Zufällige Graphen

Christoph Schumacher

09. Juli 2020

RGCN I, Kap. 8.3, 8.4, 8.5, 8.6:
Das Preferential Attachment Modell II

Zusammenfassung

Grade im Preferential Attachment Modell:

- Verteilung einzelner Grade
- Statistik der Gradfolge bzw. dei empirischen Verteilung der Grade

Auch mit dabei: Martingale

Inhaltsverzeichnis

Fragen

1. —

1 Der asymptotische Grad einzelner Knoten

1.1 $m = 1$

Theorem 8.2 beschreibt das Wachstumsverhalten von Knotengraden bei $t \to \infty$, also im Limes vieler Knoten. Dabei ist

$$\mathbb{E}[D_i(t) + \delta] = (1 + \delta)\frac{\Gamma(i - \frac{1}{2+\delta})}{\Gamma(i)} t^{\frac{1}{2+\delta}} (1 + O(\tfrac{1}{t})), \tag{1}$$

wie man aus (8.3.9) sieht. Erinnerung: $\Gamma(i) = (i - 1)!$. Man sieht wieder, wie ein kleineres δ eine krassere Konzentration der Knotengrade bewirkt: Bei beispielsweise $\delta = 0$ wachsen die Knotengrade proportional zu $\sqrt{t}$, bei $\delta = 998$ nur mit $\sqrt[1000]{t}$.

Am besten beweisen wir erst einmal (8.3.9) für $a \in \mathbb{R}$ und $t \in (\max\{1, -a\}, \infty)$ mit der Stirling-Formel (8.11.1):

$$\frac{1}{1/t}\left(\frac{\Gamma(t+a)}{t^a\Gamma(t)} - 1\right) \tag{2}$$

$$\leq t\left(\frac{\sqrt{2\pi(t+a-1)}(\frac{t+a-1}{\mathrm{e}})^{t+a-1}\exp(\frac{1}{12(t+a-1)})}{t^a\sqrt{2\pi(t-1)}(\frac{t-1}{\mathrm{e}})^{t-1}} - 1\right) \tag{3}$$

$$\leq t\left(\left(1+\frac{a}{t-1}\right)^{t+a-\frac{1}{2}}\mathrm{e}^{-a}\exp\left(\frac{1}{12(t+a-1)}\right) - 1\right) \tag{4}$$

$$= t\left(\exp\left(\left(t+a-\frac{1}{2}\right)\log\left(1+\frac{a}{t-1}\right) - a + \frac{1}{12(t+a-1)}\right) - 1\right) \tag{5}$$

$$\leq t\left(\exp\left(\left(t+a-\frac{1}{2}\right)\frac{a}{t-1} - a + \frac{1}{12(t+a-1)}\right) - 1\right) \tag{6}$$

$$= \frac{\exp\left(a\frac{a+\frac{1}{2}}{t-1} + \frac{1}{12(t+a-1)}\right) - 1}{1/t}. \tag{7}$$

Mit der Regel von l'Hospital bekommen wir

$$\limsup_{t\to\infty}\frac{1}{1/t}\left(\frac{\Gamma(t+a)}{t^a\Gamma(t)} - 1\right) \leq \lim_{t\to\infty}\frac{\exp\left(a\frac{a+\frac{1}{2}}{t-1} + \frac{1}{12(t+a-1)}\right) - 1}{1/t} \tag{8}$$

$$\overset{\text{l'H}}{=} \lim_{t\to\infty}\exp\left(a\frac{a+\frac{1}{2}}{t-1} + \frac{1}{12(t+a-1)}\right)\frac{-a\frac{a+\frac{1}{2}}{(t-1)^2}}{-1/t^2} \tag{9}$$

$$= a^2 + \frac{a}{2} < \infty. \tag{10}$$

Jetzt gehen wir in den Beweis von Theorem 8.2. Für $t \geq i$ und $m = 1$ ist

$$\mathbb{E}[D_i(t+1) + \delta \mid D_i(t)] \tag{11}$$

$$= D_i(t) + \delta + \mathbb{E}[D_i(t+1) - D_i(t) \mid D_i(t)] \tag{12}$$

$$= D_i(t) + \delta + \mathbb{P}\big(v_{t+1}^{(1)} \to v_i^{(1)} \mid D_i(t)\big) \tag{13}$$

$$\overset{(8.2.1)}{=} D_i(t) + \delta + \frac{D_i(t)+\delta}{t(2+\delta)+1+\delta} \tag{14}$$

$$= \cdots = \big(D_i(t)+\delta\big)\frac{(t+1)(2+\delta)}{t(2+\delta)+1+\delta}. \tag{8.3.4}$$

