Mechanisms for Generating Power-Law Size Distributions, Part 3

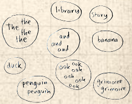
Last updated: 2025/10/08, 21:11:29 EDT

Principles of Complex Systems, Vols. 1, 2, 3D, 4 Fourever, V for Vendetta

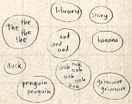
Prof. Peter Sheridan Dodds

Computational Story Lab | Vermont Complex Systems Institute
University of Vermont | Santa Fe Institute

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



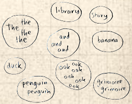
These slides are brought to you by:



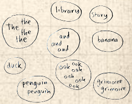
These slides are also brought to you by:

Special Guest Executive Producer

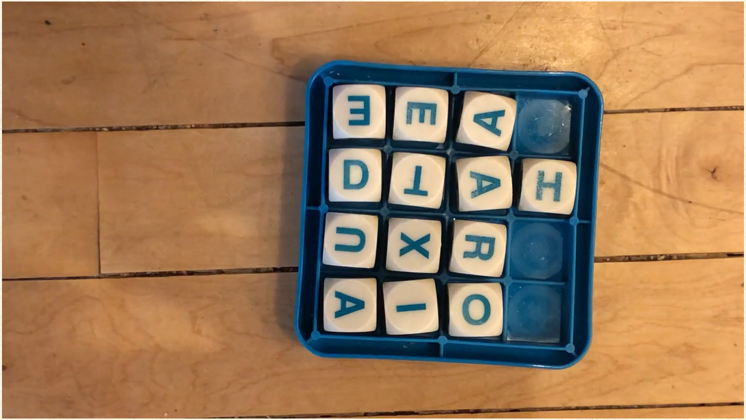
On Instagram at pratchett_the_cat

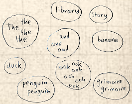


Outline

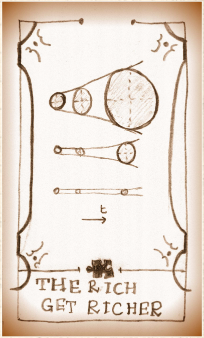

The the the the
library
story
and and and
banana
duck
ook ook ook ook ook
penguin penguin
grimoire grimoire

The Boggoracle Speaks:

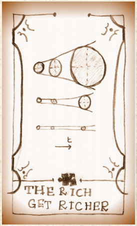





THE SUN
THE GOLDEN AGE OF REDUCTIONISM
THE MANIFESTO
WELL OF SCALES
THE LAWS OF DECAYING POWER
THE UNEXPECTED TRANSFIGURATION
MINISTRY OF RANDOM WALKS
THE DREAM OF EFFICIENT LANGUAGE
THE LAW OF DIGITS FIRST
THE SURPRISE OF BEING ROBUST YET FRAGILE
THE MISREADING OF THE BOOK OF SAND
THE MANY TAILS OF MANY TAILS
THE MEASURING
THE DATA POOR
THE POURING DATA
THE CHARMS OF THE MANY WRONG PATHS
THE BOOKS THAT SHOULD NOT BE OPENED
THE KNOWLEDGE OF DEAD MONKS
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 BIRDS OF A FEATHER
THE BIRDS OF NOT A FEATHER
THE ARK OF THE RANDOM NETWORK
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 THRESHOLDS OF THE MIND
THE SOCIAL WILD
THE SOCIAL CONTAGION
THE RATIO OF GAIN
THE MECHANICS OF FAME
THE TALK OF FAME
THE STORIES CONTAGIOUS
THE JOHN DORY
THE ADJACENT NARRATIVES
THE STORYTELLERS
THE FATED ASTRAL PATH
THE PATHS OF UNIVERSALITY
THE DEPENDENT PATHS
THE TROUSERS OF BIFURCATING REALITIES
THE UNKNOWN MECHANISM
THE LANDSCAPES OF FORKING PATHS
THE NETWORKS OF BLOOD
THE TREES OF REALITY
THE ORDERS OF STREAMS
THE LAWS OF BRANCHING
THE SCALING OF BROKEN HIERARCHIES
THE BRAID OR THE BRANCH
THE LAW OF OPTIMAL FORKS
THE CHURCH OF QUARTEROLOGY
THE TRUTHICIDE
THE FAMILY OF PANCAKE COWS
THE FAMILY OF SPHERICAL COWS
THE IMPROBABLE PACHYDERM
THE ILL-SCALING OF ALLOMETRIC BODIES
THE SUPPLY MOST FAIR
THE SUPPLY TO THE MOST
THE DARK ART OF GENERATING FUNCTIONS
THE RANDOM SUM OF RANDOMNESS
THE TWO-SIDED WORLDS
THE STORIES AND THE TROPES
THE TRUTH OF DIFFUSION
THE UNVEILING OF HIDDEN POWERS
THE SCRYING FOR THE SHAPES OF THINGS
THE LIMITS OF ORGANIZATIONS

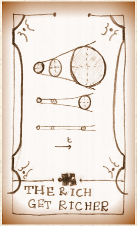

t
THE RICH GET RICHER

Outline



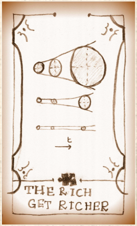
Aggregation:

- Random walks represent additive aggregation



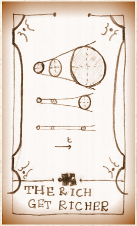
Aggregation:

- Random walks represent additive aggregation
- Mechanism: Random addition and subtraction



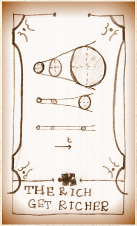
Aggregation:

- Random walks represent additive aggregation
- Mechanism: Random addition and subtraction
- Compare across realizations, no competition.



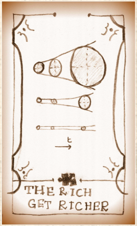
Aggregation:

- Random walks represent additive aggregation
- Mechanism: Random addition and subtraction
- Compare across realizations, no competition.
- Next: Random Additive/Copying Processes involving Competition.



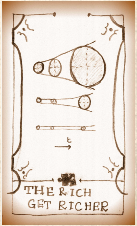
Aggregation:

- 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, ...)

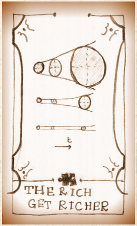


Aggregation:

- 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)

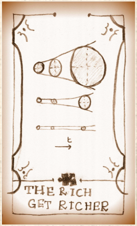


Outline



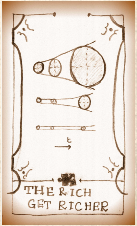
Pre-Zipf's law observations of Zipf's law

- 1910s: Word frequency examined re Stenography (or shorthand or brachygraphy or tachygraphy), Jean-Baptiste Estoup [5].



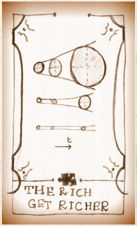
Pre-Zipf's law observations of Zipf's law

- 1910s: Word frequency examined re Stenography (or shorthand or brachygraphy or tachygraphy), Jean-Baptiste Estoup [5].
- 1910s: Felix Auerbach pointed out the Zipfitude of city sizes in
  "Das Gesetz der Bevölkerungskonzentration"
  ("The Law of Population Concentration") [1].



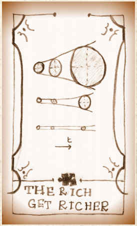
Pre-Zipf's law observations of Zipf's law

- 1910s: Word frequency examined re Stenography (or shorthand or brachygraphy or tachygraphy), Jean-Baptiste Estoup [5].
- 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 [11]:
  # Species per Genus (offers first theoretical mechanism)



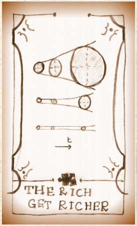
Pre-Zipf's law observations of Zipf's law

- 1910s: Word frequency examined re Stenography (or shorthand or brachygraphy or tachygraphy), Jean-Baptiste Estoup [5].
- 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 [11]:
  # Species per Genus (offers first theoretical mechanism)
- 1926: Lotka [7]:
  # Scientific papers per author (Lotka's law)



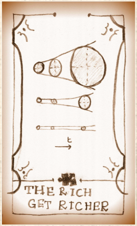
Theoretical Work of Yore:

- 1949: Zipf's "Human Behaviour and the Principle of Least-Effort" is published. [12]



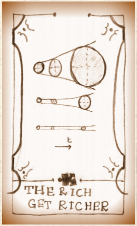
Theoretical Work of Yore:

- 1949: Zipf's "Human Behaviour and the Principle of Least-Effort" is published. [12]
- 1953: Mandelbrot [8]:
  Optimality argument for Zipf's law; focus on language.



