

Powerlaw Netzwerke



<https://iuk.one/1066-1008>

Clemens H. Cap
ORCID: 0000-0003-3958-6136

Department of Computer Science
University of Rostock
Rostock, Germany
clemens.cap@uni-rostock.de

Version 1



1. Zipfsches Gesetz

Ein erstes Beispiel für ein Powerlaw.
Aus dem Gebiet der Linguistik.

1. Zipfsches Gesetz

2. Gesetz von Pareto

3. Graphentheoretische Analyse

4. Powerlaw Verteilungen

5. Folge für Web 2.0

1. Zipfsches Gesetz

Zipfsches Gesetz

Entdeckung: Linguist George Zipf.

Empirische Beobachtung:

- Das 2-häufigste Wort einer Sprache tritt $1/2$ so oft auf wie das häufigste.
- Das 3-häufigste Wort einer Sprache tritt $1/3$ so oft auf wie das häufigste.
- Das 4-häufigste Wort einer Sprache tritt $1/4$ so oft auf wie das häufigste.
- Das k -häufigste Wort einer Sprache tritt $1/k$ so oft auf wie das häufigste.

Nutzung in Informatik: Optimierung von Speicherstrukturen in Text-Suche.

Allgemein: Relative Häufigkeit des k -häufigsten Wortes unter insgesamt N Worten:

$$f(k, s, N) = \frac{1/k^s}{\sum_{n=1}^N 1/n^s} \quad s \text{ ist Modellierungsparameter}$$

Struktur des Gesetzes: Häufigkeit ist proportional zu $1 /$ Potenz des Rangs.
Daher auch der Name **Powerlaw**.

$$f(k) = \frac{C}{k^s}$$

Bei **doppelt logarithmischer Skala** ergibt sich eine Geradengleichung:

$$\log(f) = \log(C) - s \cdot \log(k)$$

1. Zipfsches Gesetz

Beispiel: Worthäufigkeiten eines Romans

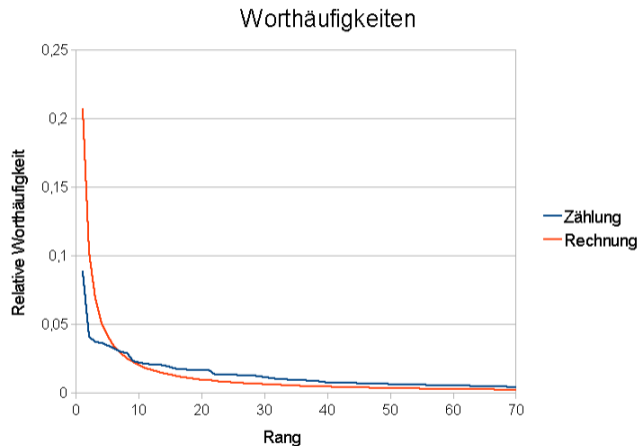


Abb. 1: Verteilung der Worthäufigkeiten im Roman Effi Briest von Theodor Fontane, im Vergleich mit dem theoretischen Modell. © Rechte siehe Anhang.

1. Zipfsches Gesetz

Beispiel: Tag-Häufigkeit bei Del.icio.us

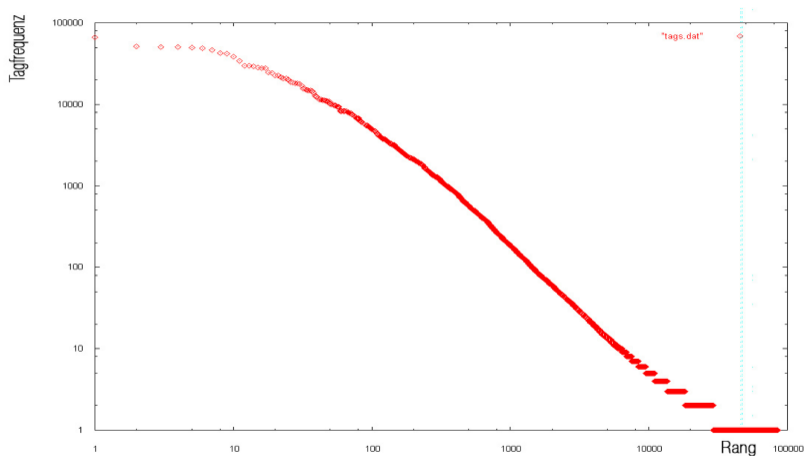


Abb. 2: Verteilung der Häufigkeiten von Tags in der Social Bookmarking Webseite <https://del.icio.us>

© Rechte siehe Anhang.

