

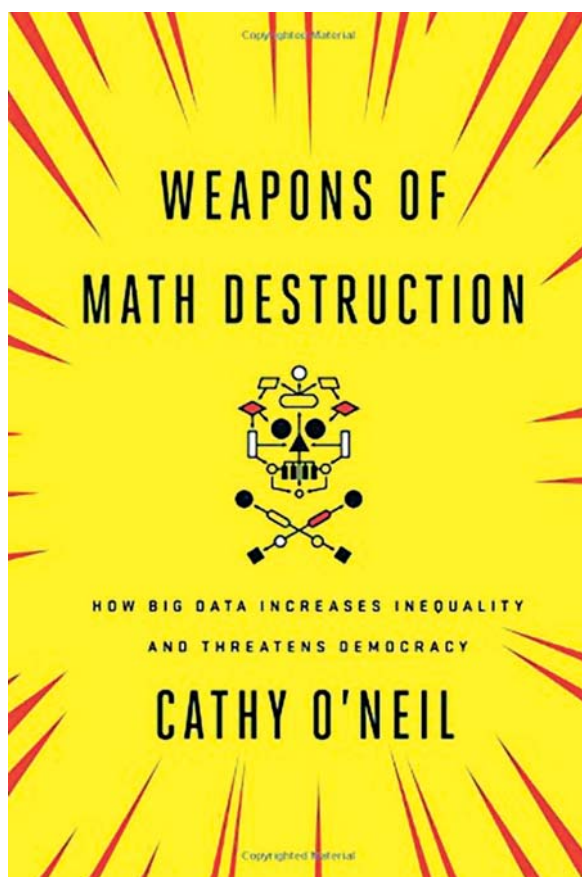
Logbuch Mathematik

Thilo Kuessner

*Big Data gibt Ihnen diese Scheingewissheit,
im Besitz objektiver Wahrheiten zu sein.*
Saatchi & Saatchi-Chef Christian Rättsch,
Mai 2015

Mathe-Vernichtungswaffen

Erst einmal gewann bisher ein Mathematikbuch den American Book Award – Douglas R. Hofstadter: *Gödel, Escher, Bach*. Immerhin in die Longlist des Awards schaffte es dieses Jahr *Weapons of Math Destruction*, Cathy O’Neils Abrechnung mit den Algorithmen der Massendatenverarbeitung.



Zahlreiche Rezensionen und Interviews greifen seitdem die zentrale These des Bestsellers auf:

Big Data is largely in the hands of large corporate self-interest groups that have no motivation to establish the fair distribution of anything. [...] Like gods, these mathematical models were opaque, their workings invisible to all but the highest priests

in their domain: mathematicians and computer scientists. And they tended to punish the poor and the oppressed in our society, while making the rich richer.

Postleitzahlen entscheiden über Versicherungsbeiträge und Kreditwürdigkeit. Verbrechensstatistiken lassen die Polizei vor allem in ärmeren Stadtvierteln Präsenz zeigen. Drogenkonsumierende College-Studenten in wohlhabenden Stadtvierteln bleiben so unter dem Radar der Polizei, Drogenkonsumenten in benachteiligten Vierteln werden umso häufiger kontrolliert und bestraft. Und wenn in staatlichen Schulen die Lehrer mittels statistischer Erhebungen bewertet werden, dann profitieren davon die teuren Privatschulen, die auf solche Erhebungen verzichten und die an staatlichen Schulen zu Unrecht gekündigten LehrerInnen übernehmen können.

„Math is racist“, fasste CNN den Buchinhalt zusammen.

Die Schlagzeilen verwässern ein wenig die eigentliche Essenz der im Buch erzählten Erfahrungen: dass manche Modelle und manche Modellierer schlicht unzuverlässiger sind als andere.

Eine transparente und zuverlässige Datenverwertung sieht O’Neil im Sport:

Baseball is an ideal home for predictive mathematical modeling. Baseball models are fair in part because they are transparent. Everyone has access to the stats and can understand more or less how they are interpreted.

Dem gegenüber stellt O’Neil mathematische Modelle in anderen, wichtigeren Bereichen. Besonders arbeitet sie sich an Schulbewertungen ab:

Now you may look at the baseball model, with its thousands of changing variables, and wonder how we could even be comparing it to the model used to evaluate teachers in Washington, D.C., schools. In one of them, an entire sport is modeled in fastidious detail and updated continuously. The other, while cloaked in mystery, appears to lean heavily on a handful of test results from one year to the next. Is that really a model?

Immer neue Werkzeuge würden entwickelt, doch niemand kontrolliere die Ergebnisse. Die Anekdote, an der sie das zum Beginn des Buches darstellt, handelt von einer Lehrerin, die entlassen wird, weil sich ihre Schüler in nur einem Jahr dramatisch verschlechtern. Die Geschichte beginnt im Jahr 2007.

In 2007, Washington D.C.'s new mayor, Adrian Fenty, was determined to turn around the city's underperforming schools. He had his work cut out for him: at the time, barely one out of every two high school students was surviving to graduation after ninth grade, and only 8 percent of eighth graders were performing at grade level in math. Fenty hired an education reformer named Michelle Rhee to fill a powerful new post, chancellor of Washinton's schools. Rhee developed a teacher assessment tool called IMPACT, and at the end of the 2009-10 school year the district fired all the tachers whose scores put them in the bottom 2 percent.

