On Power-Law Relationships of the Internet Topology

M. Faloutsos, P. Faloutsos und C. Faloutsos

Proseminar Algorithmen für Rechnernetze
Prof. Dr. Christian Schindelhauer
Johannes Wendeberg

Albert-Ludwigs-Universität Freiburg

Stefan Bodenlos
Student der Informatik
Bachelor of Science im 4. Semester

Überblick

- Motivation
- Hintergründe
- Die Gesetzmäßigkeiten des Internets
- Diskussion
- Zusammenfassung

Übersicht

Motivation

Häufige Fragen

- „Wie sieht das Internet aus?“
- „Gibt es topologische Eigenschaften, die sich zeitlich gesehen nicht verändern?“
- „Wie wird es in einem Jahr aussehen?“
- „Wie kann ich internetähnliche Graphen für meine Simulationen generieren?“ [1]

Internettopologie

- Topologie des Internets verstehen und modellieren, für ...
  - Simulationsmodelle des Internets
    - Welche Parameter sollen die Graphengeneratoren verwenden?
    - Sind die generierten Graphen realistisch?
  - Effizientere Protokolle
    - In wie vielen Hops kann ein bestimmter Router erreicht werden?
  - Zukunft des Internets voraussagen

Power-Laws

- Formale Definition:

$$y \propto x^a$$

y	Funktionswert
α	„Proportional zu“
x	Argument
a	Exponent

- Powerlaws sind bereits nachgewiesen bei …
  - Einkommensverteilung (Pareto)
  - Städtegrößen
  - Intensität von Erdbeben
  - Häufigkeit von Wörtern (Zipf) [2]

→ Die wichtige Erkenntnis ist, dass solche Powerlaws im Internet überhaupt existieren!

Übersicht

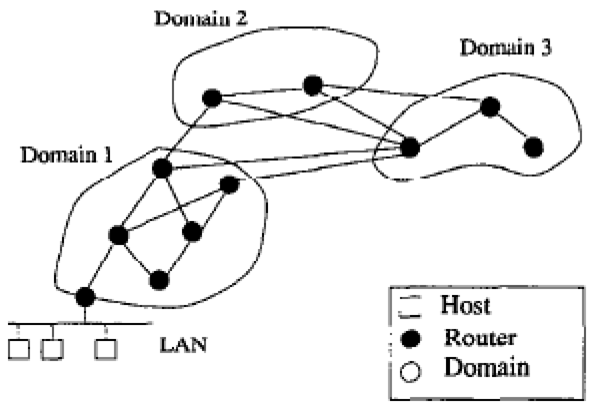
Hintergründe

Repräsentation des Internets

- Internet = Verbundene Subnetzwerke (Domänen)

- Innerhalb eines Sub-netzwerks (Router-Ebene)

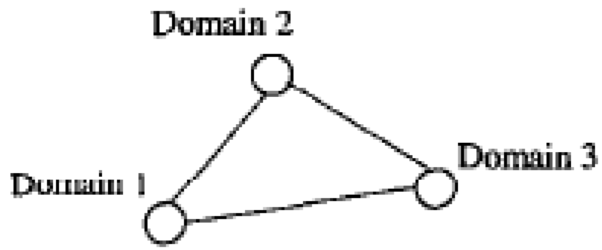
  Jeder Router ist ein Knoten



Bildquelle: [1]

- Zwischen den Subnetzwerken

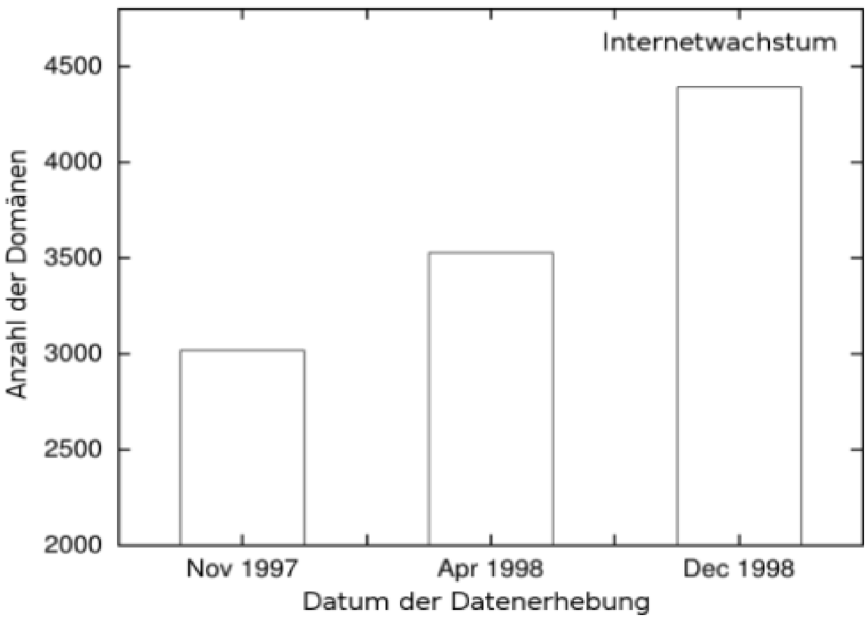
  Jede Domäne ist ein Knoten



Bildquelle: [1]

Verwendete Daten

	Nov. '97	Apr. '98	Dez. '98	1995
Knoten	3015	3530	4389	3888
Kanten	5156	6432	8256	5012
∅ Ausgangsgrad	3,42	3,65	3,76	2,57
Ebene	Router			Domain



- 6-Monate-Intervall
- Wachstum um 45%

Bildquelle: [1]

Vorhergehende Arbeiten

- Metriken für Graphen
  - Durchschnittswerte und Extrema
- Echtzeitnetzwerke
  - 1995: Ein1 Hop Wachstum in der Interdomain-Ebene ≙ zwei Hops Wachstum in der Routerebene
- Generierung von Internetmodellen
  - Transit-Stub: Guter Graphgenerator, der aber Parameter über das Internet benötigt
- Powerlaws in Kommunikationsnetzwerken
  - Beschreiben nur den Traffic in LAN und WAN
  - nicht die Topologie

