

# Cumulative Advantage in Online Markets

Stefan Wehrli

ETH Zürich

VIU, December 2006

- 1 Einführung
  - ▶ Kumulative Vorteile in den Wissenschaften
  - ▶ Beispiele: Scale-free Networks
- 2 Reputationseffekte in Online Märkten (Mikro-Ebene)
  - ▶ Games
  - ▶ Daten und Ergebnisse
- 3 Reputationssystem als Trade-Network (Makro-Ebene)
  - ▶ Verteilungen
  - ▶ Generative Prozesse
  - ▶ Degree-Verteilungsanalyse
- 4 Fazit, Kritik & Ausblick

# Cumulative Advantage (CA)

- "Cumulative advantage [...] refers to the social process through which various kinds of opportunities for scientific inquiry [...] tends to *accumulate*." (Merton 1988) → Matthäus-Effekt
- "Cumulative advantage is a general mechanism for inequality across any temporal process [...] in which a favorable relative position becomes a resource that produces further relative gains." (Diprete 2006)

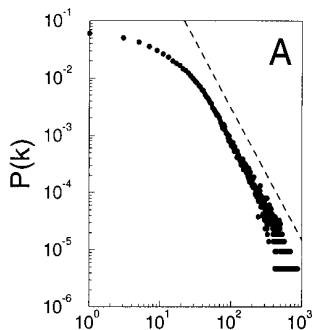
## Einfachstes Beispiel:

Entwicklung eines Kapitalstocks mit fixem Zinssatz über die Zeit

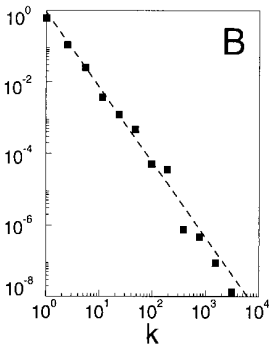
## CA auch bekannt unter:

"Preferential Attachment", "Positive Rückkopplung", "Path dependent increasing returns", "multiplier effects", "first mover advantage", "scarring effects", "the rich get richer", ...

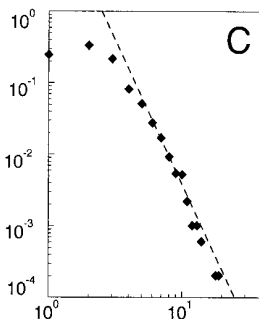
# Aktuelle Beispiele: Scale-free Networks



Movie Actors



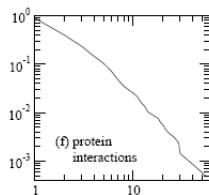
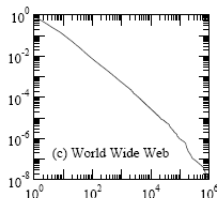
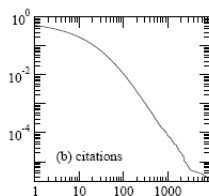
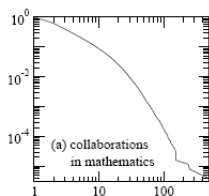
WWW



Power Grid

Albert und Barabási (1999): Emergence of Scaling in Random Networks

# Weitere skalen-invariante Netzwerke / Verteilungen



Weitere Beispiele:

- Sexual-Netzwerke \*
- Siedlungsgrößen (US)
- Unternehmensgrößen
- Vermögensverteilungen
- Anzahl verkaufte Bücher
- Kriegsintensitäten \*
- Wort-Häufigkeiten (Zipf)
- Familiennamen
- ...

Newman (2003): The Structure and Function of Complex Networks.

\* Genaue Verteilungscharakteristika noch umstritten

# Charakteristika

- Kleine Ausprägungen treten sehr häufig auf, grosse dagegen extrem selten.
- Fat Tails: Degree-Verteilungen extrem rechtsschief.
- Small Worlds: Kleiner Diameter ( $\ln \ln N$ ) und kurze durchschnittliche Pfadlängen.
- Clustering: Dichte Nachbarschaften, Communities, geschlossene Triaden
- ...