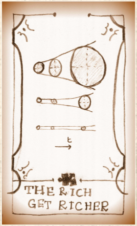
Theoretical Work of Yore:

- 1949: Zipf's "Human Behaviour and the Principle of Least-Effort" is published. [12]
- 1953: Mandelbrot [8]:
  Optimality argument for Zipf's law; focus on language.
- 1955: Herbert Simon [10, 12]:
  Zipf's law for word frequency, city size, income, publications, and species per genus.



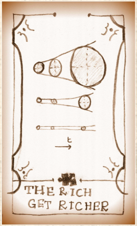
Theoretical Work of Yore:

- 1949: Zipf's "Human Behaviour and the Principle of Least-Effort" is published. [12]
- 1953: Mandelbrot [8]:
  Optimality argument for Zipf's law; focus on language.
- 1955: Herbert Simon [10, 12]:
  Zipf's law for word frequency, city size, income, publications, and species per genus.
- 1965/1976: Derek de Solla Price [3, 9]:
  Network of Scientific Citations.



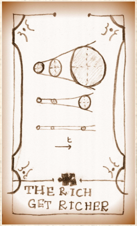
Theoretical Work of Yore:

- 1949: Zipf's "Human Behaviour and the Principle of Least-Effort" is published. [12]
- 1953: Mandelbrot [8]:
  Optimality argument for Zipf's law; focus on language.
- 1955: Herbert Simon [10, 12]:
  Zipf's law for word frequency, city size, income, publications, and species per genus.
- 1965/1976: Derek de Solla Price [3, 9]:
  Network of Scientific Citations.
- 1999: Barabasi and Albert [2]:
  The World Wide Web, networks-at-large.



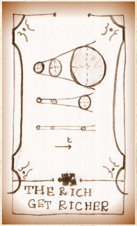
Herbert Simon (1916–2001):

- Political scientist (and much more)



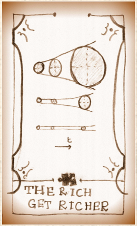
Herbert Simon (1916–2001):

- Political scientist (and much more)
- Involved in Cognitive Psychology, Computer Science, Public Administration, Economics, Management, Sociology



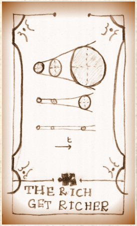
Herbert Simon (1916–2001):

- Political scientist (and much more)
- Involved in Cognitive Psychology, Computer Science, Public Administration, Economics, Management, Sociology
- Coined 'bounded rationality' and 'satisficing'



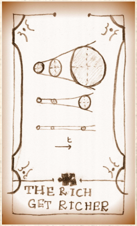
Herbert Simon (1916–2001):

- 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)



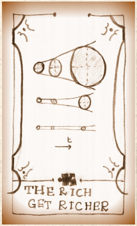
Herbert Simon (1916–2001):

- 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.

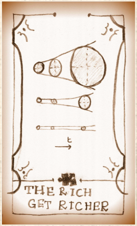


Herbert Simon (1916–2001):

- 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).



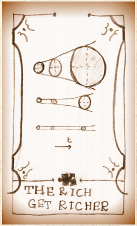
Outline

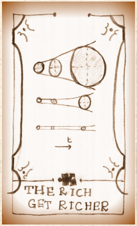


Essential Extract of a Growth Model:

Random Competitive Replication (RCR):

1. Start with 1 Moo Deng (or element) of a particular flavor at $t = 1$





Essential Extract of a Growth Model:

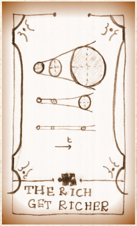
Random Competitive Replication (RCR):

1. Start with 1 Moo Deng (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

Essential Extract of a Growth Model:

Random Competitive Replication (RCR):

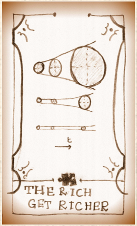
1. Start with 1 Moo Deng (or element) of a particular flavor at $t = 1$
2. At time $t = 2, 3, 4, \ldots$, add a new elephant in one of two ways:
   - With probability ρ, create a new elephant with a new flavor
   - With probability $1 - \rho$, randomly choose from all existing elephants, and make a copy.



Essential Extract of a Growth Model:

Random Competitive Replication (RCR):

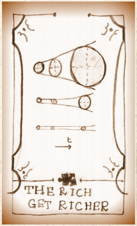
1. Start with 1 Moo Deng (or element) of a particular flavor at $t = 1$
2. At time $t = 2, 3, 4, \ldots$, add a new elephant in one of two ways:
   - With probability ρ, create a new elephant with a new flavor
   - With probability $1 - \rho$, randomly choose from all existing elephants, and make a copy.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



Essential Extract of a Growth Model:

Random Competitive Replication (RCR):

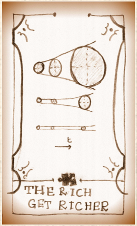
1. Start with 1 Moo Deng (or element) of a particular flavor at $t = 1$
2. At time $t = 2, 3, 4, \ldots$, add a new elephant in one of two ways:
   - With probability ρ, create a new elephant with a new flavor = Mutation/Innovation
   - With probability $1 - \rho$, randomly choose from all existing elephants,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



Essential Extract of a Growth Model:

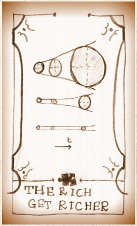
Random Competitive Replication (RCR):

1. Start with 1 Moo Deng (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—Simon's rich-get-richer model: [10]

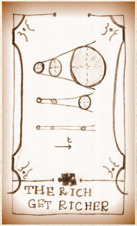
Example: Words appearing in a language



Random Competitive Replication—Simon's rich-get-richer model: [10]

Example: Words appearing in a language

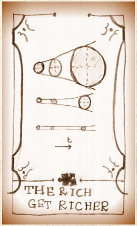
- Consider words as they appear sequentially.



Random Competitive Replication—Simon's rich-get-richer model: [10]

Example: Words appearing in a language

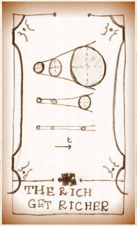
- Consider words as they appear sequentially.
- With probability ρ, the next word has not previously appeared



Random Competitive Replication—Simon's rich-get-richer model: [10]

Example: Words appearing in a language

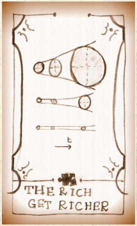
- Consider words as they appear sequentially.
- With probability ρ, the next word has not previously appeared
- With probability $1 - \rho$, randomly choose one word from all words that have come before, and reuse this word



Random Competitive Replication—Simon's rich-get-richer model: [10]

Example: Words appearing in a language

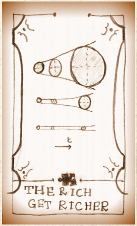
- Consider words as they appear sequentially.
- With probability ρ, the next word has not previously appeared
  = Mutation/Innovation
- With probability $1 - \rho$, randomly choose one word from all words that have come before, and reuse this word



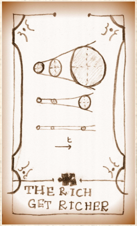
Random Competitive Replication—Simon's rich-get-richer model: [10]

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



Random Competitive Replication—Simon's rich-get-richer model: [10]



