ball to the urn. As a result, we have 2 red balls and 5 green balls. It is now even more likely to choose a green ball.

Here we have more green balls in an urn, so it is more likely to choose a green ball, as in the figure. After that, we add another green ball, so in the next round, it's even more likely to choose a green ball again. This is exactly the rich-get-richer mechanism. Pólya's urn is another example of a mathematical model that appeared before its applications. Mathematical techniques developed for Pólya's urn scheme, are often used to analyze preferential attachment dynamics in networks.

Emergence of power laws

Intuitively, it is clear that preferential attachment helps some vertices to get lots of connections, making power laws plausible. In this section we will go one step beyond intuition and see why it happens, mathematically. There are many ways to do this. In this article I use the argument in lines of [9, Section 8.3]. The way I derive it here is not strictly rigorous, we may call it a *heuristic* argument. However, it does give mathematical reasons why degrees in the preferential attachment model follow the power law distribution (10).

Suppose the network starts at time $t=0$ with vertex zero without edges. At time $t=0$, vertex 1 arrives with one edge and connects to vertex 0. After that, vertices appear in the network one by one, and we will number them $2,3,\ldots,t$, so t is our 'current time', and, including vertex 0, there are $t+1$ vertices at time t. Denote by $D_i(t)$ the average degree of vertex i at time $t \geq i$. With arrival of vertex t, the average degree of i in the previous step, $D_i(t-1)$, changes according to the preferential attachment rule. There are many versions of this rule, but we assume the simplest one: a new vertex makes only one connection, and attaches to one of the existing vertices with probability exactly proportionally to their degree. So, the probability that vertex t attaches to vertex i is,

$$\frac{D_i(t-1)}{D_0(t-1)+D_1(t-1)+\cdots+D_{t-1}(t-1)} = \frac{D_i(t-1)}{2(t-1)}, \qquad (13)$$

where the denominator equals to $2(t-1)$ because in any graph all degrees sum up to twice the number of edges, and at time $t-1$ we have $t-1$ edges because each vertex $1,2,\ldots,t-1$ arrives with one edge.

Let us now look how the average degree of vertex i changes at time t:

$$\begin{aligned} D_i(t) &= D_i(t-1) + 1 \cdot \frac{D_i(t-1)}{2(t-1)} \\ &= D_i(t-1) \cdot \left(1 + \frac{1}{2(t-1)}\right) \\ &= D_i(t-1) \cdot \frac{2(t-1)+1}{2(t-1)} \qquad (14) \\ &= D_i(t-1) \cdot \frac{(t-1)+1/2}{t-1}. \qquad (15) \end{aligned}$$

We can now iterate (15) back in time till $D_i(i)=1$:

$$\begin{aligned} D_i(t) &= D_i(t-1) \cdot \frac{(t-1)+1/2}{t-1} \\ &= D_i(t-2) \cdot \frac{(t-1)+1/2}{t-1} \cdot \frac{(t-2)+1/2}{t-2} \\ &= \cdots = D_i(i) \cdot \frac{(t-1)+1/2}{t-1} \cdot \frac{(t-2)+1/2}{t-2} \cdots \frac{i+1/2}{i} \\ &= 1 \cdot \frac{\Gamma(t+1/2)}{\Gamma(i+1/2)} \cdot \frac{\Gamma(i)}{\Gamma(t)}. \qquad (16) \end{aligned}$$

We can now use the property of the gamma function $\Gamma(\cdot)$:

$$\frac{\Gamma(x+\alpha)}{\Gamma(x)} \approx x^{\alpha} \quad \text{for large enough } x \text{ and fixed } \alpha.$$

If t and i in (16) are large, then we have an approximate expression for the average degree of vertex i at time t:

$$D_i(t) \approx \frac{t^{1/2}}{i^{1/2}} \quad \text{when } t \text{ and } i \text{ are large enough.}$$

We already see the power $1/2$ appearing, it remains only to translate this to the power law (10). This is where our derivation becomes 'heuristic'. Remember that $D_i(t)$ is the average degree of vertex i, but let us pretend that this is the exact degree (even if it is not necessarily an integer number). Then the fraction of vertices with degree k, among the total of $t+1$ vertices, is

$$\frac{1}{t+1} \cdot \big[\text{the range of } i \text{ such that } D_i(t) \in (k-0.5, k+0.5)\big].$$

Now we need to find, which i are in the required range:

$$k-0.5 < \frac{t^{1/2}}{i^{1/2}} < k+0.5.$$

Solving for i, we get:

$$\frac{t}{(k+0.5)^2} < i < \frac{t}{(k-0.5)^2}.$$

The length of the interval that contains i above, is:

$$\begin{aligned}\frac{t}{(k-0.5)^2} - \frac{t}{(k+0.5)^2} &= t \cdot \frac{1}{(k-0.5)^2} \cdot \left(\frac{(k+0.5)^2 - (k-0.5)^2}{(k+0.5)^2} \right) \\ &= t \cdot \frac{2k}{(k-0.5)^2 (k+0.5)^2} \\ &\approx 2t \cdot k^{-3}, \quad \text{when } k \text{ is large enough.}\end{aligned}$$

It remains to divide the last expression by $t+1$:

$$2 \cdot \frac{t}{t+1} \cdot k^{-3} \approx 2k^{-3},$$

and we have obtained a power law, as in (10), with $C=2$ and $\tau=3$.

We can vary τ, by adjusting the model a little bit. For instance, the attachment probability to vertex i is proportional to

$$\big[\text{degree of } i\big] + a$$

for some $a > -1$. Interestingly, this changes τ, we get $\tau = 3+a$. We may also allow vertices to arrive with $m>1$ edges, then we must take $a > -m$, and the power law exponent is $\tau = 3 + a/m$.

Finally, I want to notice that the heuristic argument above does capture the essence because the degrees in the preferential attachment model tend to stay close to their averages. Making this statement precise, however, is not easy. It requires some work and involves several advanced probabilistic techniques that have been developed only recently, starting with paper [4] that presented such rigorous analysis for the first time in 2001.

Mathematical model for triangles

Real-life networks have many triangles, as illustrated in Figure 1(d): if Anna knows Boris and Cecile, then Boris and Cecile likely know each other as well. This phenomenon is called *triangle closure*. In this section we will talk about how to capture triangle closure in a mathematical model.

Sparse Erdős–Rényi random graphs don't have many triangles

Denote by Δ_n the average number of triangles in a sparse E-R random graph, $\mathrm{ER}_n(\lambda/n)$. We will now find a formula for Δ_n to see how large this number can be. First of all, the number of possible triplets of vertices in a graph is:

$$\binom{n}{3} = \frac{n(n-1)(n-2)}{6}.$$

Next, a triplet of vertices i,j,k forms a triangle if and only if all three edges ij, jk, ki exist in the graph, as in Figure 8.

The probability of this event is

$$\left(\frac{\lambda}{n}\right)^3 = \frac{\lambda^3}{n^3}. \tag{17}$$

Then, the average number of triangles Δ_n is the fraction $\frac{\lambda^3}{n^3}$ of all $\frac{n(n-1)(n-2)}{6}$ possible triangles:

$$\Delta_n = \frac{n(n-1)(n-2)}{6}\frac{\lambda^3}{n^3} = \frac{n(n-1)(n-2)}{n^3}\frac{\lambda^3}{6} \approx \frac{\lambda^3}{6}. \tag{18}$$

For instance, if $\lambda = 9$, there are on average 121.5 triangles. Maybe, 121.5 is not a very small number, but the main implication of formula (18) is that this number does not grow when n grows. Even if the number of vertices n is very large, the average number of triangles is stuck at 121.5. In a networks of several million vertices, having only 121.5 triangles on average doesn't conform at all to the triangle closure phenomenon.

Before going any further, let us do what I always ask my students to do after answering the question: analyze the answer.

First of all, clearly, we will never see 121.5 triangles, this number is the *average*. In this context, it means the following. If we generate many graphs of size n, count the number of triangles in each of them, and then compute the average, then we will get something close to 121.5. Of course, since the edges are placed at random, the actual number of triangles will be different in every realization of $\mathrm{ER}_n(\lambda/n)$. In fact, one can prove that the number of triangles in an $\mathrm{ER}_n(\lambda/n)$ random graph converges to the Poisson distribution with parameter $\frac{\lambda^3}{6}$ when n goes to infinity.

Second, why exactly don't sparse E-R random graphs have many triangles? Which model assumptions are responsible for this result? There are in fact only two assumptions: all edges have the same probability $\frac{\lambda}{n}$, and are independent. Of course, if we increase the edge probability, then the number of triangles increases as well. Nevertheless, when the edge probability is $\frac{\lambda}{n}$, the number of triangles is bounded even for very large λ. To let the number of triangles

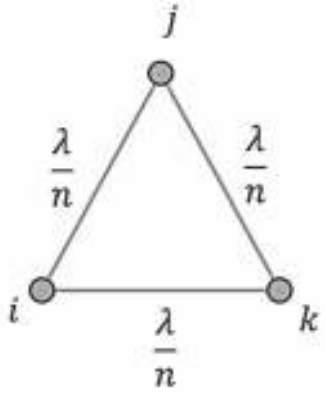


Figure 8 Triangle ijk in $\mathrm{ER}_n(\lambda/n)$. The probability of each edge is λ/n.

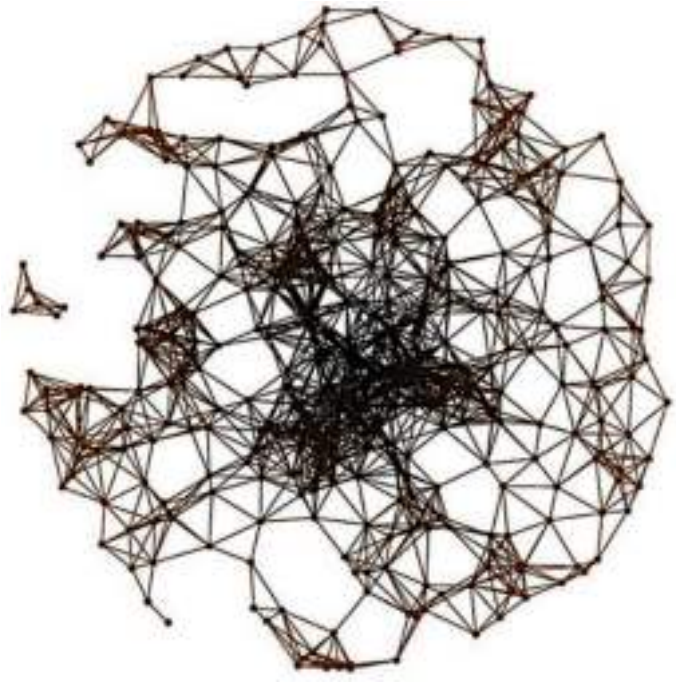


Illustration: Pim van der Hoorn

Figure 9 A geometric random graph.

grow with network size, we need a different parametrization. For instance, if $p = p(n) = \frac{\lambda}{\sqrt{n}}$, then $\Delta_n \approx \frac{\lambda^3\sqrt{n^3}}{6}$ (I leave it to the reader to verify this themselves). But then, the average degree $\approx \frac{\lambda n^{1/2}}{2}$, and the graph isn't sparse anymore. In a sparse graph, the main culprit that fails the triangle closure is the *independence* of edges. Indeed, triangle closure tell us that existence of two edges makes the third edge more likely, and this is simply not the case in the E-R model. To solve this, we could explicitly introduce complicated dependencies between edges. While such models exist, here I will explain a more elegant and now common approach: introducing geometry.

Geometric random graphs

A fundamental way to create triangle closure is to place vertices in a multi-dimensional space, and let vertices connect when they are close to each other. Sometimes such geometry is already there, for example, in the network of airports connected by direct flights, each airport has a location, and there are many quite short (so, local) direct flights. On a slightly more abstract level, a 'location' can be defined through the properties of the vertex, for example, we may 'locate' people in a multi-dimensional space depending on their age and interests. Then, in a social network, similar people are more likely to be friends. This idea is very natural and appealing. In fact, many network scientists believe that only geometric random graphs can give us realistic models for complex networks.

