Mechanisms for Generating Power-Law Size Distributions, Part 2

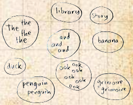
Principles of Complex Systems | @pocsvox
CSYS/MATH 300, Fall, 2017

Prof. Peter Dodds | @peterdodds

Dept. of Mathematics & Statistics | Vermont Complex Systems Center
Vermont Advanced Computing Core | University of Vermont

PoCS
Principles of Complex Systems
@pocsvox
What's the Story?

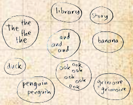
These slides are brought to you by:



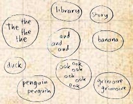
These slides are also brought to you by:

Special Guest Executive Producer: Pratchett

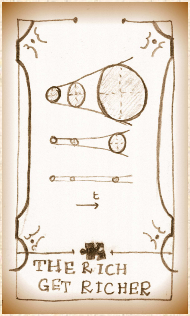
On Instagram at pratchett_the_cat



Outline

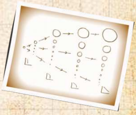

the the the the
library
story
and and and
banana
duck
ook ook ook ook ook ook
penguin penguin
grimoire grimoire


THE SUN
THE GOLDEN AGE OF REDUCTIONISM
THE MANIFESTO
THE WELL OF SCALES
THE LAWS OF DECAYING POWER
MINISTRY OF RANDOM WALKS
THE UNEXPECTED TRANSFIGURATION
THE DREAM OF EFFICIENT LANGUAGE
THE LAW OF DIGITS FIRST
THE SURPRISE OF BEING ROBUST YET FRAGILE
THE MISREADING OF THE BOOK OF SAND
THE SUPPLY MOST FAIR
THE SUPPLY TO THE MOST
THE MANY TALES OF MANY TAILS
THE MEASURING
THE DATA POOR
THE POURING DATA
THE CHARMS OF THE MANY WRONG PATHS
THE KNOWLEDGE OF DEAD MONKS
THE DOORS THAT SHOULD NOT BE OPENED
THE EMERGENCE OF STRUCTURE
THE EMERGENCE OF DESTRUCTION
THE EMERGENCE OF THINKING
It was the best of times,
THE EMERGENCE OF STORIES
THE MECHANICS STATISTICAL
THE THRESHOLDS OF PERCOLATION
THE NETWORKS COMPLEX
THE ELEMENTS OF WEBS
THE ARK OF THE RANDOM NETWORK
THE BIRDS OF A FEATHER
THE BIRDS OF NOT A FEATHER
THE STRANGENESS OF FRIENDS
THE SMALLNESS OF WORLDS
THE THEORY OF SIX DEGREES
THE NETWORKS FREE OF SCALE
CLASSIFIED
PHOENIX
BIO HAZARD
THE CONTAGION
THE CONFUSION OF CONTAGIONS
THE DISEASE OF RANDOM MIXING
THE UNENDING PANDEMICS
THE RATIO OF GAIN
THE THRESHOLDS OF THE MIND
THE SOCIAL WILD
THE SOCIAL CONTAGION
THE MECHANICS OF FAME
THE STORIES CONTAGIOUS
THE PATHS OF UNIVERSALITY
THE DEPENDENT PATHS
THE TROUSERS OF BIFURCATING REALITIES
THE UNKNOWN MECHANISM
THE JOHN DORY
THE STORYTELLERS
THE ADJACENT NARRATIVES
THE END


t
THE RICH GET RICHER

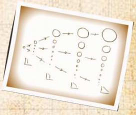
Aggregation:

- Random walks represent additive aggregation
- Mechanism: Random addition and subtraction
- Compare across realizations, no competition.
- Next: Random Additive/Copying Processes involving Competition.
- Widespread: Words, Cities, the Web, Wealth, Productivity (Lotka), Popularity (Books, People, ...)
- Competing mechanisms (trickiness)



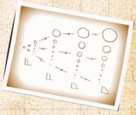
Pre-Zipf's law observations of Zipf's law

- 1910s: Word frequency examined re Stenography (or shorthand or brachygraphy or tachygraphy), Jean-Baptiste Estoup [12].
- 1910s: Felix Auerbach pointed out the Zipfitude of city sizes in "Das Gesetz der Bevölkerungskonzentration" ("The Law of Population Concentration") [1].
- 1924: G. Udny Yule [31]:
  # Species per Genus (offers first theoretical mechanism)
- 1926: Lotka [17]:
  # Scientific papers per author (Lotka's law)



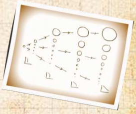
Theoretical Work of Yore:

- 1949: Zipf's "Human Behaviour and the Principle of Least-Effort" is published.[33]
- 1953: Mandelbrot[19]: Optimality argument for Zipf's law; focus on language.
- 1955: Herbert Simon[27, 33]: Zipf's law for word frequency, city size, income, publications, and species per genus.
- 1965/1976: Derek de Solla Price[8, 9]: Network of Scientific Citations.
- 1999: Barabasi and Albert[2]: The World Wide Web, networks-at-large.



Herbert Simon (1916–2001):

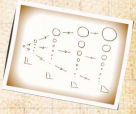
- Political scientist (and much more)
- Involved in Cognitive Psychology, Computer Science, Public Administration, Economics, Management, Sociology
- Coined ‘bounded rationality’ and ‘satisficing’
- Nearly 1000 publications (see Google Scholar)
- An early leader in Artificial Intelligence, Information Processing, Decision-Making, Problem-Solving, Attention Economics, Organization Theory, Complex Systems, And Computer Simulation Of Scientific Discovery.
- 1978 Nobel Laureate in Economics (his Nobel bio is here).



Essential Extract of a Growth Model:

Random Competitive Replication (RCR):

1. Start with 1 elephant (or element) of a particular flavor at $t = 1$
2. At time $t = 2, 3, 4, \ldots$, add a new elephant in one of two ways:
   - With probability ρ, create a new elephant with a new flavor
     = Mutation/Innovation
   - With probability $1 - \rho$, randomly choose from all existing elephants, and make a copy.
     = Replication/Imitation
   - Elephants of the same flavor form a group

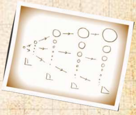


Random Competitive Replication:

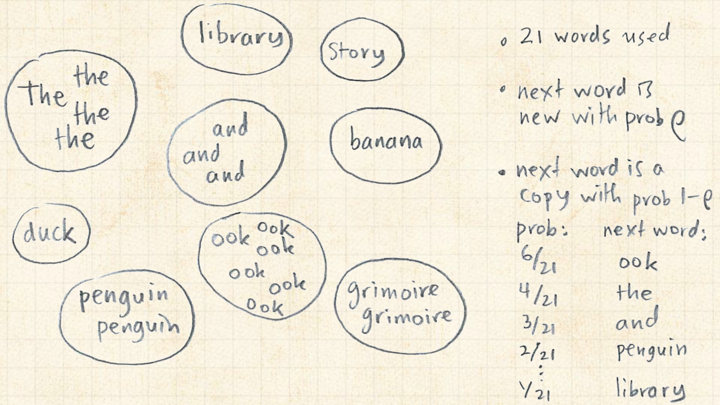
Example: Words appearing in a language

- Consider words as they appear sequentially.
- With probability ρ, the next word has not previously appeared
  = Mutation/Innovation
- With probability $1 - \rho$, randomly choose one word from all words that have come before, and reuse this word
  = Replication/Imitation

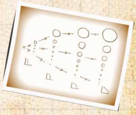
Note: This is a terrible way to write a novel.