Example: Words appearing in a language

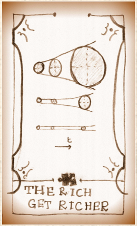
- Consider words as they appear sequentially.
- With probability ρ, the next word has not previously appeared
  = Mutation/Innovation
- With probability $1 - \rho$, randomly choose one word from all words that have come before, and reuse this word
  = Replication/Imitation

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

For example:

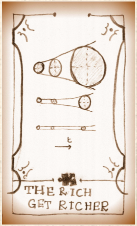
- 21 words used
- next word is new with prob ρ
- next word is a copy with prob $1-\rho$

prob:	next word:
6/21	ook
4/21	the
3/21	and
2/21	penguin
⋮	
1/21	library



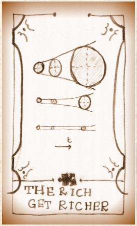
Some observations:

- Fundamental Rich-get-Richer story;



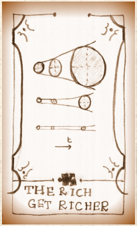
Some observations:

- Fundamental Rich-get-Richer story;
- Competition for replication between individual elephants is random;



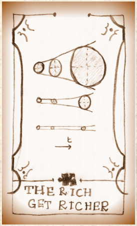
Some observations:

- Fundamental Rich-get-Richer story;
- Competition for replication between individual elephants is random;
- Competition for growth between groups of matching elephants is not random;



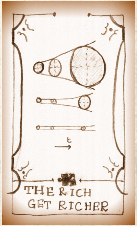
Some observations:

- 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;



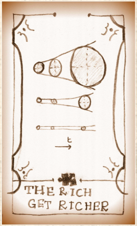
Some observations:

- 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;



Some observations:

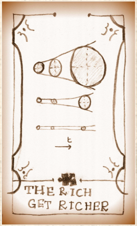
- Fundamental Rich-get-Richer story;
- Competition for replication between individual elephants is random;
- Competition for growth between groups of matching elephants is not random;
- Selection on groups is biased by size;
- Random selection sounds easy;
- Possible that no great knowledge of system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



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:

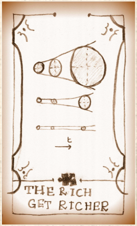


Some observations:

- 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.

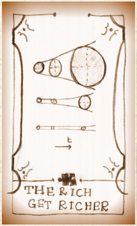


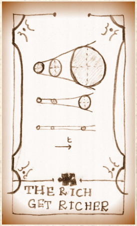
Some observations:

- 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.





Language Log Home About Comments policy

The long tail of religious studies?

August 5, 2010 @ 10:33 am · Filed by Mark Liberman under Computational linguistics

« previous post | next post »

Google Books isn't the only outfit that sometimes has trouble with metadata. I happened to notice this morning that Oxford University Press has classified Herbert A. Simon's "On a class of skew distribution functions" (*Biometrika* 43:425-440, 1955) as "Religious Studies..Death":

BIOMETRIKA

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS

Oxford Journals > Mathematics & Physical Sciences > Biometrika > Volume 42, Numb

Biometrika 1955 42(3-4):425-440; doi:10.1093/biomet/42.3-4.425
© 1955 by Biometrika Trust

ON A CLASS OF SKEW DISTRIBUTION FUNCTIONS

HERBERT A. SIMON

You have reached the most complete version of this article accessible without further authentication.
More complete versions are available.
Link to article

Article topics:

- Religious Studies..Death

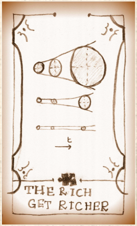
Carl Burke said,
August 5, 2010 @ 12:14 pm

If I had to guess at features that suggest 'Religious Studies — Death', I'd have to go with the word 'urn' and the suffix 'xion', almost never seen except on 'crucifixion'. Granted that Biometrika is published by Oxford University Press, and 'connexion' is a perfectly good British word, the classification algorithm might be more familiar with the American form 'connection'.

[(myl) Looking a bit further into the paper, one finds things like

> it is well known that the negative binomial and the log series distributions can be obtained as the stationary solutions of certain stochastic processes. For example, J.H. Darwin (1953) derives these from birth and death processes, with appropriate assumptions as to the birth- and death-rates and the initial conditions.

]

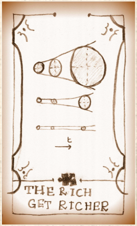


Language Log blog post, 2010/08/05

Random Competitive Replication:

Some observations:

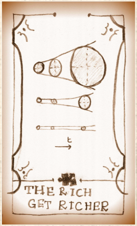
- Steady growth of system: +1 elephant per unit time.



Random Competitive Replication:

Some observations:

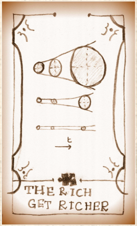
- Steady growth of system: +1 elephant per unit time.
- Steady growth of distinct flavors at rate ρ



Random Competitive Replication:

Some observations:

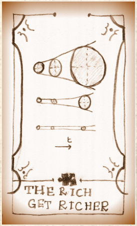
- Steady growth of system: +1 elephant per unit time.
- Steady growth of distinct flavors at rate ρ
- We can incorporate



Random Competitive Replication:

Some observations:

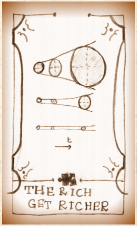
- Steady growth of system: +1 elephant per unit time.
- Steady growth of distinct flavors at rate ρ
- We can incorporate
    1. Elephant elimination



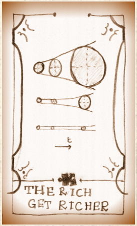
Random Competitive Replication:

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



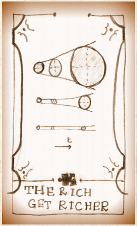
Random Competitive Replication:



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

Random Competitive Replication:



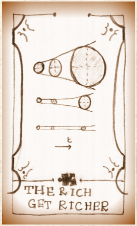
Some observations:

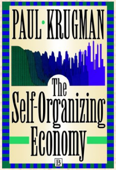
- Steady growth of system: +1 elephant per unit time.
- Steady growth of distinct flavors at rate ρ
- We can incorporate
  1. Elephant elimination
  2. Elephants moving between groups
  3. Variable innovation rate ρ
  4. Different selection based on group size

Random Competitive Replication:

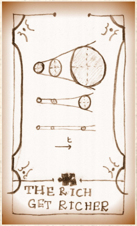
Some observations:

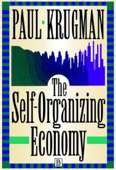
- Steady growth of system: +1 elephant per unit time.
- Steady growth of distinct flavors at rate ρ
- We can incorporate
  1. Elephant elimination
  2. Elephants moving between groups
  3. Variable innovation rate ρ
  4. Different selection based on group size
     (But mechanism for selection is not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





"The Self-Organizing Economy" by Paul Krugman (1996). [6]

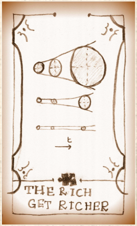


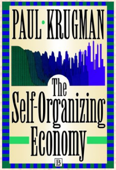


"The Self-Organizing Economy" by Paul Krugman (1996). [6]

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]

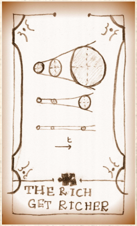




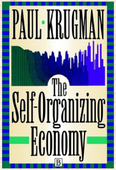
"The Self-Organizing Economy"
by Paul Krugman (1996). [6]

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.



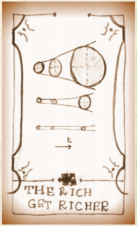
"The Self-Organizing Economy" by Paul Krugman (1996). [6]

Ch. 3: An Urban Mystery, p. 46

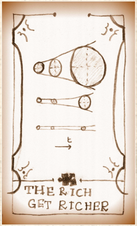
"...Simon showed—in a completely impenetrable exposition!—that the exponent of the power law distribution should be ..."[1, 2]

[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?



