

## Die Mathematik hinter Google

Prof. Dr. Michael Eisermann

Fachbereich Mathematik, Universität Stuttgart

Science Pub Stuttgart, 18. Februar 2013



[www.igt.uni-stuttgart.de/eiserm/popularisierung/#SciencePub](http://www.igt.uni-stuttgart.de/eiserm/popularisierung/#SciencePub)

## Einleitung

*„Wo simmer denn dran? Aha, heute krieje mer de Suchmaschin.  
Wat is en Suchmaschin? Da stelle mer uns janz dumm.“*

Sehr geehrte Damen und Herren, ich begrüße Sie zum heutigen Science Pub!

Ich möchte Ihnen heute Abend ein wenig über die Suchmaschine Google *erzählen*. Es wird um Zahlen und Daten gehen und wie daraus schlau wird, also um Mathematik. Sie ist heutzutage mehr denn je unverzichtbare Grundlage für Naturwissenschaft und Technik.

Sie können mich während des Vortrags gerne unterbrechen und nachfragen. Ich möchte Sie hierzu ausdrücklich ermutigen! Nach dem Vortrag wird Gelegenheit zur Diskussion sein.

Ich gehe mit gutem Beispiel voran und habe Fragen für Sie mitgebracht: Zum Science Pub gehört ein Pub Quiz, und damit möchte ich beginnen.

## Science Pub Quiz: Zahlen, bitte!

*„Data is the new oil.“*

Wie viele Suchanfragen beantwortet Google im Jahr?

- 1 Mrd       10 Mrd       100 Mrd       1 Bio

Wie viele Webseiten gibt es ungefähr im Internet?

- 1 Mrd       10 Mrd       100 Mrd       1 Bio

Wie viele Neuronen besitzt das menschliche Gehirn?

- 1 Mrd       10 Mrd       100 Mrd       1 Bio

Den Slogan *„Daten sind das neue Öl.“* hört man oft von Marketingstrategen. Der Vergleich ist griffig, auch schnell abgegriffen, und man kann hierüber geteilter Meinung sein. Wir wollen uns im folgenden Quiz erst einmal die Größenordnungen vor Augen führen.

Nach eigener Auskunft beantwortet Google 3 bis 4 Milliarden Suchanfragen pro Tag, also etwa 1 Billion Suchanfragen pro Jahr. Oder 2 Millionen pro Minute.

Wie viele Webseiten es genau gibt, vermag niemand zu sagen. Das liegt auch daran, das viele Seiten heutzutage nicht mehr „handgemacht“ sind, sondern von Computern generiert werden, z.B. [www.bahn.de](http://www.bahn.de). Bei geeigneter Zählweise kommt Google auf über eine Billion Seiten.

Das menschliche Gehirn hat schätzungsweise 100 Milliarden Nervenzellen (Neuronen). Jede ist durch Synapsen mit etwa 1000 anderen Nervenzellen verbunden. Die Vernetzung ist also ähnlich komplex wie das Web. (Die Funktionsweisen sind gänzlich unterschiedlich.)

## Science Pub Quiz: Zahlen, bitte!

*„It's the economy, stupid!“*

Wieviel Gewinn machte Google im Jahr 2012? (USD)

- 1 Mrd       11 Mrd       21 Mrd       41 Mrd

Welchen Börsenwert hat Google zur Zeit? (USD)

- 50 Mrd       100 Mrd       250 Mrd       400 Mrd

Welches Unternehmen hat zur Zeit den höchsten Börsenwert?

- Apple       Exxon       Microsoft       WalMart

Ins öffentliche Bewusstsein kam das Internet so langsam erst vor 20 Jahren.

Der seither anhaltende Boom beruht vor allem darauf, dass man damit Geld verdienen kann: Im und mit dem Internet werden bereits etwa 5% des BIP erwirtschaftet. Das ist enorm!

Googles Börsenwert am letzten Freitag: 261 Mrd USD.

Zum Vergleich: Apple 430, Exxon 400, WalMart 240, Microsoft 230.