# Relevanz von Powerlaws




## "The New Science of Networks": Barabási (2002)

- "Nature's unmistakable sign that chaos is departing in favor of order."
- "The scale-free topology is evidence of organizing principles acting at each stage of the network formation process."
- "We expect that the scale-invariant state [...] is a generic property of many complex networks, with applicability reaching far beyond the quoted examples."

## The Classics: Simon (1955)

- "The mere fact that particular data conform to the Yule distribution [...] tells little about the underlying phenomena beyond what is contained in [our] assumptions [...]"
- "The frequency with which the Yule distribution occurs in nature - particularly in social phenomena - should occasion no great surprise."

# Daten: Feedback-Profil auf eBay.de



**Bewertungsprofil: player004 (63203)**   **Power Seller**  

**Bewertungsprofil:** 63203  
**Positive Bewertungen:** 99,7%

Mitglieder, die mich positiv bewertet haben:	63376
Mitglieder, die mich negativ bewertet haben:	185

Alle positiven Bewertungen: 85131  
[Weitere Informationen](#) zur Bedeutung dieser Zahlen.

**Jüngste Bewertungen:**

	Letzter Monat	Letzte 6 Monate	Letzte 12 Monate
 positiv	1614	10277	32063
 neutral	9	40	97
 negativ	7	27	64

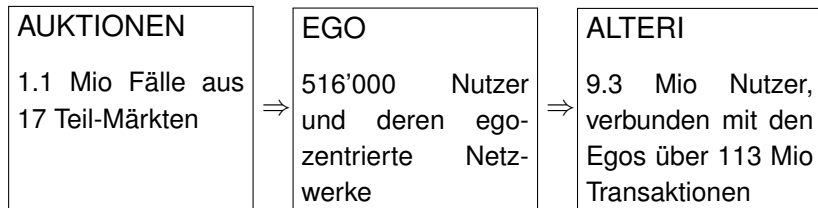
Zurückgezogene Gebote (in den letzten 6 Monaten): 0

- Prozessdaten von eBay.de, automatisierte Erhebung.
- Anzahl erhaltene Bewertungen eines eBay-Mitglieds entsprechen seinem Degree im Reputationsnetzwerk.
- Reputationsnetzwerk kann nur wachsen.
- Unterscheidung in ein Belohnungs- und Sanktionsnetzwerk.

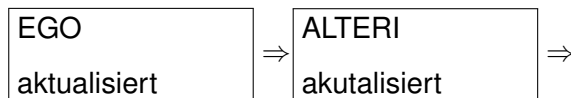


# Daten: Aufbau der Erhebung

2005



2006



Ziel: Vollerhebung des deutschen Reputationsnetzwerkes → Analyse des Netzwerkwachstums

# Mikro-Ebene: Games

## Transaktionsprozess

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	P, P	P, P

Problem:

Vertrauensspiel

Lösung:

Reputation

Payoffs:  $T > R > P > S$

## Bewertungsprozess

		Player 2		
		+	-	0
Player 1	+	$a - c, a - c$	$-a\gamma - c, a - c$	$-c, a$
	-	$a - c, -a\gamma - c$	$-a\gamma - c, -a\gamma - c$	$-c, -a\gamma$
	0	$a, -c$	$-a\gamma, -c$	$0, 0$

Problem:

Sequentielles PD

Lösung:

Reziprozität & Reputation

Payoffs:  $a > c, a > 0, c > 0$

# Reputationseffekte: empirische Evidenzen

## Transaktionsprozess

- Zeit bis erstes Gebot
- Verkaufswahrscheinlichkeit
- Endpreis
- Wahl des Startpreises
- Wahl des Zahlungsmodus

⇒ Direkte Effekte auf Fitness

## Bewertungsprozess

- Bewertungsrate: 70-80%
- Direkte Reziprozität (Second-Mover)
- Effekt der Partner-Reputation auf die Bewertungsrate (First-Mover)

⇒ Indirekte Effekte auf Fitness

## Offene Fragen:

- Geburts-, Wachstums- und Sterbeprozesse

Wehrli (2005): Alles bestens, gerne wieder. Reputation und Reziprozität in Online-Auktionen.