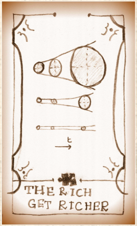
Outline



Random Competitive Replication:

Definitions:

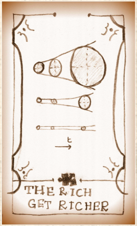
- k_i = size of a group i



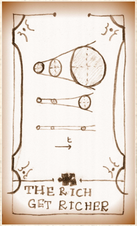
Random Competitive Replication:

Definitions:

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



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?

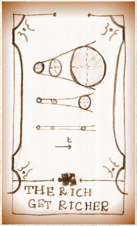
Random Competitive Replication:

Definitions:

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

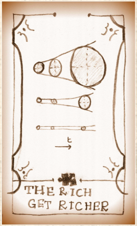
Basic question: How does $N_{k,t}$ evolve with time?

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



Random Competitive Replication:

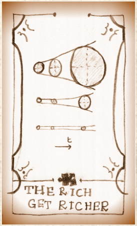
$P_k(t)$ = Probability of choosing an elephant that belongs to a group of size k:



Random Competitive Replication:

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

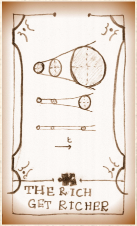
- $N_{k,t}$ size k groups



Random Competitive Replication:

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

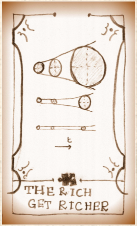
- $N_{k,t}$ size k groups
- $\Rightarrow kN_{k,t}$ elephants in size k groups



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

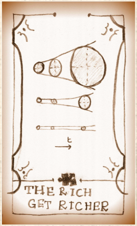


Random Competitive Replication:

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

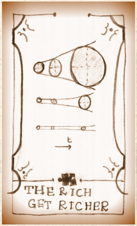
- $N_{k,t}$ size k groups
- $\Rightarrow kN_{k,t}$ elephants in size k groups
- t elephants overall

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



Random Competitive Replication:

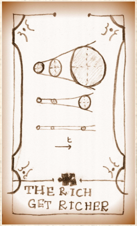
$N_{k,t}$, the number of groups with k elephants, changes at time t if



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:

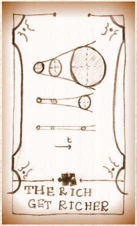


Random Competitive Replication:

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

1. An elephant belonging to a group with k elephants is replicated:

2. An elephant belonging to a group with $k-1$ elephants is replicated:

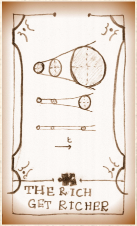


Random Competitive Replication:

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

1. An elephant belonging to a group with k elephants is replicated:
   $N_{k,t+1} = N_{k,t} - 1$

2. An elephant belonging to a group with $k-1$ elephants is replicated:

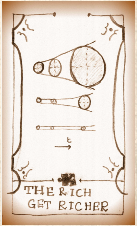


Random Competitive Replication:

$N_{k,t}$, 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:

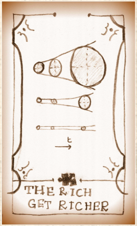


Random Competitive Replication:

$N_{k,t}$, 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:
   $N_{k,t+1} = N_{k,t} + 1$

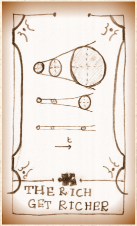


Random Competitive Replication:

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

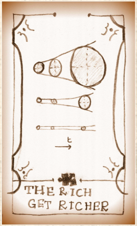
1. An elephant belonging to a group with k elephants is replicated:
   $N_{k,t+1} = N_{k,t} - 1$
   Happens with probability $(1-\rho)kN_{k,t}/t$

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

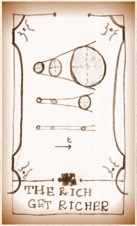


Random Competitive Replication:

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



Random Competitive Replication:



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

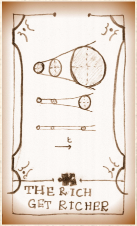
1. The new elephant is a new flavor:

Random Competitive Replication:

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

1. The new elephant is a new flavor:

2. A unique elephant is replicated:

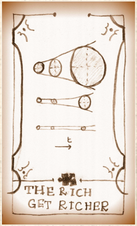


Random Competitive Replication:

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

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

2. A unique elephant is replicated:

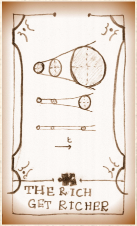


Random Competitive Replication:

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:

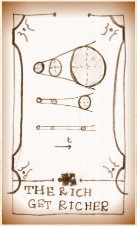


Random Competitive Replication:

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:
   $N_{1,t+1} = N_{1,t} - 1$

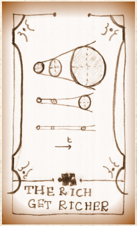


Random Competitive Replication:

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 $(1-\rho)N_{1,t}/t$

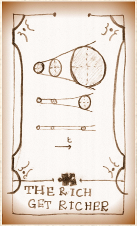


Random Competitive Replication:

Putting everything together:

For $k > 1$:

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



Random Competitive Replication:

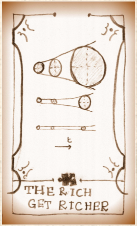
Putting everything together:

For $k > 1$:

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

For $k = 1$:

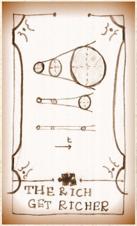
$$\left\langle N_{1,t+1} - N_{1,t} \right\rangle = (+1)\rho + (-1)(1-\rho)1 \cdot \frac{N_{1,t}}{t}$$



Random Competitive Replication:

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

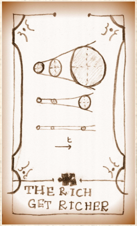
(Reasonable for t large)



Random Competitive Replication:

Assume distribution stabilizes: $N_{k,t} = n_k t$
(Reasonable for t large)

- Drop expectations

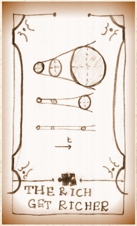


Random Competitive Replication:

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

(Reasonable for t large)

- Drop expectations
- Numbers of elephants now fractional

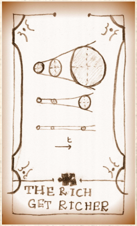


Random Competitive Replication:

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



Random Competitive Replication:

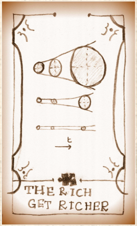
Assume distribution stabilizes: $N_{k,t} = n_k t$

(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

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



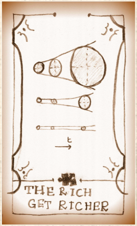
Random Competitive Replication:

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

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



Random Competitive Replication:

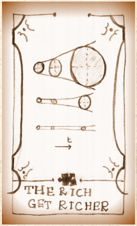
Stochastic difference equation:

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

becomes

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

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



Random Competitive Replication:

Stochastic difference equation:

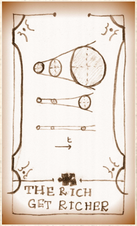
$$\left\langle N_{k,t+1} - N_{k,t} \right\rangle = (1-\rho)\left((k-1)\frac{N_{k-1,t}}{t} - k\frac{N_{k,t}}{t} \right)$$

becomes

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

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

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



Random Competitive Replication:

Stochastic difference equation:

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

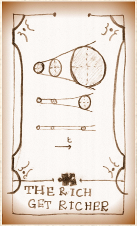
becomes

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

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

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

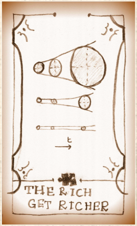
$$\Rightarrow n_k\left(1 + (1-\rho)k\right) = (1-\rho)(k-1)n_{k-1}$$



Random Competitive Replication:

We have a simple recursion:

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

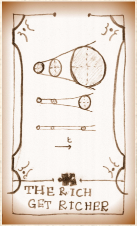


Random Competitive Replication:

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)

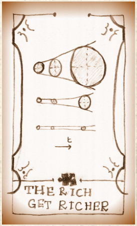


Random Competitive Replication:

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.