Googles Gewinn nach Steuern betrug 2012 etwas über 11 Mrd USD. Zum Vergleich: Apple 41, Exxon 41. Der baden-württembergische Haushalt plant Ausgaben von 41 Milliarden Euro.

## Science Pub Quiz: Zahlen, bitte!

*„Scotty, Energie!“*

Wieviel Strom verbrauchen dt. Haushalte für PC & TV? (KWh/Jahr)

- 3 Mrd       6 Mrd       12 Mrd       30 Mrd

Wieviel Strom erzeugt ein typisches Atomkraftwerk? (KWh/Jahr)

- 3 Mrd       6 Mrd       12 Mrd       30 Mrd

Wieviel Strom verbrauchen Googles Datacenter? (KWh/Jahr)

- 3 Mrd       6 Mrd       12 Mrd       30 Mrd

Bei diesen Größenordnungen ist natürlich auch der Energiebedarf ein wichtiges Thema.

Im Jahr 2012 betrug Deutschlands Nettostromverbrauch etwa 540 TWh, davon entfällt etwa die Hälfte auf die Industrie (250 TWh) und etwa ein Viertel auf die Haushalte (130 TWh). In den Haushalten sind die stromintensivsten Bereiche Computer (13%), Fernsehen (12%), die zusammen ein Viertel ausmachen, gefolgt von Licht (11%), Kochen (10%), Kühlen (10%).

Ein 2-Personen-Haushalt verbraucht um die 3000 KWh/Jahr, eine Person also um die 1500 KWh/Jahr. Bei 80 Mio Bundesbürgern sind das um die 120 Mrd KWh/Jahr. Ein Viertel davon geht auf Informations-, Kommunikations-, und Unterhaltungselektronik, also 30 KWh/Jahr.

Jedes der neun im Betrieb befindlichen deutschen Atomkraftwerke hat eine Leistung von 1300–1400 MW, also etwa 12 Mrd KWh/Jahr. Sie sehen, es ist nicht ganz unwichtig, dass man solche Zahlen kennt und sich die gesellschaftlichen Größenordnungen klar machen kann.

Googles Datacenter haben einen Energiebedarf von 300 MW, also etwa 2.5 Mrd KWh/Jahr. Wenn Google 10 Cent pro KWh zahlt, beläuft sich die Stromrechnung auf 250 Mio USD. Man kann sich vorstellen, warum Google großen Wert auf Energieeffizienz legt!

Google nutzt für seine Internetdienste ungefähr 1 Mio Server. Die genaue Anzahl wird geheim gehalten, aber aus dem Stromverbrauch kann man die Zahl recht gut schätzen.

Teilt man 1 Bio Wh für den Suchdienst durch 1 Bio Anfragen, so kostet jede Suchanfrage grob geschätzt etwa 1 Wh. Einen Liter Wasser zu kochen (von 15°C auf 100°C erhitzen) kostet etwa 100 Wh. (Ein paar Dutzend Google-Anfragen entspricht einer heißen Tasse Tee.)

## Sehr kurze Geschichte des Internets

Beim Stichwort „Internet“ denken manche an freien Informationsaustausch, Partizipation, Demokratie 2.0 und ähnlich hehre Ziele. Zahlenmäßig überwiegen vermutlich Konsum, Datenklau, Pornographie. Beinahe vergessen ist der militärische Ursprung des Internets:

- 1940 elektronische Computer
- 1957 UdSSR starten *Sputnik*
- 1958 USA gründen *ARPA*
- 1969 ARPANet verbindet 4 Unis
- 1982 Internet Protocol (TCP/IP) vereinheitlicht Netze
- 1991 erster Webserver am CERN (WWW, HTML)
- 1993 Browser *Mosaic*, Expansion des WWW
- 1998 Google indiziert 26 Millionen Webseiten
- 2000 Google indiziert 1 Milliarde Webseiten
- 2008 Google sichtet 1 Billion Webseiten



Bis in die 1940er Jahre waren die „Computer“ mit Rechnungen befasste Angestellte; das Photo zeigt einen „Computer Room“. Elektronische, frei programmierbare Rechner gibt es seit den 1940er Jahren: Zuse Z3 in Deutschland, Colossus in England, Harvard Mark 1 in den USA. Letztere wurden noch im Krieg zu Berechnungen und zur Dechiffrierung eingesetzt.