# Ergebnisse: Zeit bis zum ersten Gebot

## Proportional-Hazards-Modelle für das erste Gebot

	Verkäufer		Käufer	
Reputation (log. Score)	0.048***	(0.005)		
Positive Reputation (log.)			0.053***	(0.006)
Negative Reputation (log.)			-0.011	(0.007)
Verk. hat Shop	-0.081***	(0.024)	-0.077**	(0.024)
Verk. hat „Mich“-Seite	-0.216***	(0.023)	-0.210***	(0.023)
Verk. ist geprüftes Mitglied	-0.173***	(0.025)	-0.177***	(0.025)
Verk. ist Powerseller	-0.081**	(0.025)	-0.079**	(0.025)
Startpreis	-0.107***	(0.003)	-0.107***	(0.003)
Versandkosten	-0.098***	(0.006)	-0.098***	(0.006)
Beschreibungslänge (log.)	-0.014*	(0.007)	-0.014*	(0.007)
Galeriebild	0.057*	(0.024)	0.056*	(0.024)
Fettschrift	0.526***	(0.118)	0.529***	(0.118)
Fallzahl	34'797		34'797	
Anzahl „Ereignisse“	18'607		18'607	

Standardfehler in Klammern; Stratifiziert nach Unterkategorie und Wochentag;  
 Reputation = logarithmierte Anzahl Bewertungen

# Ergebnisse: Verkaufswahrscheinlichkeit

## Verkaufserfolg (Logistische Regression / Mobiles)

	Neu		Gebraucht	
Positive Reputation (log.)	0.142*	(0.071)	-0.014	(0.044)
Negative Reputation (log.)	-0.353*	(0.149)	-0.118	(0.117)
Startpreis	-0.032***	(0.003)	-0.035***	(0.002)
Anbieterkonkurrenz	-0.008*	(0.003)	-0.007**	(0.003)
Auktionsdauer	0.017	(0.038)	0.051*	(0.022)
Kalenderzeit	-0.011	(0.010)	-0.002	(0.006)
Beschreibungslänge (log.)	-0.018	(0.078)	-0.039	(0.060)
Produktbild	0.134	(0.387)	0.331	(0.218)
(...)				
Anteil verkauft	95.06%		95.47%	
Fallzahl	5'338		8'738	

Standardfehler in Klammern (geklumpt); Reputation = logarithmierte Anzahl Bewertungen

# Ergebnisse: Endpreis

## Effekte auf den Verkaufspreis (OLS / Mobiles)

	Neu		Gebraucht	
Positive Reputation (log.)	1.069***	(0.209)	1.047***	(0.231)
Negative Reputation (log.)	-1.995***	(0.501)	-2.738***	(0.472)
Startpreis	0.034***	(0.006)	0.127***	(0.010)
Anzahl Bieter	0.573***	(0.079)	1.495***	(0.143)
Anbieterkonkurrenz	-0.070***	(0.009)	-0.059***	(0.011)
Auktionsdauer	0.050	(0.109)	-0.105	(0.121)
Kalenderzeit	-0.312***	(0.026)	-0.260***	(0.027)
Kalenderzeit quadriert	0.023***	(0.003)	0.022***	(0.003)
Beschreibungslänge (log.)	1.001***	(0.236)	2.711***	(0.280)
Produktbild	3.736**	(1.274)	4.002*	(1.558)
(...)				
Korrigiertes R <sup>2</sup>	0.842		0.789	
Fallzahl	5'096		8'306	

Abhängige Variable: Verkaufspreis (EUR); Standardfehler in Klammern  
(geklumpt); Reputation = logarithmierte Anzahl Bewertungen