Geometric networks naturally have many triangles, as we can see in the example in Figure 9. This is easy to explain intuitively: if vertices j and k are connected to vertex i, then j and k are likely to be close to i in the space, and thus they are likely to be close to each other as well, so edge jk has a high chance to appear.

In this article we consider the simplest sparse geometric random graphs and prove that they indeed have many triangles. In this simple model, we place n vertices in a unit square (with side 1) uniformly at random, and we connect two vertices if the distance between them is at most r. We call this model $\mathrm{Geom}_n(r)$. Figure 10(a) shows an example of such model with $n = 11$.

To start, let us compute the probability of an edge. Edge ij appears if vertex j happens to be in a circle of radius r around vertex i. Since j is placed randomly on the area 1, the probability of such edge is:

$$P(j \text{ is in the circle of radius } r \text{ from } i) = \frac{\pi r^2}{1} = \pi r^2. \qquad (19)$$

Of course, we assume that the area of a circle, πr^2, is smaller than 1, otherwise most vertices are connected, and the model is not very interesting.

Looking critically at formula (19), a sharp reader may notice that the formula does not work if i is closer than r to the boundary. This does not change the essence of the results but calculations become much messier when we have to take the boundary into account. This is why geometric models are often studied not on a square, but on a *torus*, where boundaries are merged together. I recommend a nice animation of how a square becomes a torus [14]. In words, on a torus, the boundaries are merged, and therefore vertices close to the boundary become close to each other. Figure 10(b) shows that in our small example, replacing a square by a torus results in one more, 'over the boundary' edge kl. We stick to the torus in this article so that all vertices are symmetric, and the probability (19) is correct for every edge ij regardless the location of i.

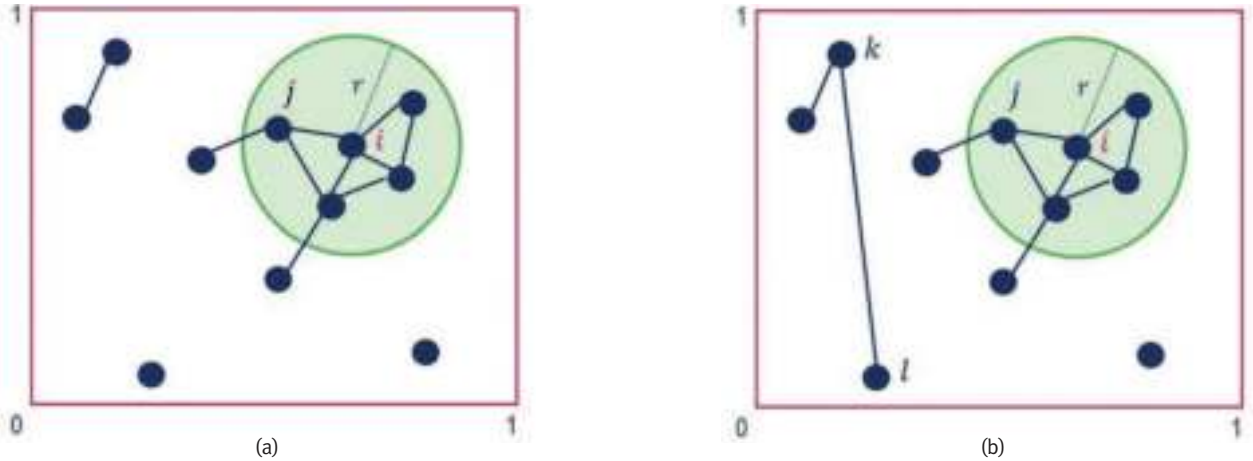


Figure 10 (a) A $\mathrm{Geom}_{11}(r)$ random graph on a unit square. Vertices within distance r from each other form an edge. (b) A $\mathrm{Geom}_{11}(r)$ random graph on a torus. The distance between k and l 'over the boundary' is less than r, therefore, on a torus, there is an edge kl.

Sparse geometric random graphs

To make the geometric random graph sparse, we again will use parametrization. As in the E-R model, every vertex has $n-1$ potential neighbors, and we have already computed the probability of connection in (19). Then the average degree of a vertex is

$$[\text{the average degree in } \mathrm{Geom}_n(r)] = (n-1)\cdot \pi r^2. \qquad (20)$$

Recall that the graph is sparse when its average degree doesn't grow with n. We can achieve this with parametrization

$$r = \frac{\mu}{\sqrt{n}}, \quad \text{where } \mu > 0.$$

Then the average degree in (20) becomes

$$(n-1)\pi\left(\frac{\mu}{\sqrt{n}}\right)^2 < \pi\mu^2.$$

We conclude that $\mathrm{Geom}_n(\mu/\sqrt{n})$ graph is sparse. And now it remains to prove that this sparse graph has many triangles.

Many triangles in sparse geometric random graphs

Calculations with circles are a little bit messy, that's why we will derive a lower bound for the average number of triangles using squares instead of circles. A lower bound is sufficient for our purposes because if the lower bound is large, then the actual number of triangles is even larger.

To get our lower bound, we assume that vertex i can be at any location, and we draw a square with diagonal r (and side $\frac{r}{\sqrt{2}}$) with i in a center, as in Figure 11. All vertices in this square are connected to i, but also to each other because they are at the distance less than r. The probability that j is in the square with center at i, is

$$\left(\frac{r}{\sqrt{2}}\right)^2,$$

and same holds for k. Since locations of j and k are independent from each other, the probability that both j and k are in the square is

$$\mathbb{P}(j \text{ and } k \text{ are in the square with side } \frac{r}{\sqrt{2}} \text{ and center at } i)$$
$$= \left(\frac{r}{\sqrt{2}}\right)^2\left(\frac{r}{\sqrt{2}}\right)^2 = \frac{r^4}{4}. \qquad (21)$$

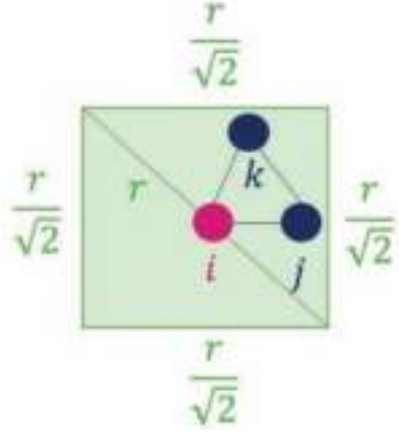


Figure 11 Vertex i is in the center of a square with diagonal r. All vertices inside this square are connected to each other.

In a sparse geometric graph, with $r = \frac{\mu}{\sqrt{n}}$, this probability is

$$\frac{\mu^4}{4n^2}.$$

The actual probability of triangle ijk is larger than that because there are many other configurations for such triangle to appear. All three vertices being in one square is only one of the possibilities.

Now recall that in any graph of size n there are $\frac{n(n-1)(n-2)}{6}$ possible triangles. Then, on average, the number of triangles in a sparse geometric graph is at least

$$\frac{n(n-1)(n-2)}{6}\cdot\frac{\mu^4}{4n^2} \approx \frac{\mu^4}{24}n.$$

The actual average number of triangles Δ_n is even larger because we computed a lower bound. However, the order of magnitude is correct because we could derive Δ_n similarly: fixing the location of i and computing the areas in which j and k should be in order to form a triangle. These areas are larger than the area of our square, but they all are proportional to r^2.

We conclude that Δ_n in $\mathrm{Geom}_n(\mu/\sqrt{n})$ grows linearly in n. This matches our intuition about, say, social networks: every new person in a network makes several friends and by that also participates in several triangles.

Looking at the formulas we can exactly pinpoint the place where sparse $\mathrm{ER}_n(\lambda/n)$ random graphs differ from the sparse $\mathrm{Geom}_n(\mu/\sqrt{n})$ random graphs. In $\mathrm{ER}_n(\lambda/n)$, in formula (17) we multiplied the probabilities of all three edges because the edges are independent. But in $\mathrm{Geom}_n(\mu/\sqrt{n})$, in formula (21), we multiplied probabilities only of two edges ij and ik, while edge jk appeared automatically, because j and k are so close to i, that they are automatically close to each other. This is exactly the mathematical representation of triangle closure.

What we learned and what's next

The three models and the four properties

Figure 12 schematically shows the four properties of real-life networks and how we can model them with random graphs. The Erdős–Rényi random graph is the simplest model where all edges have the same probability and are independent. This is already

Figure 12 Mathematical models for the properties of real-life networks. Green check mark: the property holds, and we discussed it in the article. Red stop sign: the property does not hold and we discussed it in this article. Gray check mark: the property holds, but we did not discuss it in th article. Gray stop sign: the property does not hold, but we did not discuss it in the article.

sufficient to model *sparse* networks because sparseness concerns only the average number of edges per vertex. And if we choose the probability of an edge above the critical threshold, then the random graph is also *connected*. This model however is very homogeneous: all vertices and edges have exactly the same properties, and the differences between them are only due to random fluctuations. Turns out, this is insufficient to create a *scale-free* network. Also, there are not many *triangles* because the model doesn't have a mechanism to enforce triangle closure.

In the Preferential Attachment model vertices appear one by one. Since every vertex comes with m connections, the Preferential Attachment random graph is *sparse*. It is also *connected* because new vertices attach to the existing connected graph. But, importantly, the vertices are not homogeneous: older vertices have higher chance to gain high degrees, and once they have a high degree, they are more likely to get new connections. This rich-get-richer mechanism results in the *scale-free* property, that is, huge differences between degrees of vertices. We didn't discuss the number of triangles in the Preferential Attachment model. The stop sign in gray indicates that the number of triangles is small because, similarly to the Erdős–Rényi random graph, the Preferential Attachment model does not stir towards triangle closure.

Finally, Geometric random graphs facilitate triangle closure by connecting vertices that are close to each other. Again, I included in gray the properties that we did not address in this article. If we make r large enough, geometric random graph is connected, this is indicated by the gray check mark. Further, sparse Geometric random graphs as in this article, are not scale free, this is indicated by the gray stop sign. Similarly to the Erdős–Rényi model, there is no heterogeneity between the vertices. When the graph is large, the degree of a vertex follows the $Poisson$ distribution with average degree $\pi\mu^2$.

Figure 12 may leave you with impression that it is not possible to model all four properties together. However, this is only because we considered very simple models. Recent state-of-the-art models are richer and closer to reality. For example, a very influential model that combined the scale-free property and triangles was introduced in 2010 [10]. It is in fact a geometric model, but the vertices are placed not in an Euclidean space, but in a hyperbolic space. The hyperbolic space is curved, so the size of a 'circle' around a vertex depends on the vertex's position, thus, there is high heterogeneity between vertices. Since then, many models appeared that include geometry and heterogeneity of vertices at the same time. There are also Preferential Attachment models that are either geometric or explicitly include high probability of triangle closure.

Other properties of real-life networks

There are many other very interesting and common structural properties of real-life networks that I did not mention in this article. For example, we often see *communities*. In social networks, communities can be defined by interests, language, or geography. Another famous property of real-life networks is the 'small world phenomenon': most pairs of vertices are connected by a short path of edges. In social networks, this phenomenon is also known as 'six degrees of separation' stating that "everybody on this planet is separated only by six other people" (John Guare). The communities and the small world phenomenon, of course, too, have been studied using random graphs. The research on random graphs and complex networks is happening right now, and as a mathematician I am very excited to be a part of this collective scientific effort.

I hope that this article gave you insight into mathematical tools we use to think about complex networks, and left you with motivation and curiosity to explore the endless opportunities of this quickly developing branch of modern mathematics.

References

1 Noga Alon and Joel H Spencer, *The Probabilistic Method*, Wiley, 2016.

2 Albert-László Barabasi and Reka Albert, Emergence of scaling in random networks, *Science* 286(5439) (1999), 509–512.

3 Paolo Boldi and Sebastiano Vigna, The WebGraph framework I: Compression techniques, in *Proc. of the Thirteenth International World Wide Web Conference (WWW 2004)*, *Manhattan*, ACM Press, 2004, pp. 595–601.

4 Béla Bollobás, Oliver Riordan, Joel Spencer and Gábor Tusnády, The degree sequence of a scale-free random graph process, *Random Structures & Algorithms* 18(3) (2001), 279–290.

