Ausarbeitung zum Seminarvortrag „Skalenfreie Netze“

Jens Arnold
14.07.2005

Inhaltsverzeichnis

1 Einleitung **2**

2 Skalenfreie Netze in der Realität **2**
2.1 Beschreibung komplexer Netzwerke 2
2.2 Skalenfreie Netze 3
2.3 Beispiele für reale skalenfreie Netze 3

3 Modelle **4**
3.1 Barabasi-Albert Modell 4
3.2 Verallgemeinerung des BA-Modells 5
3.2.1 Mittlerer Grad des einzelnen Knotens 7
3.2.2 Gradverteilung eines einzelnen Knotens 8
3.2.3 Die natürliche Skala eines „skalenfreien“Netwerkes 9

4 Schlussbemerkung **11**

5 Verwendete Literatur **12**

1 Einleitung

Netzwerke bestehen aus (Netzwerk-) Knoten und Verbindungen zwischen diesen (Links). Die Anzahl der Links, die mit einem einzelnen Knoten verbunden sind, wird als der Grad des Knotens bezeichnet.
Die Verteilung des Grades in realen komplexen Netztwerken folgt oft einem Potenzgesetz $P(k) \sim k^{-\gamma}$ (oder kann durch ein solches genähert werden). Solche Netze werden als „skalenfrei“ bezeichnet. Obgleich viele Knoten in einem solchen Netzwerk einen geringen Grad haben, gibt es relativ viele mit beträchtlich hohem Grad, d.h. die Gradverteilung fällt langsam ab.
Skalenfreie Netz werden in der Realität häufig beobachtet. Im Allgemeinen zeigen sie einen Small-World-Effekt (d.h. die mittlere kürzeste Weglänge zwischen zwei Knoten ist sehr gering) und eine hohe Resistenz gegenüber zufälligen Störungen bei gleichzeitig hoher Empfindlichkeit gegenüber gezielten Angriffen. D.h. das Entfernen eines zufällig ausgewählten Knotens beeinflusst kaum die mittlere kürzeste Weglänge zwischen zwei Knoten. Werden jedoch gezielt Knoten mit hohem Grad entfernt, so ist zu erwarten, dass sich die kürzeste Weglänge zwischen zwei Knoten stark erhöht.
Eine solche skalenfreie Gradverteilung kann durch den Mechanismus entstehen, durch den das Netzwerk wächst, d.h. durch den Mechanismus mit dem neue Knoten und Verbindungen dem Netzwerk hinzugefügt werden. Da Netzwerke in der Realität selten plötzlich entstehen, sind besonders Modelle von wachsenden Netzen interessant.
In dieser Ausarbeitung sollen zwei Modelle wachsender Netzwerke, das Barabasi-Albert-Modell und die Verallgemeinerung von Dorogovtsev, Mendes und Samukhin genauer untersucht werden.

2 Skalenfreie Netze in der Realität

2.1 Beschreibung komplexer Netzwerke

Die Beschreibung eines Netzwerks erfolgt, je nach Beschaffenheit des Netzwerkes, durch einen gerichteten oder ungerichteten Graphen. Ein Graph besteht aus zwei Mengen: Der Knotenmenge $V = \{v_1, v_2, \ldots, v_N\}$ und einer Kantenmenge, die sich aus Paaren der Knotenmenge zusammensetzt. $E = \{(v_i, v_j)|v_i, v_j \in V\}$. Bei einem ungerichteten Graphen ist nur wesentlich, ob zwischen zwei Knoten v_i und v_j eine Verbindung besteht, also ob $(v_i, v_j) \in E$. Ist dies der Fall, so ist auch $(v_j, v_i) \in E$, da die beiden Knoten *miteinander* verbunden sind.
Bei einem gerichteten Graphen wird die Richtung der Verbindung zusätzlich berücksichtig: Wenn es eine Verbindung von v_i nach v_j gibt, also $(v_i, v_j) \in E$ ist, folgt daraus nicht, dass auch $(v_j, v_i) \in E$ ist.
Betrachtet man die Eigenschaften eines einzelnen Knoten in diesem Netzwerk, so ist eine der am leichtesten zu ermittelnde Eigenschaft wohl der Grad des Knoten, also die Anzahl der Verbindungen, die dieser Knoten hat. Bei einem gerichteten Graphen ist dabei zwischen der Anzahl der eingehenden Verbindungen (in-degree $k_i = q$) und der