Im kalten Krieg starten 1957 die UdSSR den ersten künstlichen Erdsatelliten *Sputnik*. Der Sputnik-Schock löst tiefgreifende Reaktionen der USA aus. Sie gründen 1958 die Advanced Research Projects Agency (ARPA) für militärisch relevante Forschung und Entwicklung.

Eines der Projekte ist die Entwicklung eines Computernetzes zwischen Universitäten und Forschungslaboren. (Die Folklore besagt, dass dieses Netz im Falle eines nuklearen Angriffs wenig verwundbar sein sollte, andere Quellen bestreiten diese Zielsetzung.)

Das ARPANet verband 1969 zunächst vier Computer (UCLA, Stanford, UCSB, Utah). Parallel hierzu entstanden in den 1970 Jahre weitere Netze. Als gemeinsames Protokoll wurde 1974 das Transmission Control Program (TCP) entwickelt und 1982 das Internet Protocol (TCP/IP), das bis heute genutzt wird. Seit 1984 gibt es das Domain Name System (DNS).

1991 ging der erste Webserver online – am europäischen Forschungszentrum CERN. Der Begriff World Wide Web (WWW) und Hypertext Markup Language (HTML) entstanden hier. 1993 machte Mosaic das WWW populär und löste den bis heute andauernden Boom aus.

Mit der Explosion des WWW wurden neue Suchmechanismen notwendig. Das Unternehmen Google wurde 1998 gegründet und stellte die gleichnamige Suchmaschine zur Verfügung. Ihr Erfolg beruht auf der intelligenten Sortierung der Suchergebnisse.

## Chaos und Struktur im Web

**Dezentral:** Viele unabhängige Autoren erzeugen Inhalte.

**Heterogen:** viele Informationen aber wenig Struktur

**Syntax:** *hypertext markup language* (HTML)

- Logische Struktur (Titel, Untertitel, Paragraphen, ...)

```
<h1> Dies ist eine Überschrift. </h1>
```

```
<p> Dies ist ein Paragraph. </p>
```

- Erscheinungsbild (Schriftart, fett, kursiv, Farben, ...)

```
<b> Dieser Text erscheint fett. </b>
```

```
<i> Dieser Text erscheint kursiv. </i>
```

- Querverweise / Links (Verweis von einer Seite auf eine andere)

```
<a href="http://www.igt.uni-stuttgart.de/eiserm/">
```

```
Hier geht's zur Homepage von Michael Eisermann. </a>
```

Stellen Sie sich eine riesige Bibliothek mit einer Billion Dokumenten vor. Einen Bibliothekar gibt es nicht. Jeder darf Dokumente hinzufügen und muss niemanden darüber informieren.

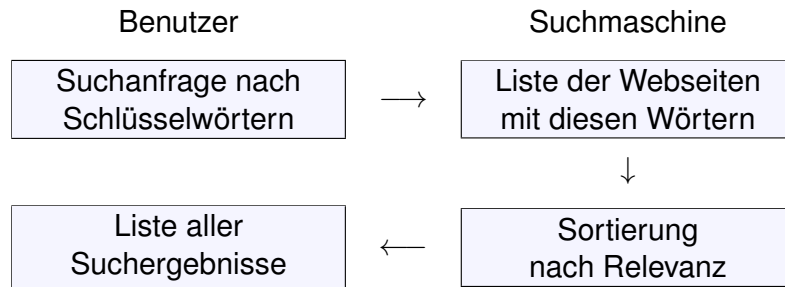
Sie suchen nun dringend nach bestimmten Informationen, und da Sie ungeduldig sind, möchten Sie das Ergebnis innerhalb von Sekunden. Wie soll das gehen? Zunächst scheint das völlig unmöglich... und doch gelingt Suchmaschinen im Internet genau das!