Damit können wir schon nachrechnen, dass $(M_i(t))_{t\geq i}$ ein Martingal ist. Wir verwenden $\sigma(M_i(t)) = \sigma(D_i(t))$:

$$\mathbb{E}[M_i(t+1) \mid M_i(t)] = \frac{\mathbb{E}[D_i(t+1)+\delta \mid D_i(t)]}{\delta+1}\prod_{s=i-1}^{t}\frac{(2+\delta)s+1+\delta}{(2+\delta)(s+1)} \tag{15}$$

$$= \frac{(D_i(t)+\delta)\frac{(t+1)(2+\delta)}{t(2+\delta)+1+\delta}}{\delta+1}\prod_{s=i-1}^{t}\frac{(2+\delta)s+1+\delta}{(2+\delta)(s+1)} \tag{16}$$

$$= \frac{D_i(t)+\delta}{\delta+1}\prod_{s=i-1}^{t-1}\frac{(2+\delta)s+1+\delta}{(2+\delta)(s+1)} \tag{17}$$

$$= M_i(t) \tag{18}$$

Der Erwartungswert ist

$$\mathbb{E}[M_i(t)] = \mathbb{E}[\mathbb{E}[M_i(t) \mid M_i(t-1)]] = \mathbb{E}[M_i(t-1)] = \cdots = \mathbb{E}[M_i(i)]. \tag{19}$$

Dafür brauchen wir (8.3.5):

$$\begin{aligned}
\mathbb{E}[D_i(i)+\delta] &= \mathbb{E}\big[\mathbb{E}[D_i(i)+\delta \mid \mathrm{PA}_{i-1}^{(1,\delta)}]\big] && (20)\\
&= \delta + \mathbb{E}\big[1\cdot\mathbb{P}(D_i(i)=1 \mid \mathrm{PA}_{i-1}^{(1,\delta)}) + 2\cdot\mathbb{P}(D_i(i)=2 \mid \mathrm{PA}_{i-1}^{(1,\delta)})\big] && (21)\\
&= \delta + \mathbb{E}\big[1-\mathbb{P}(v_i^{(1)}\to v_i^{(1)} \mid \mathrm{PA}_{i-1}^{(1,\delta)}) + 2\mathbb{P}(v_i^{(1)}\to v_i^{(1)} \mid \mathrm{PA}_{i-1}^{(1,\delta)})\big] && (22)\\
&= \delta + 1 + \mathbb{E}\big[\mathbb{P}(v_i^{(1)}\to v_i^{(1)} \mid \mathrm{PA}_{i-1}^{(1,\delta)})\big] && (23)\\
&\overset{(8.2.1)}{=} 1+\delta+\mathbb{E}\Big[\frac{1+\delta}{(i-1)(2+\delta)+1+\delta}\Big] && (24)\\
&= \cdots = (1+\delta)\frac{i(2+\delta)}{(i-1)(2+\delta)+1+\delta}. && (8.3.5)
\end{aligned}$$

Damit können wir den Erwartungswert von $M_i(t)$ überprüfen:

$$\mathbb{E}[M_t(i)] = \mathbb{E}[M_i(i)] = \frac{\mathbb{E}[D_i(i)+\delta]}{1+\delta}\frac{(2+\delta)(i-1)+1+\delta}{(2+\delta)i} = 1. \tag{25}$$

Nun zu den Γ-Funktionen. Wir nutzen $\Gamma(t+1)=t\Gamma(t)$, um in (8.3.7) mit Hilfe zweier Teleskopprodukte den zweiten Schritt zu machen:

$$\prod_{s=i-1}^{t-1}(s+x) = \prod_{s=i-1}^{t-1}\frac{\Gamma(s+x+1)}{\Gamma(s+x)} = \frac{\Gamma(t+x)}{\Gamma(i+x-1)}. \tag{26}$$

Zusätzlich mit $\frac{1+\delta}{2+\delta}-1=-\frac{1}{2+\delta}$ erhalten wir (8.3.7). Das zeigt zusammen mit (19) und (25) auch (8.3.3). In (8.3.10) muss zweimal $(1+\delta)$ im Zähler stehen.