Anzahl der ausgehenden Verbindungen (out-degree k_o) zu unterscheiden.
Will man eine Aussage über das ganze Netzwerk treffen, so ist eine einfache Möglichkeit also die Verteilung des Grades $P(k)$, also die Antwort auf die Frage „Welcher Anteil der Knoten in diesem Netzwerk hat den Grad k ?“

2.2 Skalenfreie Netze

Sind die Verbindungen völlig zufällig zwischen den Knoten des Netzwerks verteilt (hohe Unordnung), so fällt die Gradverteilung $P(k)$ für große k sehr schnell (exponentiell) ab. Bei hoher Ordnung (z.B. in einem regelmäßigem Gitter) ist der Grad überhaupt nicht zufällig verteilt.
Reale Netzwerke zeigen hingegen oft eine Gradverteilung, die einem Potenzgesetz

$$P(k) \sim k^{-\gamma}$$

folgt.
Allerdings kann ein reales Netzwerk nur näherungsweise eine solche Verteilung annehmen, da reale Netze endlich sind. D.h. die Gradverteilung verschwindet spätestens da, wo der Grad größer als die Anzahl der Verbindungen in dem gesamten Netz ist. Noch schwerwiegender ist die Tatsache, dass sich die Grad-Verteilung für große k aufgrund der statistischen Schwankungen kaum beobachten lässt. Ebenso kann bei kleinem Grad eine Abweichung vom Potenzgesetz bestehen. Diese Abweichungen können z.B. durch die Randbedingungen bei wachsenden Netzen entstehen (z.B. könnte jeder neue Knoten mit $k = 2$ erzeugt werden, also $P(1) = 0$).
Die Verbindungen zwischen den Knoten sind also weder völlig zufällig noch völlig regelmäßig festgelegt. Prozesse, die zu solchen Gradverteilungen führen, sind also komplizierter.

2.3 Beispiele für reale skalenfreie Netze

Skalenfreie Netze treten in der Realität oft auf.
Hier seien nur einige Beispiele genannt.

- World Wide Web
- Internet
- Protein-Protein-Wechselwirkung
- Netzwerk menschlicher sexueller Kontakte
- Nahrungsnetze
- Zitations- und Kollaborations-Netze
- Netz der Telefonanrufe
- email-Netzwerke

3 Modelle

Dass die Gradverteilung langsam abfällt, bedeutet in Bezug auf ein wachsendes Netzwerk (eine Eigenschaft die wohl die meisten natürlichen Netzwerke haben), dass neue Links bevorzugt mit Knoten hohen Grades verbunden werden. Es existiert also ein „Matthew-Effekt"(nach dem Bibelvers: „wer hat dem soll gegeben werden"). Die in der Literatur oft gebrauchte Wendung „popularity is attractive" drückt genau dasselbe aus.
Verschiedene Modelle versuchen diesen Effekt nachzubilden.

3.1 Barabasi-Albert Modell

Barabasi und Albert haben ein Modell für ein wachsendes Netzwerk entwickelt, dass genau diesen Effekt modelliert, indem hinzugefügte Links sich zufällig mit alten Knoten verbinden. Die Wahrscheinlichkeit, dass ein Knoten einen neuen Link erhält, ist dabei proportional zu seinem Grad. Dieser Mechanismus wird „preferential attachment" genannt.
Zu Beginn ($t = 2$) besteht der Graph, der das Netzwerk repräsentiert, aus zwei Knoten $s = 1, 2$ die durch $2m$, $m \in \mathbb{N}$ Kanten miteinander verbunden sind. Mit jedem Zeitschritt erscheint ein neuer Knoten in dem Netzwerk, der sich mittels m Kanten mit bestehenden Knoten aus dem Netzwerk verbindet. Die Wahrscheinlichkeit $p_{k->k+1}$, dass ein Knoten mit Grad k eine neue Verbindung erhält, ist proportional zu seinem Grad. Wir betrachten hier nur den Fall $m = 1$. Dann ist

$$p_{k->k+1} = \frac{k}{2t}.$$

Sei $p(k, s, t)$ die Wahrscheinlichkeit, dass der Knoten s zur Zeit t den Grad k hat.
Nun kann die Mastergleichung für dieses Netzwerk aufgestellt werden.