5 Anna D. Broido and Aaron Clauset, Scale-free networks are rare, *Nature communications* 10(1) (2019), 1017.

6 Derek J. de Solla Price, The science of science, *Bulletin of the Atomic Scientists* 21(8) (1965), 2–8.

7 Paul Erdős and Alfréd Rényi, On random graphs. I, *Publ. Math. Debrecen* 6 (1959), 290–297.

8 Michalis Faloutsos, Petros Faloutsos and Christos Faloutsos, On power-law relationships of the internet topology, *ACM SIGCOMM Computer Communication Review* 29(4) (1999), 251–262.

9 Remco van der Hofstad, *Random Graphs and Complex Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 43, Cambridge University Press, 2016.

10 Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat and Marián Boguná, Hyperbolic geometry of complex networks, *Physical Review E* 82(3) (2010), 036106.

11 M. ten Thij, T. Ouboter, D. Worm, N. Litvak, H. van den Berg and S. Bhulai, Modelling of trends in twitter using retweet graph dynamics, in *Algorithms and Models for the Web Graph: 11th International Workshop, WAW 2014, Beijing, China, December 17-18, 2014, Proceedings*, Springer, 2014, pp. 132–147.

12 Ivan Voitalov, Pim van der Hoorn, Remco van der Hofstad and Dmitri Krioukov. Scale-free networks well done, *Physical Review Research* 1(3) (2019), 033034.

13 George Udny Yule, A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis F.R.S., *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 213(402–410) (1925), 21–87.

14 Folding a torus, https://youtu.be/nkszzWYCs8c. Date accessed: 4 February 2023.

15 Hoe machtig is het superknooppunt? https://www.nrc.nl/nieuws/2019/12/20/hoe-machtig-is-het-superknooppunt-a3984586, 20 December 2019. Date accessed: 26 January 2023.

16 The network pages – the math and algorithms that keep us connected, https://www.networkpages.nl/CustomMedia/Animations/RandomGraph/ERRG/AddoneEdgepATime.html. Date accessed: 24 January 2023.

Alexander Bentkamp
Mathematisches Institut
Heinrich-Heine-Universität Düsseldorf
alexander.bentkamp@hhu.de

Jasmin Blanchette
Institut für Informatik
Ludwig-Maximilians-Universität München
jasmin.blanchette@ifi.lmu.de

Research

Finding mathematical proofs using computers

In the 1960s, some researchers believed that computers would one day replace mathematicians: Computers would autonomously suggest conjectures and prove them automatically. Despite recent progress in artificial intelligence, this vision has not yet materialized. Even if they are assisted by computers (via, e.g., computer algebra systems), mathematicians remain needed for doing mathematics. Even so, there has been substantial progress in the area of automated reasoning in the past 60 years. In this article, Alexander Bentkamp and Jasmin Blanchette describe the development of automatic proof methods and automatic theorem provers based on those methods. Broadly understood, the area also encompasses interactive theorem provers, which provide a graphical user interface for developing formal proofs.

In combination, automatic and interactive theorem provers help users develop rigorous, computer-checked proofs. The easiest parts are carried out automatically, and the more difficult parts require user intervention. For example, the following lemma, which originates from a mathematics paper, can be proved by automatic theorem provers:

$$\begin{aligned} & \gcd(a,b) = 1 \\ \wedge\ & d \mid ab \\ \wedge\ & a' \mid \gcd(d,a) \\ \wedge\ & b' \mid \gcd(d,b) \\ \Rightarrow\ & a'b' \mid d \wedge d \mid a'b' \end{aligned}$$

The symbols $\wedge$ and $\Rightarrow$ mean 'and' and 'implies', $\gcd$ is the greatest common divisor, and $\mid$ is the 'divides' relation. Such a lemma is easy for mathematicians, so it is a relief that it can be proved automatically; the alternative, a fully rigorous, interactive proof, could easily take 15 to 30 minutes.

In this article, we will review two leading proof methods implemented in automatic theorem provers: *resolution* [5] and *superposition* [2]. Although they are not based on machine learning, as descendants of the pioneering Logic Theorist system these methods constitute a form of artificial intelligence. Resolution can be used to prove problems formulated in predicate logic (also called first-order logic). Superposition is a further development of resolution with special support for the equality symbol ($=$), which is ubiquitous in mathematics. One method that we will not cover but that is also very successful is satisfiability modulo theories [3].

Automatic theorem provers based on resolution and superposition can be used on their own, but often they are invoked as backends from the comfort of an interactive theorem prover. This relies on bridges between automatic and interactive provers.

This article is based on an eponymous presentation by Blanchette at the KWG Wintersymposium 2023 on 14 January 2023 in Utrecht. That presentation in turn drew from a paper written by the present authors together with colleagues and recently published in the *Communications of the ACM*.

Our logical formalism: Predicate logic

There are hundreds, perhaps thousands of logics, but when people simply say 'logic', they often mean predicate logic. Predicate logic is parameterized by a *signature*, which consists of *function symbols* (e.g., 1, ..., gcd) and *predicate symbols* (e.g., $=$, $\mid$, prime). Thus, strictly speaking, predicate logic is a family of logics indexed by a signature. The signature also associates an *arity* with each symbol: the number of arguments the symbol may take.

Based on a fixed signature, the *terms* are defined by these two rules:

- A variable x is a term.
- If f is a function symbol of arity n and $t_1, \ldots, t_n$ are terms, then $\mathrm{f}(t_1, \ldots, t_n)$ is a term. As a special case, if $n = 0$, we write f instead of $\mathrm{f}()$ and we call f a constant.

Finally, the *formulas* are defined by these rules:

- If p is a predicate symbol of arity n and $t_1, \ldots, t_n$ are terms, then $\mathrm{p}(t_1, \ldots, t_n)$ is a formula. As a special case, if $n = 0$, we write p instead of $\mathrm{p}()$.
- If F and G are formulas and x is a variable, then $\bot$ (falsity), $\top$ (truth), $\neg F$ (negation), $F \wedge G$ (conjunction), $F \vee G$ (disjunction), $F \Rightarrow G$ (implication), $\forall x.\, F$ (universal quantification), and $\exists x.\, F$ (existential quantification) are formulas.

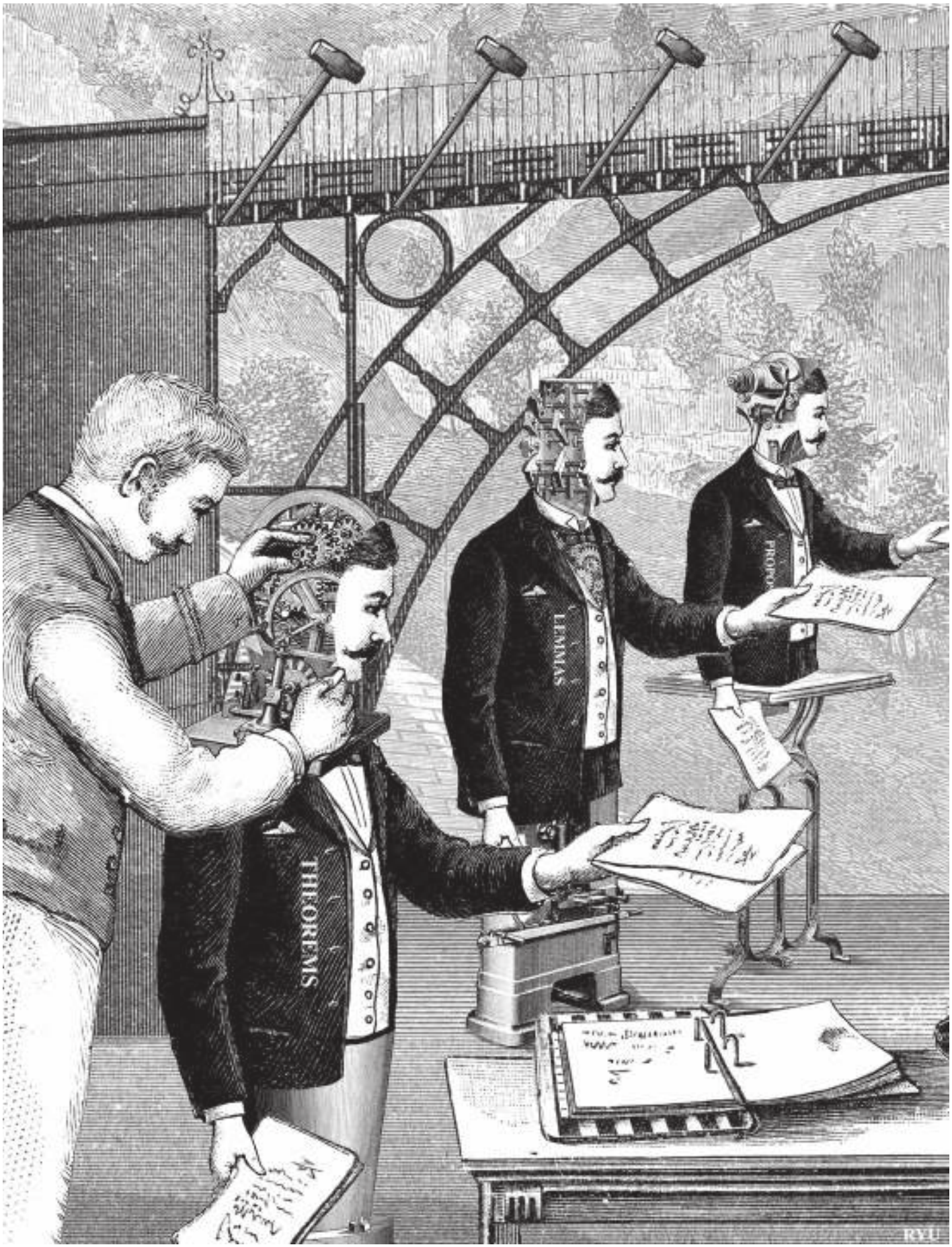


Illustration: Ryu Tajiri

Terms represent mathematical objects, whereas formulas represent mathematical statements. An example of a complex formula follows:

$$\begin{aligned} & (\forall x.\, \mathrm{food}(x) \Rightarrow \mathrm{likes}(\mathrm{johanna}, x)) \\ \wedge\ & (\forall x.\, (\exists y.\, \mathrm{eats}(y,x) \wedge \neg \mathrm{wasKilledBy}(y,x)) \Rightarrow \mathrm{food}(x)) \\ \wedge\ & \mathrm{eats}(\mathrm{bill}, \mathrm{peanuts}) \\ \wedge\ & \mathrm{alive}(\mathrm{bill}) \\ \wedge\ & (\forall y.\, \mathrm{alive}(y) \Rightarrow \forall x.\, \neg \mathrm{wasKilledBy}(y,x)) \\ \Rightarrow\ & \mathrm{likes}(\mathrm{johanna}, \mathrm{peanuts}) \end{aligned}$$

It can be a useful exercise to try to identify the function symbols and the predicate symbols above.

A general framework: Proof by saturation

Resolution and superposition are instances of a general framework called 'proof by saturation', which is a form of proof by refutation. It works as follows. Given a problem $F_1 \wedge \cdots \wedge F_n \Rightarrow G$, perform these steps:

1. Put $F_1, \ldots, F_n$ and the negation of G in clausal form, yielding a set $\mathcal{F}$ of clauses (i.e., of formulas in clausal form).
2. Repeat forever:
   2.1. Apply an inference rule to $\mathcal{F}$ and add the conclusion to $\mathcal{F}$. Stop if no new inference is possible.
   2.2. Possibly remove redundant clauses. For example, $\mathrm{p} \vee \mathrm{q}$ is redundant in the presence of the clause p.
   2.3. Stop if $\bot$ (falsity) is in $\mathcal{F}$.

Both resolution and superposition work with clauses. These are made of *literals*, which are themselves made of *atoms*:

- If p is a predicate symbol and $t_1, \ldots, t_n$ are terms, then $\mathrm{p}(t_1, \ldots, t_n)$ is an atom.
- If A is an atom, then A and $\neg A$ are literals.