For example:



the
The the
the
library
Story
and
and
and
banana
duck
ook ook
ook
ook
ook
ook
penguin
penguin
grimoire
grimoire
• 21 words used
• next word is
new with prob ρ
• next word is a
copy with prob 1−ρ
prob: next word:
6/21 ook
4/21 the
3/21 and
2/21 penguin
1/21 library

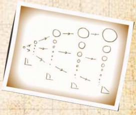


Some observations:

- Fundamental Rich-get-Richer story;
- Competition for replication between individual elephants is random;
- Competition for growth between groups of matching elephants is not random;
- Selection on groups is biased by size;
- Random selection sounds easy;
- Possible that no great knowledge of system needed (but more later ...).

Your free set of tofu knives:

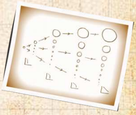
- Related to Pólya's Urn Model, a special case of problems involving urns and colored balls.
- Sampling with super-duper replacement and sneaky sneaking in of new colors.

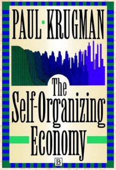


Random Competitive Replication:

Some observations:

- Steady growth of system: +1 elephant per unit time.
- Steady growth of distinct flavors at rate ρ
- We can incorporate
  1. Elephant elimination
  2. Elephants moving between groups
  3. Variable innovation rate ρ
  4. Different selection based on group size
     (But mechanism for selection is not as simple...)

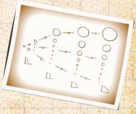




"The Self-Organizing Economy" by Paul Krugman (1996). [16]

Ch. 3: An Urban Mystery, p. 46

"...Simon showed—in a completely impenetrable exposition!—that the exponent of the power law distribution should be ..."[1, 2]



[1] Krugman's book was handed to the Deliverator by a certain Álvaro Cartea many years ago at the Santa Fe Institute Summer School.

[2] Let's use π for probability because π's not special, right guys?

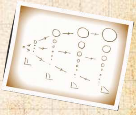
Random Competitive Replication:

Definitions:

- k_i = size of a group i
- $N_{k,t}$ = # groups containing k elephants at time t.

Basic question: How does $N_{k,t}$ evolve with time?

First: $\sum_k kN_{k,t} = t =$ number of elephants at time t

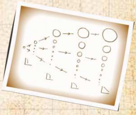


Random Competitive Replication:

$P_k(t)$ = Probability of choosing an elephant that belongs to a group of size k:

- $N_{k,t}$ size k groups
- $\Rightarrow kN_{k,t}$ elephants in size k groups
- t elephants overall

$$P_k(t) = \frac{kN_{k,t}}{t}.$$

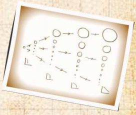


Random Competitive Replication:

$N_{k,t}$, the number of groups with k elephants, changes at time t if

1. An elephant belonging to a group with k elephants is replicated:
   $N_{k,t+1} = N_{k,t} - 1$
   Happens with probability $(1-\rho)kN_{k,t}/t$

2. An elephant belonging to a group with $k-1$ elephants is replicated:
   $N_{k,t+1} = N_{k,t} + 1$
   Happens with probability $(1-\rho)(k-1)N_{k-1,t}/t$

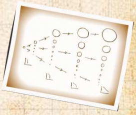


Random Competitive Replication:

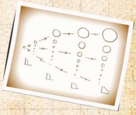
Special case for $N_{1,t}$:

1. The new elephant is a new flavor:
   $N_{1,t+1} = N_{1,t} + 1$
   Happens with probability ρ

2. A unique elephant is replicated:
   $N_{1,t+1} = N_{1,t} - 1$
   Happens with probability $(1-\rho)N_{1,t}/t$



Random Competitive Replication:



Putting everything together:

For $k > 1$:

$$\langle N_{k,t+1} - N_{k,t} \rangle = (1-\rho)\left((+1)(k-1)\frac{N_{k-1,t}}{t} + (-1)k\frac{N_{k,t}}{t} \right)$$

For $k = 1$:

$$\langle N_{1,t+1} - N_{1,t} \rangle = (+1)\rho + (-1)(1-\rho)1 \cdot \frac{N_{1,t}}{t}$$

Random Competitive Replication:

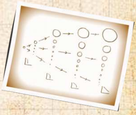
Assume distribution stabilizes: $N_{k,t} = n_k t$

(Reasonable for t large)

- Drop expectations
- Numbers of elephants now fractional
- Okay over large time scales

- For later: the fraction of groups that have size k is n_k/ρ since

$$\frac{N_{k,t}}{\rho t} = \frac{n_k \cancel{t}}{\rho \cancel{t}} = \frac{n_k}{\rho}.$$



Random Competitive Replication:

Stochastic difference equation:

$$\left\langle N_{k,t+1} - N_{k,t} \right\rangle = (1-\rho)\left((k-1)\frac{N_{k-1,t}}{t} - k\frac{N_{k,t}}{t} \right)$$

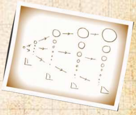
becomes

$$n_k(t+1) - n_k t = (1-\rho)\left((k-1)\frac{n_{k-1}t}{t} - k\frac{n_k t}{t} \right)$$

$$n_k(\cancel{t}+1-\cancel{t}) = (1-\rho)\left((k-1)\frac{n_{k-1}\cancel{t}}{\cancel{t}} - k\frac{n_k \cancel{t}}{\cancel{t}} \right)$$

$$\Rightarrow n_k = (1-\rho)\left((k-1)n_{k-1} - kn_k\right)$$

$$\Rightarrow n_k\left(1 + (1-\rho)k\right) = (1-\rho)(k-1)n_{k-1}$$



Random Competitive Replication:

We have a simple recursion:

$$\frac{n_k}{n_{k-1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k}$$

- Interested in k large (the tail of the distribution)
- Can be solved exactly.

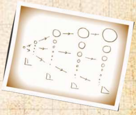
  Insert question from assignment 4
- For just the tail: Expand as a series of powers of $1/k$

  Insert question from assignment 4

  We (okay, you) find

$$n_k \propto k^{-\frac{(2-\rho)}{(1-\rho)}} = k^{-\gamma}$$

$$\boxed{\gamma = \frac{(2-\rho)}{(1-\rho)} = 1 + \frac{1}{(1-\rho)}}$$



- Micro-to-Macro story with ρ and γ measurable.

$$\boxed{\gamma = \frac{(2-\rho)}{(1-\rho)} = 1 + \frac{1}{(1-\rho)}}$$

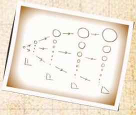
- Observe $2 < \gamma < \infty$ for $0 < \rho < 1$.
- For $\rho \simeq 0$ (low innovation rate):

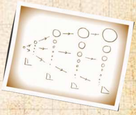
$$\gamma \simeq 2$$

- 'Wild' power-law size distribution of group sizes, bordering on 'infinite' mean.
- For $\rho \simeq 1$ (high innovation rate):

$$\gamma \simeq \infty$$

- All elephants have different flavors.
- Upshot: Tunable mechanism producing a family of universality classes.