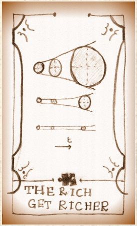
Random Competitive Replication:

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 assignment question



Random Competitive Replication:

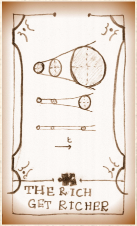
We have a simple recursion:

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

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

  Insert assignment question

- For just the tail: Expand as a series of powers of $1/k$



Random Competitive Replication:

We have a simple recursion:

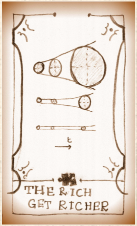
$$\frac{n_k}{n_{k-1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k}$$

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

  Insert assignment question

- For just the tail: Expand as a series of powers of $1/k$

  Insert assignment question



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 assignment question

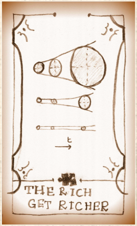
- For just the tail: Expand as a series of powers of $1/k$

  Insert assignment question

  We (okay, you) find

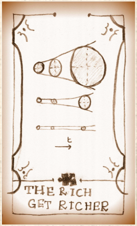
$$n_k \propto k^{-\frac{(2-\rho)}{(1-\rho)}} = k^{-\gamma}$$

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



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

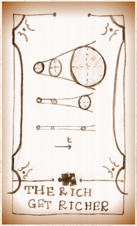
$$\boxed{\gamma = \frac{(2-\rho)}{(1-\rho)} = 1 + \frac{1}{(1-\rho)}}$$



- 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 $0 < \rho < 1$.

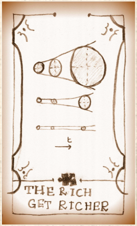


- Micro-to-Macro story with ρ and γ measurable.

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

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

$$\gamma \simeq 2$$



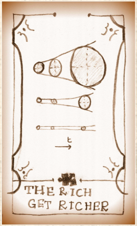
- Micro-to-Macro story with ρ and γ measurable.

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

- 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.



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

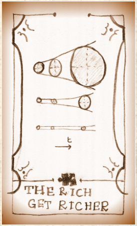
$$\boxed{\gamma = \frac{(2-\rho)}{(1-\rho)} = 1 + \frac{1}{(1-\rho)}}$$

- 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):

$$\gamma \simeq \infty$$



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

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

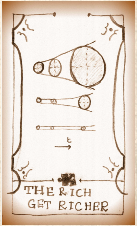
- Observe $2 < \gamma < \infty$ for $0 < \rho < 1$.
- For $\rho \simeq 0$ (low innovation rate):

$$\gamma \simeq 2$$

- '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.



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

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

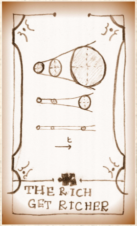
- Observe $2 < \gamma < \infty$ for $0 < \rho < 1$.
- For $\rho \simeq 0$ (low innovation rate):

$$\gamma \simeq 2$$

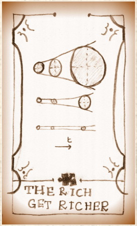
- 'Wild' power-law size distribution of group sizes, bordering on 'infinite' mean.
- For $\rho \simeq 1$ (high innovation rate):

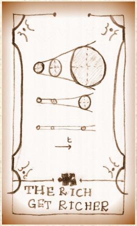
$$\gamma \simeq \infty$$

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



Recall size-ranking law: $s_r \sim r^{-\alpha}$
(s_r = size of the rth largest group of elephants)





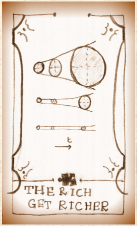
- Recall size-ranking law: $s_r \sim r^{-\alpha}$
  (s_r = size of the rth largest group of elephants)
- We found $\alpha = 1/(\gamma - 1)$ so:

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

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

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

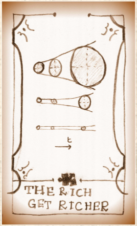
- $\gamma = 2$ corresponds to $\alpha = 1$



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

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

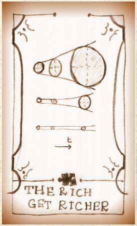
- $\gamma = 2$ corresponds to $\alpha = 1$
- We (roughly) see Zipfian exponent [12] of $\alpha = 1$ for many real systems: city sizes, word distributions, ...



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

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

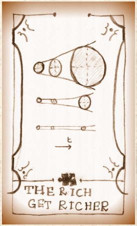
- $\gamma = 2$ corresponds to $\alpha = 1$
- We (roughly) see Zipfian exponent [12] of $\alpha = 1$ for many real systems: city sizes, word distributions, ...
- Corresponds to $\rho \to 0$, low innovation.



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

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

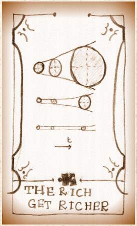
- $\gamma = 2$ corresponds to $\alpha = 1$
- We (roughly) see Zipfian exponent [12] of $\alpha = 1$ for many real systems: city sizes, word distributions, ...
- Corresponds to $\rho \rightarrow 0$, low innovation.
- Still, other quite different mechanisms are possible...



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

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

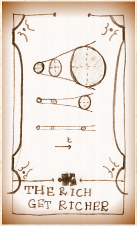
- $\gamma = 2$ corresponds to $\alpha = 1$
- We (roughly) see Zipfian exponent [12] of $\alpha = 1$ for many real systems: city sizes, word distributions, ...
- Corresponds to $\rho \to 0$, low innovation.
- Still, other quite different mechanisms are possible...
- Must look at the details to see if mechanism makes sense... more later.



What about small k?:

We had one other equation:

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

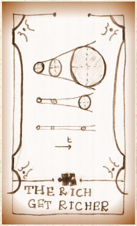


What about small k?:

We had one other equation:

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

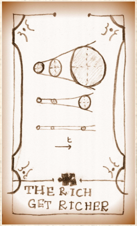
- As before, set $N_{1,t} = n_1 t$ and drop expectations



What about small k?:

We had one other equation:

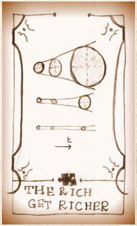
- $$\langle N_{1,t+1} - N_{1,t} \rangle = \rho - (1-\rho) 1 \cdot \frac{N_{1,t}}{t}$$
- As before, set $N_{1,t} = n_1 t$ and drop expectations
- $$n_1(t+1) - n_1 t = \rho - (1-\rho) 1 \cdot \frac{n_1 t}{t}$$
- $$n_1 = \rho - (1-\rho) n_1$$



What about small k?:

We had one other equation:

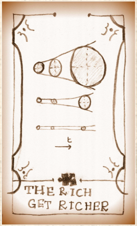
- $$\langle N_{1,t+1} - N_{1,t} \rangle = \rho - (1-\rho)1 \cdot \frac{N_{1,t}}{t}$$
- As before, set $N_{1,t} = n_1 t$ and drop expectations
- $$n_1(t+1) - n_1 t = \rho - (1-\rho)1 \cdot \frac{n_1 t}{t}$$
- $$n_1 = \rho - (1-\rho)n_1$$
- Rearrange:
$$n_1 + (1-\rho)n_1 = \rho$$



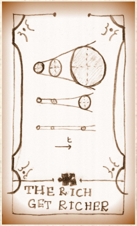
What about small k?:

We had one other equation:

- $$\langle N_{1,t+1} - N_{1,t} \rangle = \rho - (1-\rho)1 \cdot \frac{N_{1,t}}{t}$$
- 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:
  $$n_1 + (1-\rho)n_1 = \rho$$
- $$\boxed{n_1 = \frac{\rho}{2-\rho}}$$

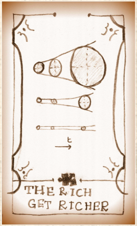


So... $$N_{1,t} = n_1 t = \frac{\rho t}{2-\rho}$$



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

- Recall number of distinct elephants = ρt.



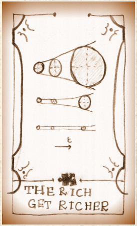
So...