# Ergebnisse: Bewertungsrate

## Proportional-Hazards-Modelle für die Vergabe von Bewertungen

	Verkäufer		Käufer	
Partner zuerst (zeitveränd.)	2.021***	(0.118)	0.617***	(0.011)
Positive Partnerreputation	0.088***	(0.006)	0.098***	(0.004)
Negative Partnerreputation	-0.184***	(0.016)	-0.085***	(0.006)
Partner zuerst X pos. P.rep.	-0.099***	(0.010)	-0.029***	(0.005)
Partner zuerst X neg. P.rep.	0.163***	(0.023)	-0.117***	(0.010)
Eigene positive Reputation	0.050**	(0.016)	0.141***	(0.004)
Eigene negative Reputation	-0.199***	(0.029)	-0.253***	(0.010)
Wiederholte Interaktion	-0.281***	(0.077)	-0.467***	(0.033)
Rollenwechsel	0.175***	(0.051)	0.185***	(0.022)
Geprüftes Mitglied	0.210*	(0.084)	0.190***	(0.037)
Preis	-0.002	(0.001)	-0.000	(0.000)
(...)				
Fallzahl (Klumpen)	177'561	(29'816)	177'561	(99'139)
Anzahl „Ereignisse“	146'693		146'300	

Bootstrap-Standardfehler in Klammern (geklumpt); Reputation = logarithmierte Anzahl Bewertungen

# Mikro-Makro-Verbindung: Generative Prozesse

## Yule Prozess

- Geburtsprozess
- Preferential Attachment

⇒ Yule Verteilung

⇒ Pareto Verteilung

## Gibrat Prozess

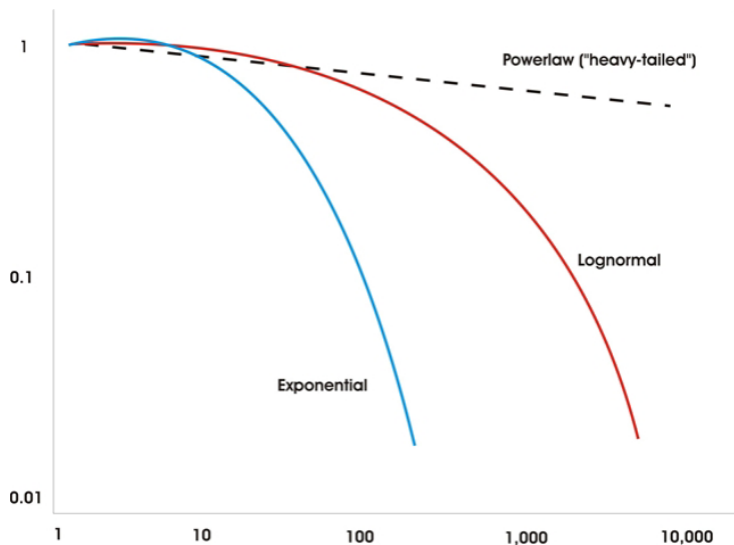
- Grosse Zahl kleiner multiplikativer Einflüsse
- Proportionales Wachstum, unabhängig von Grösse

⇒ Lognormale Verteilung

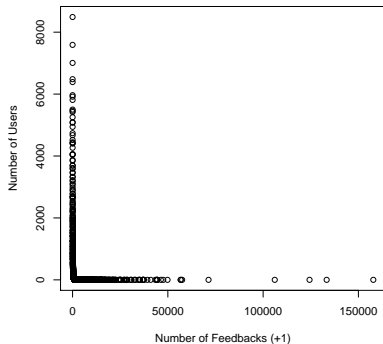
Übersicht in Simkin und Roychowdhury (2006): Re-inventing Willis.