- Recall Zipf's law: $s_r \sim r^{-\alpha}$
  (s_r = size of the rth largest group of elephants)
- We found $\alpha = 1/(\gamma - 1)$ so:

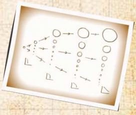
$$\boxed{\alpha = \frac{1}{\gamma - 1} = \frac{1}{\cancel{1} + \frac{1}{(1-\rho)} - \cancel{1}} = 1 - \rho.}$$

- $\gamma = 2$ corresponds to $\alpha = 1$
- We (roughly) see Zipfian exponent [33] of $\alpha = 1$ for many real systems: city sizes, word distributions, ...
- Corresponds to $\rho \rightarrow 0$, low innovation.
- Krugman doesn't like it [16] but it's all good.
- Still, other quite different mechanisms are possible...
- Must look at the details to see if mechanism makes sense... more later.

What about small k?:

We had one other equation:

- $$\left\langle N_{1,t+1} - N_{1,t} \right\rangle = \rho - (1-\rho) 1 \cdot \frac{N_{1,t}}{t}$$
- As before, set $N_{1,t} = n_1 t$ and drop expectations
- $$n_1 (t+1) - n_1 t = \rho - (1-\rho) 1 \cdot \frac{n_1 t}{t}$$
- $$n_1 = \rho - (1-\rho) n_1$$
- Rearrange:
$$n_1 + (1-\rho) n_1 = \rho$$
- $$\boxed{n_1 = \frac{\rho}{2-\rho}}$$



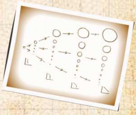
So... $$N_{1,t} = n_1 t = \frac{\rho t}{2-\rho}$$

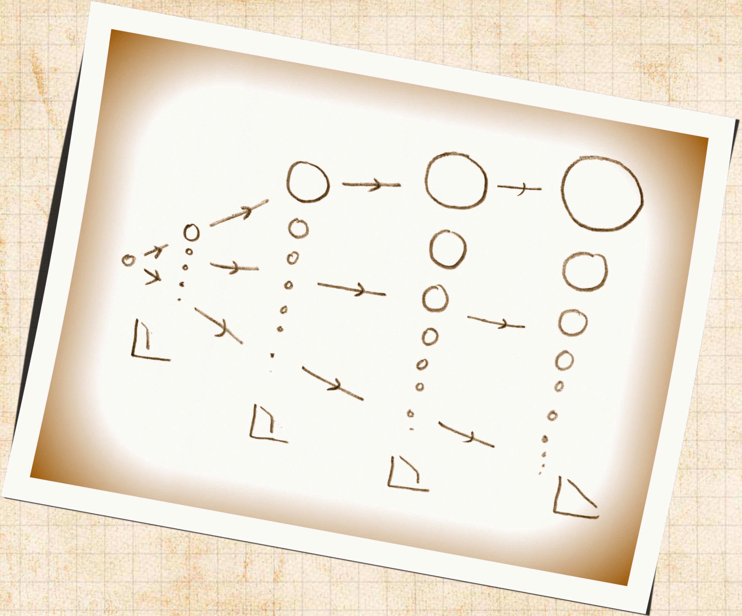
- Recall number of distinct elephants = ρt.
- Fraction of distinct elephants that are unique (belong to groups of size 1):

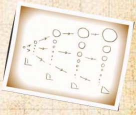
$$\frac{1}{\rho t} N_{1,t} = \frac{1}{\cancel{\rho t}} \frac{\cancel{\rho t}}{2-\rho} = \frac{1}{2-\rho}$$

(also = fraction of groups of size 1)

- For ρ small, fraction of unique elephants $\sim 1/2$
- Roughly observed for real distributions
- ρ increases, fraction increases
- Can show fraction of groups with two elephants $\sim 1/6$
- Model works well for large and small k #awesome







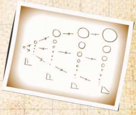
Words:

From Simon [27]:

Estimate $\rho_{\text{est}} = \#$ unique words$/\#$ all words

For Joyce's Ulysses: $\rho_{\text{est}} \simeq 0.115$

N_1 (real)	N_1 (est)	N_2 (real)	N_2 (est)
16,432	15,850	4,776	4,870



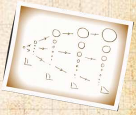
Evolution of catch phrases:

- Yule's paper (1924) [31]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."
- Simon's paper (1955) [27]:
  "On a class of skew distribution functions" (snore)

From Simon's introduction:

It is the purpose of this paper to analyse a class of distribution functions that appear in a wide range of empirical data—particularly data describing sociological, biological and economic phenomena.

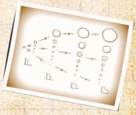
Its appearance is so frequent, and the phenomena so diverse, that one is led to conjecture that if these phenomena have any property in common it can only be a similarity in the structure of the underlying probability mechanisms.



Evolution of catch phrases:

Derek de Solla Price:

- First to study network evolution with these kinds of models.
- Citation network of scientific papers
- Price's term: Cumulative Advantage
- Idea: papers receive new citations with probability proportional to their existing # of citations
- Directed network
- Two (surmountable) problems:
  1. New papers have no citations
  2. Selection mechanism is more complicated

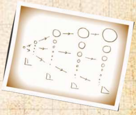


Evolution of catch phrases:

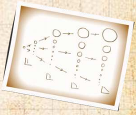
Robert K. Merton: the Matthew Effect

- Studied careers of scientists and found credit flowed disproportionately to the already famous

  From the Gospel of Matthew:
  "For to every one that hath shall be given...
  (Wait! There's more....)
  but from him that hath not, that also which he seemeth to have shall be taken away.
  And cast the worthless servant into the outer darkness; there men will weep and gnash their teeth."
- (Hath = suggested unit of purchasing power.)
- Matilda effect: women's scientific achievements are often overlooked



Evolution of catch phrases:



Merton was a catchphrase machine:

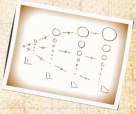
1. Self-fulfilling prophecy
2. Role model
3. Unintended (or unanticipated) consequences
4. Focused interview $\rightarrow$ focus group

And just to be clear...

Merton's son, Robert C. Merton, won the Nobel Prize for Economics in 1997.