1.2 $m \geq 2$

Gleichung (8.3.11) folgt direkt aus der Definition des Preferential Attachment Modells. Beim Übergang von $\mathrm{PA}_{mt}^{(1,\delta/m)}$ zu $\mathrm{PA}_t^{(m,\delta)}$ fasst man ja die Knoten $v_i^{(1)}$ mit $i\in\{(j-1)m+1,\dots,jm\}$ zum Knoten $v_j^{(m)}$ zusammen. Dabei kumulieren sich die Grade:

$$\begin{aligned}
\mathbb{E}_m^{\delta}[D_j(t)+\delta] &= \sum_{i=(j-1)m+1}^{jm}\mathbb{E}_1^{\delta/m}[D_i(mt)+\delta/m] && (8.3.11)\\
&= \Big(1+\frac{\delta}{m}\Big)\frac{\Gamma(mt+1)}{\Gamma(mt+\frac{1+\delta/m}{2+\delta/m})}\sum_{i=(j-1)m+1}^{jm}\frac{\Gamma\big(i-\frac{1}{2+\delta/m}\big)}{\Gamma(i)}, && (27)
\end{aligned}$$

und

$$\frac{D_i(t)}{(mt)^{\frac{1}{2+\delta/m}}} = \sum_{i=(j-1)m+1}^{jm}\frac{D_i(mt)}{(mt)^{\frac{1}{2+\delta/m}}} \xrightarrow[t\to\infty]{\text{f. s.}} \sum_{i=(j-1)m+1}^{jm}\xi_i \tag{28}$$

mit ξ_i aus Theorem 8.3 (mit δ/m statt δ).

2 Die Gradfolge im Preferential Attachment Modell

Zu (8.4.6):

$$\frac{\frac{\Gamma(k+a)}{\Gamma(k-1+b)}-\frac{\Gamma(k+1+a)}{\Gamma(k+b)}}{b-a-1} = \frac{\Gamma(k+a)}{\Gamma(k+b)}\frac{(b-1+k)-(k+a)}{b-a-1} = \frac{\Gamma(k+a)}{\Gamma(k+b)} \tag{8.4.6}$$

—

Wir zeigen jetzt: Aus Theorem 8.3 folgt, dass der Knotengrad eines uniform aus $[t]$ gewählten Knoten bei $t \to \infty$ gegen die Verteilung $(p_k)_k$ konvergiert. Der uniform gewählte Knoten $U_t \in [t]$ hat bei gegebenem Graph die Verteilung

$$\mathbb{P}(D_{U_t} = k \mid \mathrm{PA}_t^{(m,\delta)}) = P_k(t). \tag{29}$$

(Beim GRG haben wir diesen Zusammenhang rückwärts benutzt und aus der Verteilung des uniformen Knoten auf die empirische Verteilung des Grades geschlossen.) Mit dem Ereignis $A_t := \{\max_k |P_k(t) - p_k| < C\sqrt{\log t/t}\}$ aus (8.4.5) aus Theorem 8.3:

$$\begin{aligned}
|\mathbb{P}(D_{U_t} = k) - p_k| &= \left|\mathbb{E}[\mathbb{P}(D_{U_t} = k \mid \mathrm{PA}_t^{(m,\delta)})] - p_k\right| && (30)\\
&\le \mathbb{E}[|P_k(t) - p_k|] && (31)\\
&= \mathbb{E}[|P_k(t) - p_k|(\mathbf{1}_{A_t} + \mathbf{1}_{A_t^c})] && (32)\\
&\le C\sqrt{\log t/t} + \mathbb{P}(A_t^c) \xrightarrow[t\to\infty]{\text{Thm. 8.3}} 0. && (33)
\end{aligned}$$

Für diskrete ZVn zeigt das schon $D_{U_t} \xrightarrow[t\to\infty]{d} (p_k)_k$.

—

Zur Darstellung

$$p_k = \mathbb{P}(X = k - m) = \mathbb{E}[\mathbb{P}(X = k - m \mid U)], \tag{8.4.10}$$

wobei, bedingt auf $U \sim \mathcal{U}([0,1])$, $X \sim \mathrm{NB}(r = m + \delta, p = U^{\frac{1}{2+\frac{\delta}{m}}})$ negativ binomialverteilt ist:

$$\mathbb{P}(X = k) = \frac{\Gamma(r+k)}{k!\,\Gamma(r)} p^r (1-p)^k, \tag{8.4.9}$$

da brauchen wir die Betafunktion

$$[0,1]^2 \ni (x,y) \mapsto B(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \tag{34}$$