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

- 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)



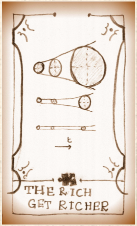
So... $$N_{1,t} = n_1 t = \frac{\rho t}{2-\rho}$$

- 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 $\sim 1/2$



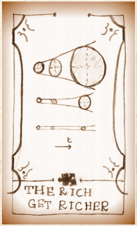
So... $$N_{1,t} = n_1 t = \frac{\rho t}{2-\rho}$$

- 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



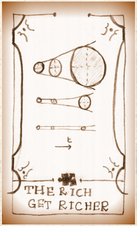
So... $$N_{1,t} = n_1 t = \frac{\rho t}{2-\rho}$$

- 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



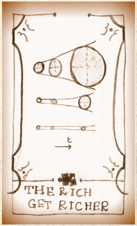
So... $$N_{1,t} = n_1 t = \frac{\rho t}{2-\rho}$$

- 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 $\sim 1/6$



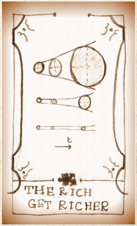
$$\text{So...} \qquad N_{1,t} = n_1 t = \frac{\rho t}{2-\rho}$$

- 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



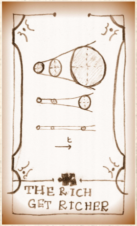
So... $$N_{1,t} = n_1 t = \frac{\rho t}{2-\rho}$$

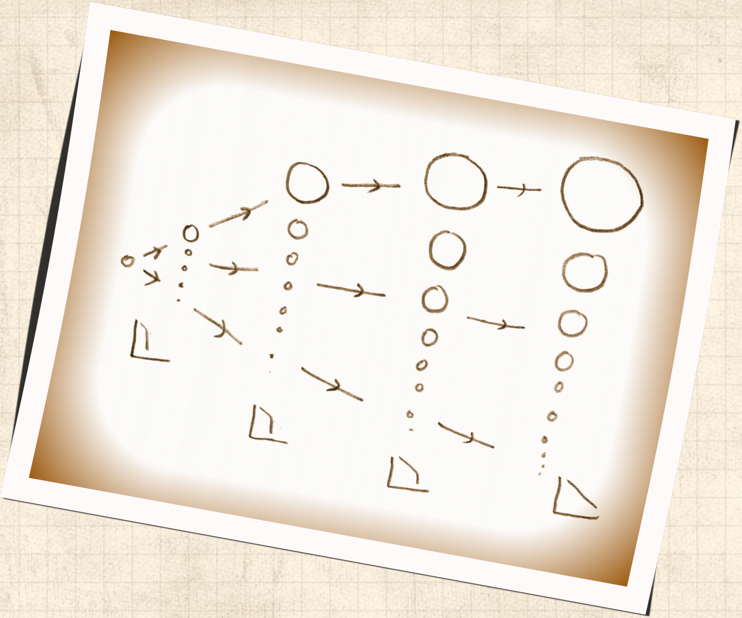
- Recall number of distinct elephants = ρt.
- Fraction of distinct elephants that are unique (belong to groups of size 1):

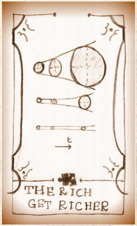
$$\frac{1}{\rho t} N_{1,t} = \frac{1}{\cancel{\rho t}} \frac{\cancel{\rho t}}{2-\rho} = \frac{1}{2-\rho}$$

(also = fraction of groups of size 1)

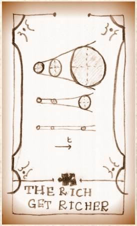
- For ρ small, fraction of unique elephants $\sim 1/2$
- Roughly observed for real distributions
- ρ increases, fraction increases
- Can show fraction of groups with two elephants $\sim 1/6$
- Model works well for large and small k #awesome






t
THE RICH
GET RICHER

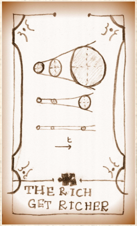
Outline



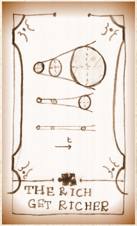
Words:

From Simon [10]:

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



Words:

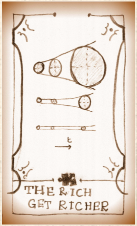


From Simon [10]:

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

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

Words:

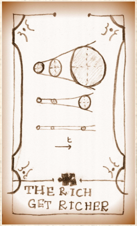


From Simon [10]:

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

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

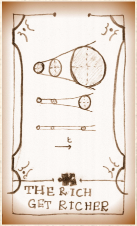
N_1 (real)	N_1 (est)	N_2 (real)	N_2 (est)
16,432	15,850	4,776	4,870



Outline

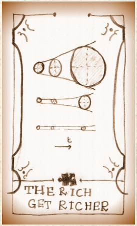
Evolution of catch phrases:

- Yule's paper (1924) [11]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."



Evolution of catch phrases:

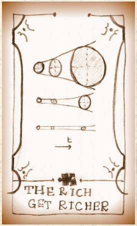
- Yule's paper (1924) [11]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."
- Simon's paper (1955) [10]:
  "On a class of skew distribution functions" (snore)



Evolution of catch phrases:

- Yule's paper (1924) [11]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."
- Simon's paper (1955) [10]:
  "On a class of skew distribution functions" (snore)

From Simon's introduction:

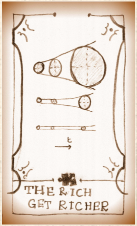


Evolution of catch phrases:

- Yule's paper (1924) [11]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."
- Simon's paper (1955) [10]:
  "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

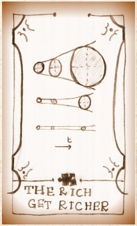


Evolution of catch phrases:

- Yule's paper (1924) [11]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."
- Simon's paper (1955) [10]:
  "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.



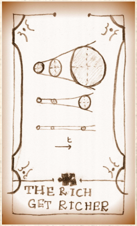
Evolution of catch phrases:

- Yule's paper (1924) [11]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."
- Simon's paper (1955) [10]:
  "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,



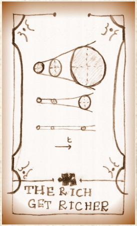
Evolution of catch phrases:

- Yule's paper (1924) [11]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."
- Simon's paper (1955) [10]:
  "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



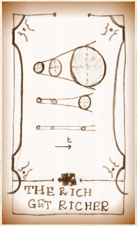
Evolution of catch phrases:

- Yule's paper (1924) [11]:
  "A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S."
- Simon's paper (1955) [10]:
  "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.

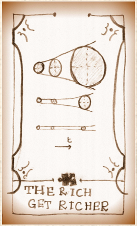
Its appearance is so frequent, and the phenomena so diverse, that one is led to conjecture that if these phenomena have any property in common it can only be a similarity in the structure of the underlying probability mechanisms.



Evolution of catch phrases:

Derek de Solla Price:

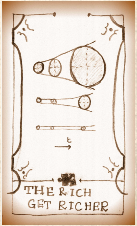
- First to study network evolution with these kinds of models.



Evolution of catch phrases:

Derek de Solla Price:

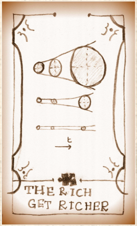
- First to study network evolution with these kinds of models.
- Citation network of scientific papers



Evolution of catch phrases:

Derek de Solla Price:

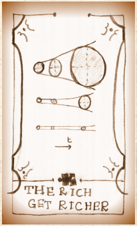
- First to study network evolution with these kinds of models.
- Citation network of scientific papers
- Price's term: Cumulative Advantage



Evolution of catch phrases:

Derek de Solla Price:

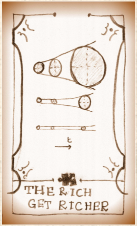
- First to study network evolution with these kinds of models.
- Citation network of scientific papers
- Price's term: Cumulative Advantage
- Idea: papers receive new citations with probability proportional to their existing # of citations



Evolution of catch phrases:

Derek de Solla Price:

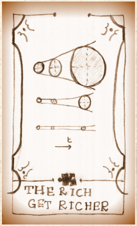
- First to study network evolution with these kinds of models.
- Citation network of scientific papers
- Price's term: Cumulative Advantage
- Idea: papers receive new citations with probability proportional to their existing # of citations
- Directed network



Evolution of catch phrases:

Derek de Solla Price:

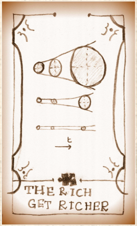
- First to study network evolution with these kinds of models.
- Citation network of scientific papers
- Price's term: Cumulative Advantage
- Idea: papers receive new citations with probability proportional to their existing # of citations
- Directed network
- Two (surmountable) problems:
  1. New papers 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

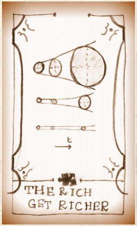


Evolution of catch phrases:

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...