Evolution of catch phrases:

- Barabasi and Albert [2]—thinking about the Web
- Independent reinvention of a version of Simon and Price's theory for networks
- Another term: "Preferential Attachment"
- Considered undirected networks (not realistic but avoids 0 citation problem)
- Still have selection problem based on size (non-random)
- Solution: Randomly connect to a node (easy) ...
- ...and then randomly connect to the node's friends (also easy)
- "Scale-free networks" = food on the table for physicists

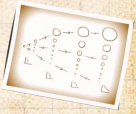


Another analytic approach: [10]

- Focus on how the nth arriving group typically grows.
- Analysis gives:

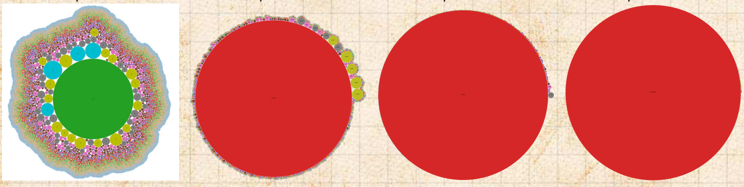
$$S_{n,t} \sim \begin{cases} \frac{1}{\Gamma(2-\rho)} \left[\frac{1}{t}\right]^{-(1-\rho)} & \text{for } n = 1, \\ \rho^{1-\rho} \left[\frac{n-1}{t}\right]^{-(1-\rho)} & \text{for } n \geq 2. \end{cases}$$

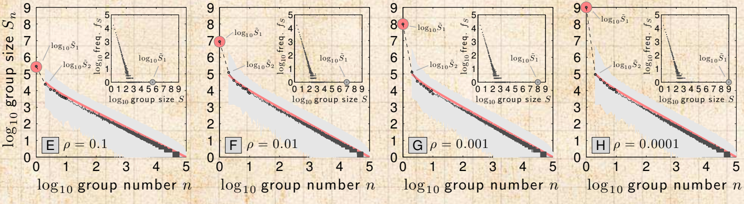
- First mover is a factor $1/\rho$ greater than expected.
- Because ρ is usually close to 0, the first element is truly an elephant in the room.
- Appears that this has been missed for 60 years ...



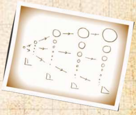
"Simon's fundamental rich-gets-richer model entails a dominant first-mover advantage"

Dodds et al.,
Available online at
http://arxiv.org/abs/0909.1104, 2016. [10]





See visualization at paper's online app-endices



Alternate analysis:

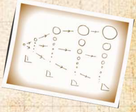
- Focus on the nth arriving group:

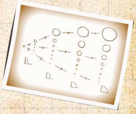
$$\left\langle S_{n,t+1} - S_{n,t} \right\rangle = (1-\rho_t) \cdot \frac{S_{n,t}}{t} \cdot (+1).$$

- For $t \geq t_n^{\textrm{init}}$, fix $\rho_t = \rho$ and shift t to $t-1$:

$$S_{n,t} = \left[1 + \frac{(1-\rho)}{t-1} \right] S_{n,t-1}.$$

where $S_{n,t_n^{\textrm{init}}} = 1$.





Betafication ensues:

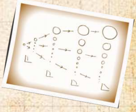
$$
\begin{aligned}
S_{n,t} &= \left[1 + \frac{(1-\rho)}{t-1}\right]\left[1 + \frac{(1-\rho)}{t-2}\right]\cdots\left[1 + \frac{(1-\rho)}{t_n^{\text{init}}}\right]\cdot 1 \\
&= \left[\frac{t+1-\rho}{t-1}\right]\left[\frac{t-\rho}{t-2}\right]\cdots\left[\frac{t_n^{\text{init}}+1-\rho}{t_n^{\text{init}}}\right] \\
&= \frac{\Gamma(t+1-\rho)\Gamma(t_n^{\text{init}})}{\Gamma(t_n^{\text{init}}+1-\rho)\Gamma(t)} \\
&= \frac{\mathrm{B}(t_n^{\text{init}}, 1-\rho)}{\mathrm{B}(t, 1-\rho)}.
\end{aligned}
$$

The first mover is really different:

- The issue is t_n^{init} in

$$S_{n,t} = \frac{\mathrm{B}(t_n^{\text{init}}, 1-\rho)}{\mathrm{B}(t, 1-\rho)}$$

- For $n \geq 2$ and $\rho \ll 1$, the nth group typically arrives at $t_n^{\text{init}} \simeq [\frac{n-1}{\rho}]$
- But $t_1^{\text{init}} = 1$ and the scaling is distinct in form.
- Simon missed the first mover by working on the size distribution.
- Contribution to $P_{k,t}$ of the first element vanishes as $t \to \infty$.
- Note: Does not apply to Barabási-Albert model.

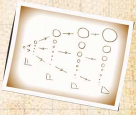


Variability:

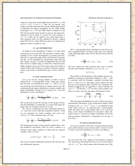
- The probability that the nth arriving group, if of size $S_{n,t} = k$ at time t, first replicates at time $t+\tau$:

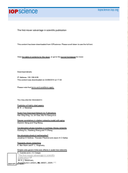
$$\mathbf{Pr}(S_{n,t+\tau} = k+1 \,|\, S_{n,t+i} = k) \text{ for } i = 0, \ldots, \tau - 1$$
$$= \prod_{i=0}^{\tau-1} \left[1 - (1-\rho)\frac{k}{t+i}\right] \cdot (1-\rho)\frac{k}{t+\tau}$$
$$= k\frac{B(\tau, t)}{B\left(\tau, t - (1-\rho)\right)}\frac{1-\rho}{t+\tau} \propto \frac{\tau^{-(1-\rho)k}}{t+\tau}.$$

- Upshot: nth arriving group starting at size 1 will on average wait for an infinite time to replicate.



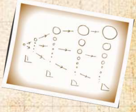
Related papers:



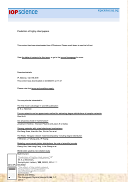

IOPscience

The first-mover advantage in scientific publication

"Organization of Growing Random Networks"
Krapivsky and Redner,
Phys. Rev. E, **63**, 066123, 2001.[15]



Related papers:


IOPscience iopscience.iop.org

Prediction of highly cited papers

This content has been downloaded from IOPscience. Please scroll down to see the full text.

View the table of contents for this issue, or go to the journal homepage for more

Download details:

Please note that terms and conditions apply.

You may also be interested in:

The first-mover advantage in scientific publication
M. E. J. Newman

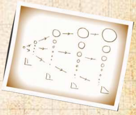
Are physicists afraid of mathematics?

Growing networks with mixed attachment mechanisms

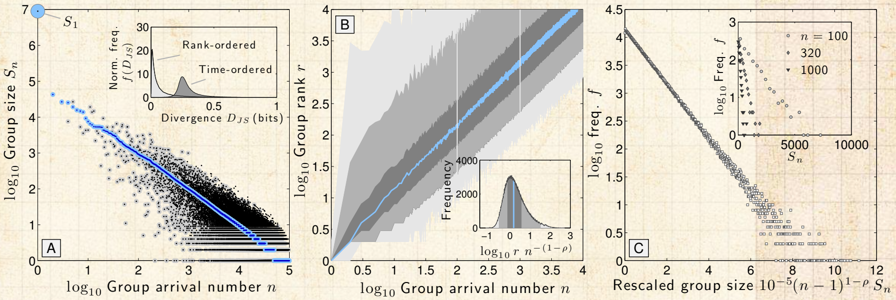
World order upset by new citation study

"Prediction of highly cited papers"
M. E. J. Newman,
Europhysics Letters, **105**, 28002, 2014.