Damit kann man alles ausrechnen:

$$\begin{aligned}
&\mathbb{P}(X = k - m) && (35)\\
&= \mathbb{E}[\mathbb{P}(X = k - m \mid U)] && (36)\\
&\overset{(8.4.9)}{=} \mathbb{E}\left[\frac{\Gamma(k+\delta)}{(k-m)!\,\Gamma(m+\delta)} U^{\frac{m+\delta}{2+\frac{\delta}{m}}}\left(1 - U^{\frac{1}{2+\frac{\delta}{m}}}\right)^{k-m}\right] && (37)\\
&\overset{U\sim\mathcal{U}([0,1])}{=} \frac{\Gamma(k+\delta)}{(k-m)!\,\Gamma(m+\delta)} \int_0^1 u^{\frac{m+\delta}{2+\frac{\delta}{m}}}\left(1 - u^{\frac{1}{2+\frac{\delta}{m}}}\right)^{k-m} \mathrm{d}u && (38)\\
&\overset{u:=t^{2+\frac{\delta}{m}}}{=} \frac{\Gamma(k+\delta)}{(k-m)!\,\Gamma(m+\delta)} \int_0^1 t^{m+\delta}(1-t)^{k-m}\left(2 + \frac{\delta}{m}\right) t^{1+\frac{\delta}{m}}\,\mathrm{d}t && (39)\\
&\overset{(34)}{=} \frac{\Gamma(k+\delta)}{(k-m)!\,\Gamma(m+\delta)}\left(2 + \frac{\delta}{m}\right) B\left(m + \delta + 2 + \frac{\delta}{m}, k - m + 1\right) && (40)\\
&\overset{(34)}{=} \left(2 + \frac{\delta}{m}\right)\frac{\Gamma(k+\delta)}{(k-m)!\,\Gamma(m+\delta)} \frac{\Gamma(m+\delta+2+\frac{\delta}{m})\Gamma(k-m+1)}{\Gamma(k+\delta+3+\frac{\delta}{m})} && (41)\\
&\overset{\Gamma(n+1)=n!}{=} \left(2 + \frac{\delta}{m}\right)\frac{\Gamma(k+\delta)\Gamma(m+\delta+2+\frac{\delta}{m})}{\Gamma(m+\delta)\Gamma(k+\delta+3+\frac{\delta}{m})} && (42)\\
&\overset{(8.4.2)}{=} p_k. && (43)
\end{aligned}$$

3 Konzentration der empirischen Verteilung der Grade

(8.5.2) kann man auch so schreiben:

$$\mathbb{P}\left(\max_{k\in\mathbb{N}_0}\left|P_k(t)-\mathbb{E}[P_k(t)]\right|\geq C\sqrt{\frac{\log t}{t}}\right)=o(1). \tag{44}$$

Das ist etwas näher an (8.4.5).

4 Die erwartete empirische Gradverteilung

Der Beweis von Theorem 8.3 geht etwas direkter: Setze $C_3 := C_1 + \frac{C_2}{\sqrt{2\log 2}} \geq C_1 + \frac{C_2}{\sqrt{t\log t}}$, wobei $C_1 > 8\sqrt{m}$ die Konstante aus Proposition 8.4, C_2 die Konstante aus Proposition 8.7 und $t \geq 2$ ist. Dann folgt für alle $t \geq 2$

$$\mathbb{P}\left(\max_k |P_k(t)-p_k|\geq C_3\sqrt{\tfrac{\log(t)}{t}}\right) \tag{45}$$

$$\leq \mathbb{P}\left(\max_k |P_k(t)-\mathbb{E}[P_k(t)]|\geq C_1\sqrt{\tfrac{\log(t)}{t}}\right)+\mathbb{P}\left(\max_k |\mathbb{E}[P_k(t)]-p_k|\geq \tfrac{C_2}{t}\right) \tag{46}$$

$$\xrightarrow{t\to\infty} 0. \tag{47}$$

—

Stimmt (8.6.14) mit (8.6.12) für $k=1$ überein?

$$(2+\delta)\frac{\Gamma(1+\delta)\Gamma(3+2\delta)}{\Gamma(4+2\delta)\Gamma(1+\delta)} \overset{(8.3.2)}{=} (2+\delta)\frac{\Gamma(3+2\delta)}{(3+2\delta)\Gamma(3+2\delta)} = \frac{2+\delta}{3+2\delta} \tag{48}$$

Zu (8.6.16): Mit Induktion ist das übersichtlicher.