Zunächst einmal sichtet jede Suchmaschine die vorhandenen Daten. Hierzu läuft unablässig im Hintergrund eine automatische Surfsoftware (Spider, Webrobots oder kurz Bots genannt), die das Internet permanent durchforstet. Mit den angesteuerten Seiten geschieht zweierlei:

- Erstens speichert die Suchmaschine eine Kopie der Seite im hauseigenen Rechenzentrum. Dabei gibt sie jeder katalogisierten Seite eine Nummer.
- Zweitens wird ein Index erstellt: Dieser ist wie das Schlagwortregister eines Buches eine lange Liste von Wörtern und dazu die Nummern der Seiten, auf denen diese vorkommen.

Bei einer Suchabfrage schaut die Suchmaschine in ihrem vorbereiteten Index nach, auf welchen Seiten der gesuchte Begriff vorhanden ist. Für den Nutzer muss die Liste der Suchergebnisse dann nach Relevanz sortiert werden, damit das Wichtigste ganz oben steht.

## Was leistet eine Suchmaschine?



Kernprobleme:

- Informatik: Wie verarbeitet man enorme Datenmengen?
- Mathematik: Wie misst man Relevanz von Informationen?
- Finanzstrategie: Wie verdient man an einem Gratisprodukt?

Als das Web Mitte der 1990er noch klein war, genügte es, zu einer Suchanfrage einfach alle Treffer aufzulisten. Die Liste war noch kurz, der Nutzer konnte sie noch überblicken.

Das Internet blieb jedoch nicht lange so klein und überschaubar.

Die Suchmaschine Google ist seit 1998 in Betrieb und dominiert seither den Markt. Ihre Stärke liegt in der intelligenten Sortierung der Suchergebnisse. Bei vorherigen Suchmaschinen musste man endlose Trefferlisten durchforsten, bis man auf die ersten interessanten Ergebnisse stieß. Bei Google stehen sie auf wundersame Weise ganz oben auf der Liste. Wie ist das möglich? Die Antwort liegt (zu einem großen Teil) in der folgenden Formel, die ich hier erklären will:

$$p_i = \frac{q}{n} + \sum_{j \rightarrow i} \frac{1-q}{\ell_j} p_j$$

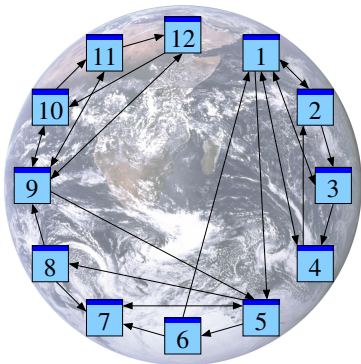
Keine Angst, die Formel sieht nur auf den ersten Blick kompliziert aus. Ich werde sie anhand von Beispielen Schritt für Schritt erläutern. Wer sowas schon gesehen hat, weiß, dass es sich um eine besonders einfache Formel handelt, nämlich eine *lineare* Gleichung, die keine Quadrate oder komplizierteres enthält. (Schon die Formel von Pythagoras  $a^2 + b^2 = c^2$  ist komplizierter. ;-)

📖 Sergey Brin, Larry Page (1998)

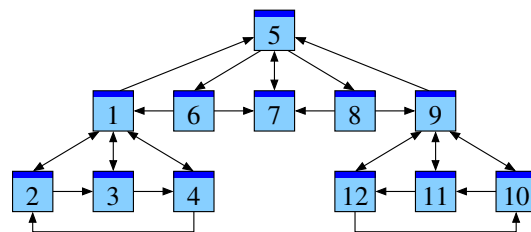
*The anatomy of a large-scale hypertextual web search engine.*

<http://infolab.stanford.edu/pub/papers/google.pdf>

## Chaos und Struktur im Web



Versuch einer hierarchischen Anordnung:



Eine Seite ist populär, wenn viele Seiten auf sie verweisen. Zu naiv!  
 Eine Seite ist populär, wenn viele **populäre** Seiten auf sie verweisen.