1. Zipfsches Gesetz

Beispiel: Worthäufigkeiten in der Wikipedia

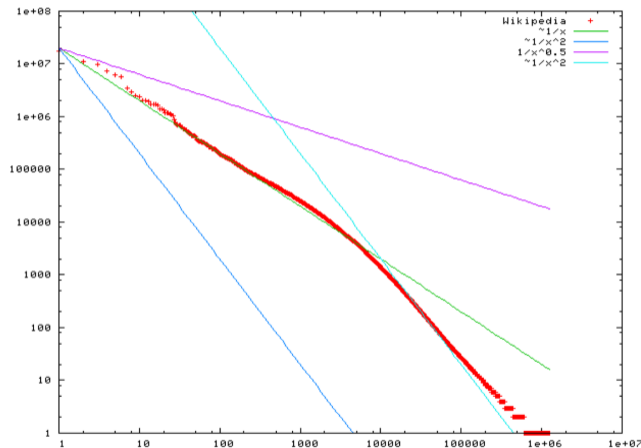


Abb. 3: Verteilung der Worthäufigkeiten in der englischen Wikipedia © Rechte siehe Anhang.

2. Gesetz von Pareto

Ein zweites Beispiel für ein Powerlaw.
Aus der ökonomischen Modellierung.

1. Zipfsches Gesetz
2. **Gesetz von Pareto**
3. Graphentheoretische Analyse
4. Powerlaw Verteilungen
5. Folge für Web 2.0

Pareto Modell (1)

Empirische Beobachtung: 80 – 20 Regel

- 80% der Einkommens in Italien werden von den 20% Reichsten erzielt.
- Die ersten 20% im Rang haben 80% im Wert.

Wahrscheinlichkeit, mehr als den **Minimalbesitz** zu besitzen ist 1:

$$P(X > x_{\min}) = 1$$

Für **größere Werte**: Kumulative Wahrscheinlichkeit invers proportional zu x .

$$P(X > x) = \left(\frac{x}{x_{\min}} \right)^{-k} = Cx^{-k} \quad \text{kist Modellierungsparameter}$$

Bei **doppelt logarithmischer Skala** ergibt sich eine Geradengleichung:

$$\log(P) = \log(C) - k \cdot \log(x)$$

Pareto Modell (2)

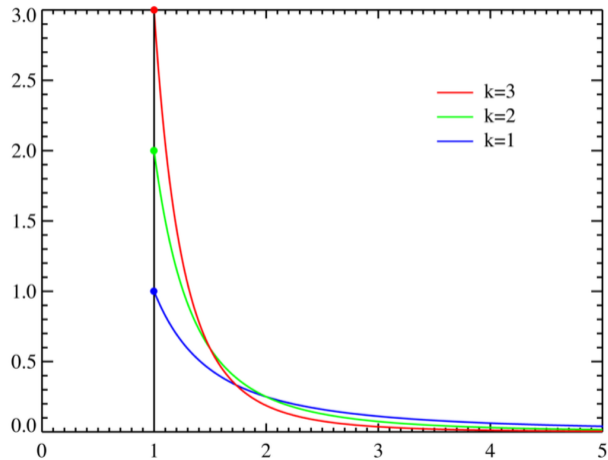


Abb. 4: Modellierung von Pareto für verschiedene Exponenten

2. Gesetz von Pareto

Beispiel: Forbes Top-500 Liste (1)