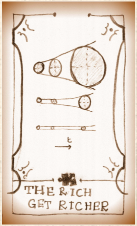


Evolution of catch phrases:

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....)

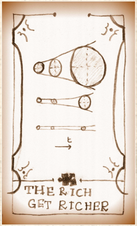


Evolution of catch phrases:

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.

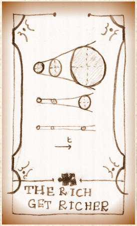


Evolution of catch phrases:

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."



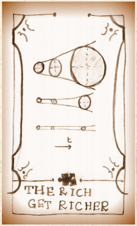
Evolution of catch phrases:

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.)

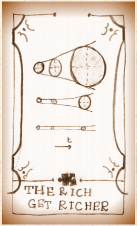


Evolution of catch phrases:

Robert K. Merton: the Matthew Effect

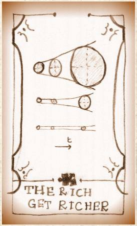
- Studied careers of scientists and found credit flowed disproportionately to the already famous

  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:

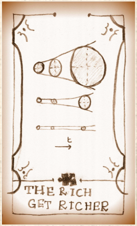
Merton was a catchphrase machine:



Evolution of catch phrases:

Merton was a catchphrase machine:

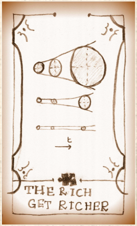
1. Self-fulfilling prophecy



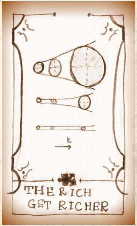
Evolution of catch phrases:

Merton was a catchphrase machine:

1. Self-fulfilling prophecy
2. Role model



Evolution of catch phrases:



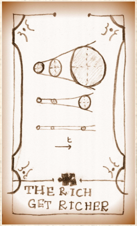
Merton was a catchphrase machine:

1. Self-fulfilling prophecy
2. Role model
3. Unintended (or unanticipated) consequences

Evolution of catch phrases:

Merton was a catchphrase machine:

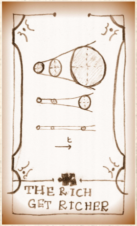
1. Self-fulfilling prophecy
2. Role model
3. Unintended (or unanticipated) consequences
4. Focused interview → focus group

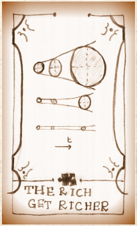


Evolution of catch phrases:

Merton was a catchphrase machine:

1. Self-fulfilling prophecy
2. Role model
3. Unintended (or unanticipated) consequences
4. Focused interview → focus group
5. Obliteration by incorporation (includes above examples from Merton himself)





Evolution of catch phrases:

Merton was a catchphrase machine:

1. Self-fulfilling prophecy
2. Role model
3. Unintended (or unanticipated) consequences
4. Focused interview → focus group
5. Obliteration by incorporation (includes above examples from Merton himself)

And just to be clear...

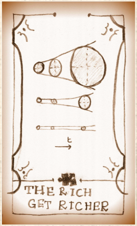
Evolution of catch phrases:

Merton was a catchphrase machine:

1. Self-fulfilling prophecy
2. Role model
3. Unintended (or unanticipated) consequences
4. Focused interview → focus group
5. Obliteration by incorporation (includes above examples from Merton himself)

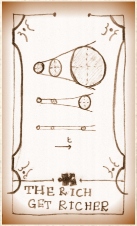
And just to be clear...

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



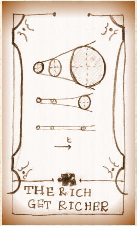
Evolution of catch phrases:

- Barabasi and Albert [2]—thinking about the Web



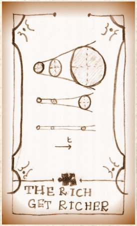
Evolution of catch phrases:

- Barabasi and Albert [2]—thinking about the Web
- Independent reinvention of a version of Simon and Price's theory for networks



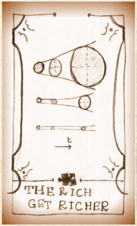
Evolution of catch phrases:

- 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"



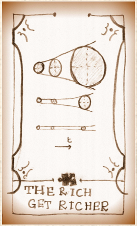
Evolution of catch phrases:

- 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)



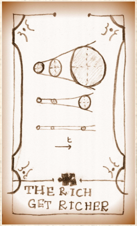
Evolution of catch phrases:

- 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)



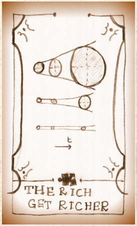
Evolution of catch phrases:

- 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) ...



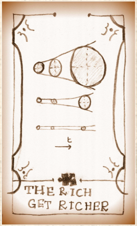
Evolution of catch phrases:

- 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)



Evolution of catch phrases:

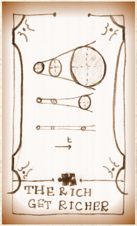
- Barabasi and Albert [2]—thinking about the Web
- Independent reinvention of a version of Simon and Price's theory for networks
- Another term: "Preferential Attachment"
- Considered undirected networks (not realistic but avoids 0 citation problem)
- Still have selection problem based on size (non-random)
- 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



More mattering:

Rich-get-richerness in social contagion:

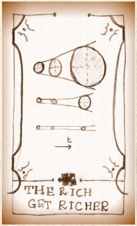
- We love to rank everyone, everything: Top n lists.



More mattering:

Rich-get-richerness in social contagion:

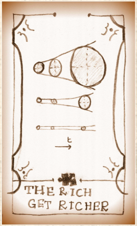
- We love to rank everyone, everything: Top n lists.
- People, wealth, sports, music, movies, books, schools, cities, countries, dogs (13/10), ...



More mattering:

Rich-get-richerness in social contagion:

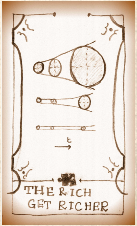
- We love to rank everyone, everything: Top n lists.
- People, wealth, sports, music, movies, books, schools, cities, countries, dogs (13/10), ...
- Gameable: payola, astroturfing, sockpuppetry, John Barron (the sockpuppet hype man), ...



More mattering:

Rich-get-richerness in social contagion:

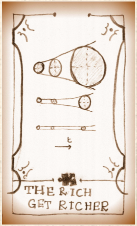
- We love to rank everyone, everything: Top n lists.
- People, wealth, sports, music, movies, books, schools, cities, countries, dogs (13/10), ...
- Gameable: payola, astroturfing, sockpuppetry, John Barron (the sockpuppet hype man), ...
- Black-box ranking algorithms make ranking opaque.



More mattering:

Rich-get-richerness in social contagion:

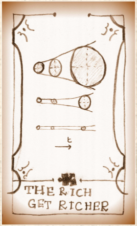
- We love to rank everyone, everything: Top n lists.
- People, wealth, sports, music, movies, books, schools, cities, countries, dogs (13/10), ...
- Gameable: payola, astroturfing, sockpuppetry, John Barron (the sockpuppet hype man), ...
- Black-box ranking algorithms make ranking opaque.
- Black boxes are gameable but takes money and commensurate skill.