Klassische Texte sind linear: Ein Buch hat einen Anfang und ein Ende. Typischerweise liest man es von vorne bis hinten in der Reihenfolge der Seiten. (Meist gibt es zudem ein Inhaltsverzeichnis oder auch einen Index, und bei einer Doktorarbeit sogar Fußnoten.)

Webseiten bilden hingegen eine gänzlich andere Struktur. Niemand würde auf die Idee kommen, das Internet von Anfang bis Ende durchzulesen: Es hat keine lineare Struktur, keine erste und keine letzte Seite, es ist viel zu groß, und das meiste ist ohnehin uninteressant.

Die Webseiten verweisen gegenseitig aufeinander und bilden einen *Hypertext*. Zur Illustration betrachten wir ein Miniaturbeispiel bestehend aus 12 Webseiten. Unter den Seiten 1, 2, 3, 4 wird 1 am häufigsten zitiert. Die Seite 1 scheint daher besonders relevant oder populär. Gleiches gilt für 9, 10, 11, 12 mit 9 an der Spitze. Die Struktur von 5, 6, 7, 8 ist ähnlich mit 7 an der Spitze. Aber die Seiten 1 et 9, die wir schon als relevant erkannt haben, verweisen beide auf die Seite 5. Diese scheint daher wichtig und für die Suche besonders relevant.

Diese Anordnung war Handarbeit. Lässt sie sich automatisieren? Nach welchen Regeln?

Erster Versuch: Eine Seite ist populär, wenn viele Seiten auf sie verweisen.

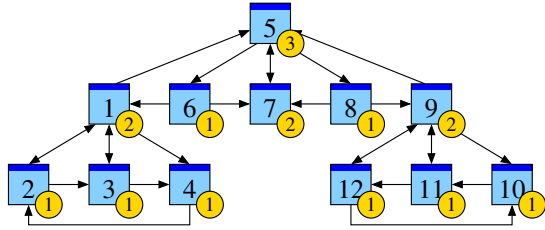
Diese Linkzählung ist zu naiv und anfällig für Manipulationen.

Zweiter Versuch: Eine Seite ist populär, wenn viele populäre Seiten auf sie verweisen.

Das klingt zunächst zirkulär, lässt sich aber in eine einfache Gleichung fassen und lösen.

Ich erläutere die besonders anschauliche Betrachtungsweise des zufälligen Surfers.

## Irrfahrt eines zufälligen Surfers



Aufenthaltswahrscheinlichkeit bei Start auf Seite 7:

	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t = 1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t = 2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t = 3$	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
$t = 4$	.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042
$t = 5$	.118	.021	.021	.021	.111	.139	.250	.139	.118	.021	.021	.021
...												
$t = 29$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t = 30$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Wir folgen einem Surfer im Internet, der auf jeder Seite zufällig einen Link anklickt.

Als Beispiel starten wir zur Zeit  $t = 0$  auf Seite 7. Hier gibt es nur einen Link, nämlich nach 5.

Zur Zeit  $t = 1$  befinden wir uns also auf Seite 5. Von hier gehen drei Links weg.

Zur Zeit  $t = 2$  befinden wir uns auf Seite 6, 7, 8 jeweils mit Wahrscheinlichkeit  $1/3$ .

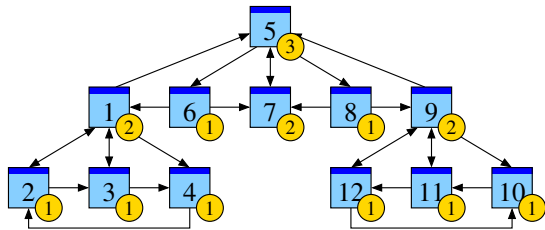
Wir beobachten eine Art Diffusion. Sie konvergiert gegen eine Gleichgewichtsverteilung.