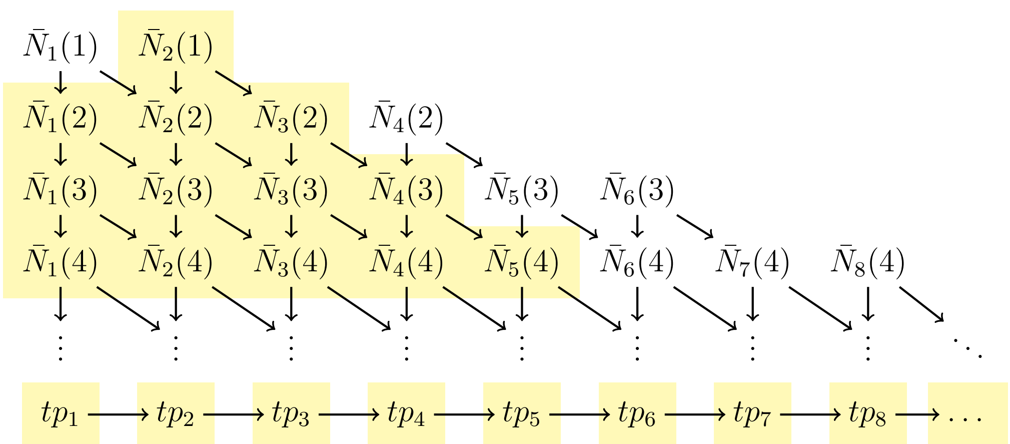
$$p_k \overset{(8.6.15)}{=} \frac{k+\delta-1}{k+2+2\delta}p_{k-1} \tag{49}$$

$$\overset{\text{IV}}{=} \frac{k+\delta-1}{k+2+2\delta}\cdot(2+\delta)\frac{\Gamma(k-1+\delta)\Gamma(3+2\delta)}{\Gamma(k+2+2\delta)\Gamma(1+\delta)} \tag{50}$$

$$\overset{(8.3.2)}{=} (2+\delta)\frac{\Gamma(k+\delta)\Gamma(3+2\delta)}{\Gamma(k+3+2\delta)\Gamma(1+\delta)} \tag{51}$$

—

Der Vergleich von $\bar{N}_k(t)$ mit tp_k wird über die Rekursionen (8.6.10) bzw. (8.6.11) geführt. Wie die Werte rekursiv in spätere Werte eingehen, ist hier dargestellt. Die gelb hinterlegten Werte sind positiv.



Die Rekursionen für $\bar{N}_k(t)$ und tp_k führen auf die Rekursionsformel (8.6.22) für $\varepsilon_k(t)$. In Lemma 8.9 wird mit Induktion über t

$$\sup_t |\varepsilon_1(t)| < \infty. \tag{52}$$

gezeigt. In Lemma 8.10 kommt dann eine weitere Induktion über k und beendet den Beweis dass

$$\sup_{k,t} |\varepsilon_1(t)| < \infty. \tag{53}$$

—

Zu den Abschätzungen in (8.6.29): Wir haben aus (8.6.15) für alle $k \geq 2$:

$$\frac{(k+\delta)p_k}{(k-1+\delta)p_{k-1}} = \frac{k+\delta}{k-1+\delta}\frac{k-1+\delta}{k+2+2\delta} = \frac{k+\delta}{k+2+2\delta} < 1. \tag{54}$$

Das heißt für alle $k \geq 1$

$$\begin{aligned}
|\kappa_k(t)| &= \Big(\frac{1}{2+\delta} - \frac{t}{t(2+\delta)+1+\delta}\Big)\big|(k-1+\delta)p_{k-1} - (k+\delta)p_k\big| && (55)\\
&\leq \frac{1}{2+\delta}\Big(1 - \frac{t}{t+\frac{1+\delta}{2+\delta}}\Big)\sup_{k\in\mathbb{N}}(k+\delta)p_k && (56)\\
&= \frac{1}{2+\delta}\frac{\frac{1+\delta}{2+\delta}}{t+\frac{1+\delta}{2+\delta}}(1+\delta)p_1 && (57)\\
&= \frac{1+\delta}{2+\delta}\frac{1}{t\frac{2+\delta}{1+\delta}+1}\frac{2+\delta}{3+2\delta} && (58)\\
&\leq \frac{1+\delta}{3+2\delta}\frac{1}{t+1}. && (59)
\end{aligned}$$

Die andere Abschätzung ist einfacher:

$$|\gamma_k(t)| \leq \frac{1+\delta}{t(2+\delta)+1+\delta} = \frac{1}{t\frac{2+\delta}{1+\delta}+1} \leq \frac{1}{t+1}. \tag{60}$$