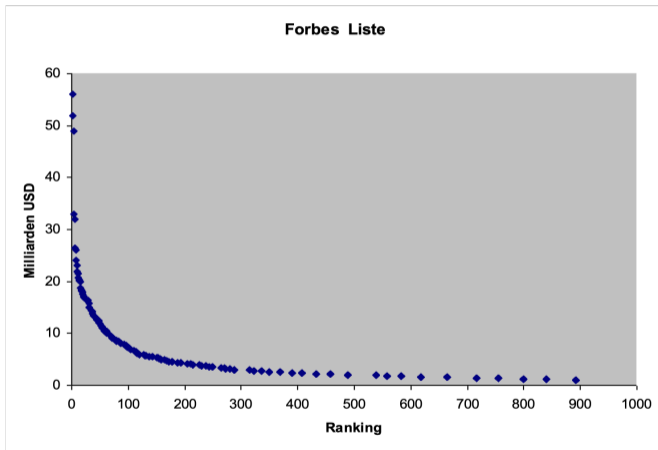


Abb. 5: Analyse der Reichtumsverteilung nach der Forbes Top-500 Liste

2. Gesetz von Pareto

Beispiel: Forbes Top-500 Liste (2)

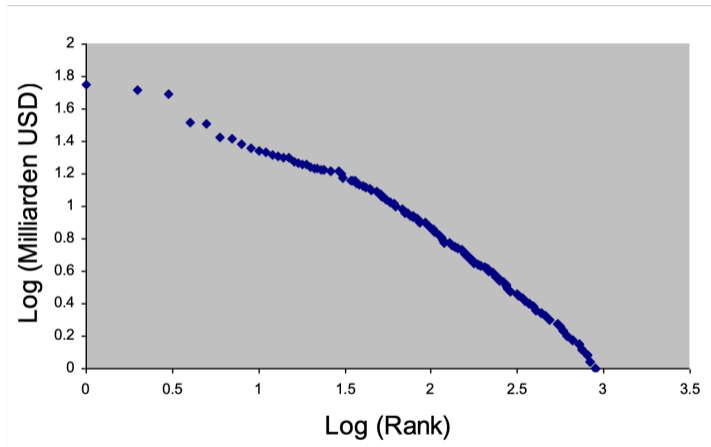


Abb. 6: Analyse der Reichtumsverteilung nach der Forbes Top-500 Liste, in doppelt logarithmischer Skala angegeben. Eine über weite Strecken lineare Kurve ist erkennbar.

2. Gesetz von Pareto

Beispiel: Forbes Top-500 Liste (3)

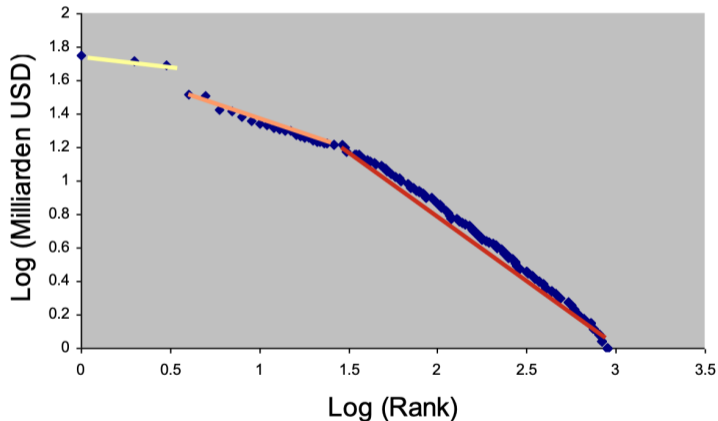


Abb. 7: Analyse der Reichtumsverteilung nach der Forbes Top-500 Liste, in doppelt logarithmischer Skala angegeben. Eine genauere Analyse lässt drei Linearitätsbereiche erkennen.

3. Graphentheoretische Analyse

3.1. Erdős und Renyi

3.2. Albert und Barabasi

Wie sehen Wachstumsprozesse
in vernetzten Strukturen aus?

1. Zipfsches Gesetz

2. Gesetz von Pareto

3. Graphentheoretische Analyse

4. Powerlaw Verteilungen

5. Folge für Web 2.0

Allgemeine Fragestellung

Idee:

- Es gibt einen initialen Graphen.
- Ein Wachstumsprozess erzeugt neue Knoten und Kanten.
- Meist ist das ein Zufallsprozess, von dem man “ein wenig” weiß.

Fragen:

- Wie sieht der “typische” Zufallsgraph aus?
- Welche “interessanten” Zufallsgraphen gibt es?
- Welche Konsequenzen hat das für soziale Netze?