Konvergenz bedeutet: Die Verteilung nähert sich dem Gleichgewicht beliebig genau an.

Beim Start in 1 finden wir eine Diffusion, und diese konvergiert zum selben Gleichgewicht!

Dank dieser Betrachtungsweise löst sich unser Problem sozusagen von allein!

## Zufall und Notwendigkeit



Aufenthaltswahrscheinlichkeit bei Start auf Seite 1:

	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t = 1$	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t = 2$	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
$t = 3$	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
$t = 4$	.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010
$t = 5$	.233	.126	.126	.126	.118	.050	.109	.050	.045	.005	.005	.005
...												
$t = 69$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t = 70$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Auch der Zufall gehorcht Gesetzen: Man kann damit rechnen!

Zur Zeit  $t$  ist unser Surfer auf der Seite  $j$  mit Wahrscheinlichkeit  $p_j$ .

Er wählt zufällig einen der  $\ell_j$  ausgehenden Links.

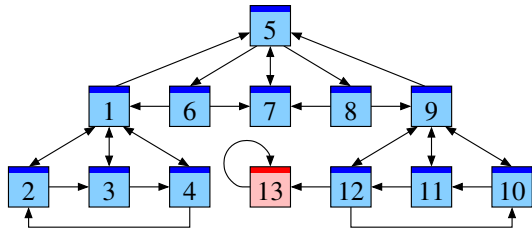
Also folgt er dem Link  $j \rightarrow i$  mit Wahrscheinlichkeit  $\frac{1}{\ell_j} p_j$ .

Zur Zeit  $t + 1$  landet er auf der Seite  $i$  mit Wahrscheinlichkeit

$$p'_i = \sum_{j \rightarrow i} \frac{1}{\ell_j} p_j.$$

Wenn man die Aufenthaltswahrscheinlichkeiten zur Zeit  $t$  kennt, so kann man hieraus leicht die Aufenthaltswahrscheinlichkeiten zur Zeit  $t + 1$  berechnen.

## Vorsicht vor schwarzen Löchern!



Zum Schutz vor schwarzen Löchern nutzt Google folgendes Modell:

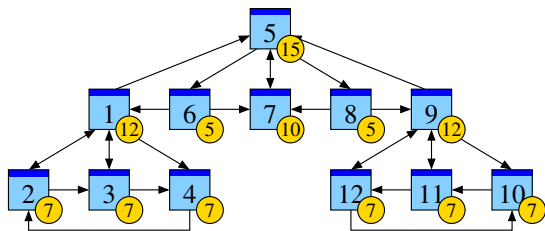
- Teleportation: Mit Wkt  $q = 15\%$  beginnt der Surfer irgendwo neu.
- Irrfahrt: Andernfalls folgt er zufällig einem Link der aktuellen Seite.

Was passiert, wenn es (Gruppen von) Seiten ohne ausgehende Links gibt? Im Beispiel ist anschaulich klar: Unser Surfer landet früher oder später auf der Seite 13, wo er den Rest seines Lebens verbringt. Hier ist unser Modell nicht realistisch!

Das verfeinerte Modell mit Teleportation löst dieses Problem auf erstaunlich einfache Weise. Die Aufenthaltswahrscheinlichkeiten sind ebenso leicht zu definieren und zu berechnen. Die so berechnete Wahrscheinlichkeitsverteilung entspricht recht gut der Nutzererwartung. Zudem ist sie recht robust gegenüber Manipulationen.

Die Konstante  $q$  können wir frei wählen. Bei  $q = 0$  erhalten wir die Irrfahrt wie zuvor, ohne Teleportation. Bei  $q = 1$  springt der Surfer willkürlich, ohne Ansehen der Links. Ein geeignete Wahl von  $q$  liegt zwischen 0 und 1. Zum Beispiel entspricht  $q = 0.15$  dem Besuch von durchschnittlich etwa 7 aufeinanderfolgenden Seiten. Das entspricht ungefähr dem beobachteten Nutzerverhalten. . . und lässt sich empirisch anpassen.