5 8.6.2: $m \geq 1$

Um (8.6.48) zu verstehen, verwenden wir die direkte Konstruktion des Preferential Attachment Modells. Dafür notiere $\mathrm{PA}_{e,t}^{(m,\delta)}$ den Graphen nach dem Hinzufügen der e-ten Kante des Knoten $t+1$. Weiter sei $D_i(e,t)$ der Grad des Knoten i nach dem Hinzufügen der e-ten Kante des Knoten $t+1$. Die Übergangswahrscheinlichkeiten sind dann mit $\delta' := \frac{\delta}{m}$

$$\mathbb{P}\big(v_{t+1}^{(m)} \xrightarrow{e} v_i^{(m)} \mid \mathrm{PA}_{e-1,t}^{(m,\delta)}\big) = \begin{cases} \dfrac{D_i(e-1,t)+\delta}{(mt+e-1)(2+\delta')+1+\delta'} & i \leq t, \\ \dfrac{D_{t+1}(e-1,t)+1+e\delta'}{(mt+e-1)(2+\delta')+1+\delta'} & i = t+1. \end{cases} \tag{61}$$

Sicherheitshalber überprüfen wir

$$\begin{aligned}
\sum_{i=1}^{t+1}\mathbb{P}\big(v_{t+1}^{(m)} \xrightarrow{e} v_i^{(m)} \mid \mathrm{PA}_{e-1,t}^{(m,\delta)}\big) &= \frac{2(mt+e-1)+t\delta+1+e\delta'}{(mt+e-1)(2+\delta')+1+\delta'} && (62)\\
&= \frac{2mt+2e-1+t\delta+e\delta'}{2mt+2e-1+mt\delta'+(e-1)\delta'+\delta'} = 1. && (63)
\end{aligned}$$

Damit können wir abschätzen:

$$\sum_{k=m+1}^{2m} \alpha_k(t) = \sum_{k=m+1}^{2m} \mathbb{P}(D_{t+1}(t+1) = k) \tag{64}$$

$$= \mathbb{P}(D_{t+1}(t+1) \geq k) \tag{65}$$

$$\leq \sum_{e=1}^{m} \mathbb{E}\big[\mathbb{P}(v_{t+1}^{(m)} \xrightarrow{e} v_{t+1}^{(m)} \mid \mathrm{PA}_{e-1,t}^{(m,\delta)})\big] \tag{66}$$

$$= \sum_{e=1}^{m} \frac{\mathbb{E}[D_{t+1}(e-1,t)] + 1 + e\delta'}{(mt+e-1)(2+\delta') + 1 + \delta'} \tag{67}$$

$$D_{t+1}(e-1,t) \leq 2(e-1) \implies \leq \sum_{e=1}^{m} \frac{2(e-1) + 1 + e\delta'}{mt(2+\delta') + 1 + \delta'} \tag{68}$$

$$= \frac{(2+\delta')\sum_{e=1}^{m} e - m}{mt(2+\delta') + 1 + \delta'} \tag{69}$$

$$= \frac{(2+\delta')\frac{1}{2}m(m+1) - m}{mt(2+\delta') + 1 + \delta'} \tag{70}$$

$$= \frac{m}{2} \cdot \frac{(2+\delta')m + \delta'}{mt(2+\delta') + 1 + \delta'}. \tag{71}$$

Das ist nicht ganz (8.6.48), und wir können (8.6.48) von hier aus auch nicht mehr erreichen, wie man sieht, wenn man sich die Grenzwerte $\delta \searrow -1$ überlegt. Es sieht aus, als wäre die Abschätzung in (66) zu grob. Wichtig ist hier aber das Verhalten für $t \to \infty$ und die Konsequenz für $\gamma_k(t) := \alpha_k(t) - \delta_{k,m}$:

$$\frac{|\gamma_k(t)|}{\frac{1}{t+1}} = \begin{cases} (t+1)(1-\alpha_m(t)) & k = m \\ (t+1)\alpha_k(t) & k \neq m \end{cases} \tag{72}$$

$$\leq (t+1) \sum_{k=m+1}^{2m} \alpha_k(t) \tag{73}$$

$$\leq (t+1) \frac{\mathrm{const.}_{m,\delta}}{mt(2+\delta') + 1 + \delta'} \tag{74}$$

$$\xrightarrow{t\to\infty} \frac{\mathrm{const.}_{m,\delta}}{m(2+\delta')} < \infty. \tag{75}$$

Das etabliert (8.6.50).

—