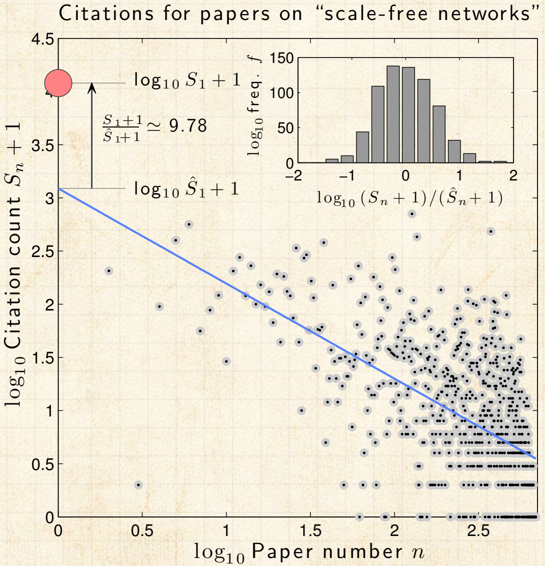
"The effect of the initial network configuration on preferential attachment"
Berset and Medo,
The European Physical Journal B, **86**, 1–7, 2013.



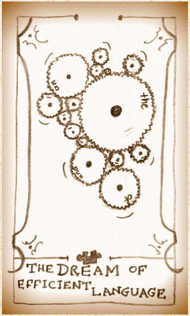
Arrival variability:



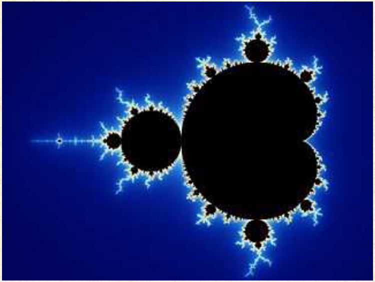
Self-referential citation data:




THE SUN
THE GOLDEN AGE OF REDUCTIONISM
THE MANIFESTO
THE WELL OF SCALES
THE LAWS OF DECAYING POWER
MINISTRY OF RANDOM WALKS
THE UNEXPECTED TRANSFIGURATION
THE RICH GET RICHER
THE LAW OF DIGITS FIRST
THE SURPRISE OF BEING ROBUST YET FRAGILE
THE MISREADING OF THE BOOK OF SAND
THE SUPPLY MOST FAIR
THE SUPPLY TO THE MOST
THE MANY TALES OF MANY TAILS
THE MEASURING
THE DATA POOR
THE POURING DATA
THE CHARMS OF THE MANY WRONG PATHS
THE KNOWLEDGE OF DEAD MONKS
THE DOORS THAT SHOULD NOT BE OPENED
THE EMERGENCE OF STRUCTURE
THE EMERGENCE OF DESTRUCTION
THE EMERGENCE OF THINKING
THE EMERGENCE OF STORIES
THE MECHANICS STATISTICAL
THE THRESHOLDS OF PERCOLATION
THE NETWORKS COMPLEX
THE ELEMENTS OF WEBS
THE ARK OF THE RANDOM NETWORK
THE BIRDS OF A FEATHER
THE BIRDS OF NOT A FEATHER
THE STRANGENESS OF FRIENDS
THE SMALLNESS OF WORLDS
THE THEORY OF SIX DEGREES
THE NETWORKS FREE OF SCALE
THE CONTAGION
THE CONFUSION OF CONTAGIONS
THE DISEASE OF RANDOM MIXING
THE UNENDING PANDEMICS
THE RATIO OF GAIN
THE THRESHOLDS OF THE MIND
THE SOCIAL WILD
THE SOCIAL CONTAGION
THE MECHANICS OF FAME
THE STORIES CONTAGIOUS
THE PATHS OF UNIVERSALITY
THE DEPENDENT PATHS
THE TROUSERS OF BIFURCATING REALITIES
THE UNKNOWN MECHANISM
THE JOHN DORY
THE STORYTELLERS
THE ADJACENT NARRATIVES
THE END


THE DREAM OF EFFICIENT LANGUAGE

Benoît Mandelbrot

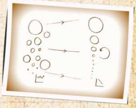


- Mandelbrot = father of fractals
- Mandelbrot = almond bread
- Bonus Mandelbrot set action: here.

Another approach:

Benoît Mandelbrot

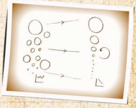
- Derived Zipf's law through optimization [19]
- Idea: Language is efficient
- Communicate as much information as possible for as little cost
- Need measures of information (H) and average cost (C)...
- Language evolves to maximize H/C, the amount of information per average cost.
- Equivalently: minimize C/H.
- Recurring theme: what role does optimization play in complex systems?



The Quickening—Mandelbrot v. Simon:

There Can Be Only One:

- Things there should be only one of: Theory, Highlander Films.
- Feel free to play Queen's It's a Kind of Magic in your head (funding remains tight).



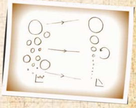
Now let us enjoy the Trailer for Highlander:

We were born to be Princes of the Universe

vs.

Mandelbrot vs. Simon:

- Mandelbrot (1953): "An Informational Theory of the Statistical Structure of Languages" [19]
- Simon (1955): "On a class of skew distribution functions" [27]
- Mandelbrot (1959): "A note on a class of skew distribution functions: analysis and critique of a paper by H.A. Simon" [20]
- Simon (1960): "Some further notes on a class of skew distribution functions" [28]

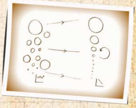


I have no rival, No man can be my equal

 vs. 

Mandelbrot vs. Simon:

- Mandelbrot (1961): "Final note on a class of skew distribution functions: analysis and critique of a model due to H.A. Simon" [21]
- Simon (1961): "Reply to 'final note' by Benoit Mandelbrot" [30]
- Mandelbrot (1961): "Post scriptum to 'final note'" [22]
- Simon (1961): "Reply to Dr. Mandelbrot's post scriptum" [29]



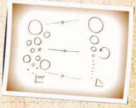
I am immortal, I have inside me blood of kings

Mandelbrot:

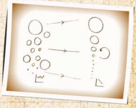
"We shall restate in detail our 1959 objections to Simon's 1955 model for the Pareto-Yule-Zipf distribution. Our objections are valid quite irrespectively of the sign of p-1, so that most of Simon's (1960) reply was irrelevant." [21]

Simon:

"Dr. Mandelbrot has proposed a new set of objections to my 1955 models of the Yule distribution. Like his earlier objections, these are invalid." [30]



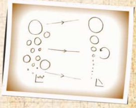
Two theories enter, one theory leaves



Zipfarama via Optimization:

Mandelbrot's Assumptions:

- Language contains n words: $w_1, w_2, \ldots, w_n$.
- ith word appears with probability p_i
- Words appear randomly according to this distribution (obviously not true...)
- Words = composition of letters is important
- Alphabet contains m letters
- Words are ordered by length (shortest first)



Zipfarama via Optimization:

Word Cost

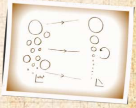
- Length of word (plus a space)
- Word length was irrelevant for Simon's method

Objection

- Real words don't use all letter sequences

Objections to Objection

- Maybe real words roughly follow this pattern (?)
- Words can be encoded this way
- Na na na-na naaaaa...