Graphklassen von Erdős und Renyi

Idee: Zufälliger Graph (Gleichverteilung)

- 2 mögliche Modelle: Zufälliger Graph und Zufällige Kante

Zufälliger Graph:

- Parameter: n Knoten, e Kanten.
- Unter allen Graphen mit n Knoten und e Kanten wählt man einen zufällig aus.
- Auswahl geschieht gleichverteilt.

Zufällige Kante:

- Parameter: n Knoten, p Kanten-Wahrscheinlichkeit.
- Graphen mit n Knoten konstruieren.
- Für jede mögliche Kante: Wahrscheinlichkeit ihrer Realisierung ist p .
- Wahrscheinlichkeit für je zwei Kanten voneinander unabhängig.

Unterschiede in den Modellen:

- Die zwei Konstruktions-Varianten unterscheiden sich “nicht wesentlich”.
- Genauer: Wenn

$$e \approx \binom{n}{2} p$$

dann sehen die Graphklassen bei großem n “ähnlich” aus.

Verteilung der Knotengrade:

- **Binomial**-Verteilung, bei großen Graphen näherungsweise Gauß-Verteilung.
- Viele Knoten haben eine mittlere Kantenzahl.
- Wenig Knoten sind sehr gut oder sehr schlecht vernetzt.

Graphklasse von Erdős und Barabasi

Idee: Preferential Attachment Modell.

- Netz beginnt mit n Knoten.
- Es kommen neue Knoten hinzu.
- Jeder neue Knoten verbindet sich zufällig mit alten Knoten.
- Wahrscheinlichkeit proportional zur Zahl der Kanten, die alter Knoten bereits hat.

Verteilung der Knotengrade:

- **Powerlaw**-Verteilung
- Fast alle Knoten haben sehr wenig Kanten.
- Sehr wenig Knoten sind außergewöhnlich gut vernetzt.

Anschaulich:

- Wer schon viele Verbindungen hat kriegt noch besonders viele weitere dazu.
- Auch als “Mattheus Effekt” bezeichnet. (“Wer hat dem wird gegeben”).

Vergleich der Graphenklassen

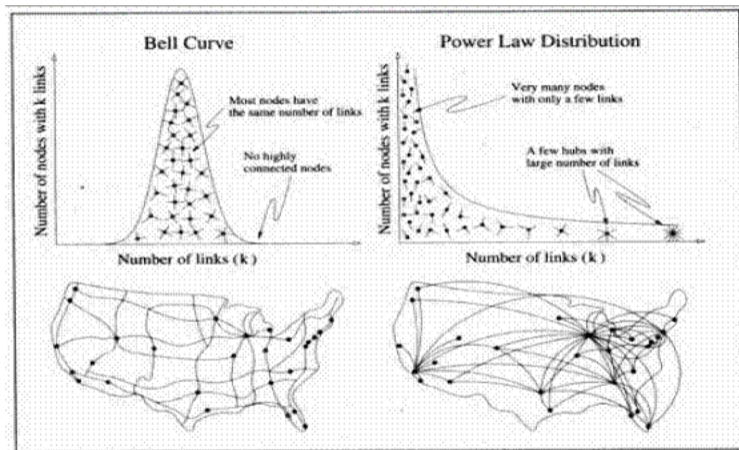


Abb. 8: Vergleich der Graphenklasse von Erdős und Renyi (links) mit der Graphenklasse von Albert und Barabasi (rechts).

Typischer Graph der Klasse von Albert und Barabasi

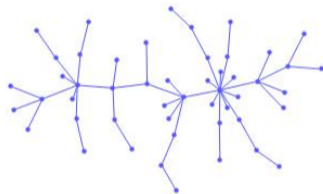


Abb. 9: Ein typisches Netzwerk der Netzklasse von Albert und Barabasi © Rechte siehe Anhang.

4. Powerlaw Verteilungen

Wo beobachten wir Powerlaw Verteilungen?

Überall dort, wo wir
Preferential Attachment finden!