$$p(k, s, t + 1) = p_{k-1->k}p(k - 1, s, t) + (1 - p_{k->k+1})p(k, s, t)$$

also

$$p(k, s, t + 1) = \frac{k - 1}{2t}p(k - 1, s, t) + (1 - \frac{k}{2t})p(k, s, t)$$

mit den Anfangsbedingungen $p(k, s \in 1, 2, t = 2) = \delta_{k,2}$ und der Randbedingung $p(k, s = t, t > 2) = \delta_{k,1}$. Wir suchen die stationäre Verteilung des Grades

$$P(k) := \lim_{t\to\infty} P(k, t)$$

mit $P(k, t) := \frac{1}{t} \sum_{s=1}^{t} p(k, s, t)$.
Es folgt

$$\sum_{s=1}^{t} p(k, s, t + 1) = (t + 1)P(k, t + 1) - \delta_{k,1} = \frac{k - 1}{2}P(k - 1, t) - \frac{k}{2}P(k, t) + tP(k, t).$$

Für die stationäre Verteilung gilt $P(k,t+1) = P(k,t) = P(k)$. Es folgt

$$P(k) - \delta_{k,1} = \frac{k-1}{2}P(k-1) - \frac{k}{2}P(k).$$

Der geringste auftretende Grad ist $k = 1$

$$P(1) - 1 = -\frac{1}{2}P(1)$$

also

$$P(1) = \frac{2}{3}.$$

Für $k > 1$ folgt

$$P(k) = \frac{k-1}{k+2}P(k-1) = \frac{(k-1)(k-2)\dots 3\cdot 2\cdot 1}{(k+2)(k+1)k(k-1)\dots 4}P(1)$$

also

$$P(k) = \frac{4}{(k+2)(k+1)k}.$$

Insbesondere geht die Verteilung für $k \gg 1$ wie

$$P(k) \sim k^{-\gamma}$$

mit $\gamma = 3$. Das Barabasi-Albert Modell modelliert also ein wachsendes Zufallsnetzwerk mit einer skalenfreien Gradverteilung. Der γ-Exponent nimmt allerdings nur den Wert 3 an. Der Grund hierfür liegt an dem Prozess, durch den neue Links sich mit alten Knoten mit einer Wahrscheinlichkeit, die proportional zu ihrem Grad ist, verbinden.

3.2 Verallgemeinerung des BA-Modells

Dorogovtsev, Mendes und Samukhin haben das BA-Modell, welches recht strenge Bedingungen an den Wachstumsprozess stellt, verallgemeinert. In diesem Modell ist die Wahrscheinlichkeit mit der ein Knoten mit in-degree $k_i = q$ einen der neuen Links erhält, proportional zu $q + A$

$$p_{q->q+1} \sim q + A = A(s).$$

Der Parameter A wird hiernach auch „zusätzliche Attraktivität“ genannt, $A(s)$ entsprechend die Attraktivität des Knotens.
Betrachtet wird hier nur der in-degree q, d.h. es wird keine Aussage darüber getroffen, woher die neuen Links kommen. Diese können also alle aus dem neuen Knoten kommen (wie im BA Modell), zwischen alten Knoten geknüpft werden, oder sogar von ausserhalb des Netzwerkes kommen.
Es war von in-degree die Rede. Wir betrachten also einen gerichteten Graphen, der zur Zeit $t = 1$ aus einem Knoten mit in-degree $m \in \mathbb{N}$ besteht. Mit jedem Zeitschritt wird

ein neuer Knoten hinzugefügt und es erscheinen m weitere Links. Die Wahrscheinlichkeit, dass ein Knoten mit in-degree q einen neuen Link bekommt, ist

$$\frac{q+A}{\sum_{s=0}^{t} A(s)} = \frac{q+am}{tm(a+1)} \quad [A = am].$$

Die Wahrscheinlichkeit, dass ein Knoten mit in-degree q genau l der m neuen Links bekommt, beträgt

$$\binom{m}{l} \left(\frac{q+am}{tm(a+1)}\right)^{l} \left(1 - \frac{q+am}{tm(a+1)}\right)^{m-l} .$$