The clauses are then defined by this rule:

- If $L_1, \ldots, L_n$ are literals, then $L_1 \vee \cdots \vee L_n$ is a clause. Clauses are considered equal up to associativity and commutativity of $\vee$. If $n = 0$, we write $\bot$ (which is appropriate since $\bot$ is the neutral element for $\vee$).

For example, by commutativity, $\mathrm{p} \vee \mathrm{q}(x)$ is the same clause as $\mathrm{q}(x) \vee \mathrm{p}$. It may help to think of clauses as multisets of literals. In keeping with this view, we call $\bot$ the empty clause.

Note that the logical symbols $\top$, $\wedge$, $\Rightarrow$, $\forall$, and $\exists$ cannot occur in a clause. Any variable occurring in a clause is understood as a $\forall$ variable; for example, the clause $\mathrm{p}(x) \vee \mathrm{q}(x)$ is understood to mean the same as the formula $\forall x.\, \mathrm{p}(x) \vee \mathrm{q}(x)$.

Any problem can be transformed into a set of clauses. For example, the problem

$$\begin{aligned} & (\forall x.\, \mathrm{human}(x) \Rightarrow \mathrm{mortal}(x)) \\ \wedge\ & \mathrm{human}(\mathrm{anne}) \\ \Rightarrow\ & \mathrm{mortal}(\mathrm{anne}) \end{aligned}$$

can be translated to the following three clauses:

$$\begin{aligned} & \neg \mathrm{human}(x) \vee \mathrm{mortal}(x) \\ & \mathrm{human}(\mathrm{anne}) \\ & \neg \mathrm{mortal}(\mathrm{anne}) \end{aligned}$$

The original problem is clearly provable. Correspondingly, its translation to clauses is inconsistent, meaning that it admits a proof by refutation, as we will see below.

It may seem surprising that even formulas containing $\exists$ can be converted to the clausal format. This is possible thanks to a technique called *Skolemization* [1], whereby the $\exists$ variables are replaced by new symbols representing unknown witnesses. For example, $\forall x.\, \exists y.\, \mathrm{p}(x,y)$ is translated to $\mathrm{p}(x, \mathrm{wit}(x))$, where $\mathrm{wit}(x)$ represents the witness associated with x ("for every x, there exists a witness $\mathrm{wit}(x)$ such that $\mathrm{p}(x, \mathrm{wit}(x))$").

A first proof method: Resolution

Resolution is an instance of the saturation framework that consists of one main inference rule (and of a side rule, which we will not cover). Ignoring for a moment that clauses may contain variables, we can state the rule as follows:

> If the clauses $C \vee A$ and $\neg A \vee D$ are contained in $\mathcal{F}$, then add the clause $C \vee D$ to $\mathcal{F}$.

Here, C and D stand for clauses, and A stands for an atom. Because clauses are defined up to associativity and commutativity of $\vee$, the literal A may actually occur anywhere in $C \vee A$, and similarly for $\neg A$ in $\neg A \vee D$. Thus, $C \vee A$ denotes the clause that contains the literal A and all the literals from C, whereas $\neg A \vee D$ denotes the clause that contains the literal $\neg A$ and all the literals from D.

Suppose $\mathcal{F}$ consists of the following clauses:

$$\begin{aligned} & \neg \mathrm{human}(\mathrm{anne}) \vee \mathrm{mortal}(\mathrm{anne}) \\ & \mathrm{human}(\mathrm{anne}) \\ & \neg \mathrm{mortal}(\mathrm{anne}) \end{aligned}$$

This is the same set as above, except that we instantiated x with anne to avoid variables. A first inference is possible involving the first and second clauses, taking C to be $\bot$ (the empty clause), A to be $\mathrm{human}(\mathrm{anne})$, and D to be $\mathrm{mortal}(\mathrm{anne})$. The conclusion, $C \vee D$, consists of the literals of $\bot$ and those of $\mathrm{mortal}(\mathrm{anne})$. Thus, the conclusion is

$$\mathrm{mortal}(\mathrm{anne})$$

and we add it to $\mathcal{F}$.

Next, we can perform a second inference involving the newly added clause with the clause $\neg \mathrm{mortal}(\mathrm{anne})$. This time, C and D are both $\bot$, and the conclusion to add to $\mathcal{F}$ is $\bot$. Once the empty clause is added to $\mathcal{F}$, the saturation loop stops and the contradiction is reported. The proof by refutation is successful.

The resolution inference would appear to be working correctly on this example, but is it correct in general? And is it complete, meaning: Will saturation always find a contradiction from an inconsistent clause set? The answer to both questions is yes.

We start with correctness. We will assume the two premises $C \vee A$ and $\neg A \vee D$ are true and show that $C \vee D$ is true. We proceed by case distinction on A:

- If A is true, then $\neg A$ is false, and $\neg A \vee D$ can be true only if D is true. If D is true, then the conclusion $C \vee D$ is true as well, as desired.
- If A is false, then $C \vee A$ can be true only if C is true. If C is true, then the conclusion $C \vee D$ is true as well, as desired.

Resolution (as presented in the literature) is also complete in the following sense: If the clause set $\mathcal{F}$ is initially inconsistent and

inferences are performed fairly, then $\mathcal{F}$ will eventually contain $\bot$. The fairness requirement is vital; without it, a necessary inference might be delayed forever, preventing the derivation of $\bot$.

So far, we have ignored the difficulties arising from the presence of variables in clauses, but resolution can actually cope with variables. Suppose we have the two clauses $C(y) \vee \mathrm{p}(\mathrm{a},y)$ and $\neg\mathrm{p}(x,\mathrm{b}) \vee D(x)$, where $C(y)$ denotes a clause that depends on the variable y and $D(x)$ a clause that depends on x. The atoms $\mathrm{p}(\mathrm{a},y)$ and $\mathrm{p}(x,\mathrm{b})$ are not syntactically identical, but they can be made identical by taking x to be a and y to be b, following the German saying "was nicht passt, wird passend gemacht". This process of instantiating variables to make atoms identical is called *unification* [5]. Once we unify the two atoms, we get the two clauses

$$C(\mathrm{b}) \vee \mathrm{p}(\mathrm{a},\mathrm{b}) \qquad \neg\mathrm{p}(\mathrm{a},\mathrm{b}) \vee D(\mathrm{a})$$

From these, a resolution inference derives $C(\mathrm{b}) \vee D(\mathrm{a})$.

We can now consider the 'Anne is mortal' example in its full generality:

$$\neg\mathrm{human}(x) \vee \mathrm{mortal}(x)$$
$$\mathrm{human}(\mathrm{anne})$$
$$\neg\mathrm{mortal}(\mathrm{anne})$$

From $\mathrm{human}(\mathrm{anne})$ and $\neg\mathrm{human}(x) \vee \mathrm{mortal}(x)$, we instantiate x with anne and compute the conclusion $\mathrm{mortal}(\mathrm{anne})$. From this conclusion and $\neg\mathrm{mortal}(\mathrm{anne})$, we derive $\bot$.

Although resolution is complete, it is not always efficient. Consider the inconsistent set $\mathcal{F}$ consisting of the six clauses

$$\mathrm{a} \vee \mathrm{b} \vee \mathrm{c} \vee \mathrm{d} \vee \mathrm{e} \qquad \neg\mathrm{a} \qquad \neg\mathrm{b} \qquad \neg\mathrm{c} \qquad \neg\mathrm{d} \qquad \neg\mathrm{e}$$

A *particularly inefficient strategy* would first derive all four-literal clauses that can be derived from this set. There are five of them:

$$\mathrm{b} \vee \mathrm{c} \vee \mathrm{d} \vee \mathrm{e} \qquad \mathrm{a} \vee \mathrm{c} \vee \mathrm{d} \vee \mathrm{e} \qquad \mathrm{a} \vee \mathrm{b} \vee \mathrm{d} \vee \mathrm{e}$$
$$\mathrm{a} \vee \mathrm{b} \vee \mathrm{c} \vee \mathrm{e} \qquad \mathrm{a} \vee \mathrm{b} \vee \mathrm{c} \vee \mathrm{d}$$

Then it would derive the three-literal clauses:

$$\mathrm{c} \vee \mathrm{d} \vee \mathrm{e} \qquad \mathrm{b} \vee \mathrm{d} \vee \mathrm{e} \qquad \mathrm{b} \vee \mathrm{c} \vee \mathrm{e} \qquad \mathrm{b} \vee \mathrm{c} \vee \mathrm{d} \qquad \mathrm{a} \vee \mathrm{d} \vee \mathrm{e}$$
$$\mathrm{a} \vee \mathrm{c} \vee \mathrm{e} \qquad \mathrm{a} \vee \mathrm{c} \vee \mathrm{d} \qquad \mathrm{a} \vee \mathrm{b} \vee \mathrm{e} \qquad \mathrm{a} \vee \mathrm{b} \vee \mathrm{d} \qquad \mathrm{a} \vee \mathrm{b} \vee \mathrm{c}$$

Then the one-literal clauses:

$$\mathrm{e} \qquad \mathrm{d} \qquad \mathrm{c} \qquad \mathrm{b} \qquad \mathrm{a}$$

Finally, from any one-literal clause (e.g., a) and its negation (e.g., $\neg\mathrm{a}$), it would derive $\bot$.

A *more efficient strategy* would derive, from the initial set $\mathcal{F}$, a single four-literal clause (e.g., $\mathrm{a} \vee \mathrm{b} \vee \mathrm{c} \vee \mathrm{d}$), then a single three-literal clause (e.g., $\mathrm{a} \vee \mathrm{b} \vee \mathrm{c}$), then a single two-literal clause (e.g., $\mathrm{a} \vee \mathrm{b}$), then a single one-literal clause (e.g., a), and finally $\bot$. Such a strategy can be programmed in the automatic provers, but we can do better: We can enforce it in the proof method itself by introducing an order.

Specifically, *ordered resolution* is a variant of resolution that is parameterized by an order on literals. For our example, we will simply take the alphabetical order, requiring $\mathrm{e} > \mathrm{d} > \mathrm{c} > \mathrm{b} > \mathrm{a}$. Ignoring that clauses may contain variables, we can state the main inference rule of ordered resolution as follows:

> If the clauses $C \vee A$ and $\neg A \vee D$ are contained in $\mathcal{F}$, A is maximal in $C \vee A$, and $\neg A$ is maximal in $\neg A \vee D$, then add the clause $C \vee D$ to $\mathcal{F}$.

This rule is less explosive than the standard resolution rule. It is also complete; for example, the 'more efficient strategy' above would be allowed, whereas the 'particularly inefficient strategy' would be disallowed. Intuitively, ordered resolution focuses on the largest literals first and tries to eliminate them by performing inferences. If they cannot be eliminated, the other literals are never considered.

The situation is analogous to a scenario often encountered by mathematicians. Suppose we want to use a lemma of the form "If F_1, F_2, and F_3, then G" in a proof, we can without loss of generality focus on F_1 and try to prove it first, then proceed with F_2, then with F_3, and finally retrieve the conclusion G. If we fail at proving the condition F_1 (i.e., at 'eliminating' F_1), we can immediately give up; there is no point in trying to prove F_2 and F_3.

How to deal with equality

Equality is ubiquitous in mathematical problems. With resolution, to reason about equality we need to specify its properties as axioms to be included in the problem:

reflexivity:	$\forall x.\ x = x$
symmetry:	$\forall x,y.\ x = y \Rightarrow x = z$
transitivity:	$\forall x,y,z.\ (x = y \wedge y = z) \Rightarrow x = z$
congruence:	$\forall x,y.\ x = y \Rightarrow \mathrm{f}(x) = \mathrm{f}(y)$

Congruence is shown for a single unary function symbol f, but it needs to be repeated for all function and predicate symbols.

An alternative, which is the approach taken by superposition, is to treat equality specially in the proof method. Reflexivity, symmetry, transitivity, and congruence are then not axiomatized. Moreover, once equality is available, we can eliminate all other predicates. Specifically, for any predicate symbol p other than equality, we can use a function symbol f instead that returns a truth value. A literal $\mathrm{p}(x)$ can be coded as $\mathrm{f}(x) = \mathrm{true}$, and a literal $\neg\mathrm{p}(x)$ can be coded as $\neg(\mathrm{f}(x) = \mathrm{true})$. Literals then have the form $t = t'$ and $\neg(t = t')$, and we write $t \neq t'$ to abbreviate $\neg(t = t')$. Finally, we consider literals equal up to commutativity of $=$; thus, $\mathrm{b} = \mathrm{a}$ and $\mathrm{a} = \mathrm{b}$ are the same literal.