1. Zipfsches Gesetz
2. Gesetz von Pareto
3. Graphentheoretische Analyse
4. Powerlaw Verteilungen
5. Folge für Web 2.0

4. Powerlaw Verteilungen

Beispiele für Powerlaw Verteilungen (1)

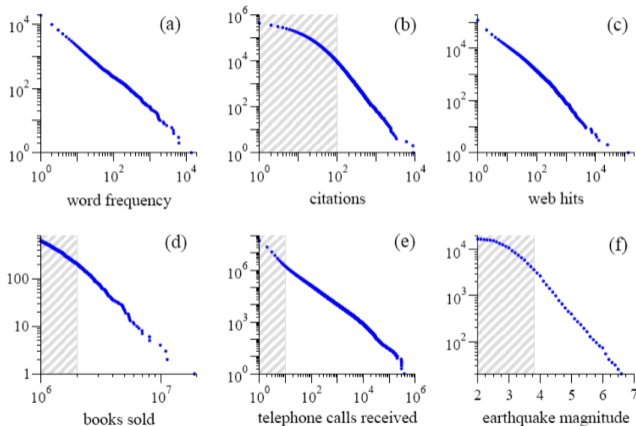


Abb. 10: Phänomene, bei denen Powerlaws empirisch beobachtet wurden und die verschieden gut durch Preferential Attachment Modell erklärt werden können. © Rechte siehe Anhang.

4. Powerlaw Verteilungen

Beispiele für Powerlaw Verteilungen (2)

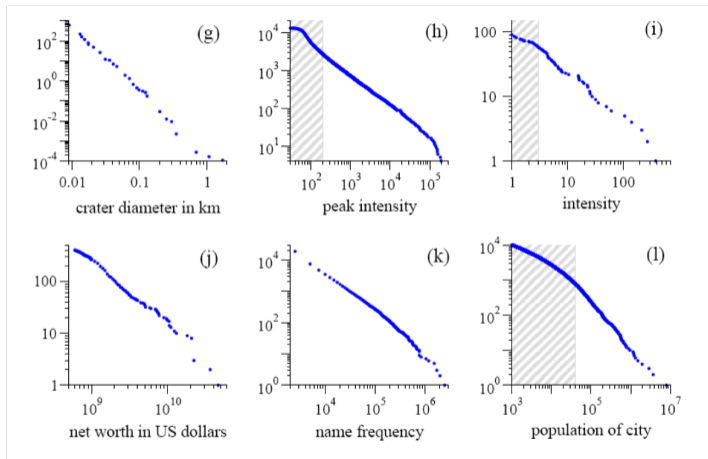


Abb. 11: Phänomene, bei denen Powerlaws empirisch beobachtet wurden und die verschieden gut durch Preferential Attachment Modell erklärt werden können. © Rechte siehe Anhang.

4. Powerlaw Verteilungen

Beispiel: Wikipedia Linkstruktur

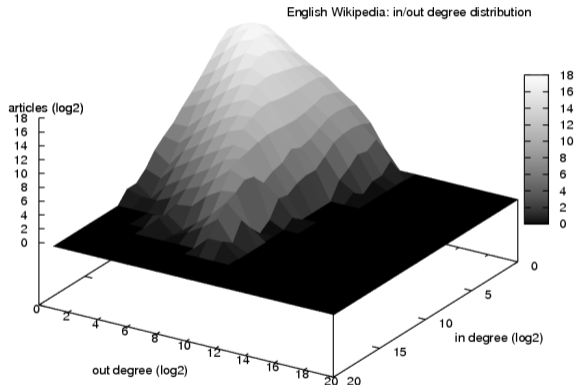


Abb. 12: Verteilung von Indegree und Outdegree im Artikel-Netzwerk der englischen Wikipedia auf doppelt-logarithmischer Skala dargestellt © Rechte siehe Anhang.

4. Powerlaw Verteilungen

Beispiel: Verlinkung universitärer Webseiten

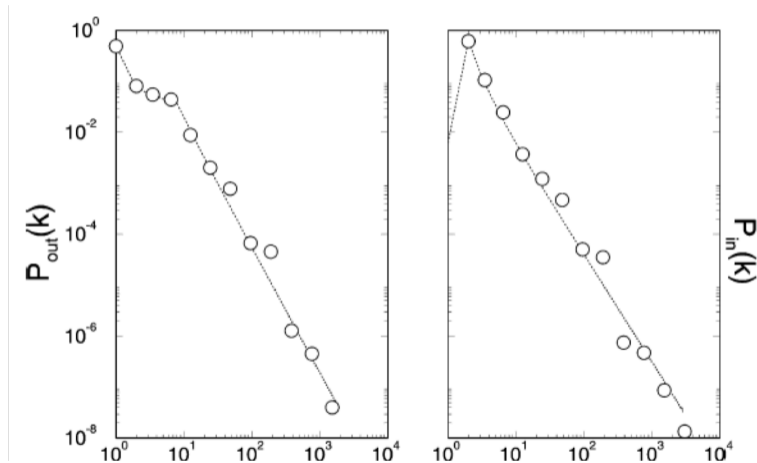


Abb. 13: Verteilung von Indegree und Outdegree bei der Verlinkung universitärer Webseiten

Scale Free Laws

Motivation: File System Skalierung.