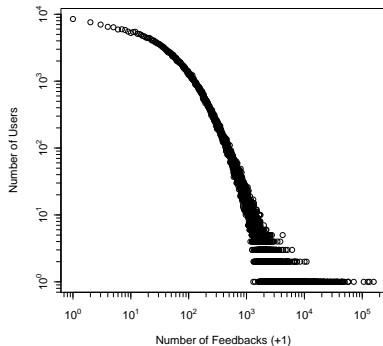
# Verteilungen



# Verteilung der positiven Bewertungen (2005)

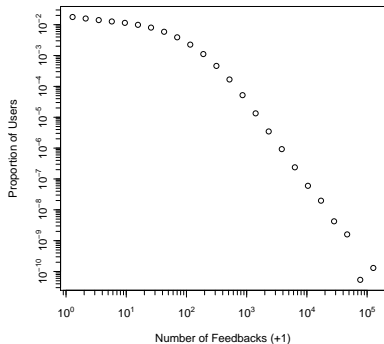


(a) normal

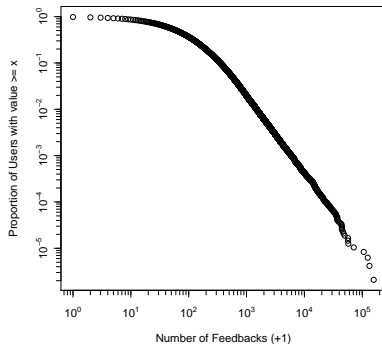


(b) Loglog-Plot

# Verteilung der positiven Bewertungen (2005)

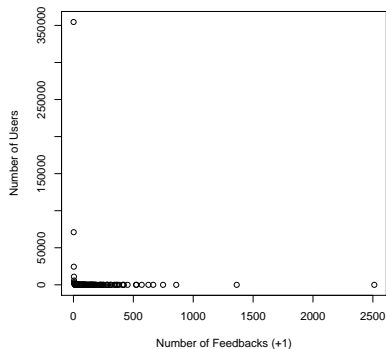


(c) log bins

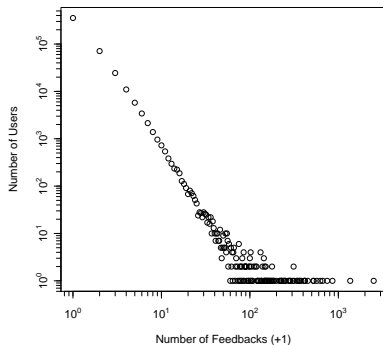


(d) ccdf (1-cdf)

# Verteilung der negativen Bewertungen (2005)

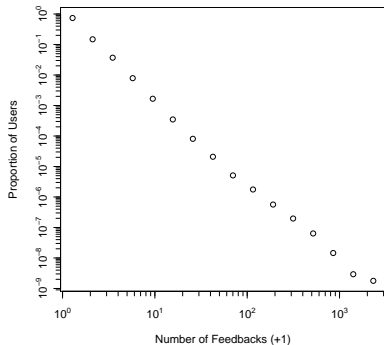


(a) normal

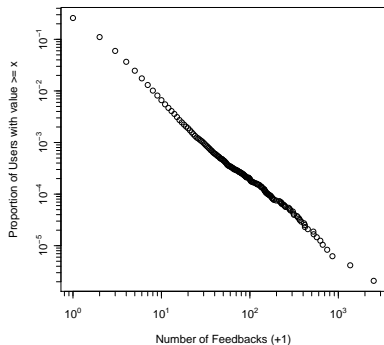


(b) Loglog-Plot

# Verteilung der negativen Bewertungen (2005)



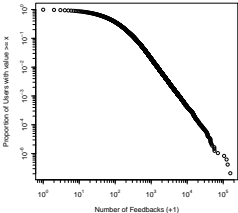
(c) log bins



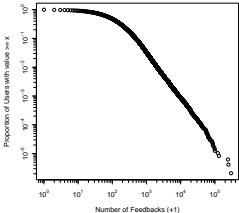
(d) ccdf (1-cdf)

# Degree-Verteilungen im Vergleich

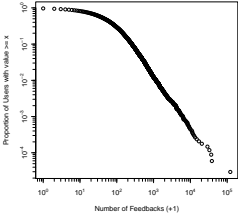
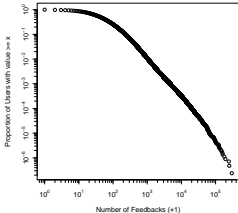
2005