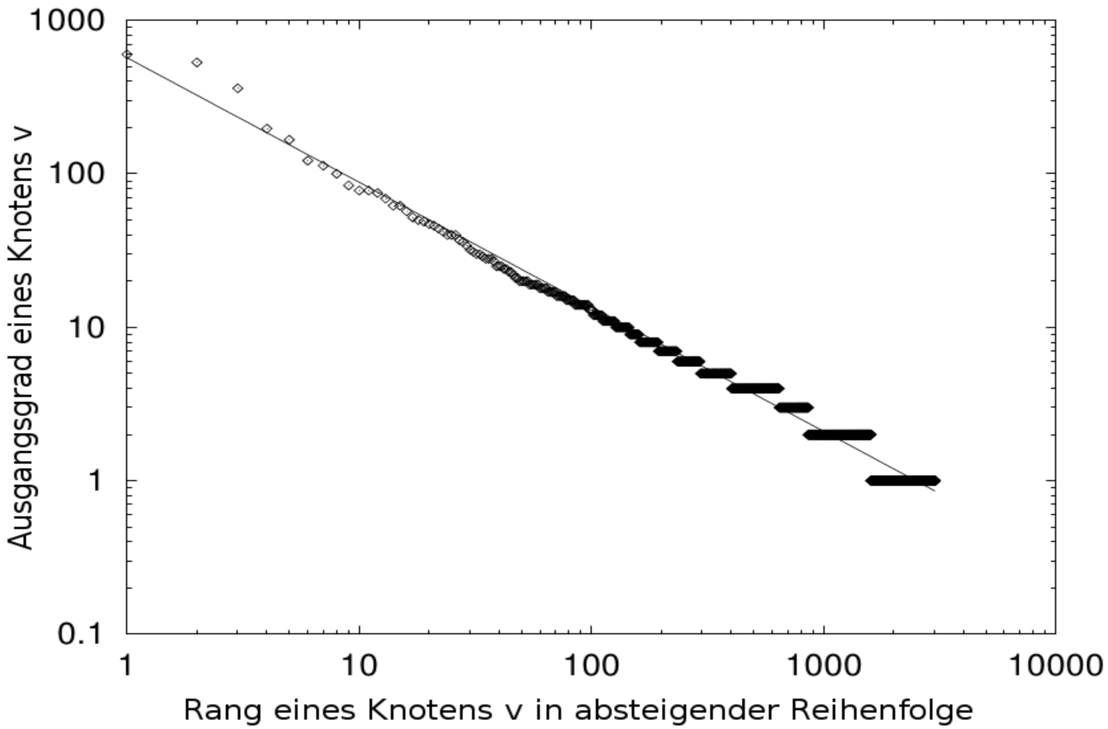
Übersicht

Die Gesetzmäßigkeiten des Internets

1. Power-Law (Rang-Exponent)

Symbol	Definition
d_v	Ausgangsgrad eines Knoten v
r_v	Der Rang eines Knoten v ist sein Index, wenn man alle Knoten absteigend nach ihren Ausgangsgraden ordnet.

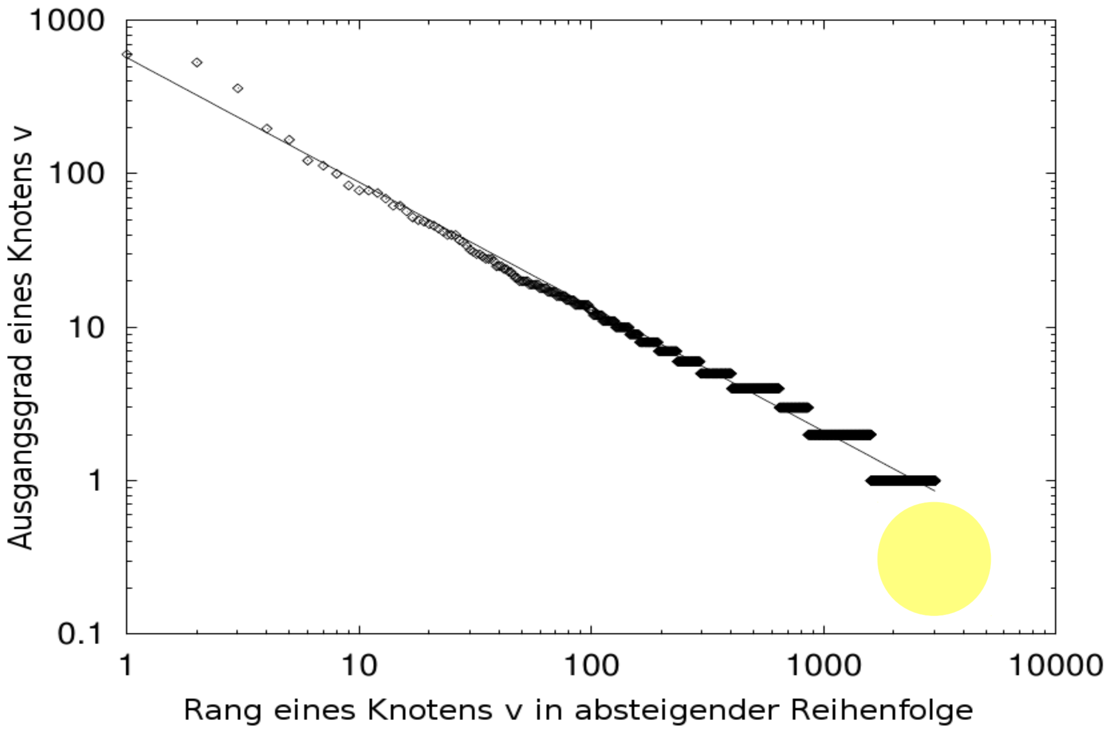
1. Power-Law (Rang-Exponent)



Bildquelle: [1]

- Log-Log-Skaliereung
  - Graph der Funktion ist eine Gerade gdw. die Funktion polynomiell wächst