A more advanced proof method: Superposition

Superposition resembles resolution, but it works on clauses whose atoms are equalities $t = t'$, and it performs inferences that embody the four characteristic properties of equality (reflexivity, symmetry, transitivity, and congruence). It consists of three rules, of which we will review two.

Ignoring variables and multiple-literal clauses, we can state the main inference rule as follows:

> If the clauses $t = t'$ and $L[t]$ are contained in $\mathcal{F}$, then add the clause $L[t']$ to $\mathcal{F}$.

Here, $L[t]$ denotes a literal that contains t as a subterm, and $L[t']$ denotes the same literal in which the singled-out occurrence of t is replaced by t'. The rule essentially allows the replacement of equals with equals.

Clause sets rarely consist exclusively of single-literal clauses, so we need to generalize the above inference rule to allow multiple literals:

> If the clauses $C \vee t = t'$ and $D \vee L[t]$ are contained in $\mathcal{F}$, then add the clause $C \vee D \vee L[t']$ to $\mathcal{F}$.

The C and D components play a similar role as in resolution.

Next to the main rule, the following side rule invariably appears in successful refutations:

If the clause $C \vee t \neq t$ is contained in $\mathcal{F}$, then add the clause C to $\mathcal{F}$.

Both rules are stated above without worrying about variables, but superposition, like resolution, unifies terms as necessary to make them syntactically equal, as we will see with an example. Informally, the problem is as follows:

Assuming that $\pi \neq 0$ and that $x^{-1} = 1/x$ for all $x \neq 0$, we have $|\pi^{-1}| = |1/\pi|$.

Expressed as a formula, the problem becomes

$$\begin{aligned} & (\forall x.\, x \neq \mathrm{zero} \Rightarrow \mathrm{inv}(x) = \mathrm{div}(\mathrm{one}, x)) \\ & \wedge\, \mathrm{pi} \neq \mathrm{zero} \\ \Rightarrow\quad & \mathrm{abs}(\mathrm{inv}(\mathrm{pi})) = \mathrm{abs}(\mathrm{div}(\mathrm{one}, \mathrm{pi})) \end{aligned}$$

Conversion into clausal form yields three clauses:

$$\begin{aligned} & x = \mathrm{zero} \vee \mathrm{div}(\mathrm{one}, x) = \mathrm{inv}(x) \\ & \mathrm{pi} \neq \mathrm{zero} \\ & \mathrm{abs}(\mathrm{div}(\mathrm{one}, \mathrm{pi})) \neq \mathrm{abs}(\mathrm{inv}(\mathrm{pi})) \end{aligned}$$

Looking at the first and third clauses, we notice that we can unify the term $\mathrm{div}(\mathrm{one}, x)$ in the first clause with the subterm $\mathrm{div}(\mathrm{one}, \mathrm{pi})$ in the third clause, by taking x to be pi. The main inference rule is applicable and adds the clause

$$\mathrm{pi} = \mathrm{zero} \vee \mathrm{abs}(\mathrm{inv}(\mathrm{pi})) \neq \mathrm{abs}(\mathrm{inv}(\mathrm{pi}))$$

to $\mathcal{F}$. At this point, the side rule applies to eliminate the second literal, resulting in

$$\mathrm{pi} = \mathrm{zero}$$

Now, we can apply the main rule on this clause and on the clause $\mathrm{pi} \neq \mathrm{zero}$, resulting in

$$\mathrm{zero} \neq \mathrm{zero}$$

Finally, an application of the side rule eliminates the literal, yielding $\bot$.

Superposition is correct and complete. Moreover, like resolution, superposition can take an order into account to restrict its search space. The main inference rule then focuses on the larger side of the largest literal of each of the two premises, trying to rewrite larger clauses into smaller clauses. That superposition is complete despite such drastic restrictions is far from obvious [2, Section 4].

Although superposition is not strictly speaking a generalization of resolution, in practice superposition provers have replaced resolution provers. Resolution is nowadays seen mostly as a stepping stone, as within this article.

Bridges between automatic and interactive provers

Automatic theorem provers, including those based on superposition, are integrated in proof assistants via bridges. These bridges are called 'hammers' [4] in honor of Sledgehammer, possibly the most successful such bridge.

Automatic provers work best when their input does not contain too many axioms. If all of the proof assistant's definitions, lemmas, theorems, and actual axioms were exported as axioms to the automatic provers, these would have to find their way among perhaps 10 000 formulas and would not perform very well. Hence, the first step of a hammer is to filter the available facts (definitions, lemmas, et cetera) to a reasonable number — typically less than a thousand.

The second step is to translate the problem. Interactive provers typically work in a richer logic (such as higher-order logic and dependent type theory) than automatic provers. There is work on reducing the gap between the two types of systems, including by the present authors, but for most combinations of interactive and automatic theorem provers a translation is necessary.

At this point, the automatic provers run on the translated problem. Sledgehammer works with a default time limit of 30 seconds. If a proof is found within that time, the last step is to import this proof into the interactive proof assistant, where it is independently rechecked.

By building on the strengths of automatic theorem provers, hammers make users more productive. One user claims he is three to five times more productive thanks to Sledgehammer. Another compared working with it to running as opposed to walking. As developers of automatic provers further improve their systems, hammers become even stronger, benefiting users of interactive provers.

Acknowledgment

We thank Mark Summerfield and Mark Timmer for suggesting several textual improvements.

References

1 M. Baaz, U. Egly and A. Leitsch, Normal form transformations, in J.A. Robinson and A. Voronkov, eds., *Handbook of Automated Reasoning* (in 2 volumes), Elsevier and MIT Press, 2001, pp. 273–333.

2 L. Bachmair and H. Ganzinger, Rewrite-based equational theorem proving with selection and simplification, *J. Log. Comput.* 4(3) (1994), 217–247.

3 C.W. Barrett, R. Sebastiani, S.A. Seshia and C. Tinelli, Satisfiability modulo theories, in A. Biere, M. Heule, H. van Maaren and T. Walsh, eds., *Handbook of Satisfiability*, Second Edition, Vol. 336 of Frontiers in Artificial Intelligence and Applications, IOS Press, 2021, pp. 1267–1329.

4 J.C. Blanchette, C. Kaliszyk, L.C. Paulson and J. Urban, Hammering towards QED, *J. Formaliz. Reason.* 9(1) (2016), 101–148.

5 J.A. Robinson, A machine-oriented logic based on the resolution principle, *J. ACM* 12(1) (1965), 23–41.

Clara Stegehuis
Faculteit EWI
Universiteit Twente
c.stegehuis@utwente.nl

Francesca Arici
Mathematisch Instituut
Universiteit Leiden
f.arici@math.leidenuniv.nl

Proof by example Portraits of women in Dutch mathematics

Fulya Kula

In 'Proof by example', Clara Stegehuis and Francesca Arici portray women in Dutch mathematics. This edition portrays Fulya Kula, lecturer and researcher on mathematics didactics at the University of Twente. In this interview she tells about her research and her motivation to pursue a career in mathematics.

How did you first become interested in mathematics?
"I actually did not really like mathematics in primary school. I found it difficult to memorize all multiplication tables for example, as I did not really understand the concept behind them. However, during high school, I had a great teacher, who could explain really well. She introduced us to theorems and proofs, and I found this challenging and rewarding. I still remember the formula of Menelaus's theorem in geometry and enjoyed working through the theorems to figure out the proof as homework.

After that, I did my BSc in mathematics, but I was also very intrigued by the way my professors were teaching, maybe because of my experience in primary school. All were very talented mathematicians, but some of them were not explaining very well, while others were. This motivated me to do my undergraduate and PhD level in the didactics of mathematics. In my PhD for example, I focused on the concept of the derivative. What prior knowledge is necessary to fully understand the concept of the derivative? And what happens when some of that knowledge is missing?"


Fulya Kula

What are you now working on?
"I am now still working in the field of mathematics and statistics didactics. I investigate how we can improve the teaching and learning of mathematical and statistical concepts. This combines my pedagogical skills and scholarly knowledge. I try to gain a better understanding into how people learn, and how this knowledge can improve teaching."

What is the most interesting thing that you have been working on recently?
"I am currently working on my project funded by 4TU.CEE that aims to make the transition from high school math to college-level math easier for students. This means that students should have a better understanding of several mathematical concepts and skills when they are at university. To achieve this, I investigate best practices in curriculum development. I will also create videos and practice material on topics that many students are struggling with. I find this project particularly exciting because it can make a real difference in students' academic lives, as I often see them struggling in the first year during my teaching."

Can you give an example of how your research shows how teaching can be improved?

"I had very interesting results on teaching statistical inference. In statistics, you often make probabilistic statements about an entire population while you only investigate at a small sample of it. This concept is often very difficult to grasp for students. Usually, during a course students are first told about the sample (for example the sample mean), and are then told what this sample statistic tells about the entire population. My research shows that it is actually better to start discussing the population first, and how you actually create a sample from this entire population. After that, you can teach what this then tells you about the entire population that we started with.

My research endorses that this second way of teaching makes students grasp the concept of statistical inference more easily. I would really like to now investigate the most common statistics textbooks to compare their way of explaining to my proposed model. Doing so will help me to slowly but surely change the way statistics is taught. This model offers a unique opportunity to teach inference not through the way it is applied, but through the way it is constructed."

Is there anything else that you would like to achieve in the future?

"I would really like to make sure that research in the didactics of mathematics is actually applied in mathematical teaching. Despite the fact that there is plenty of research that could be useful, the connection between research and practical teaching is not fully realized. My goal is to make research in the didactics of mathematics more accessible and engaging for teachers at all levels. This would allow teachers to have a quick overview of existing research, with tips on how to apply this in their practice.

I would love to create a course on didactics for mathematics teachers at universities as well. I feel that most people at the university really like their teaching, and are also interested in my didactical research, but it is difficult and time-consuming for them to get a good overview of the literature and existing knowledge. In such a course, we could go over this together, and together discuss how we can implement this in practice. In this way, mathematics education research really would have an impact on the way mathematics is taught at the university."

What do you like the most about your job?

"I really enjoy teaching and find it very motivating. My favorite moments are when a student has an 'A-Ha' moment and gains a better understanding of a concept. This is also very rewarding for myself, as I managed to make an impact on the student by teaching them a topic that they first did not fully understand. It also shows you the beauty of mathematics: if a student understands all single, small concepts, they can understand a much bigger problem."

And what do you like the least?

"The workload is sometimes high. This means that I cannot always change my courses in the way that didactical research shows is best, because you simply do not always have the time to do so on top of your current workload. So I am now changing my courses one by one, and year by year, one step at a time."

Boekbesprekingen

| Book Reviews

Redactie: Hans Cuypers en Hans Sterk

Review Editors NAW - MF 5.092
Faculteit Wiskunde & Informatica
Technische Universiteit Eindhoven
Postbus 513
5600 MB Eindhoven

reviews@nieuwarchief.nl
www.win.tue.nl/wgreview

Tom H. Koornwinder, Jasper V. Stokman (eds.)