Nun lässt sich wieder die Mastergleichung aufstellen

$$p(q,s,t+1) = \sum_{l=0}^{m} \binom{m}{l} \left(\frac{q-l+am}{tm(a+1)}\right)^{l} \left(1 - \frac{q-l+am}{tm(a+1)}\right)^{m-l} p(q-l,s,t).$$

Wir interessieren uns für große Netzwerke: $t \gg 1$, also

$$p(q,s,t+1) = \left(1 - m\frac{q+am}{tm(a+1)} + O(t^{-2})\right) p(q,s,t) + \left(m\frac{q-1+am}{tm(a+1)} + O(t^{-2})\right) p(q-1,s,t)$$

also

$$p(q,s,t+1) = \left(1 - \frac{q+am}{t(a+1)}\right) p(q,s,t) + \left(\frac{q-1+am}{t(a+1)}\right) p(q-1,s,t) + O(t^{-2}).$$

Natürlich gilt $p(q<0,s,t) = 0$, aber auch $p(q,s,t=s) = \delta(q)$, denn ein neuer Knoten wird ohne Links in das Netzwerk eingefügt.
Mit $P(q,t) := \frac{1}{t}\sum_{s=1}^{t} p(q,s,t)$ ergibt Summation über $s = 1 \dots t$ und Vernachlässigung von Termen der Ordnung $O(t^{-2})$

$$(t+1)P(q,t+1) - p(q,t+1,t+1) = tP(q,t) + \frac{1}{1+a}\left[(q-1+am)P(q-1,t) - (q+am)P(q,t)\right].$$

Für die stationäre Verteilung $P(q)$ gilt dann

$$P(q) - \delta(q) = \frac{1}{1+a}\left((q-1+am)P(q-1) - (q+am)P(q)\right).$$

Für $q \gg 1$ kann man die endliche Differenz $P(q) - P(q-1)$ durch $\partial_q P(q)$ ersetzen. Dies ist besonders dann sinnvoll (oder einfacher) wenn man nur an dem Verhalten der Verteilung für große q (also am γ-Exponenten) interessiert ist.
Es gilt

$$(1+a)P(q) = -\partial_q\left[(q+am)P(q)\right] = -P(q) - (am+q)\partial_q P(q)$$