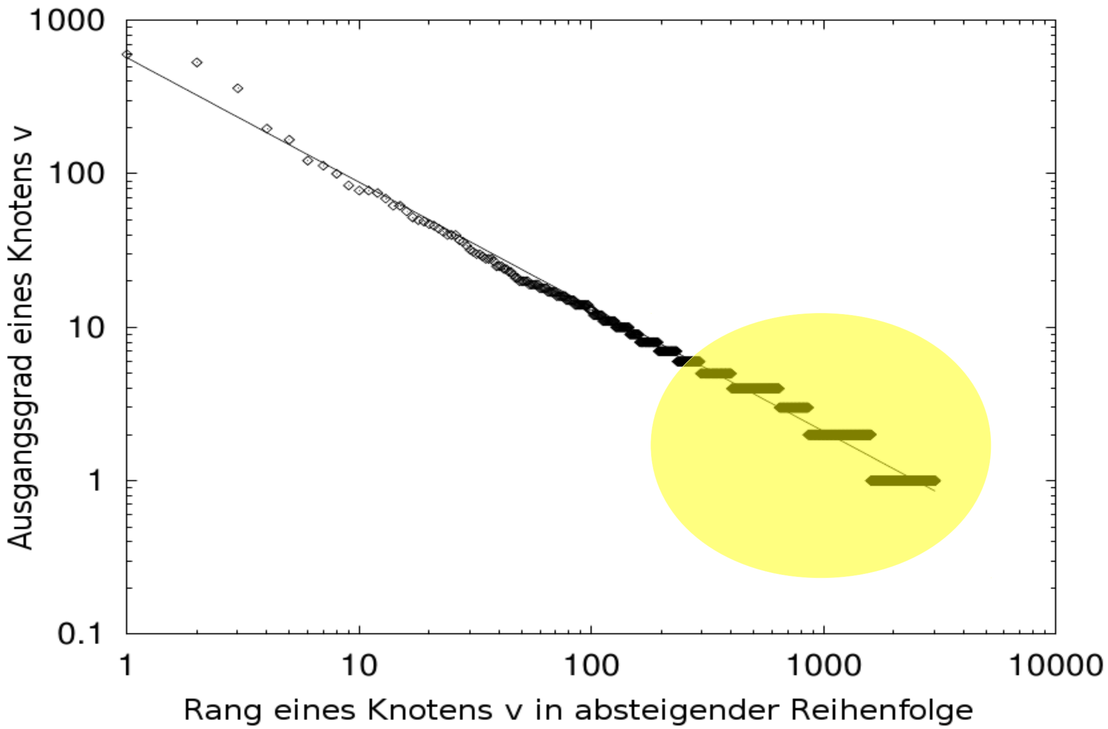
1. Power-Law (Rang-Exponent)



Bildquelle: [1]

- Ausgangsgrad
  - Kann nicht zwischen 0 und 1 kiegen

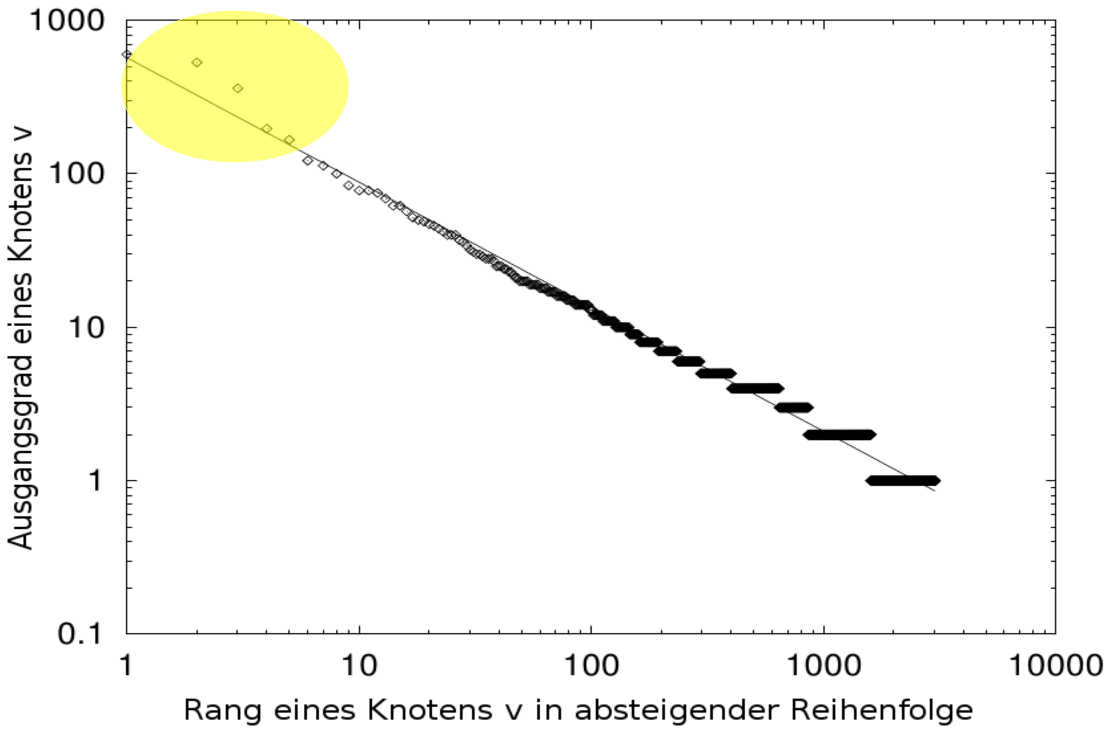
1. Power-Law (Rang-Exponent)



Bildquelle: [1]