Encyclopedia of Special Functions: The Askey–Bateman Project

Volume II Multivariable Special Functions

Cambridge University Press, 2020
xii + 427 p., prijs £59.99
ISBN 9781107003736

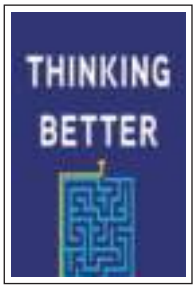
The three volume series *Higher Transcendental Functions* published in the 1950s resulted from the initiative by Harry Bateman (1882–1946) to have an extensive compilation of special functions and their properties. Bateman was a professor at California Institute of Technology, and after his death it was decided to finish his project. Arthur Erdélyi (1908–1977) was in charge of a team which eventually published the three volumes *Higher Transcendental Functions* as well as the two volume set *Tables of Integral Transforms*, collectively known as the Bateman project. Since the publication of the Bateman project enormous progress in the field of special functions has been made, and an update of the Bateman project was required. Askey gives more details on the background of the Bateman project, and the history of other handbooks (see: R. Askey, 'Handbooks of Special Functions', pp. 369–391 in *'A Century of Mathematics in America, Part III*, P. Duren, ed., AMS, 1989). Since Richard Askey (1933–2019) was very influential in the developments in special functions (see: H.S. Cohl, M.E.H. Ismail, H.-H. Wu, 'The Legacy of Dick Askey (1933–2019)', *Notices AMS*, January 2022, 59–75), it was decided to name the update the Askey–Bateman project. The book under review concerns the second volume of the Askey–Bateman project, which deals with multivariable special functions. In a sense, this topic was already mentioned by Askey in his 'Handbooks of Special Functions', pp. 382–383, as one of the important developments. In the original Bateman project multivariable special functions were essentially restricted to the study of some 2-variable orthogonal polynomials and 2-variable hypergeometric functions introduced by Appell, Kampé de Fériet and Horn.

The current volume under review has contributions by many of the experts that have made fundamental contributions to the developments of multivariable special functions over the last decades or longer. Each of these chapters gives an introduction to the topic of the chapter and next explains the main results in this topic. Some of the statements are provided with proofs, but there are references to the literature as well, and each chapter has a list of pointers to the main references. For instance, the chapter by Matsumoto on Appell and Lauricella hypergeometric functions extends the information of the Bateman project. In its 22 pages, the chapter brings the reader's knowledge up to the current state of affairs, including links to (co)homology groups. All the other chapters discuss new developments that were not foreseen in the Bateman project. Some of the chapters deal with topics which are more well established than others. For instance, the chapter by Heckman and Opdam on 'their' polynomials and hypergeometric functions is very well established and so is the chapter by Dunkl on Dunkl operators, whereas the introduction of multivariable elliptic hypergeometric series is much more recent and develops very quickly. The chapter on elliptic hypergeometric series is written by

Rosengren and Warnaar. The topics covered in the volume also include applications to other fields, such as Lie theory, combinatorics and mathematical physics.

Having so many authors working on one volume requires an effort to make the volume coherent and the chapters of constant quality. There are many cases known where these goals have not been achieved, a famous one being the classic Abromowitz–Stegun *Handbook of Mathematical Functions* (supposedly the mathematics all times bestseller), see also Askey's 'Handbooks of Special Functions', or even the first volume of this new Askey–Bateman project. The editors of the current volume, Tom Koornwinder and Jasper Stokman, have done a superb job in making this volume much more than just a collection of twelve chapters on multivariable special functions. Moreover, the introductory chapter by the editors gives an excellent overview of the book as a whole.

The twelve chapters contain a wealth of information, and for many topics the book gives a great introduction to various new and exciting developments in the theory of multivariable special functions. It makes this volume of the Askey–Bateman project an excellent point of departure for a study in multivariable special functions. It is a pleasure to read the chapters, and learn more about the fascinating developments in these topics. *Erik Koelink*



Marcus du Sautoy

Thinking Better

The Art of the Shortcut in Math and Life

Basic Books (Hachette Book Group), 2021
iiv + 326 p., prijs $30.00
ISBN 9781541600362

De bekende auteur Marcus du Sautoy (1965) is hoogleraar wiskunde in Oxford en tevens benoemd op de zogenaamde Simonyi leerstoel voor de 'Public Understanding of Science'. Een leerstoel waarvan wij er in ons land voor verschillende disciplines ook enkele hebben. We kennen hem van boeken zoals *Finding Moonshine* (in het Nederlands uitgegeven onder de titel *Het symmetrie-monster*), *The Creativity Code* (*De code van creativiteit*), en diverse andere. Maar ook van zijn veelvuldige optredens op de Engelse tv. Kortom, hij beoefent een mooie mix van zuiver wiskundig werk en werk dat gericht is op 'Public Understanding'. Het boek dat nu voor mij ligt past helemaal in deze mix.

De uitgever prijst het boek aan met de woorden: "One of the world's great mathematicians shows why math is the ultimate timesaver — and how everyone can make their lives easier with a few simple shortcuts." Nu moet je altijd uitkijken met datgene wat een uitgever aanprijst, want zijn financiële uitgeversbelang zal hij moeilijk kunnen onderdrukken. We zijn dus op onze hoede.

Het woord 'shortcut' is wellicht niet bij ieder bekend. Denk in dit verband gewoon aan een sneltoets op de computer. In feite heb je dan de essentie te pakken. Du Sautoy begint met het (wellicht apocriefe) verhaal van de negenjarige Gauss, als leerling van de basisschool. De onderwijzer liet zijn leerlingen de natuurlijke getallen van 1 tot 100 optellen. Binnen enkele minuten wist de jonge Gauss het antwoord 5050. De verbaasde onderwijzer werd geconfronteerd met Gauss' verklaring 50×101. Hier hebben we in feite een concretisering van een shortcut: een 'timesaver' die het leven gemakkelijker maakt.

Wat hier in het klein is verteld wordt op allerlei plaatsen door ons allen gepraktiseerd. Het construeren van formules is in feite het maken van shortcuts. Een waarschuwing is op z'n plaats zoals wij telkens weer merken. Kijk naar eenvoudige getalpatronen waarbij het vervolg verraderlijk kan zijn: $1, 2, 4, 8, 16$ suggereert een vervolg van 32, maar kan net zo goed 31 zijn. Toepassingen van wiskundige patronen in de sterrenkunde, morfologieën in de natuur en dergelijke kunnen daardoor tot volkomen onjuiste conclusies leiden. En hebben dat in het verleden ook gedaan.

Shortcuts in andere gebieden? Bijvoorbeeld in de muziek? Het zal duidelijk zijn dat shortcuts tot het ontwikkelen van pianovirtuositeit ondenkbaar is. Oefenen, oefenen is daarbij het recept. Maar dat is bepaald niet het hele verhaal. Het op de juiste wijze oefenen levert eveneens een vorm van shortcut op, want op een bepaald niveau gekomen ervaart de cellist, violist, pianist, voetballer zeker dat er sprake is van een ontwikkelde shortcut. Tot dusver is dit alles nog vrij triviaal, maar het verhaal gaat verder.

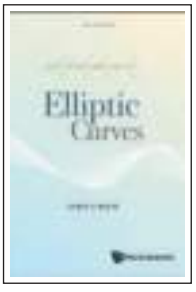
Het is gebleken dat het ontwikkelen en gebruiken van een aan problemen aangepaste taal eveneens kan werken als shortcut. Problemen worden door het veranderde taalgebruik dikwijls veel doorzichtiger. Zelfs (en vooral ook in de wiskunde) is dit op veel plaatsen duidelijk. Du Sautoy werkt dit mooi uit in zijn boek. Op allerlei verschillende terreinen verschijnen onverwachte shortcuts. Voor de hand liggend zijn bijvoorbeeld geheugentrainingen. Meetkundige en geodetische problematieken liggen al minder voor de hand, maar de schrijver weet ook deze overtuigend tot leven te wekken. Ook reisproblemen en het maken van kaarten zijn niet vrij van shortcuts. Diagrammen spelen daarbij een belangrijke rol. Economie, kunst, statistiek, waarschijnlijkheidsrekening, psychotherapie, het casino, financiële problematieken, grafen en netwerken, neurowetenschappen, overal komen shortcuts te voorschijn. Het boek begint heel elementair, maar door de grote verscheidenheid raakt het boek steeds dieper wordende lagen van shortcuts aan.

Maar het is niet allemaal shortcuts wat de klok slaat. De langdurige zoektocht naar de exacte oplossing van bijvoorbeeld het handelsreizigersprobleem heeft ons de laatste tijd met de neus op de feiten gedrukt. Namelijk op de vraag of bij ieder probleem wel een shortcut mogelijk is. En hier doemt een nieuw theoretisch probleem op: bestaan er problemen die geen shortcuts toelaten? Zou dat het geval zijn met het vierkleurenprobleem en/of met het handelsreizigersprobleem? En zou het werk van Andrew Wiles ooit met een shortcut ingekort kunnen worden? Tot nu toe zijn dit onopgeloste vragen. Zelfs weten we niet hoe we dit probleem in zijn algemeenheid zouden moeten aanpakken. Du Sautoy heeft hiermee een rijk gevarieerd boek geschreven. Lees door het triviale begin heen en zie hoe hij een grote diepte probeert te vatten in dit populair bedoelde boek.

Hij sluit het boek af met een uitvoerige index die op zichzelf al een aanwijzing vormt voor de brede visie die in dit boek wordt ontwikkeld. Kortom, een populair boek waarin onnoemelijk veel voorbeelden benoemd worden, waar wiskunde op veel plaatsen een rol speelt, maar waarin geen diepe wiskunde bedreven wordt. Maar dat was ook niet de intentie van de schrijver. De intentie ligt

vooral hierin om in brede kring door middel van een overvloed aan voorbeelden bekendheid te geven aan het wezen van een van de kernen van ons vak. Vooral voor leraren in het voortgezet onderwijs is het boek een juweel. Bij vrijwel alle onderwerpen kan men in dit boek instructieve voorbeelden vinden. Deze zullen ongetwijfeld sterk motivatieverhogend werken bij hun leerlingen. Bovendien werken de voorbeelden inzichtbevorderend. Misschien neemt de wetenschap dat er nu ook een Nederlandse vertaling van het boek is verschenen (onder de titel *Beter denken*, bij Uitgeverij Nieuwezijds) een drempel weg om het boek aan te schaffen.

Tot slot: alles overziend kan ik de eerder genoemde uitgever gelijk geven. Wat mij betreft: aanbevolen! *Wim Kleijne*



James S. Milne

Elliptic Curves

World Scientific, 2020, tweede druk
x + 308 p., prijs $85.00
ISBN 9789811221835

In de hoofdstukken 1–2–3 wordt standaardtheorie over elliptische krommen beschreven, met in hoofdstuk 3 de theorie over $\mathbb{C}$. We zien dat de affiene ruimte $\mathbb{A}^n$ over een willekeurig lichaam niet wordt gedefinieerd. Ook wordt het geslacht van een algebraïsche kromme over een willekeurig lichaam niet gegeven. Veel details worden terzijde en onvoldoende behandeld, bijvoorbeeld: "E arises from an elliptic curve over k if and only if $j(E) \in k$." Voor iemand die dan al begrepen heeft hoe je een kromme over een willekeurig lichaam definieert, is het duister wat de auteur hier bedoelt te zeggen. Iemand die de theorie kent weet dat dit voor de eerste keer door Deuring bewezen is, en dat de mededeling niet een definitie maar een stelling is. Dit is kenmerkend voor heel veel details in dit boek: je zoekt het maar uit wat de auteur bedoelt, en hoe je de tekst interpreteert.

Hoofdstuk 4, 'The arithmetic of elliptic curves' en hoofdstuk 5, 'Elliptic curves and modular formes', bevatten interessant materiaal. Bijvoorbeeld laat de auteur zien dat voor een priemgetal p de normalisatie van de projectieve kromme gegeven als nulpunten van $Y^2Z^2 + 2pX^4 - 2Z^4$ niet aan Hasse's lokaal–globaal principe voldoet (maar een van de essentiële punten wordt aan de lezer overgelaten).

Wat de 'Riemann-hypothese' voor een elliptische kromme over een eindig lichaam is, vind ik moeilijk te begrijpen uit deze tekst (plus de verwarring dat dit boek de klassieke Riemann-hypothese en de Riemann-hypothese over een eindig lichaam met dezelfde naam beschrijft, de ene keer als stelling, de andere keer als vermoeden). In Hoofdstuk 5 wordt in acht pagina's doorgenomen hoe het bewijs van Wiles van de laatste stelling van Fermat en voorlopers van dit bewijs in elkaar zitten. Gelukkig heb ik het niet hoeven te leren uit dit boek.