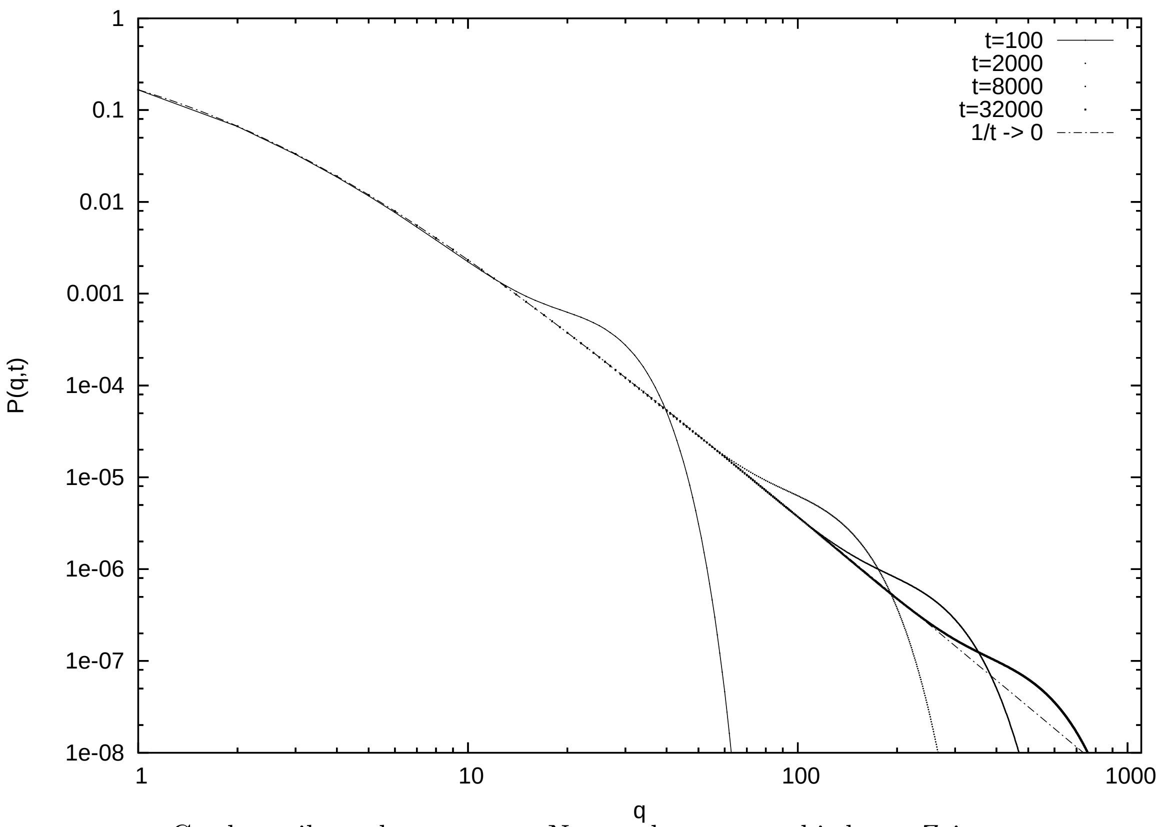
also

$$-(2+a)P(q) = (q+am)\partial_q P(q) \approx q\partial P(q).$$

Die Gradverteilung nähert sich also für große q an

$$P(q) \sim q^{-\gamma} \quad [\gamma = 2 + a].$$

Der γ-Exponent kann in diesem Modell also Werte zwischen 2 und ∞ annehmen. Für $a = 1$ ist $\gamma = 3$ und das Modell ist tatsächlich äquivalent zum BA-Modell. Für $A = am \to \infty$ gibt es keine Präferenz mehr und es liegt der Fall des exponentiellen Zufallsgraphen vor, dessen Gradverteilung, wie der Name schon sagt, schneller als jede Potenz von q abfällt.



Gradverteilung der gesamten Netzwerkes zu verschiedenen Zeiten

3.2.1 Mittlerer Grad des einzelnen Knotens

Eine exakte Lösung $p(q, s, t)$ der Mastergleichung zu erhalten ist i.A. sehr schwierig und auch wenig aussagekräftig, da diese Ein-Knoten-Eigenschaft praktisch nicht beobachtet werden kann. Das liegt daran, dass in einem gegebenen Netzwerk ein Knoten nur *einen* Grad annimmt und man nur selten Informationen darüber hat, wann dieser Knoten erzeugt wurde.
Bevor nun einige generelle Eigenschaften der Gradverteilung von einzelnen Knoten betrachtet werden sollen, berechnen wir zunächst das erste Moment dieser Verteilung, den mittleren Grad $\bar{q}(s, t) = \sum_{q=0}^{\infty} qp(q, s, t)$.

Aufsummieren der Mastergleichung ergibt

$$\bar{q}(s,t+1) = \bar{q}(s,t) + \frac{1}{t(a+1)}\left(-\sum_{q=0}^{\infty} q(q+am)p(q,s,t) + \sum_{q=1}^{\infty} q(q-1+am)p(q-1,s,t)\right)$$

also

$$\bar{q}(s,t+1) = \bar{q}(s,t) + \frac{1}{t(a+1)}\left(-\sum_{q=0}^{\infty} q(q+am)p(q,s,t) + \sum_{q=0}^{\infty}(q+1)(q+am)p(q,s,t)\right).$$