- Schwanz der Verteilung
  - Sehr viele Knoten
  - Abstände wegen Skalierung

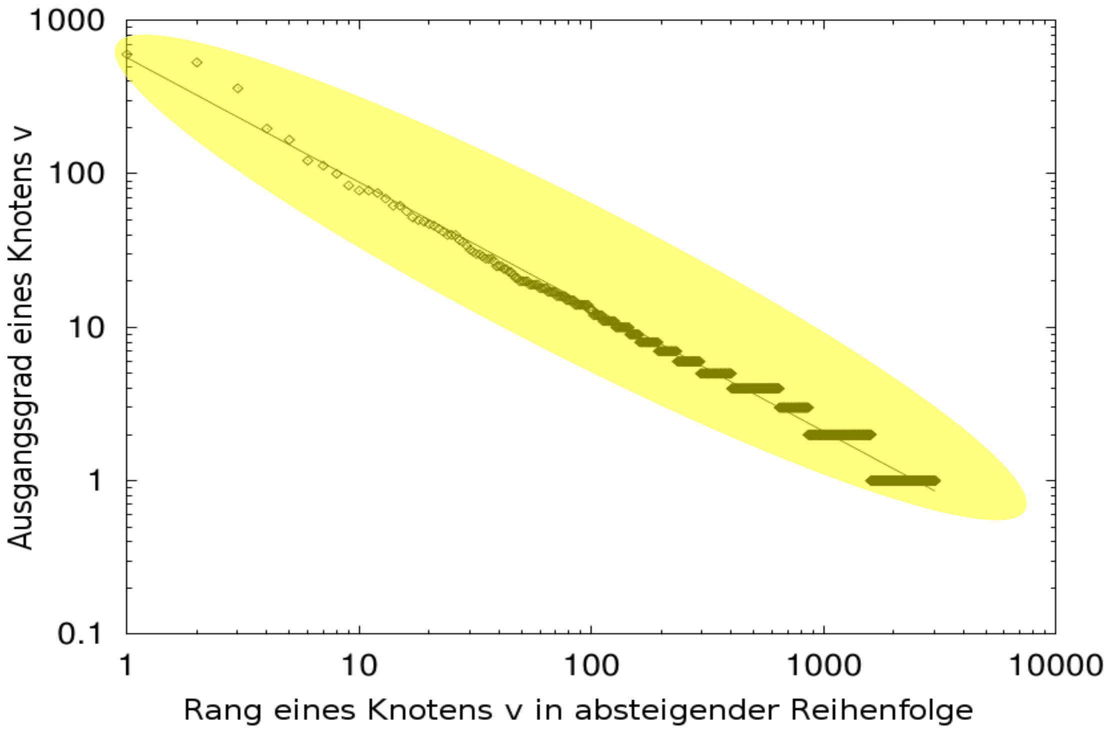
1. Power-Law (Rang-Exponent)



Bildquelle: [1]

- Knoten mit geringem Rang
  - geringe Anzahl

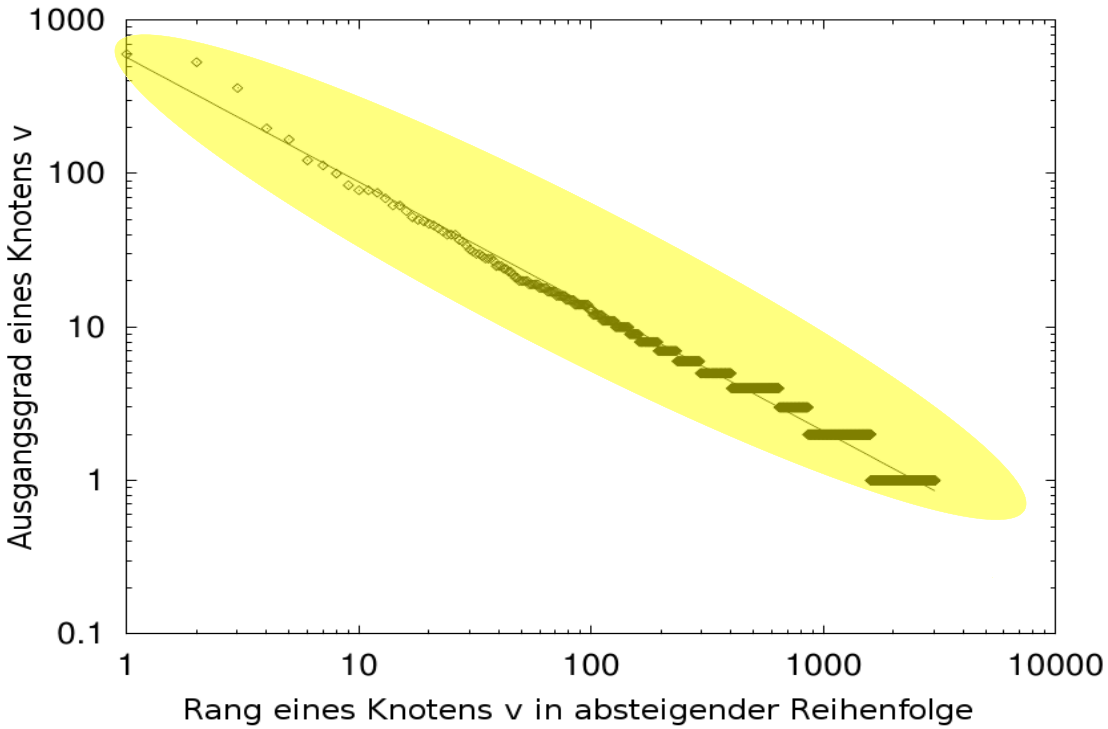
1. Power-Law (Rang-Exponent)



Bildquelle: [1]

- Regressionsgerade
  - Um so präziser, je mehr Knoten es pro Rang gibt
  - Die Steigung ist der Wert des Exponenten

1. Power-Law (Rang-Exponent)



Bildquelle: [1]

- Regressionsgerade
  - Korrelationskoeffizient > 0,974 (Router Ebene)
  - Korrelationskoeffizient = 0,948 (Domain Ebene)

1. Power-Law (Rang-Exponent)

Symbol	Definition
d_v	Ausgangsgrad eines Knoten v
r_v	Der Rang eines Knoten v ist sein Index, wenn man alle Knoten absteigend nach ihren Ausgangsgraden ordnet.
R	Der Rang-Exponent

$$d_v \propto r_v^R$$

1. Power-Law (Rang-Exponent)

- Ergebnisse der linearen Regression:

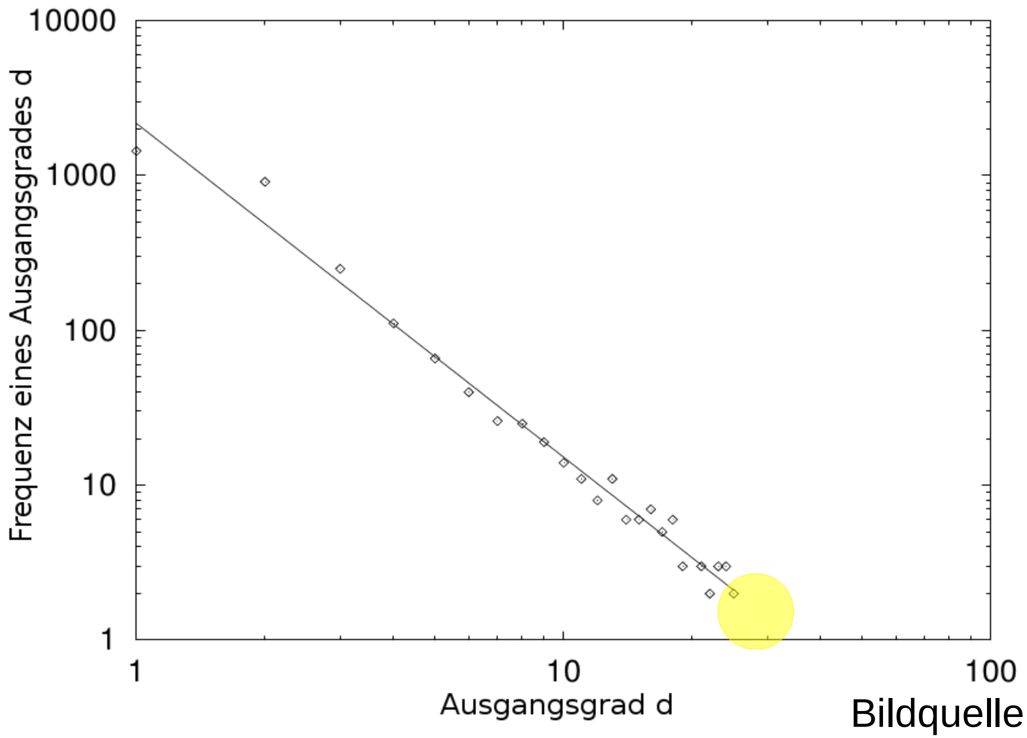
Datensatz	Nov. '97	Apr. '98	Dez. '98	1995
Rang-Exponent R	-0,81	-0,82	-0,74	-0,48

- Der Rangexponent kann Graphenfamilien kennzeichnen
- Gleichgewicht zwischen Nutzen und Kosten einer neuen Kante, aus ökonomischer und funktionaler Perspektive
- Daraus kann man Abschätzungen herleiten

2. Power-Law (Ausgangsgrad-Exponent)

Symbol	Definition
f_d	Die Frequenz, ist die Anzahl der Knoten mit dem Ausgangsgrad d
d	Ausgangsgrad eines Knotens

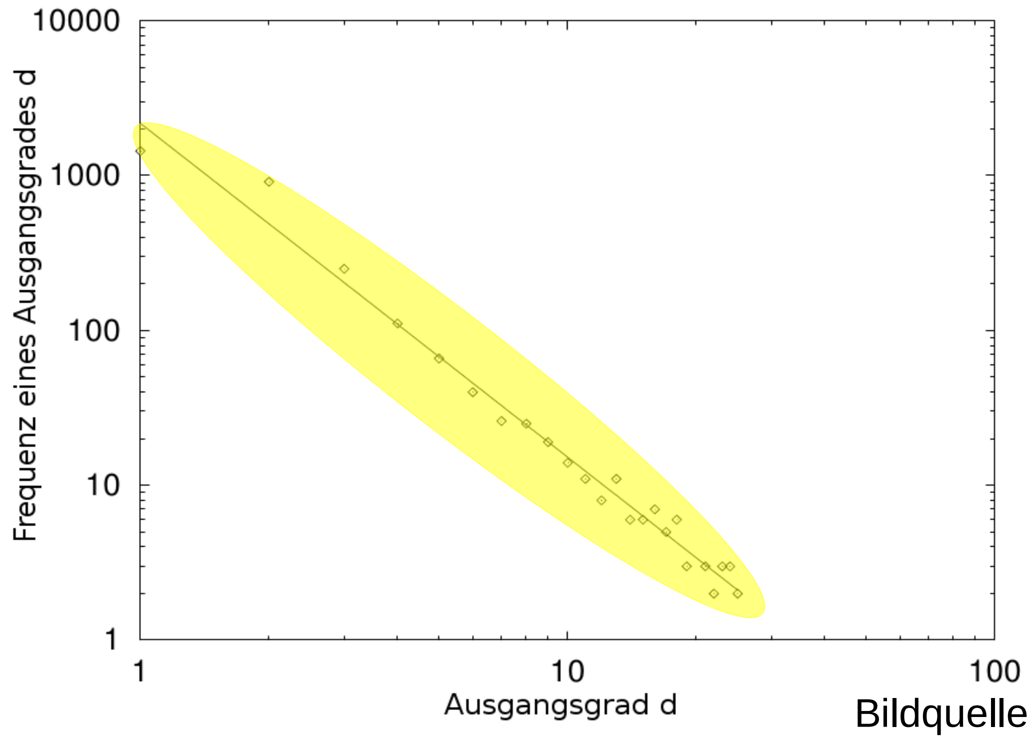
2. Power-Law (Ausgangsgrad-Exponent)



Bildquelle: [1]

- Höhere Ausgangsgrade mit einer Frequenz von 1 sind ausgeschlossen (2% der Knoten)
- Endlastige Verteilung

2. Power-Law (Ausgangsgrad-Exponent)



Bildquelle: [1]

- Regressionsgerade
  - Korrelationskoeffizient 0,968-0,99 (Router Ebene)
  - Korrelationskoeffizient = 0,966 (Domain Ebene)

2. Power-Law (Ausgangsgrad-Exponent)

Symbol	Definition
f_d	Die Frequenz, ist die Anzahl der Knoten mit dem Ausgangsgrad d
d	Ausgangsgrad eines Knotens
O	Der Ausgangsgrad-Exponent

$$f_d \propto d^O$$

2. Power-Law (Ausgangsgrad-Exponent)

- Ergebnisse der linearen Regression:

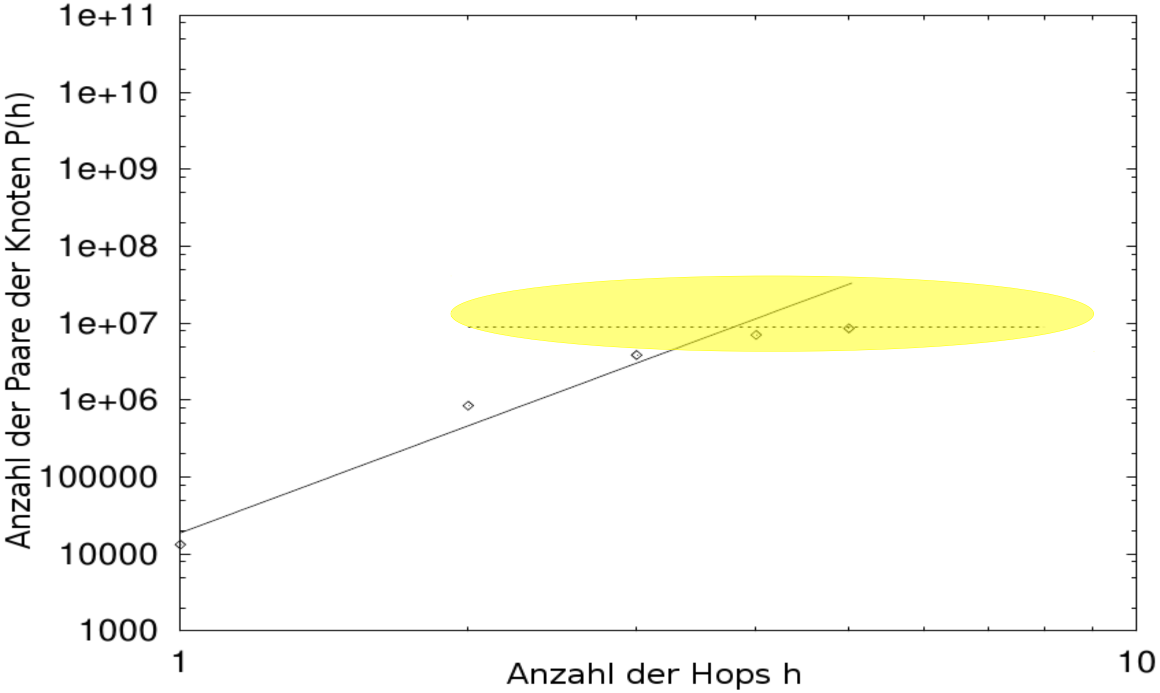
Datensatz	Nov. '97	Apr. '98	Dez. '98	1995
Ausg.-Exponent O	-2,15	-2,16	-2,2	-2,48

- O ist in der Domain- und der Router-Ebene ähnlich
  - Fundamentale Eigenschaft des Internets?
- Mehr Knoten mit niedrigem Ausgangsgrad
  - Keine Zufälligkeit

Approximation (Hop-Plot-Exponent)

- Paare eines Knotens P(h)
  - Sei P(h) die Gesamtzahl aller Knoten die innerhalb h Hops erreicht werden können
  - Inklusive der Selbst-Paare
  - Paare verschiedener Knoten zählen doppelt
- Beispiele
  - $P(0) = N$
  - $P(\delta) = N^2$, δ ist der Durchschnitt des Graphen (Maximum, da alle möglichen Paare)

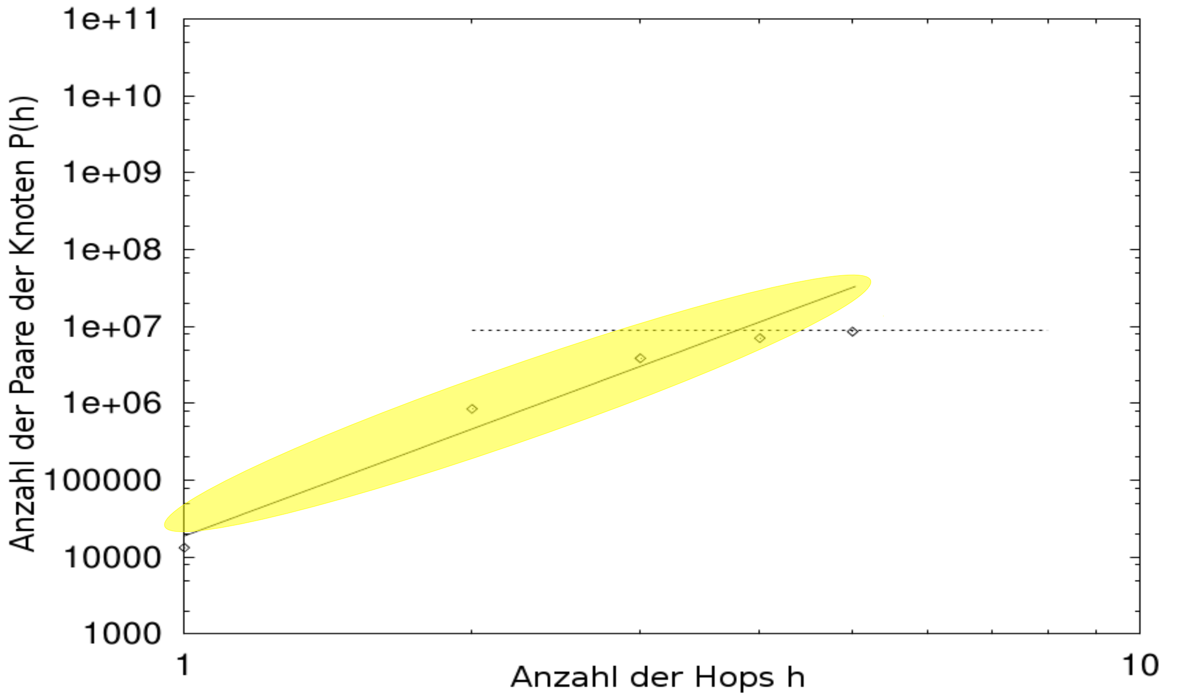
Approximation (Hop-Plot-Exponent)