Conclusie: dit boek is niet geschikt als lesmateriaal voor een beginnend student. Voor wie de theorie al kent staan er interessante dingen in dit boek. *Frans Oort*

Jean-Marie De Koninck, Nicolas Doyon

The Life of Primes in 37 Episodes

American Mathematical Society, 2021
xiv + 329 p., prijs $65.00
ISBN 9781470464899

Waarom nog een boek over priemgetallen? Zoals de schrijvers zelf uitleggen, om de chronologische geschiedenis van (de wiskunde achter) de priemgetallen en gerelateerde zaken te vertellen en een overzicht te geven van de problemen waar tegenwoordige getaltheoretici zich mee bezighouden. Een waarschuwing vooraf (van de schrijvers zelf, maar van harte door mij ondersteund) is wellicht dat het niveau dat nodig is om het gehele boek te kunnen volgen dat van bij vlagen zeer geavanceerde wiskunde is, maar dat de grote lijnen goed te volgen zijn voor een — laten we zeggen — tweedejaars student wiskunde.

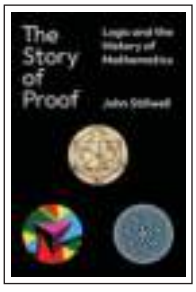
Dit — toegegeven, zeer rijke en onderhoudende — boek behandelt (ik doe slechts een kleine greep) zeer ingenieuze factorisatiemethoden, priemtesten (statistische of deterministische), priemclusters, diverse priemzeven, de vele verschillende (bekende en minder bekende) types priemgetallen (onder andere de tot nu toe enige bekende, dus zeer zeldzame *Wilson-priemgetallen* 5, 13 en 563 die voldoen aan $(p-1)! = -1 \pmod{p^2}$, de *narcistische priemgetallen* (vrij schaars, bijvoorbeeld het 14-cijferige $28116440335967 = 2^{14} + 8^{14} + \cdots + 7^{14}$), de *Wieferich-priemgetallen*, de *Carmichael-priemgetallen*), de zogenaamde *prime gaps* (de soms kleine, soms verrassend grote ruimte tussen twee opeenvolgende priemgetallen), priem-tel-functies (die bijvoorbeeld tellen hoeveel priemgetallen of priemparen er onder een zeker getal zijn), de tientallen spectaculaire functies waarmee je de snelheid kan bepalen van factorisatie-algoritmes, de voor onze toekomst belangrijke cryptografie (van Caesar tot RSA) en nog veel meer (inmiddels bewezen of nog onbewezen; een aardige opmerking in dit verband is dat er nog zoveel open vragen zijn — alleen al op dit terrein van de getaltheorie — dat er nog voor honderden jaren werk zal zijn voor de wiskundigen). Maar, veel van deze zaken zijn wellicht ook in andere boeken te vinden, dus wat is de meerwaarde van dit boek?

Heel handig is bijvoorbeeld een zeer compacte tijdlijn (zes met zo'n honderd jaartallen met baanbrekende resultaten op het gebied van priemgetallen en getaltheorie volgepakte pagina's) die begint in circa 300 voor Christus (*Eratosthenes starts the show*) en gaat tot (uiteraard) het heden (het welbekende grote belang benadrukkend van priemgetallen voor veilige encryptie, maar ook gerelateerd aan wat eventuele quantumcomputers nog mogelijk gaan maken). Verder voor de geïnteresseerde vele meestal kleine maar effectieve voorbeelden van in Mathematica geschreven software. De tientallen door het hele boek heen verspreide biografieën van sleutelfiguren en de — apart vermelde — daarvan bekende (of minder bekende) anekdotes geven het boek een prettig menselijk tintje, een aspect dat ook volgens de schrijvers vaak ontbreekt in veel wiskundeboeken. Aan het einde een handige appendix met een overzicht van de belangrijkste resultaten in de getaltheorie en een wel zeer uitgebreide (bijna 240 boeken of artikelen) bibliografie voor verdere studie. De opbouw in 37 korte (van drie tot

zes pagina's) zogenaamde episoden leest prettig. Verder bevat het boek honderden opgaven (van eenvoudig tot extreem moeilijk).

Dat laatste gegeven brengt mij tot mijn enige echte bezwaar. Namelijk dat slechts van een frustrerend klein (mijns inziens te klein) deel van de — soms zeer intrigerende en/of zeer pittige — opgaven de volledige uitwerking wordt gegeven en dat van een ander — ook klein — deel slechts een hint of soms zelfs alleen het antwoord. Deze zijn ook niet te vinden op een website om — zeer begrijpelijk — te vermijden dat het boek veel te dik zou worden (het is nu al een hele klus om de 329 pagina's door te werken, hoe boeiend ook). Een gemiste kans.

Samenvattend, een feest om dit alles bij elkaar te zien staan en weer eens te lezen hoe ongelofelijk slim heel veel grote geesten, maar vooral Euler (hij weer) is geweest om de diverse takken van wiskunde te verbinden en daardoor doorbraken te forceren. Ik eindig met een van die vele tientallen spectaculaire weetjes waardoor je je soms eventjes afvraagt (tot realiteitszin je weer met beide voeten op de grond brengt) waarom niet elke vwo-leerling wiskunde gaat studeren: Bewijs dat $f(x) = x^9 - x^3 + 2520$ weliswaar irreducibel is, maar dat $f(x)$ samengesteld is voor elke gehele positieve x. De gegeven hint — en daar zal ook u het mee moet doen — is dat elk getal $f(x)$ deelbaar is door 504. *Joop van der Vaart*



John Stillwell

The Story of Proof

Logic and the History of Mathematics

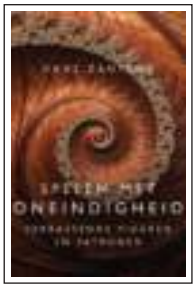
Princeton University Press, 2022
xiv + 441 p., prijs $45.00
ISBN 9780691234366

"Proof is the glory of mathematics — and its most characteristic feature — yet proof itself is not considered an interesting topic by many mathematicians." In deze eerste zin uit het voorwoord gaat het om bewijs als logisch of wiskundig concept. De wiskunde beperkt zich tot (meestal elementaire) meetkunde, algebra, getaltheorie, analyse, topologie en verzamelingenleer. Dit boek is geen 'inleiding logica' en uiteraard ook geen leerboek voor één van de hierboven genoemde domeinen van wiskunde, maar een "panoramic view of proof in basic mathematics" schrijft Stillwell in het voorwoord. Het boek eindigt met een viertal hoofdstukken over axiomatiek, bewijsbaarheid en (on)beslisbaarheid.

Het boek is in grote lijnen chronologisch opgebouwd. Het eerste hoofdstuk 'Before Euclid' introduceert thema's als irrationaliteit, inklemmen van reële getallen door rationale getallen, limietprocessen (Eudoxos) en bewijzen uit het ongerijmde. In het tweede hoofdstuk worden aspecten van Euclides' *Elementen* aangestipt: de invoering van axiomatiek, constructie versus existentie, equivalenten van het befaamde parallellenaxioma, 'infinitesimale methoden' en getaltheorie. Verwante thema's blijven terugkomen in Stillwells boek. Het derde hoofdstuk behandelt de perfectionering van Euclides' axiomatische opzet van meetkunde (in dimensies 2 en 3) door Hilbert, de ontdekking van projectieve meetkunde en de connecties tussen de stellingen van Pappos en Desargues enerzijds en commutativiteit en associativiteit anderzijds. Verderop in het boek wordt het meetkundethema afgerond met een hoofdstuk over niet-euclidische meetkunde. Beltrami's constructie met euclidische middelen van een model voor hyperbolische meetkunde bewees meer dan tweeduizend jaar na Euclides eindelijk dat het parallellenaxioma onafhankelijk is van Euclides' andere axioma's.

De overige hoofdstukken gaan merendeels over reële getallen. Stillwell beschrijft hoe in de negentiende eeuw met de supremum-eigenschap van reële getallen het laatste gat in het bewijs van de hoofdstelling van de algebra werd gedicht. Met behulp van Dedekindsneden (vergelijk met de insluiting van reële getallen door rationale getallen in Euclides!) kunnen we reële getallen beschrijven als *verzameling* natuurlijke getallen. Zelfs continue functies zijn zo te coderen en daarmee is een deel van de elementaire analyse te 'aritmetiseren'. En zo is Stillwell aangekomen bij verzamelingenleer. Hij behandelt axiomatiek en in het bijzonder keuzeaxioma's en hun soms wonderlijke consequenties. Het boek eindigt met een korte introductie op de beroemde resultaten van Gödel, Turing en Gentzen over de grenzen van bewijsbaarheid en berekenbaarheid.

De boeken van Stillwell zijn soepel geschreven en goed leesbaar voor een 'breed wiskundig publiek'. Hij schetst grote lijnen in de geschiedenis van de wiskunde en legt verbindingen die ik nog niet kende. Hij kan in een paar regels de kernideeën van een technisch bewijs schetsen en hij strooit met pareltjes. Hoewel soms iets meer precisie of detail wenselijk zou zijn, heb ik met plezier veel van zijn boeken gelezen. *Jeroen Spandaw*



Hans Zantema

Spelen met oneindigheid: verrassende figuren en patronen

Noordboek, 2023
240 p., prijs € 22,99
ISBN 9789464710212

In dit boek analyseert Hans Zantema stapfiguren die je kunt maken bij oneindige symboolrijen, zoals de rij alt die begint met 01 en dit patroon eindeloos herhaalt. Het recept voor een stapfiguur is eenvoudig. Voor elk symbool s leg je een draaihoek $h(s)$ vast, bijvoorbeeld $h(0) = 0^\circ$, $h(1) = 90^\circ$. Vervolgens laat je een (virtuele) schildpad bij de symboolrij een figuur tekenen. Bij elk symbool s beweegt de schildpad een vaste afstand vooruit en draait daarna over de bijbehorende hoek $h(s)$. De figuur bij de voorbeeldrij alt is een vierkant, dat eindeloos opnieuw getekend wordt. De oneindige rij levert dus een eindige figuur.

Om het boek opzichzelfstaand te maken, bouwt Zantema de benodigde voorkennis in de eerste vijf (van in totaal veertien) hoofdstukken van de grond af op. Dit gebeurt op een speelse maar degelijke wijze, inclusief stellingen en bewijzen, afgewisseld met persoonlijke beschouwingen. Het boek is zo voor iedereen toegankelijk, ook voor (goede vwo-) leerlingen. Die basis vergt echter tachtig bladzijden, waardoor de lezer misschien het zicht op het uiteindelijke doel verliest en een aardige dosis doorzettingsvermogen dient te hebben.

Elk hoofdstuk sluit af met een uitdaging. Als oud-deelnemer aan de (internationale) wiskundeolympiade houdt Zantema van pittige problemen. Je hoeft je niet te schamen als je daar op moet ploeteren. Het leesadvies om pen en papier te gebruiken (en zelfs een computer) is niet misplaatst. Het is ook goed om in het achterhoofd te houden dat hij zijn boek ziet als 'een wetenschappelijk avontuur', waarbij niet alleen een mooi resultaat (lees: stelling) telt, maar waarbij de weg om er te komen zeker zo belangrijk is.

Een groot deel van de rijen die in het boek aan bod komen, betreffen zogenaamde *morfische* rijen. Laat daartoe $f : A \to A^+$ een functie (morfisme) van symbolen naar eindige niet-lege symboolrijen zijn. Zo'n morfisme f kun je optillen naar (on)eindige symboolrijen a: $f(a)$ is dan de (on)eindige rij die je krijgt als je elk symbool s in a vervangt door $f(s)$. Definieer je bijvoorbeeld $f(0) = f(1) = 01$, dan is $f(01) = 0101$. Stel nu dat voor symbool $s \in A$ geldt dat $f(s)$ begint met s. In ons voorbeeld is dat zo voor $s = 0$. Dan zegt Stelling 19 dat er precies één oneindige rij a bestaat die begint met s en waarvoor geldt $f(a) = a$. Die laatste eis drukt uit dat a een *dekpunt* is van f. Zo is rij alt het dekpunt van voornoemde f dat met 0 begint. Morfische rijen krijg je door de symbolen in een dekpunt van een morfisme te herbenoemen met een functie $c : A \to B$. De stapfiguren van zulke morfische rijen hebben allerlei mooie eigenschappen, die zorgvuldig geformuleerd en bewezen worden.