- Files der Größe 1 KB sind 4 mal so häufig wie Files der Größe $2 \times 1 \text{ KB} = 2 \text{ KB}$
- Files der Größe 1 MB sind 4 mal so häufig wie Files der Größe $2 \times 1 \text{ MB} = 2 \text{ MB}$
- Files der Größe 1 GB sind 4 mal so häufig wie Files der Größe $2 \times 1 \text{ GB} = 2 \text{ GB}$

Die Verdoppelung der Filegröße reduziert die Häufigkeit um den Faktor 4. Die Beobachtung gilt unabhängig vom Ort auf der Skala.

Definition: Eine Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ heißt **skalenfrei**, wenn es eine Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ gibt mit

$$f(a \cdot x) = g(a) \cdot f(x)$$

Beispiel: Power Laws sind skalenfrei.

$$f(a \cdot x) = C \cdot (a \cdot x)^s = \underbrace{C \cdot a^s}_{g(a)} \cdot \underbrace{x^s}_{f(x)} = g(a) \cdot f(x)$$

5. Folge für Web 2.0

Was lernen wir aus dem Preferential Attachment Modell?

1. Zipfsches Gesetz
2. Gesetz von Pareto
3. Graphentheoretische Analyse
4. Powerlaw Verteilungen
5. Folge für Web 2.0

Theorie der Web-Entwicklung

Annahme:

- Netzwerk aus Knoten und Kanten.
- Knoten sind Sites / Blogs / Artikel.
- Kanten sind Links.
- Netz wächst: Neue Knoten, neue Kanten.
- Hoher Wert = gut verlinkt = hoher Google Page-Rank.

Modell von Erdős und Renyi

- Kann das beobachtete Power Law nicht erzeugen.
- Klassischer Fall von Falsifikation einer Theorie.

Modell von Barabasi und Albert

- Preferential Attachment Modelle erzeugen Powerlaw Netzwerke.
- Empirische Beobachtung findet Powerlaw Netzwerke.

Beobachtung und Strategien

Beobachtung:

- Netzwerke mit Power Law Verhalten werden von 2 – 3 Big Players monopolisiert.
- Die Masse wird nicht gesehen und geht in Bedeutungslosigkeit unter.
- Wichtigste Strategie für Wachstum ist Sichtbarkeit.

Strategien, die sich daraus ableiten lassen:

- **Technologie:** Maximale automatische Verlinkung (Wikis, Blogs).
- **Inhalte:** Maximal aufreißerische Inhalte.
- **Prozesse:** Maximale Interaktivität (Kommentare, Feedback).

Folgen und Probleme:

- Hype-driven Content niederen Anspruchs im Boulevard-Stil für den Massenmarkt.
- Führt zu Monopolpositionen.
- Zeichnet sich ein Monopolist ab, so ist Konkurrenz fast zwecklos.
- Wer aus den Regeln ausschert bleibt unsichtbar.
- Wer die Regeln beachtet wird sichtbar.
- Statistische Modelle kennen keine Medienethik.

High Peak und Long Tail



Abb. 14: Eine Powerlaw Verteilung ist charakterisiert durch einen High Peak (grün) und einen Long Tail (gelb).

© Rechte siehe Anhang.

Anderson und Pareto

Frage: Was ist besser?

- Sehr selten sehr viel verdienen. Modell "Ferrari-Händler".
- Sehr oft sehr wenig verdienen. Modell "Amazon" Jede Nische bedienen.

These von Anderson: Long Tail ist besser: Sell more of less.

- Konzentriere Dich auf den Long Tail.
- Das Integral über den Long Tail enthält sehr viele individuelle Nischen.
- Digitale kann diesen Long Tail erschließen.