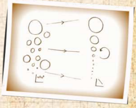


Zipfarama via Optimization:

Binary alphabet plus a space symbol

i	1	2	3	4	5	6	7	8
word	1	10	11	100	101	110	111	1000
length	1	2	2	3	3	3	3	4
$1+\log_2 i$	1	2	2.58	3	3.32	3.58	3.81	4

- Word length of 2^kth word: $= k+1 = 1+\log_2 2^k$
- Word length of ith word $\simeq 1+\log_2 i$
- For an alphabet with m letters, word length of ith word $\simeq 1+\log_m i$.



Zipfarama via Optimization:

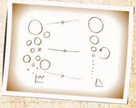
Total Cost C

- Cost of the ith word: $C_i \simeq 1 + \log_m i$
- Cost of the ith word plus space: $C_i \simeq 1 + \log_m (i+1)$
- Subtract fixed cost: $C_i' = C_i - 1 \simeq \log_m (i+1)$
- Simplify base of logarithm:

$$C_i' \simeq \log_m (i+1) = \frac{\log_e (i+1)}{\log_e m} \propto \log_e (i+1)$$

- Total Cost:

$$C \sim \sum_{i=1}^{n} p_i C_i' \propto \sum_{i=1}^{n} p_i \log_e (i+1)$$



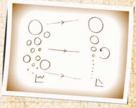
Zipfarama via Optimization:

Information Measure

- Use Shannon's Entropy (or Uncertainty):

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i$$

- (allegedly) von Neumann suggested 'entropy'...
- Proportional to average number of bits needed to encode each 'word' based on frequency of occurrence
- $-\log_2 p_i = \log_2 1/p_i$ = minimum number of bits needed to distinguish event i from all others
- If $p_i = 1/2$, need only 1 bit ($\log_2 1/p_i = 1$)
- If $p_i = 1/64$, need 6 bits ($\log_2 1/p_i = 6$)



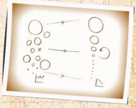
Zipfarama via Optimization:

Information Measure

- Use a slightly simpler form:

$$H = -\sum_{i=1}^{n} p_i \log_e p_i / \log_e 2 = -g \sum_{i=1}^{n} p_i \log_e p_i$$

where $g = 1/\log_e 2$



Zipfarama via Optimization:

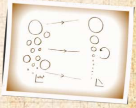
- Minimize

$$F(p_1, p_2, \ldots, p_n) = C/H$$

subject to constraint

$$\sum_{i=1}^{n} p_i = 1$$

- Tension:
(1) Shorter words are cheaper
(2) Longer words are more informative (rarer)



Zipfarama via Optimization:

Time for Lagrange Multipliers:

- Minimize

$$\Psi(p_1, p_2, \ldots, p_n) =$$

$$F(p_1, p_2, \ldots, p_n) + \lambda G(p_1, p_2, \ldots, p_n)$$

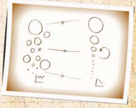
where

$$F(p_1, p_2, \ldots, p_n) = \frac{C}{H} = \frac{\sum_{i=1}^{n} p_i \log_e(i+1)}{-g \sum_{i=1}^{n} p_i \log_e p_i}$$

and the constraint function is

$$G(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} p_i - 1 (= 0)$$

Insert question from assignment 5

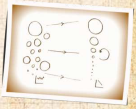


Zipfarama via Optimization:

Some mild suffering leads to:

- $$p_j = e^{-1-\lambda H^2/gC}(j+1)^{-H/gC} \propto (j+1)^{-H/gC}$$
- A power law appears [applause]: $\alpha = H/gC$
- **Next:** sneakily deduce λ in terms of g, C, and H.
- Find

$$p_j = (j+1)^{-H/gC}$$



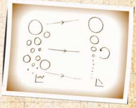
Zipfarama via Optimization:

Finding the exponent

- Now use the normalization constraint:

$$1 = \sum_{j=1}^{n} p_j = \sum_{j=1}^{n} (j+1)^{-H/gC} = \sum_{j=1}^{n} (j+1)^{-\alpha}$$

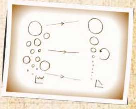
- As $n \to \infty$, we end up with $\zeta(H/gC) = 2$ where ζ is the Riemann Zeta Function
- Gives $\alpha \simeq 1.73$ (> 1, too high) or $\gamma = 1 + \frac{1}{\alpha} \simeq 1.58$ (very wild)
- If cost function changes ($j + 1 \to j + a$) then exponent is tunable
- Increase a, decrease α

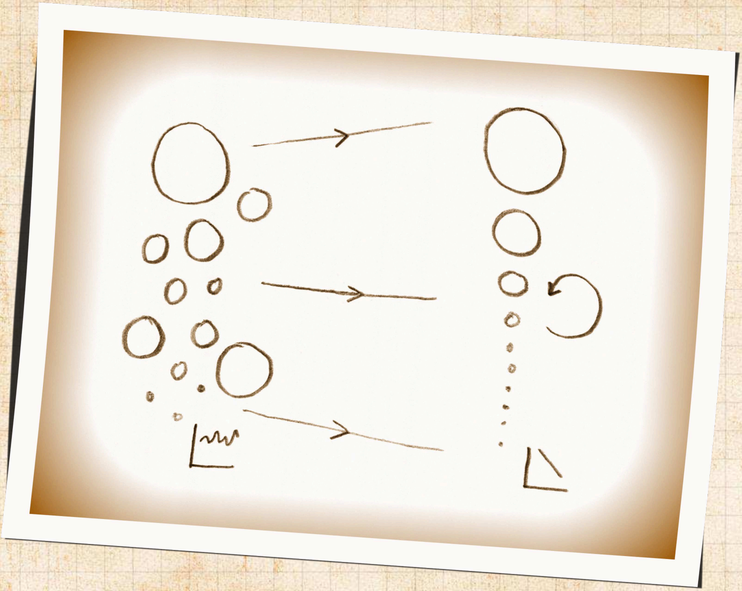


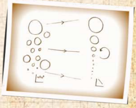
Zipfarama via Optimization:

All told:

- Reasonable approach: Optimization is at work in evolutionary processes
- But optimization can involve many incommensurate elephants: monetary cost, robustness, happiness,...
- Mandelbrot's argument is not super convincing
- Exponent depends too much on a loose definition of cost

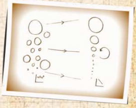






From the discussion at the end of Mandelbrot's paper:

- A. S. C. Ross: "M. Mandelbrot states that 'the actual direction of evolution (sc. of language) is, in fact, towards fuller and fuller utilization of places'. We are, in fact, completely without evidence as to the existence of any 'direction of evolution' in language, and it is axiomatic that we shall remain so. Many philologists would deny that a 'direction of evolution' could be theoretically possible; thus I myself take the view that a language develops in what is essentially a purely random manner."
- Mandelbrot: "As to the 'fundamental linguistic units being the least possible differences between pairs of utterances' this is a logical consequence of the fact that two is the least integer greater than one."

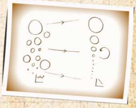


More:

Reconciling Mandelbrot and Simon

- Mixture of local optimization and randomness
- Numerous efforts...