## Schnelle Diffusion zum Gleichgewicht



Aufenthaltswahrscheinlichkeit für  $q = 0.15$ :

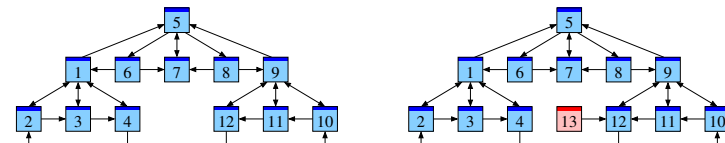
	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t = 1$	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
$t = 2$	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
$t = 3$	.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028
$t = 4$	.180	.105	.105	.105	.140	.057	.075	.057	.057	.040	.040	.040
$t = 5$	.171	.095	.095	.095	.126	.052	.101	.052	.087	.042	.042	.042
...												
$t = 29$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066
$t = 30$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066

Unsere Beobachtungen sind nicht zufällig sondern beruhen auf mathematischen Gesetzmäßigkeiten. Diese kann man beweisen und darf sich darauf verlassen:

Es gibt immer genau eine Gleichgewichtsverteilung  $p$ .

Für jede Anfangsverteilung konvergiert die Diffusion gegen  $p$ .

Die Konvergenz ist mindestens so schnell wie die der geometrischen Folge  $(1 - q)^n$  gegen 0.



Zudem ist das Modell robust gegenüber Manipulationen. Im ersten Beispiel haben wir:

1	2	3	4	5	6	7	8	9	10	11	12
.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066

Einfügen einer Spamseite ändert hieran wenig:

1	2	3	4	5	6	7	8	9	10	11	12	13
.113	.062	.062	.062	.145	.053	.097	.053	.126	.071	.069	.077	.012

## Fazit

Webinhalte sind dezentral, heterogen und meist **wenig strukturiert**.

Eine Suchmaschine soll Suchergebnisse **auflisten**.

Hierzu muss sie bewerten und nutzergerecht **sortieren**.

Aus Webseiten und Links ermittelt Google die **Popularität**.

$$p_i = \frac{q}{n} + \sum_{j \rightarrow i} \frac{1-q}{\ell_j} p_j$$

Einfach aber erfolgreich. Verfeinerungen bleiben Betriebsgeheimnis.

Vielen Dank für Ihre Aufmerksamkeit!

[www.igt.uni-stuttgart.de/eiserm/popularisierung/#SciencePub](http://www.igt.uni-stuttgart.de/eiserm/popularisierung/#SciencePub)

## Fazit

Ich habe hier als einfaches Beispiel die Grundidee der Suchmaschine Google skizziert. Die Besonderheit von Hypertext sind die gegenseitigen Links: Millionen von Autoren lesen gegenseitig ihre Webseiten, und ihre Bewertung schlägt sich in den Links nieder. Das Modell der Irrfahrt berechnet hieraus ein Maß der Popularität. Hauptargument für das Modell ist sein Erfolg: Die entstehende Sortierung scheint den Nutzererwartungen recht nahe zu kommen.

Zu Beginn sah sich Google rein deskriptiv: Wenn eine Seite relevant ist, dann steht sie oben auf der Liste. Sein überwältigender Erfolg macht Google normativ: Wenn eine Seite oben auf der Liste steht, dann ist sie relevant. Für kommerzielle Seiten ist die Optimierung daher inzwischen unerlässlich und zu einer eigenen Industrie geworden (*search engine optimization*, SEO).

Offensichtliche Strategie: Viele Links anlocken, am besten von anderen populären Seiten, und selbst nur gut gewählte Links setzen. Somit verändert die Allgegenwart von Google das Verhalten der Autoren . . . und damit die Grundannahme des Modells! Modelle und Werkzeuge werden ständig weiterentwickelt, leider wuchert auch der Webspam. Nutzer, Autoren und Suchmaschinen durchlaufen eine Art Evolution. . . Es bleibt spannend.