Bei der entlassenen Lehrerin war es dann so, dass offensichtlich manipulierte Prüfungen mit zu guten Resultaten bei einem anderen Lehrer im Jahr zuvor zu einem relativen Einbruch der Klausurnoten im Folgejahr und damit zur Kündigung der neuen Lehrerin führten. O'Neil schlussfolgert dann aber nicht, dass etwa Erziehungswissenschaftler anders arbeiten würden als Baseballtrainer, oder dass – anders als beim Baseball – kein kontinuierliches Hinterfragen und Bewerten der jeweils neuen Bewertungsansätze möglich sei, sondern dass es auch hier darum ginge, die Unterprivilegierten zu benachteiligen:

So, thanks to a highly questionable model, a poor school lost a good teacher, and a rich school, which didn't fire people on the basis of their students' scores, gained one.

Die entlassene Lehrerin wurde nämlich von einer reicheren Schule wieder eingestellt.

Anscheinend braucht es solche griffigen Deutungen, um mit einem Mathematikbuch ins Fernsehen zu kommen. Ein Skeptiker würde hier, und bei manchen anderen Geschichten im Buch, wohl eher Hanlons Rasiermesser zitieren: *Schreibe nichts der Böswilligkeit zu, was durch Dummheit hinreichend erklärbar ist.*

Die größte Familie

Man kann einwenden, dass die Probleme der Schulbewertungen nichts mit *Big Data* zu tun haben, denn diese Bezeichnung meint ja eigentlich Datenmengen, die sich selbst wenn man es wollte und könnte, nicht mehr manuell verarbeiten ließen. Dieser Einwand trifft auch auf das folgende Beispiel zu, das aber trotzdem zeigt, warum man sich auf die reine Datenanalyse nicht verlassen sollte.

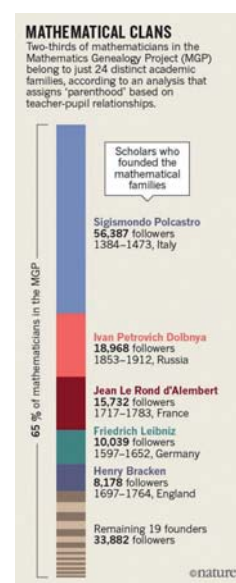
In der Genetik wird bekanntlich die These diskutiert, dass alle menschlichen mütterlichen Linien auf eine Frau zurückgehen, die vor etwa 200 000 Jahren in Afrika lebte. Ein ähnlicher Effekt in sehr viel kürzeren Zeiträumen ist in Mathematiker-Stammbäumen zu beobachten, also wenn man zurückverfolgt, wer zum Beispiel der Doktorvater des Doktorvaters des Doktorvaters eines Mathematikers gewesen ist.

Viele Mathematiker landen bei David Hilbert oder Solomon Lefschetz, wenn sie in der Online-Datenbank *Math Genealogy* ihren Stammbaum zurückverfolgen. Das ist nicht überraschend: Hilbert dominierte lange Zeit die Mathematik in Göttingen und Lefschetz sorgte dafür, dass nach 1933 Princeton die Rolle Göttingens übernahm. Hilbert hatte 75 Doktoranden und insgesamt 27 276 in der Datenbank erfasste Nachkommen, bei Lefschetz sind es 26 Doktoranden und 8305 Nachkommen. Geht man weiter zurück, so werden die Nachkommenschaften natürlich größer: Carl Friedrich Gauß kommt mit nur 10 in der Datenbank erfassten Doktoranden auf 75 055 Nachkommen, darunter Hilbert und Lefschetz samt deren Nachkommenschaft.

Eine in *EPJ Data Science* erschienene und Ende August von *Nature News* aufgegriffene Studie „The classical origin of modern mathematics“ geht nun noch weiter zurück und findet heraus, dass zwei Drittel aller Mathematiker aus 24 „Familien“ stammen, und dass die mit Abstand größte die bis ins 15. Jahrhundert zurückgehende Familie von Sigismondo Polcastro sei.

Sie kennen Sigismondo Polcastro nicht? Er war auch kein Mathematiker, sondern Mediziner und laut Datenbank der Zweitgutachter von Pietro Roccabonella. Dessen Student Niccolò Leoniceno wird in der Datenbank unter *Biology and other natural sciences* klassifiziert, ebenso wie die folgenden zehn Generationen bis zu Johann Segner im Jahr 1728. Die erste mathematische Arbeit in diesem Teilbaum stammt vom Nationalökonom Johann Büsch, der laut Datenbank 1752 in Göttingen eine Magisterarbeit über „General algebraic systems“ schrieb. (Ich nehme an, das ist die englische Übersetzung eines lateinischen Titels?) Büschs Doktorand Johann Bode promovierte in Astronomie und erst mit Johann Pfaff, bei dem Bode 1786 Zweitgutachter war, haben wir einen richtigen Mathematiker im Stammbaum. Pfaff wiederum war dann der Doktorvater von Carl Friedrich Gauß und damit gehören Klein, Hilbert, Lefschetz und alle ihre Nachkommen zur Familie.