Bildquelle: [1]

- Maximale Anzahl der Paare (N^2)

Approximation (Hop-Plot-Exponent)



Bildquelle: [1]

- Regressionsgerade, für h ≪ δ
  - Nur vier Punkte! (Router-Ebene)
  - Zwölf Punkte bei der Domainebene

Approximation (Hop-Plot-Exponent)

Symbol	Definition
P(h)	Alle Knoten innerhalb h Hops
h	Anzahl der Hops
H	Der Hop-Plot-Exponent
δ	Der Durchschnitt des Graphen

$$P(h) \propto h^{H}, h \ll \delta$$

Approximation (Hop-Plot-Exponent)

- Ergebnisse der linearen Regression:

Datensatz	Nov. '97	Apr. '98	Dez. '98	1995
Hop-Plot-Exponent H	4,6	4,7	4,86	2,8

- Kann Graphenfamilien exakt kennzeichnen
- Wegen vier Punkten kann dies nur eine Abschätzung sein
- Stärke der Verbundenheit als eine Zahl beschreibbar
- Die Approximation kann verfeinert werden

Approximation (Hop-Plot-Exponent)

- Netzwerk: Ziel erreichen ohne die genaue Position zu kennen
  - Wie groß soll der Broadcast sein?
  - Zu klein: Nicht alle Ziele werden erreicht
  - Zu Groß: Zu viel Traffic und Zeit

Symbol	Definition
δ_{ef}	Effektiver Durchschnitt zweier Knoten e und f
H	Der Hop-Plot-Exponent
N	Anzahl der Knoten in einem Graphen
E	Anzahl der Kanten in einem Graphen

$$\delta_{ef} = \left(\frac{N^2}{N + 2E} \right)^{1/H}$$

Approximation (Hop-Plot-Exponent)

- Effektiver Durchschnitt
  - Die Knoten e und f mit hoher Wahrscheinlichkeit innerhalb δ_{ef} Hops erreichbar
  - Leicht berechenbar
    - Wenn N bekannt (Power-Law 2), dann ist der effektive Durchschnitt des zukünftigen Internets abschätzbar
- Bei unseren Datensätzen
  - $\delta_{ef} = 4$ sind 80% der Knoten mit hoher Wahrscheinlichkeit erreichbar
  - $\delta_{ef} = 5$ sind 95% der Knoten

Approximation (Hop-Plot-Exponent)

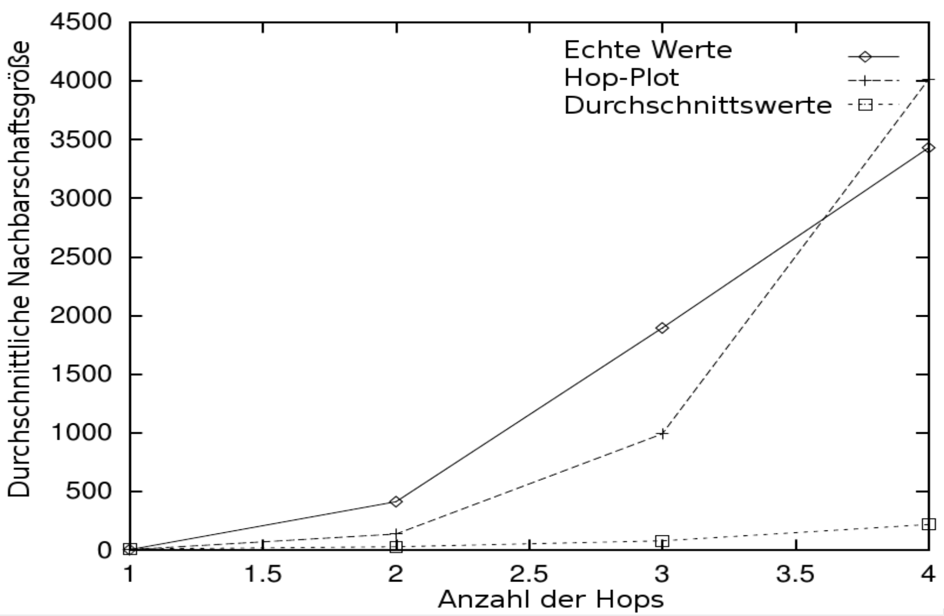
- Durchschnitttliche Nachbarschaftsgröße

Symbol	Definition
NN(h)	Die durchschnittliche Anzahl der Knoten, die in h Sprüngen erreicht werden können.
h	Anzahl der Hops
H	Der Hop-Plot-Exponent
N	Anzahl der Knoten in einem Graphen
E	Anzahl der Kanten in einem Graphen
c	N + 2 E

$$NN(h) = \frac{c}{N} h^H - 1$$

Approximation (Hop-Plot-Exponent)

- Durchschnitttliche Nachbarschaftsgröße
  - Wichtiger Parameter der Protokoll-Performance
  - Besser als Abschätzung anhand Durchschnittswerte
    - Grund: Verteilung der Ausgangsgrade ist sehr schief