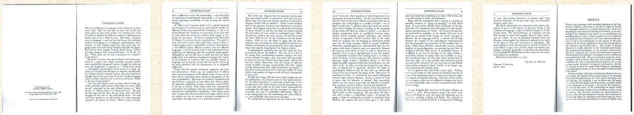
1. Carlson and Doyle, 1999:
   Highly Optimized Tolerance
   (HOT)—Evolved/Engineered Robustness [6, 7]
2. Ferrer i Cancho and Solé, 2002:
   Zipf's Principle of Least Effort [14]
3. D'Souza et al., 2007:
   Scale-free networks [11]



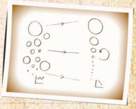
More

Other mechanisms:

- Much argument about whether or not monkeys typing could produce Zipf's law... (Miller, 1957) [23]
- Miller gets to slap Zipf rather rudely in an introduction to a 1965 reprint of Zipf's "Psycho-biology of Language" [24, 32]
- Let us now slap Miller around by simply reading his words out (see next slides):



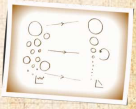
- Side note: Miller mentions "Genes of Language."
- Still fighting: "Random Texts Do Not Exhibit the Real Zipf's Law-Like Rank Distribution" [13] by Ferrer-i-Cancho and Elvevåg, 2010.



INTRODUCTION

The Psycho-Biology of Language is not calculated to please every taste. Zipf was the kind of man who would take roses apart to count their petals; if it violates your sense of values to tabulate the different words in a Shakespearean sonnet, this is not a book for you. Zipf took a scientist's view of language — and for him that meant the statistical analysis of language as a biological, psychological, social process. If such analysis repels you, then leave your language alone and avoid George Kingsley Zipf like the plague. You will be much happier reading Mark Twain: "There are liars, damned liars, and statisticians." Or W. H. Auden: "Thou shalt not sit with statisticians nor commit a social science."

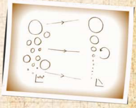
However, for those who do not flinch to see beauty murdered in a good cause, Zipf's scientific exertions yielded some wonderfully unexpected results to boggle the mind and tease the imagination. Language *is* — among other things — a biological, psychological, social process; to apply statistics to it merely acknowledges its essential unpredictability, without which it would be useless. But who would have thought that in the very heart of all the freedom language allows us Zipf would find an invariant as solid and reliable as the law of gravitation?



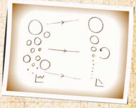
Put it this way. Suppose that we acquired a dozen monkeys and chained them to typewriters until they had produced some very long and random sequence of characters. Suppose further that we defined a "word" in this monkey-text as any sequence of letters occurring between successive spaces. And suppose finally that we counted the occurrences of these "words" in just the way Zipf and others counted the occurrences of real words in meaningful texts. When we plot our results in the same manner, we will find exactly the same "Zipf curves" for the monkeys as for the human authors. Since we are not likely to argue that the poor monkeys were searching for some equilibrium between uniformity and diversity in expressing their ideas, such explanations seem equally inappropriate for human authors.

A mathematical rationalization for this result has been provided by Benoit Mandelbrot. The crux of it is that if we assume that word-boundary markers (spaces) are scattered randomly through a text, then there will necessarily be more occurrences of short than long words. Add to this fact the further observation that the variety of different words available increases exponentially with their length and the phenomenon Zipf reported becomes inescapable: a few short words will be used an enormous number of times while a vast number of longer words will occur infrequently or not at all.

So Zipf was wrong. His facts were right enough, but not his explanations. In a broader sense he was right, however, for he called attention to a stochastic process that is frequently seen in the social sciences, and by accumulating statistical data that cried out for some better explanation he challenged his colleagues and his successors to explore an important new type of probability distribution. Zipf belongs among those rare but stimulating men whose failures are more profitable than most men's successes.



So who's right?



Bornholdt and Ebel (PRE), 2001:
"World Wide Web scaling exponent from Simon's 1955 model" [5].

- Show Simon's model fares well.
- Recall ρ = probability new flavor appears.
- Alta Vista crawls in approximately 6 month period in 1999 give $\rho \simeq 0.10$
- Leads to $\gamma = 1 + \frac{1}{1-\rho} \simeq 2.1$ for in-link distribution.
- Cite direct measurement of γ at the time: 2.1 ± 0.1 and 2.09 in two studies.

So who's right?

Recent evidence for Zipf's law...

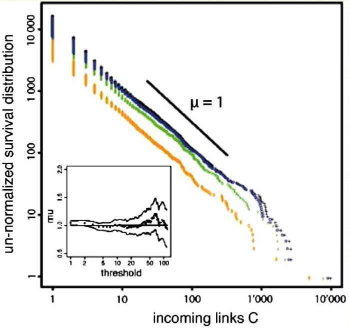
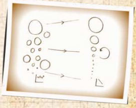


FIG. 1 (color online). (Color Online) Log-log plot of the number of packages in four Debian Linux Distributions with more than C in-directed links. The four Debian Linux Distributions are Woody (19.07.2002) (orange diamonds), Sarge (06.06.2005) (green crosses), Etch (15.08.2007) (blue circles), Lenny (15.12.2007) (black+'s). The inset shows the maximum likelihood estimate (MLE) of the exponent μ together with two boundaries defining its 95% confidence interval (approximately given by $1 \pm 2/\sqrt{n}$, where n is the number of data points using in the MLE), as a function of the lower threshold. The MLE has been modified from the standard Hill estimator to take into account the discreteness of C.

Maillart et al., PRL, 2008:
"Empirical Tests of Zipf's Law Mechanism in Open Source Linux Distribution" [18]



So who's right?

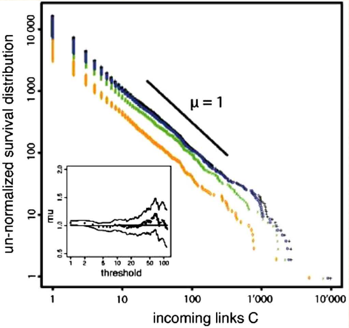


FIG. 1 (color online). (Color Online) Log-log plot of the number of packages in four Debian Linux Distributions with more than C in-directed links. The four Debian Linux Distributions are Woody (19.07.2002) (orange diamonds), Sarge (06.06.2005) (green crosses), Etch (15.08.2007) (blue circles), Lenny (15.12.2007) (black+'s). The inset shows the maximum likelihood estimate (MLE) of the exponent μ together with two boundaries defining its 95% confidence interval (approximately given by $1 \pm 2/\sqrt{n}$, where n is the number of data points using in the MLE), as a function of the lower threshold. The MLE has been modified from the standard Hill estimator to take into account the discreteness of C.

Maillart et al., PRL, 2008:
"Empirical Tests of Zipf's Law Mechanism in Open Source Linux Distribution" [18]

So who's right?