Im Kontinuumslimes $\partial_t \bar{q}(s,t) = \bar{q}(s,t+1) - \bar{q}(s,t)$ für $t \gg 1$ folgt

$$t(a+1)\partial_t \bar{q}(s,t) = \bar{q}(s,t) + am$$

mit der Randbedingung $\bar{q}(s,t=s) = 0$, da neue Knoten keine Links haben. Die allg. Lösung dieser Differentialgleichung lautet

$$\bar{q}(s,t) = C(s)t^{\beta} - am \quad \left[\beta := \frac{1}{1+a}\right].$$

Auswerten der Randbedingungen gibt $C(s) = ams^{-\beta}$ und somit

$$\bar{q}(s,t) = am\left(\frac{s}{t}\right)^{-\beta} - am.$$

Für $1 \ll s \ll t$ geht $\bar{q}(s,t)$ asymptotisch wie

$$\bar{q}(s,t) \sim \left(\frac{s}{t}\right)^{-\beta}.$$

Zwischen den Exponenten $\beta = \frac{1}{a+1}$ und $\gamma = 2 + a$ besteht damit der Zusammenhang

$$\gamma = 1 + \frac{1}{\beta}.$$

3.2.2 Gradverteilung eines einzelnen Knotens

In der Kontinuumsnäherung für $1 \ll q, t$ wird die Mastergleichung zu

$$(1+a)t\partial_t p(q,s,t) = -\partial_q\left[(q+am)p(q,s,t)\right]$$

mit der Randbedingung $p(q,s,t=s) = \delta(q)$.
Geht man über zu Variablen q und $\bar{q}$, so gilt mit

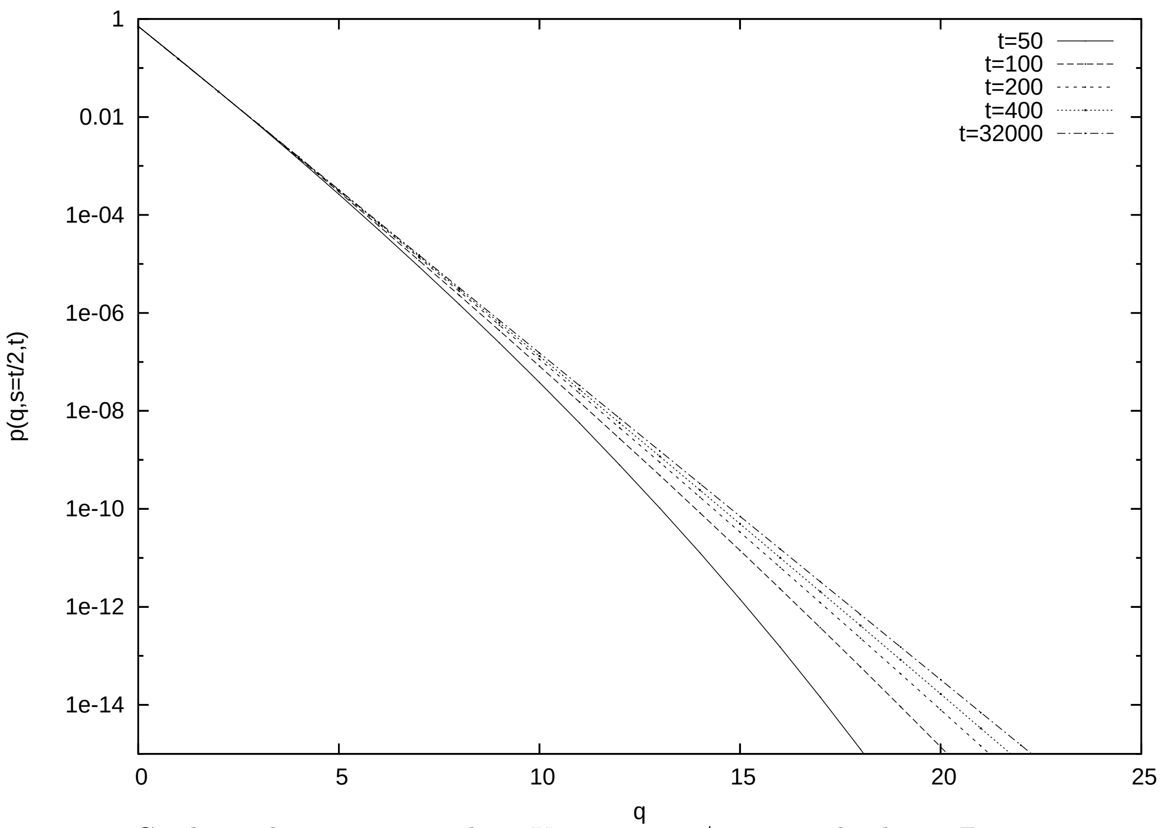
$$t(a+1)\partial_t \bar{q}(s,t) = \bar{q}(s,t) + am$$

$$t(1+a)\partial_t p(q,s,t) = (\bar{q}(s,t) + am)\partial_q p(q,\bar{q}) = -\partial_q\left[(q+am)p(q,\bar{q})\right].$$