Antithese von Pareto: Big Peak ist besser: 80% des Effekts in 20% der Produkte.

Stand der Diskussion: Es dürfte eher die Pareto These stimmen.

- Empirisch: Verkaufskurven steilen sich zunehmend noch auf.
- Erziele 90% des Verkaufserfolgs mit 10% der Produkte.

Anhang

Übersicht

Rechtsnachweise



Rechtliche Hinweise



Zitierweise dieses Dokuments



Verzeichnis aller Folien



Verzeichnis aller Abbildungen

Abb. 1 Quelle: <https://commons.wikimedia.org/wiki/File:Zipf-Verteilungn.png>, Berklas, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>

Abb. 2 Lukas Zielinski, Diplomarbeit am Lehrstuhl IuK.

Abb. 3 Quelle: https://en.wikipedia.org/wiki/Talk%3AZipf's_law und <http://oc-co.org/?p=79>, Victor Grishchenko

Abb. 9 Quelle: https://commons.wikimedia.org/wiki/File:Barabasi_Albert_generated_network.jpg, Yves berset, CC BY-SA 3.0, <https://creativecommons.org/licenses/by-sa/3.0>

Abb. 10 Newman, Power Laws, Pareto Distributions and Zipf's Law

Abb. 11 Newman, Power Laws, Pareto Distributions and Zipf's Law

Abb. 12 Quelle: <https://commons.wikimedia.org/wiki/File:Enwiki-degree-distribution.png>, Gritzko at English Wikipedia, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>

Abb. 14 Quelle: https://commons.wikimedia.org/wiki/File:Long_tail.svg, User:Husky, Public domain, via Wikimedia Commons

Rechtliche Hinweise (1)

Die hier angebotenen Inhalte unterliegen deutschem Urheberrecht. Inhalte Dritter werden unter Nennung der Rechtsgrundlage ihrer Nutzung und der geltenden Lizenzbestimmungen hier angeführt. Auf das Literaturverzeichnis wird verwiesen. Das **Zitatrecht** in dem für wissenschaftliche Werke üblichen Ausmaß wird beansprucht. Wenn Sie eine Urheberrechtsverletzung erkennen, so bitten wir um Hinweis an den auf der Titelseite genannten Autor und werden entsprechende Inhalte sofort entfernen oder fehlende Rechtsnennungen nachholen. Bei Produkt- und Firmennamen können Markenrechte Dritter bestehen. Verweise und Verlinkungen wurden zum Zeitpunkt des Setzens der Verweise überprüft; sie dienen der Information des Lesers. Der Autor macht sich die Inhalte, auch in der Form, wie sie zum Zeitpunkt des Setzens des Verweises vorlagen, nicht zu eigen und kann diese nicht laufend auf Veränderungen überprüfen.

Alle sonstigen, hier nicht angeführten Inhalte unterliegen dem Copyright des Autors, Prof. Dr. Clemens Cap, ©2020. Wenn Sie diese Inhalte nützlich finden, können Sie darauf verlinken oder sie zitieren. Jede weitere Verbreitung, Speicherung, Vervielfältigung oder sonstige Verwertung außerhalb der Grenzen des Urheberrechts bedarf der schriftlichen Zustimmung des Rechteinhabers. Dieses dient der Sicherung der Aktualität der Inhalte und soll dem Autor auch die Einhaltung urheberrechtlicher Einschränkungen wie beispielsweise **Par 60a UrhG** ermöglichen.

Die Bereitstellung der Inhalte erfolgt hier zur persönlichen Information des Lesers. Eine Haftung für mittelbare oder unmittelbare Schäden wird im maximal rechtlich zulässigen Ausmaß ausgeschlossen, mit Ausnahme von Vorsatz und grober Fahrlässigkeit. Eine Garantie für den Fortbestand dieses Informationsangebots wird nicht gegeben.

Die Anfertigung einer persönlichen Sicherungskopie für die private, nicht gewerbliche und nicht öffentliche Nutzung ist zulässig, sofern sie nicht von einer offensichtlich rechtswidrig hergestellten oder zugänglich gemachten Vorlage stammt.