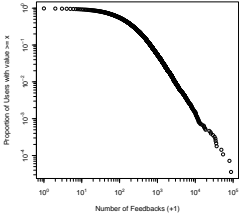
2006



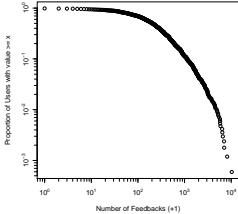
Alteri



Mobiles



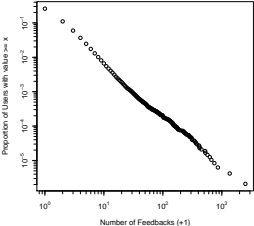
DVDs



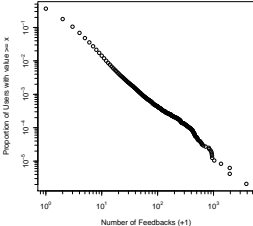
Muenzen

# Degree-Verteilungen im Vergleich

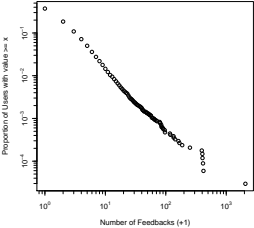
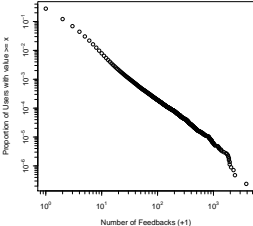
2005



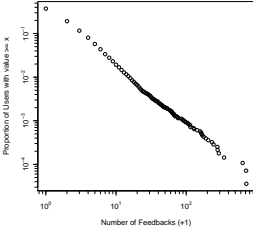
2006



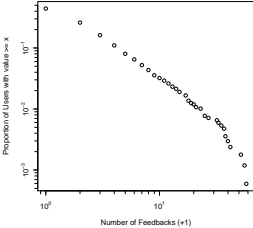
Alteri



Mobiles



DVDs



Muenzen

# Schätzung Skalenparameter $\gamma$

Hypothese:  $2 \leq \gamma \leq 3$

(1) Graphische Extraktion am Loglog-Plot oder (2) OLS bei Linearität

(3) Maximum Likelihood

Pareto (PDF und CDF):

$$\begin{aligned}Pr[X = x] &= \alpha a^\alpha x^{-(\alpha+1)} \\Pr[X \leq x] &= 1 - (a/x)^\alpha\end{aligned}$$

Power law (PDF):

$$\begin{aligned}Pr[K = k] &\sim k^{-\gamma} \\ \Rightarrow \gamma &= \alpha + 1\end{aligned}$$

Hill's Estimator von  $\alpha$

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \log\left(\frac{x_i}{a}\right)}$$



# Schätzung Skalenparameter $\gamma$

	$\alpha(a = 1)$	$n$	$\alpha(a = 100)$	$n$
2005	<b>0.257</b>	457'255	<b>1.017</b>	157'925
2006	<b>0.225</b>	475,311	<b>0.988</b>	228'723
Alteri	<b>0.266</b>	3'280'884	<b>1.201</b>	873'525

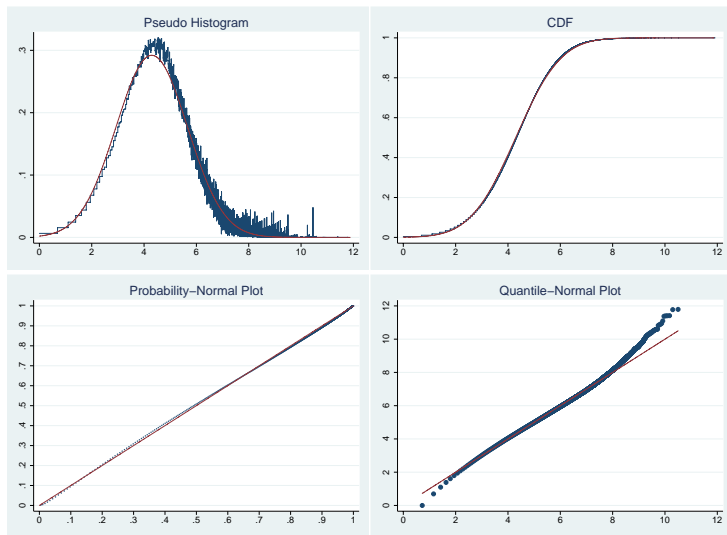
Verteilung der positiven Feedbacks  
Bootstrap-Schätzungen des Pareto-Index  $\alpha$