Dieses Randwertproblem wird durch die δ-Distribution gelöst

$$p(q,s,t) = \delta(q - \bar{q}) = \frac{1}{\bar{q}}\delta(\frac{q}{\bar{q}} - 1).$$

Dabei ist $\bar{q}$ die natürliche Skala der Gradverteilung.
Mit Sicherheit ist die δ-Distribution nur eine grobe Beschreibung der Gradverteilung. Allerdings gibt sie Aufschluss über das Skalenverhalten der Ein-Knoten-Gradverteilung und über das Skalenverhalten der Gradverteilung eines Netzwerks endlicher Größe.
Eine weitere wichtige Eigenschaft ist, dass $p(q,s,t)$ sehr schnell abfällt.



Gradverteilung eines einzelnen Knoten $s = t/2$ zu verschiedenen Zeiten

3.2.3 Die natürliche Skala eines „skalenfreien“ Netwerkes

Wir haben gesehen, dass die Gradverteilung eines Netzwerkes, das unter dem Prozess des „preferential attachment“ wächst, sich für $t \to \infty$ asymptotisch an

$$P(q) \sim q^{-\gamma}$$

nähert, also skalenfrei wird.
Für endliches t gilt jedoch mit Sicherheit $P(q > mt) = 0$, da kein Knoten mehr als alle Links des Netzwerkes auf sich vereinen wird.

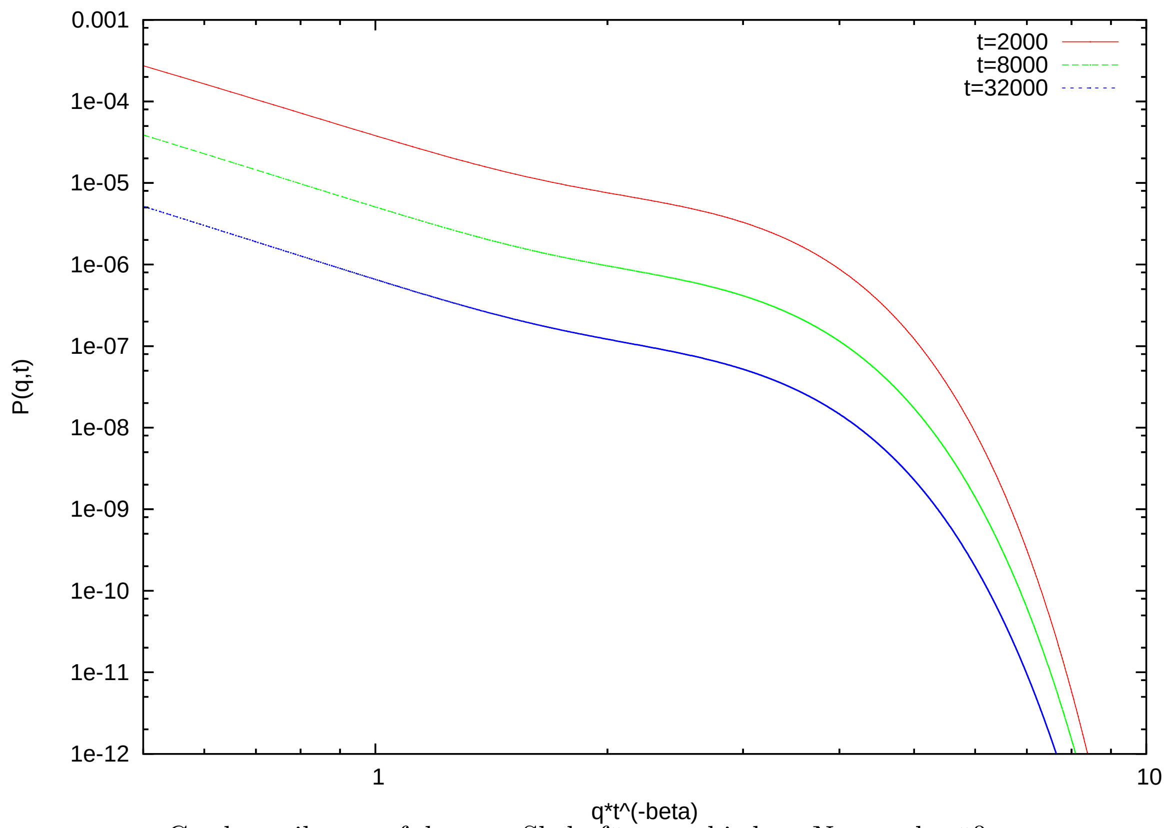
Tatsächlich wird bei einem Netzwerk endlicher Größe die Gradverteilung i.d.R. viel früher abgeschnitten. Der Grad, bei dem dies geschieht wird mit q_{cut} bezeichnet. Es gilt