Use of Logos and Trademark Symbols: The logos and trademark symbols used here are the property of their respective owners. The YouTube logo is used according to brand request 2-9753000030769 granted on November 30, 2020. The GitHub logo is property of GitHub Inc. and is used in accordance to the GitHub logo usage conditions <https://github.com/logos> to link to a GitHub account. The Tweedback logo is property of Tweedback GmbH and here is used in accordance to a cooperation contract.

Disclaimer: Die sich immer wieder ändernde Rechtslage für digitale Urheberrechte erzeugt für mich ein nicht unerhebliches Risiko bei der Einbindung von Materialien, deren Status ich nicht oder nur mit unverhältnismäßig hohem Aufwand abklären kann. Ebenso kann ich den Rechteinhabern nicht auf sinnvolle oder einfache Weise ein Honorar zukommen lassen, obwohl ich – und in letzter Konsequenz Sie als Leser – ihre Leistungen nutzen.

Daher binde ich gelegentlich Inhalte nur als Link und nicht durch Framing ein. Lt EuGH Urteil 13.02.2014, C-466/12 ist das unbedenklich, da die benutzten Links ohne Umgehung technischer Sperrern auf im Internet frei verfügbare Inhalte verweisen.

Wenn Sie diese Rechtslage stört, dann setzen Sie sich für eine Modernisierung des völlig veralteten Vergütungssystems für urheberrechtliche Leistungen ein. Bis dahin klicken Sie bitte auf die angegebenen Links und denken Sie darüber nach, warum wir keine für das digitale Zeitalter sinnvoll angepaßte Vergütungssysteme digital erbrachter Leistungen haben.

Zu Risiken und Nebenwirkungen fragen Sie Ihren Rechtsanwalt oder Gesetzgeber.

Weitere Hinweise finden Sie im Netz [hier](#) und [hier](#) oder [hier](#).

Zitierweise dieses Dokuments

Wenn Sie Inhalte aus diesem Werk nutzen oder darauf verweisen wollen, zitieren Sie es bitte wie folgt:

Clemens H. Cap: Powerlaw Netzwerke. Electronic document. <https://iuk.one/1066-1008> 6. 6. 2021.

Bibtex Information: <https://iuk.one/1066-1008.bib>

```
@misc{doc:1066-1008,  
  author      = {Clemens H. Cap},  
  title       = {Powerlaw Netzwerke},  
  year        = {2021},  
  month       = {6},  
  howpublished = {Electronic document},  
  url         = {https://iuk.one/1066-1008}  
}
```

Typographic Information:

Typeset on June 6, 2021

This is pdfTeX, Version 3.14159265-2.6-1.40.21 (TeX Live 2020) kpathsea version 6.3.2




This is pgf in version 3.1.5b

This is preamble-slides.tex myFormat©C.H.Cap

Titelseite	1
1. Zipfsches Gesetz	
Zipfsches Gesetz	3
Mathematik des Zipfschen Gesetzes	4
Beispiel: Worthäufigkeiten eines Romans	5
Beispiel: Tag-Häufigkeit bei Del.icio.us	6
Beispiel: Worthäufigkeiten in der Wikipedia	7
2. Gesetz von Pareto	
Pareto Modell (1)	9
Pareto Modell (2)	10
Beispiel: Forbes Top-500 Liste (1)	11
Beispiel: Forbes Top-500 Liste (2)	12
Beispiel: Forbes Top-500 Liste (3)	13
3. Graphentheoretische Analyse	
Allgemeine Fragestellung	15
3.1. Erdős und Renyi	
Graphklassen von Erdős und Renyi	16
Analyse	17

3.2. Albert und Barabasi	
Graphklasse von Erdős und Barabasi	18
Vergleich der Graphenklassen	19
Typischer Graph der Klasse von Albert und Barabasi	20
4. Powerlaw Verteilungen	
Beispiele für Powerlaw Verteilungen (1)	22
Beispiele für Powerlaw Verteilungen (2)	23
Beispiel: Wikipedia Linkstruktur	24
Beispiel: Verlinkung universitärer Webseiten	25
Scale Free Laws	26
5. Folge für Web 2.0	
Theorie der Web-Entwicklung	28
Beobachtung und Strategien	29
Folgen und Probleme	30
High Peak und Long Tail	31
Anderson und Pareto	32

Legende:

-  Fortsetzungsseite
-  Seite ohne Überschrift
-  Bildseite