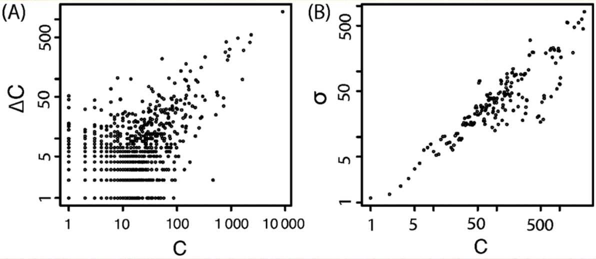


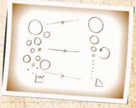
FIG. 2. Left panel: Plots of ΔC versus C from the Etch release (15.08.2007) to the latest Lenny version (05.05.2008) in double logarithmic scale. Only positive values are displayed. The linear regression $\Delta C = R \times C + C_0$ is significant at the 95% confidence level, with a small value $C_0 = 0.3$ at the origin and $R = 0.09$. Right panel: same as left panel for the standard deviation of ΔC.

- Rough, approximately linear relationship between C number of in-links and ΔC.

So who's right?

Nutshell:

- Simonish random 'rich-get-richer' models agree in detail with empirical observations.
- Power-lawfulness: Mandelbrot's optimality is still apparent.
- Optimality arises for free in Random Competitive Replication models.



Others are also not happy:

Krugman and Simon

- "The Self-Organizing Economy" (Paul Krugman, 1996) [16]
- Krugman touts Zipf's law for cities, Simon's model
- "Déjà vu, Mr. Krugman" (Berry, 1999)
- Substantial work done by Urban Geographers

Who needs a hug?

From Berry [3]

- Déjà vu, Mr. Krugman. Been there, done that. The Simon-Ijiri model was introduced to geographers in 1958 as an explanation of city size distributions, the first of many such contributions dealing with the steady states of random growth processes, ...
- But then, I suppose, even if Krugman had known about these studies, they would have been discounted because they were not written by professional economists or published in one of the top five journals in economics!

Who needs a hug?

From Berry [3]

- ... [Krugman] needs to exercise some humility, for his world view is circumscribed by folkways that militate against recognition and acknowledgment of scholarship beyond his disciplinary frontier.
- Urban geographers, thank heavens, are not so afflicted.

References I

[1] F. Auerbach.
Das gesetz der bevölkerungskonzentration.
Petermanns Geogr. Mitteilungen, 59:73–76, 1913.

[2] A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286:509–511, 1999. pdf

[3] B. J. L. Berry.
Déjà vu, Mr. Krugman.
Urban Geography, 20:1–2, 1999. pdf

[4] Y. Berset and M. Medo.
The effect of the initial network configuration on preferential attachment.
The European Physical Journal B, 86(6):1–7, 2013. pdf

References II

[5] S. Bornholdt and H. Ebel.
World Wide Web scaling exponent from Simon's 1955 model.
Phys. Rev. E, 64:035104(R), 2001. pdf

[6] J. M. Carlson and J. Doyle.
Highly optimized tolerance: A mechanism for power laws in designed systems.
Phys. Rev. E, 60(2):1412–1427, 1999. pdf

[7] J. M. Carlson and J. Doyle.
Complexity and robustness.
Proc. Natl. Acad. Sci., 99:2538–2545, 2002. pdf

[8] D. J. de Solla Price.
Networks of scientific papers.
Science, 149:510–515, 1965. pdf

References III

[9] D. J. de Solla Price.
A general theory of bibliometric and other cumulative advantage processes.
J. Amer. Soc. Inform. Sci., 27:292–306, 1976. pdf

[10] P. S. Dodds, D. R. Dewhurst, F. F. Hazlehurst, C. M. Van Oort, L. Mitchell, A. J. Reagan, J. R. Williams, and C. M. Danforth.
Simon's fundamental rich-gets-richer model entails a dominant first-mover advantage, 2016.
Available online at http://arxiv.org/abs/0909.1104. pdf

[11] R. M. D'Souza, C. Borgs, J. T. Chayes, N. Berger, and R. D. Kleinberg.
Emergence of tempered preferential attachment from optimization.
Proc. Natl. Acad. Sci., 104:6112–6117, 2007. pdf

References IV

[12] J.-B. Estoup.
Gammes sténographiques: méthode et exercices pour l'acquisition de la vitesse.
Institut Sténographique, 1916.

[13] R. Ferrer-i-Cancho and B. Elvevåg.
Random texts do not exhibit the real Zipf's law-like rank distribution.
PLoS ONE, 5:e9411, 03 2010.

[14] R. Ferrer-i-Cancho and R. V. Solé.
Zipf's law and random texts.
Advances in Complex Systems, 5(1):1–6, 2002.

[15] P. L. Krapivsky and S. Redner.
Organization of growing random networks.
Phys. Rev. E, 63:066123, 2001. pdf

References V

[16] P. Krugman.
The Self-Organizing Economy.
Blackwell Publishers, Cambridge, Massachusetts, 1996.

[17] A. J. Lotka.
The frequency distribution of scientific productivity.
Journal of the Washington Academy of Science, 16:317–323, 1926.

[18] T. Maillart, D. Sornette, S. Spaeth, and G. von Krogh.
Empirical tests of Zipf's law mechanism in open source Linux distribution.
Phys. Rev. Lett., 101(21):218701, 2008. pdf

References VI

[19] B. B. Mandelbrot.
An informational theory of the statistical structure of languages.
In W. Jackson, editor, Communication Theory, pages 486–502. Butterworth, Woburn, MA, 1953.
pdf

[20] B. B. Mandelbrot.
A note on a class of skew distribution function. Analysis and critique of a paper by H. A. Simon.
Information and Control, 2:90–99, 1959.

[21] B. B. Mandelbrot.
Final note on a class of skew distribution functions: analysis and critique of a model due to H. A. Simon.
Information and Control, 4:198–216, 1961.

References VII

[22] B. B. Mandelbrot.
Post scriptum to 'final note'.
Information and Control, 4:300–304, 1961.

[23] G. A. Miller.
Some effects of intermittent silence.
American Journal of Psychology, 70:311–314, 1957. pdf

[24] G. A. Miller.
Introduction to reprint of G. K. Zipf's "The Psycho-Biology of Language." MIT Press, Cambridge MA, 1965. pdf

[25] M. E. J. Newman.
The first-mover advantage in scientific publication.
Europhysics Letters, 86:68001, 2009. pdf

References VIII

[26] M. E. J. Newman.
Prediction of highly cited papers.
Europhysics Letters, 105:28002, 2014. pdf

[27] H. A. Simon.
On a class of skew distribution functions.
Biometrika, 42:425–440, 1955. pdf

[28] H. A. Simon.
Some further notes on a class of skew distribution functions.
Information and Control, 3:80–88, 1960.

[29] H. A. Simon.
Reply to Dr. Mandelbrot's post scriptum.
Information and Control, 4:305–308, 1961.

References IX

[30] H. A. Simon.
Reply to 'final note' by Benoît Mandelbrot.
Information and Control, 4:217–223, 1961.

[31] G. U. Yule.
A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S.
Phil. Trans. B, 213:21–87, 1925. pdf

[32] G. K. Zipf.
The Psycho-Biology of Language.
Houghton-Mifflin, New York, NY, 1935.

[33] G. K. Zipf.
Human Behaviour and the Principle of Least-Effort.
Addison-Wesley, Cambridge, MA, 1949.