$$tP(q,t) \approx \sum_{s=1}^{t} p(q,s,t) = \sum_{s=1}^{t} \delta\left(q - \left(\frac{s}{t}\right)^{-\beta} am + am\right).$$

Für große q wird also die Grad-Verteilung des gesamten Netzes bestimmt durch die Gradverteilungen der ältesten Knoten. q_{cut} wird damit durch $p(q, s = 1, t)$ festgelegt. Es folgt

$$q_{cut} = amt^{\beta} - am \sim t^{\beta} \quad [t \gg 1].$$

Damit wird q_{cut} zur natürlichen Skala der Gradverteilung des endlichen Netzwerkes.



Gradverteilung auf der q_{cut}-Skala für verschiedene Netzwerkgrößen

4 Schlussbemerkung

Die beiden vorgestellten Modelle beschreiben wachsende Netzwerke mit einer skalenfreien Gradverteilung

$$p \sim k^{-\gamma}.$$

Es zeigt sich, dass solche Gradverteilungen durch den Prozess des „prefential attachment“ zustande kommen können. Die Präferenz-Funktion hat dabei die Form

$$f(q) = q + A.$$

Grundsätzlich könnten in dieser Funktion auch Terme höherer Ordnung in q auftreten. Die Grad-Verteilung wird dann aber nicht mehr durch ein Potenzgesetz beschrieben. Im allgemeinsten Fall lautet die Präferenzfunktion die solche Verteilungen erzeugt

$$f(q, s, t) = g(s, t) + qh(s, t).$$

Mittels dieser Modelle lassen sich skalenfreie Netze beschreiben, deren Struktur durch den Wachstumsprozess bestimmt wird. Das Barabasi-Albert-Modell sagt dabei einen γ-Exponenten von 3 vorher. In der Verallgemeinerung nimmt der γ-Exponent einen Wert zwischen 2 und ∞ an. Einmal erzeugte Verbindungen und Knoten werden nicht wieder entfernt. Spielen also andere Prozesse, durch die Links gesetzt werden, eine wichtige Rolle, so sind diese Modelle keine gute Repräsentation der Realität.

5 Verwendete Literatur

- M. E. J. Newman
  The structure and function of complex networks
  SIAM Review 45, 167-256 (2003)

- S. N. Dorogovtsev, J. F. F. Mendes
  Evolution of Networks: From Biological Nets to the Internet and WWW
  Oxford University Press (2003)

- S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin
  Size-dependent degree distribution of a scale-free growing network
  Physical Review E, Volume 63 062101 (2001)

- S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin
  Structure of Growing Networks with Preferential Linking
  Physical Review Letters, Volume 85, Number 21, p. 4633 (2000)

- A.-L. Barabási, R. Albert and H. Jeong
  Emergence of Scaling in random networks
  Science 286 509 (1999)

- R. Albert, H. Jeong, and A.-L. Barabási
  Error and attack tolerance of complex networks
  Nature 406 378 (2000)

- M. E. J. Newman
  The structure of scientific collaboration networks
  Proc. Natl. Acad. Sci. USA 98, 404-409 (2001)