	$\alpha(a = 1)$	$n$
2005	<b>2.012</b>	124,011
2006	<b>1.643</b>	174'811
Alteri	<b>1.874</b>	933'398

Verteilung der negativen Feedbacks  
Bootstrap-Schätzungen des Pareto-Index  $\alpha$

$$\gamma = \alpha + 1, N_{ego} = 478,614, N_{alteri} = 3'290'613$$

# Oder: Positive Feedbacks sind lognormal verteilt



Nur Verkäufer! Pareto und lognormale Verteilung lassen sich kaum unterscheiden.

# Schlussfolgerungen

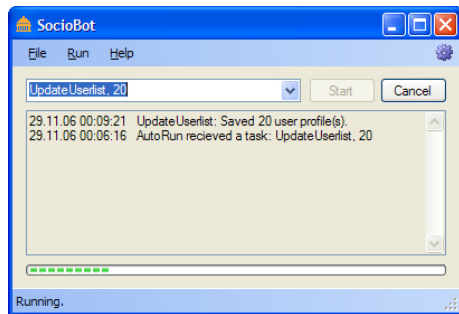
- Powerlaw-Verteilungen sind weder neu noch besonders!
- Aus Verteilungen kann man nicht auf die zugrundeliegenden Mikro-Prozesse schliessen.
- CA-Effekte von Reputation können auf der Mikro-Ebene empirisch gezeigt werden.
- Auf der Makro-Ebene zeigen sich charakteristische Verteilungen (Lognormal, Pareto), die auf einen kumulativen Prozess hinweisen.
- Geburts- und Sterbeprozesse sind zentral für die Prüfung des generativen Prozessmodells. Ein integrales CA-Modell für eBay fehlt noch!
- Sollten Reputationsprämien auf der Mikro-Ebene tatsächlich eine Form von Sozialkapitalrendite darstellen, dann sei mit der lognormalen Verteilung ein Indiz der Kapitallogik von Reputation gefunden.

# Reputation@Office Reloaded

Verteiltes Beobachten. Helft mir das Netzwerk zu vervollständigen!

Mail:

wehrli@soz.gess.ethz.ch



# Literatur

- Axtell, R.L. (2001): Zipf Distribution of U.S. Firm Sizes, *Science* 293: 1818-1820.
- Barabasi, A-L., Albert, R. (1999): Emergence of Scaling in Random Networks, *Science* 286: 509-512.
- Barabasi, A-L. (2002): *Linked: The new science of networks*. Cambridge: Perseus Publishing.
- DiPrete, T., Eirich, G. (2006): Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments. *American Sociological Review* 32: 271-297.
- Gibrat, R. (1931): Les inégalités économiques; applications: aux inégalités des richesses, à la concentrations des entreprise, aux populations des villes, aux statistiques des familles, etc., d'une nouvelle, la loi de l'effet proportionnel. Paris: Librairie du Recueil Sirey.
- Merton, R.K. (1988): The Matthew Effect in Science, II. Cumulative Advantage and the Symbolism of Intellectual Property. *ISIS* 79: 606-623.
- Newman, M.E.J. (2003): The Structure and Function of Complex Networks. *SIAM Review* 45.
- Pareto, V. (1896): *Cours d'Economie Polotique*. Genf: Droz.
- Simkin und Roychowdhury (2006): *Re-inventing Willis*. University of California, Los Angeles.
- Simon, H.A. (1955): On a Class of Skew Distribution Functions, *Biometrika* 42:425-440.
- Zipf, G.K. (1949): *Human Behaviour and the Principle of Least-Effort*. Addison-Wesley, Cambridge MA.