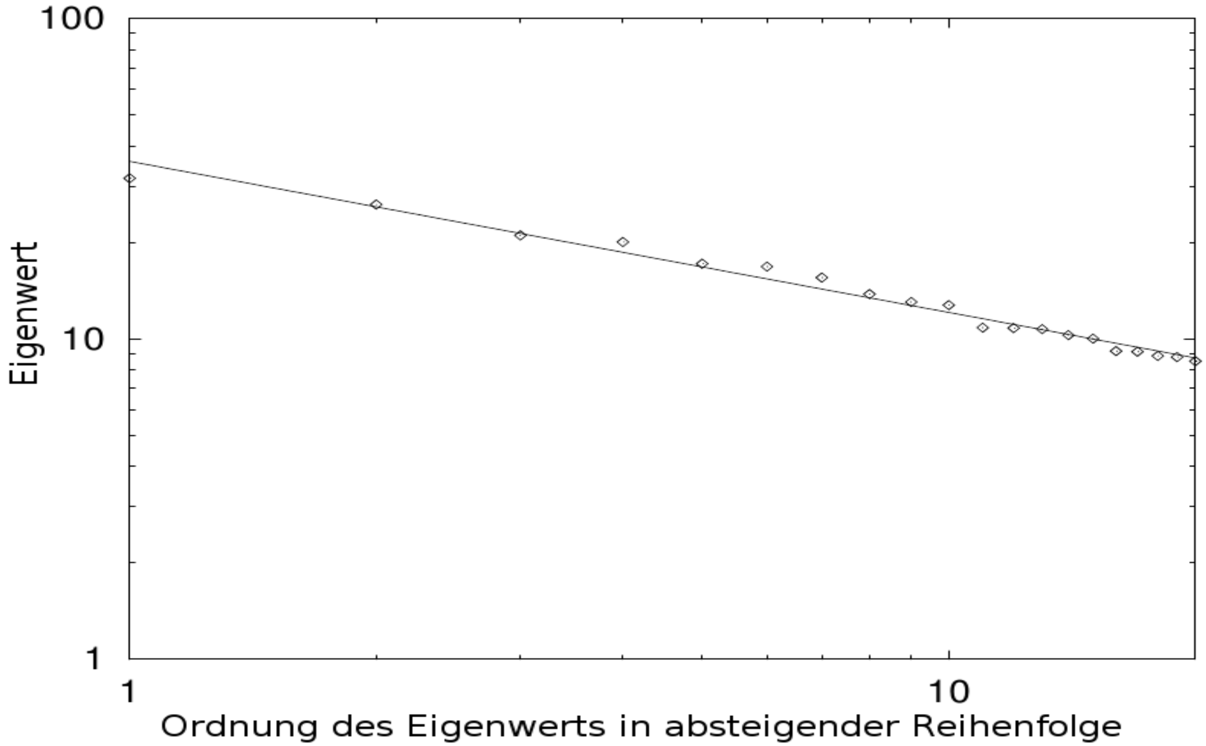


Bildquelle: [1]

Crashkurs: Eigenwerte von Graphen

- Der Graph wir in eine Adjazenzmatrix überführt
- Ein Eigenvektor ist ein Vektor (außer Nullvektor) dessen Richtung sich nicht durch die Abbildung mit einer Matrix verändert
- Der Streckungsfaktor des Vektors ist der Eigenwert der Abbildung
- Aussagekräftig über grundlegende topologische Eigenschaften
  - Anzahl der Kanten, Durchschnitt, Anzahl der verbundenen Knoten uvm.

Power-Law 3 (Eigen-Exponent)



Bildquelle: [1]

- Nur die ersten 20 Eigenwerte
- Korrelationskoeffizient: 0,99 (Bei allen Graphen)

3. Powerlaw (Eigen-Exponent)

Symbol	Definition
λ	Der Eigenwert einer quadratischen Matrix A
i	Die Ordnung von λ_i in $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$
E	Der Eigen-Exponent

$$\delta_i \, \alpha \, i^{\varepsilon}$$

3. Powerlaw (Eigen-Exponent)

- Ergebnisse der linearen Regression:

Datensatz	Nov. '97	Apr. '98	Dez. '98	1995
Eigen-Exponent E	-0,47	-0,50	-0,48	-0,177

- Kennzeichnet unterschiedliche Graphfamilien
- Hoher Korrelationskoeffizient: Fundamentale Eigenschaft des Internets
- Sehr ähnliche Eigenexponenten: Eigenschaft vermutlich unabhängig vom Wachstum

Übersicht

Diskussion

Diskussion

- Internettopologie mithilfe Exponenten statt Durchschnittswerten beschreiben
  - Durchschnittswerte können irreführenden sein (Bsp. Ausgangsgrad von 85% der Knoten < durchschnittlicher Ausgangsgrad)
  - Eigenschaften mit einer Zahl beschreibbar
- Effizienz der Protokolle verbessern
  - Genauere Abschätzungen

Diskussion

- Vorhersagen des Internetwachstums
  - Keine Vorhersage der Anzahl der Knoten
  - Keine explosionsartige Vergrößerung vohrersehbar
    - Soziale und technische Umbrüche?
  - Keine Übergangsphase erkennbar

→ Bestmögliche Vorhersage

Diskussion

- Künstliche Graphen generieren und bewerten
  - Großer, aber zeitlich sinkender Prozentsatz an Knoten gehört zu Bäumen
  - Über 80% der Bäume haben eine Tiefe von 1 bis 3
  - Ausgangsgrad muss den Powerlaws 1 und 2 folgen
  - Prüfen der Graphen mithilfe der Powerlaws
- Die chaotische Struktur des Internets ordnen
  - Intuitive Erklärung der Gesetzmäßigkeiten
    - Neue Knoten erhöhen den Traffic → Reorganisierung
    - Analogie zum Sandhaufen

Kritik

- Analyse der Methodik
  - Fehlende Beweise und Unklarheiten bei der Datenauswertung (z. B. Wie wurden die Eigenwerte berechnet?)
- Ausgangsmaterial
  - 3 Datensätze in 6-Monats-Intervallen für Vorhersagen mehrerer Jahrzehnte
  - Domainebene wir nur mit einer Instanz berücksichtigt

Zusammenfassung

Zusammenfassung

- Gesetzmäßigkeiten der Internettopologie
  - Existenz der Powerlaws
  - Konkrete Formeln
- Praktische Einsatzmöglichkeiten
  - Synthetische Graphen der Internettopologie erstellen und verifizieren
  - Protokollperformance verbessern
  - Vorhersagen der Entwicklung

Fragen

Zeit für Nachfragen

Quellnachweise

[1] Michalis Faloutsos, Petros Faloutsos and Christos Faloutsos (1999). On Power-Law Relationships of the Internet Topology. SIGCOMM ’99 8/99. Cambridge MA, USA

[2] Adamic, L. (2000). Zipf, Power-laws, and Pareto - a ranking tutorial. Retrieved from http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html