De stof is allemaal elementair, maar het resulterende bouwwerk is behoorlijk hoog opgestapeld. Een aantal analyses is bovendien tamelijk lang, waardoor de lezer goed moet opletten de draad niet kwijt te raken. Zantema raadt aan om zulke passages over te slaan. Ik kan beamen dat je tot het einde toe zaken blijft tegenkomen die met minder inspanning te begrijpen zullen zijn.

Het boek heeft een index, maar geen literatuurlijst. In de tekst wordt onder andere verwezen naar het boek *Automatic Sequences: Theory, Applications, Generalizations* van Jean-Paul Allouche en Jeffrey Shallit (Cambridge University Press, 2003), dat duidelijk als inspiratiebron heeft gefungeerd. Zantema heeft destijds dat boek zelf voor NAW gerecenseerd in serie 5, deel 11, september 2010, pp. 218–219. *Tom Verhoeff*

Recent verschenen publicaties. Als u een van deze boeken wilt bespreken of als u suggesties heeft voor andere boeken voor deze rubriek, laat dit dan per e-mail weten aan reviews@nieuwarchief.nl.

Persi Diaconis, Jason Fulman

The Mathematics of Shuffling Cards

American Mathematical Society, 2023
ISBN 9781470463038
bookstore.ams.org/mbk-146

Samuel Coskey, Paul Ellis, Japheth Wood

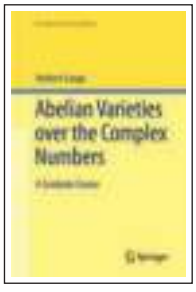
5 Fabulous Activities for Your Math Circle

American Mathematical Society, 2023
ISBN 9781945899089
bookstore.ams.org/nmath-10

Jesper Lützen

A History of Mathematical Impossibility

Oxford University Press, 2023
ISBN 9780192867391
global.oup.com/academic/product/9780192867391



Herbert Lange

Abelian Varieties over the Complex Numbers

Springer, 2023
ISBN 9783031255694
springer.com/978-3-031-25570-0

Chris Budd

Climate, Chaos and COVID
How Mathematical Models Describe the Universe

World Scientific, 2023
ISBN 9781800613041
worldscientific.com/worldscibooks/10.1142/q0385

András Juhász

Differential and Low-Dimensional Topology

Cambridge University Press, 2023
ISBN 9781009220576
cambridge.org/9781009220576

Problemen

| Problem Section

This Problem Section is open to everyone; everybody is encouraged to send in solutions and propose problems. Group contributions are welcome. We will select the most elegant solutions for publication. For this, solutions should be received before **15 Juli 2023**. The solutions of the problems in this issue will appear in one of the subsequent issues.

Problem A (proposed by Hendrik Lenstra)

Let R be a ring. We say $x \in R$ is *central* if $xy = yx$ for all $y \in R$. Suppose that for every $x \in R$ the element $x^2 - x$ is central in R. Show that R is commutative.

Problem B (proposed by Onno Berrevoets)

Isaac really likes apples, but does not like pears. He does not have any fruit now. Each time he visits

- Andrea, he gets 3 apples in exchange for 2 pears;
- Bob, he gets 3 pears;
- Caroline, he gets 1 apple and 1 pear.

Prove that the maximum number of apples Isaac can have after n visits equals $\lfloor 5n/9 \rfloor$.

Problem C (proposed by Daan van Gent)

For $S \subseteq \mathbb{Z}_{>0}$ write $\langle S \rangle$ for the submonoid of $\mathbb{Z}_{>0}$ generated by S. For $S \subseteq \mathbb{Z}_{>0}$ a *foundation* for S is a subset $C \subseteq \mathbb{Z}_{>0}$ for which $\langle C \rangle$ is minimal with respect to inclusion such that the elements of C are pairwise coprime and $S \subseteq \langle C \rangle$. For example, a foundation for $\{150, 180\}$ is $\{5, 6\}$.

a. Show that all subsets of $\mathbb{Z}_{>0}$ have a unique foundation.

Write $w(a,b)$ for the cardinality of the foundation for $\{a,b\}$ and let

$$f(n) = \min\{ab \mid a,b \in \mathbb{Z}_{>0}, w(a,b) = n\}.$$

b. Compute $f(11)$.

c* What is the asymptotic behaviour of f?

Edition 2023-1 We received solutions from Michel Bel, Carsten Dietzel and Albert Visser.

Problem 2023-1/A

Does there exist a partitioning X of $\mathbb{R}$ into infinite sets such that for every *choice map* $c : X \to \mathbb{R}$, i.e. a map c such that $c(S) \in S$ for all $S \in X$, the image of c is dense in $\mathbb{R}$?

Solution We received correct solutions from Michel Bel, Carsten Dietzel and Albert Visser. We show here the solution from Albert Visser: Let's say that the triple (m,n,p) is *adequate* if m is an odd integer, n is a natural number, and p is an odd prime. We define

$$A_{m,n,p} = \left\{ \frac{m}{2^n} + \frac{1}{p^k} \;\middle|\; k \in \mathbb{Z}_{\geq 0} \right\} \quad \text{where } (m,n,p) \text{ is adequate,}$$

and write A_∞ for the complement in $\mathbb{R}$ of the union of all such $A_{m,n,p}$. Consider $\mathcal{X} = \{A_{m,n,p} \mid (m,n,p) \text{ is adequate}\} \cup \{A_\infty\}$. Clearly, all elements of $\mathcal{X}$ are infinite and the union of $\mathcal{X}$ is $\mathbb{R}$. We show that the elements of $\mathcal{X}$ are pairwise disjoint. By definition, A_∞ is disjoint from the $A_{m,n,p}$. We note that for adequate (m,n,p) and $k \in \mathbb{Z}_{\geq 0}$ the fraction

Redactie: Onno Berrevoets en Daan van Gent
problems@nieuwarchief.nl
www.nieuwarchief.nl/problems

Oplossingen | Solutions

$$\frac{m}{2^n}+\frac{1}{p^k}=\frac{mp^k+2^n}{2^n p^k},$$

cannot be further simplified. This means that, if

$$\frac{2^{n_1}+m_1p_1^{k_1}}{2^{n_1}p_1^{k_1}}=\frac{2^{n_2}+m_2p_2^{k_2}}{2^{n_2}p_2^{k_2}}$$

for adequate (m_i,n_i,p_i), we deduce from the denominators that $n_1=n_2$, $p_1=p_2$ and $k_1=k_2$, and finally $m_1=m_2$. Hence the $A_{m,n,p}$ are pairwise disjoint.

Finally, consider any real r and any $\delta>0$. We can find $m\in\mathbb{Z}$ and $n\in\mathbb{Z}_{\geq 0}$ such that $|r-m2^{-n}|<\delta/2$ by considering the binary expansion of r, and an odd prime p with $1/p<\delta/2$. It follows that for all $q\in A_{m,n,p}$ we have $|r-q|<\delta$. In particular, this holds for $q=c(A_{m,n,p})$, so the image of c is dense.

Problem 2023-1/B

Show that for all $k\in\mathbb{Z}$ there exists an $x\in\mathbb{Q}$ for which there are at least two subsets $S\subseteq\mathbb{Z}_{\geq 1}$ such that $\sum_{s\in S}s^k=x$.

Solution We received a correct solution from Carsten Dietzel: First assume that $k\geq 0$. Let n be an integer with $2^n>n^{k+1}+1$. Denote by P_n the power set of $\{1,\ldots,n\}$ and by $\mathbb{N}$ the set of positive integers. Define the map

$$\begin{aligned}\sigma : P_n &\to \mathbb{N},\\ S &\mapsto \sum_{s\in S}s^k.\end{aligned}$$

For $S\in P_n$ we have $\sigma(S)\leq n\cdot n^k=n^{k+1}$. Therefore, the range of σ is a subset of $\{0,\ldots,n^{k+1}\}$. By the pigeonhole principle, σ is not injective, i.e., there are distinct $S_1,S_2\in P_n$ with $\sigma(S_1)=\sigma(S_2)$.

Now let $k<0$. From the case $k\geq 0$ we know that there are finite subsets $T_1,T_2\subset\mathbb{N}$ such that $\sum_{t\in T_1}t^{-k}=\sum_{t\in T_2}t^{-k}$. Let N be the least common multiple of the elements of $T_1\cap T_2$. Define $S_i=\{N/t\mid t\in T_i\}$ for $i=1,2$. Then for $i=1,2$ we have

$$\sum_{s\in S_i}s^k=\sum_{t\in T_i}(t/N)^{-k}=N^k\sum_{t\in T_i}t^{-k}$$

and thus $\sum_{s\in S_1}s^k=\sum_{s\in S_2}s^k$. Hence, S_1 and S_2 are as desired!

Problem 2023-1/C (proposed by Daan van Gent)

For a group G and $g\in G$ write $c(g)=\{hgh^{-1}\mid h\in G\}$ and $G^\circ=\{g\in G\mid \#c(g)<\infty\}$.

a. Show that G° is a normal subgroup of G and that $G^{\circ\circ}=G^\circ$.

b. Now define $G_\circ=G/G^\circ$. Show that there exists a group G for which the sequence $G,G_\circ,G_{\circ\circ},\ldots$ does not stabilize, i.e. for none of the groups H in the sequence we have $H^\circ=1$.

Solution We received a correct solution from Carsten Dietzel. The solution shown here is by Daan van Gent.

a. For all $g,h\in G$ we have $c(g^{-1})=\{a^{-1}\mid a\in c(g)\}$ and $c(gh)\subseteq\{ab\mid a\in c(g),b\in c(h)\}$. Thus for $g,h\in G^\circ$ both $c(g^{-1})$ and $c(gh)$ are finite and $c(1)=\{1\}$, so G° is a subgroup of G. Then note that $c(g)=c(h)$ for conjugates g and h, so G° is normal.

Clearly $G^{\circ\circ}\subseteq G^\circ$. Suppose $g\in G^\circ$. Since g has only finitely many conjugates in G, it certainly has only finitely many conjugates in a subgroup, so $g\in G^{\circ\circ}$.

b. For a family of groups $(G_i)_{i\in I}$ we define the *restricted product*

$$\bigoplus_{i\in I}G_i=\left\{(x_i)_{i\in I}\in\prod_{i\in I}G_i\mid x_i=1 \text{ for all but finitely } i\right\}.$$

Oplossingen

| Solutions

We inductively define $P_0 = \{1\}$ and for $k \geq 0$ define

$$F_k = \bigoplus_{\substack{H \lhd P_k \\ \#(P_k/H) < \infty}} \mathbb{Z}_2^{P_k/H} \quad \text{and} \quad P_{k+1} = F_k \rtimes P_k,$$

where $\mathbb{Z}_2$ is the group of 2-adic integers and $\pi \in P_k$ acts on each factor $\mathbb{Z}_2^{P_k/H}$ as $(\varphi_{H,x})_{x \in P_k/H} \mapsto (\varphi_{H,\pi x})_{x \in P_k/H}$.

Note that every $x \in F_k$ has a finite orbit under P_k, because all P_k/H are finite and each $x \in F_k$ has finite support. Hence $F_k \subseteq P_{k+1}^\circ$. Conversely, suppose $(\varphi,\pi) \in P_{k+1}^\circ$. For $(\psi,1) \in P_{k+1}$ we have $(\psi,1)(\varphi,\pi)(\psi,1)^{-1} = ((\psi_{H,x} + \varphi_{H,x} - \psi_{H,\pi x})_{H,x},\pi)$. If $\pi x \neq x$ for some (H,x), then with $\psi_{H,x} = 0$ and $\psi_{H,\pi x}$ ranging over $\mathbb{Z}_2$ we obtain infinitely many conjugates of (φ,π). Hence $\pi x = x$, and π acts trivially on every finite quotient of P_k. One shows inductively that P_k is profinite, hence $\pi = 1$. Thus $P_{k+1}^\circ = F_k$ and $(P_{k+1})_\circ = P_k$.

Finally, let $G = \bigoplus_{k=0}^\infty P_k$. Observe that $G^\circ = \bigoplus_{k=0}^\infty (P_k^\circ) = \bigoplus_{k=1}^\infty F_{k-1} \neq 1$, while $G_\circ \cong G$. Hence the sequence never stabilizes.