More mattering:

Rich-get-richerness in social contagion:

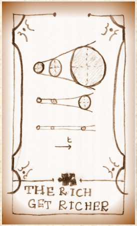
- We love to rank everyone, everything: Top n lists.
- People, wealth, sports, music, movies, books, schools, cities, countries, dogs (13/10), ...
- Gameable: payola, astroturfing, sockpuppetry, John Barron (the sockpuppet hype man), ...
- Black-box ranking algorithms make ranking opaque.
- Black boxes are gameable but takes money and commensurate skill.
- Black box algorithms can make things spread rampantly.[1]



More mattering:

Rich-get-richerness in social contagion:

- We love to rank everyone, everything: Top n lists.
- People, wealth, sports, music, movies, books, schools, cities, countries, dogs (13/10), ...
- Gameable: payola, astroturfing, sockpuppetry, John Barron (the sockpuppet hype man), ...
- Black-box ranking algorithms make ranking opaque.
- Black boxes are gameable but takes money and commensurate skill.
- Black box algorithms can make things spread rampantly.[1]



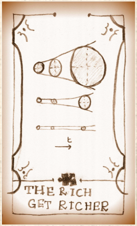
[1] "With great power comes great responsibility." –S. Man.

More mattering:

Rich-get-richerness in social contagion:

- We love to rank everyone, everything: Top n lists.
- People, wealth, sports, music, movies, books, schools, cities, countries, dogs (13/10), ...
- Gameable: payola, astroturfing, sockpuppetry, John Barron (the sockpuppet hype man), ...
- Black-box ranking algorithms make ranking opaque.
- Black boxes are gameable but takes money and commensurate skill.
- Black box algorithms can make things spread rampantly.[1]
- No "regramming" is a positive feature of Instagram (also: Pratchett the Cat)

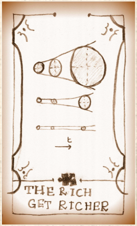
[1] "With great power comes great responsibility." –S. Man.



More mattering:

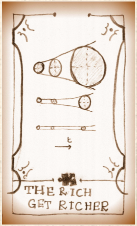
Rich-get-richerness in social contagion:

- We love to rank everyone, everything: Top n lists.
- People, wealth, sports, music, movies, books, schools, cities, countries, dogs (13/10), ...
- Gameable: payola, astroturfing, sockpuppetry, John Barron (the sockpuppet hype man), ...
- Black-box ranking algorithms make ranking opaque.
- Black boxes are gameable but takes money and commensurate skill.
- Black box algorithms can make things spread rampantly.[1]
- No "regramming" is a positive feature of Instagram (also: Pratchett the Cat)
- What if a healthier Facebook is just ... Instagram? (hahahhaaha)



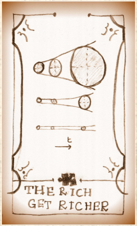
[1] "With great power comes great responsibility." –S. Man.

Rich-get-richerness is everywhere in reality



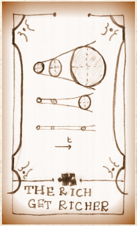
Rich-get-richerness is everywhere in reality

But:



Rich-get-richerness is everywhere in reality

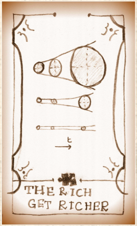
But:



Rich-get-richerness is everywhere in reality

But:

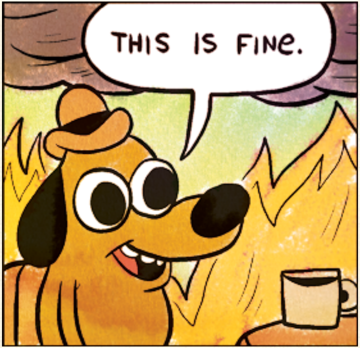
Simon's Model

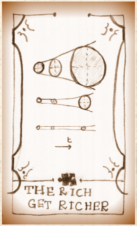


Rich-get-richerness is everywhere in reality

But:

Simon's Model





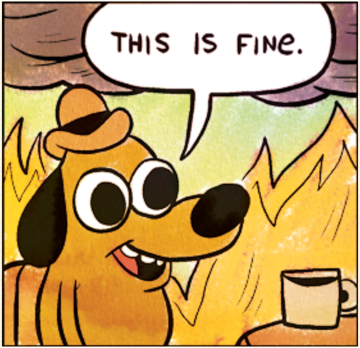
Burn it all down.

Rich-get-richerness is everywhere in reality

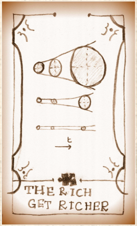
But:

Simon's Model

Is



Burn it all down.

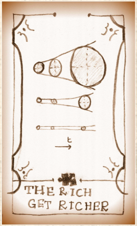


Rich-get-richerness is everywhere in reality

But:

Simon's Model

Is



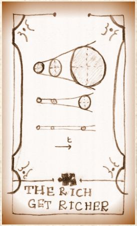
Burn it all down.

Rich-get-richerness is everywhere in reality

But:

Simon's Model

Is



Burn it all down.

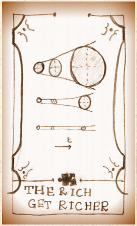
Rich-get-richerness is everywhere in reality

But:

Simon's Model

Is

WRONG.



Burn it all down.

Rich-get-richerness is everywhere in reality

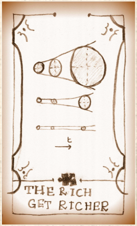
But:

Simon's Model

Is

WRONG.

Burn it all down.



Rich-get-richerness is everywhere in reality

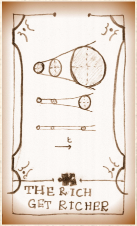
But:

Simon's Model

Is

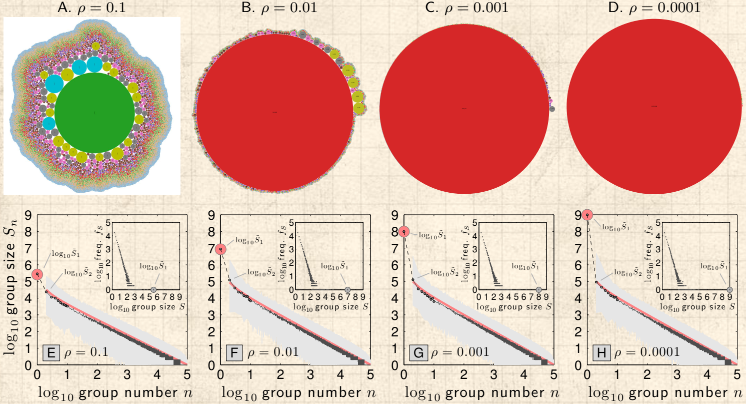
WRONG.

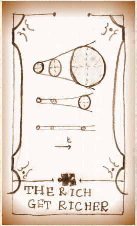
To be continued ...



Burn it all down.

"Simon's fundamental rich-get-richer model entails a dominant first-mover advantage"
Dodds et al.,
Physical Review E, **95**, 052301, 2017. [4]





See Fletcher Hazlehurst's visualization at paper's online app-endices

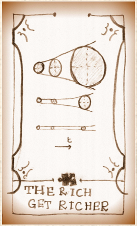
References I

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

[4] 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-get-richer model entails a dominant first-mover advantage.
Physical Review E, 95:052301, 2017. pdf

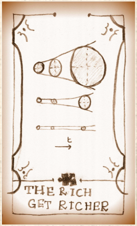


References II

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

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

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

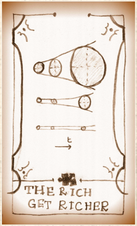


References III

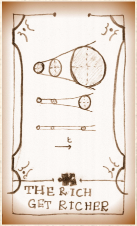
[8] B. B. Mandelbrot.
An informational theory of the statistical structure of languages.
In W. Jackson, editor, Communication Theory, pages 486–502. Butterworth, Woburn, MA, 1953. pdf

[9] D. D. S. Price.
A general theory of bibliometric and other cumulative advantage processes.
Journal of the American Society for Information Science, pages 292–306, 1976. pdf

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



References IV



[11] 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

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