Kurz: man hätte nicht nur Maschinen, sondern auch mal einen Mathematiker auf die Stammbäume schauen lassen sollen. Beim größten Teilbaum handelt es sich nicht um die Polcastro-Familie, sondern um die Pfaff-Gauß-Familie. Deren Zuordnung zur Polcastro-Familie ist einfach nur ein Artefakt der Datenbank, die für die Gauß-Vorfahren



Abdruck mit freundlicher Genehmigung von Macmillan Publishers Ltd: *Nature* 537, 20-21

eben bis zum Beginn des 15. Jahrhunderts zurückgeführt wurde, während bei anderen Familien schon früher, also in einem späteren Jahrhundert, die Erfassung endet.

Endlich viele Galaxien

Je mehr Daten man hat, desto bewusster wird einem die eigene Endlichkeit. Das dachte sich wohl der Redakteur bei Focus Online, als er einen Artikel über die „Entdeckung“ neuer Galaxien mit dieser sinnentstellenden Überschrift versah.

Forscher widerlegen bisherige Meinung: Das Universum ist gar nicht unendlich

Das Universum hat mindestens zehnmal mehr Galaxien als angenommen. Zu diesem Schluss kommen Astronomen nach der Analyse des Hubble-Weltraumteleskops und anderer Beobachtungen. Der überwältigende Teil der Galaxien ist demnach mit heutigen Instrumenten gar nicht zu sehen.

(Focus Online, 17. Oktober 2016. Die Überschrift wurde nach einigen Stunden korrigiert.)

Was ist Topologie?

Und dann war da noch der Physik-Nobelpreis, dank dem es die Topologie – wahrscheinlich zum ersten Mal? – in die Meldungen der Tagesschau schaffte. Gezeigt wurde Laudator Thors Hansson, wie er mit Boller, Bagel und Brezel die Topologie von Flächen erklärt.

Das Aufreißen der Brezel sollte dann wohl die sprunghafte Zunahme des elektrischen Widerstandes veranschaulichen.

Kosterlitz und Thouless entdeckten bereits in den 1970er Jahren, dass sich mit der Topologie ungewöhnliche Veränderungen von Materie erklären lassen. Bei normalen Phasenübergängen, beispielsweise zwischen Eis und Wasser, verändern sich die Symmetrieeigenschaften, bei topologischen Phasenübergängen trifft das nicht zu. Die Akademie würdigt,

dass Thouless, Haldane und Kosterlitz der Menschheit eine Tür zu einer *unbekannten Welt* eröffnet haben. Dafür haben die Quantenforscher *fortschrittliche mathematische Methoden* benutzt – beispielsweise Superkonduktoren, Superfluide oder dünne magnetische Schichten.

Dank ihrer Pionierarbeit ist die Jagd auf neue und exotische Zustände von Materie eröffnet, heißt es in der Begründung des Nobelpreiskomitees. [...] Die Pionierarbeit von Thouless, Haldane und Kosterlitz motivierte bereits in den letzten Jahrzehnten zahlreiche andere Forschungsgruppen, sich für die Topologie zu interessieren – und diese in physikalischen Systemen aufzuspüren. Eine direkte Anwendung haben diese Arbeiten bisher zwar noch nicht erfahren. Doch *viele Menschen erhoffen sich davon künftige Anwendungen sowohl in der Materialforschung als auch in der Elektronik*, erklärte das Komitee.

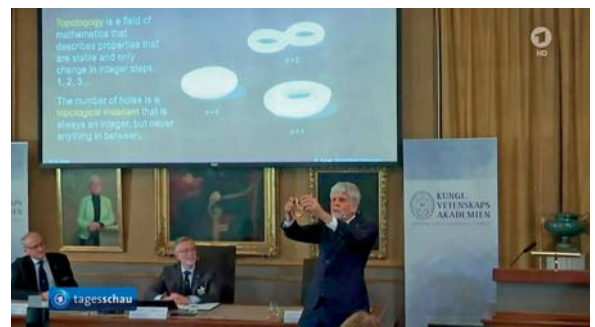
So die Erläuterung bei tagesschau.de. Was ich – auch anderswo – nicht gefunden habe: eine Erklärung, die auch wir Topologen verstehen, welche Topologie dort eigentlich verwendet wird.

Dr. Thilo Kuessner, Korea Institute for Advanced Study, 85 Hoegi-ro, Dongdaemun-gu, 130-722 Seoul, Korea kuessner@kias.re.kr

Thilo Kuessner forscht am Korea Institute for Advanced Study zur Topologie und Geometrie niedrig-dimensionaler Mannigfaltigkeiten. In dieser Kolumne zweitverwertet er seine Beiträge aus dem Mathlog: <http://scienceblogs.de/mathlog/>



Norwegisches Frühstücksgebäck erklärt die Topologie von Flächen.



Die Brezel zerreißt – fließt der Strom jetzt langsamer?