The Scaling of the Numbers of Types and the Numbers of Things

Last updated: 2025/09/23, 07:58:58 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

Outline

A key framing from language:

Types and Tokens:

- In linguistics, words are described on the two levels of types and tokens [4].
- In semiotics, signs can be thought of having two components of the signified and the signifier.

Example:

- Types are 1-grams, e.g., '!', 'the', 'love', and 'spork'.[1]
- Tokens are 1-grams as written down.
- In "Pride and Prejudice", for example, there are 498 '!'s, 4,058 'the's, 90 'love's, and 0 'spork's.

[1] Linguists have a long history of not agreeing on what a word is.

Types and Things and Measures, Oh My!

Beyond language:

Lift out and expand the type-token framing to complex systems in general.

~~Three~~ Four possible parts:

1. Type: A kind or class of category of individual things based on shared characteristics.
2. Thing: An individual manifestation of a type.
3. Measure: A quantification of the manifestation of things.
4. Experience: An interaction of any kind with a manifestation of a type.[2]

[2] Fame.

Language:

1. Type: A defined word.
2. Thing (token): An instance of spoken or printed word.
3. Number or Frequency (counts of tokens).
4. Experience: Listening to others, reading a book.

Atoms:

1. Type: Atom
2. Thing: Element (stuff made of a given atom; e.g., gold)
3. Measure: Mass; could be Number.
4. Experience: Atomic bonds.

Water:

1. Type: Water molecule, H^2O.
2. Thing: Water.
3. Measure: Volume (liters, gallons); given pressure and temperature, equivalent to Number (counts of molecules) and then Mass.
4. Experience: Rain.

Biology:

- Example type: The species Ornithorhynchus anatinus, the platypus.
- Thing: Any given platypus.
- Measure: The number of platypuses ('instances' of the species) living in Australia in the wild.
- Experience: Seeing a platypus in the wild; being hunted by a platypus.

Moneyspace:

- Example type: Corporation.
- Things: The publicly traded companies of Apple and Microsoft.
- Measure: Market capitalization.
- Experience: Being sued by Microsoft.

- Apple and Microsoft may be viewed as components of the publicly-owned corporate world.
- The sizes of corporations may be broken down into many rankable dimensions such as annual revenue or number of employees worldwide.
- In principle, market capitalization represents a kind of current collective belief in terms of money.

Sizes and Rankings:

- We will often consider systems where each component type τ has at least one measurable—and hence rankable—'size' s_τ.
- Perceived size is a combination of Measure (what exists) and Experience (what is measured).
- Important: We may also have rankings where we do not know the underlying 'size' (e.g., book/thing sales on Amazon).

Three examples which show some of the range of what 'size' can mean:

1. Size for a word in a corpus means the number of indistinguishable instances of that word (many identical entites—tokens);
2. Size for species means the number of 'biological replications' of an individual type (many genetically similar entities of varying ages); and
3. Size for a corporation might mean monetary value (market cap, one entity).
4. May have more than one measure of a system:
   - Total biomass of a species.[3]
   - Number of employees in a corporation.
   - Number of stars in a galaxy.[3]
5. Measure of size allows for rankings.
6. Again, sizes may be hidden.

[3] Somewhat hard to estimate.

When tokens are fungible:

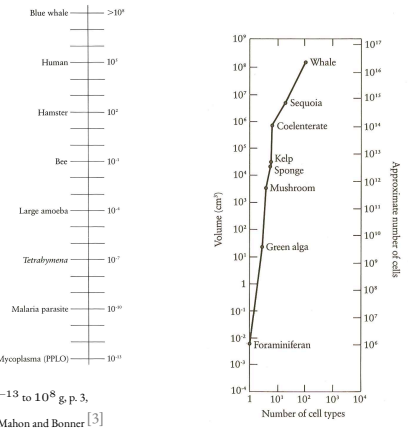
- Randomly permute all of the words (tokens) of the same type in Pride and Prejudice.
- Measure and Experience will be unchanged.
- NFTs: Non-fungible tokens.
- Tricking people into thinking tokens are types.
- "The Oxymoron for Morons."

When tokens are funguses:

- NFF: Non-fungible fungus (from a sentient fungus's point of view).
- But in cooking, funguses are fungible.
- Lack of exposure leads to fungibility of "the other."[4]

[4] Universal: Identical twins look the same until they don't.

From scaling: Size range (in grams) and cell differentiation:



10^{-13} to 10^{8} g, p. 3,
McMahon and Bonner [3]

Scaling of Specialization:

"Scaling of Differentiation in Networks: Nervous Systems, Organisms, Ant Colonies, Ecosystems, Businesses, Universities, Cities, Electronic Circuits, and Legos"
Changizi, McDannald, and Widders,
J. Theor. Biol, **218**, 215–237, 2002. [1]

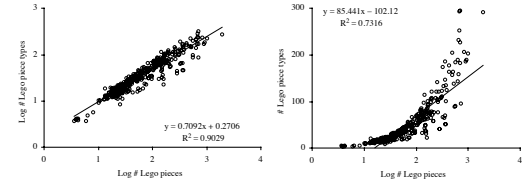


FIG. 3. Log–log (base 10) (left) and semi-log (right) plots of the number of Lego piece types vs. the total number of parts in Lego structures ($n = 391$). To help to distinguish the data points, logarithmic values were perturbed by adding a random number in the interval $[-0.05, 0.05]$, and non-logarithmic values were perturbed by adding a random number in the interval $[-1, 1]$.

2012 wired.com write-up

Independent observation:

$C \sim N^{1/d}, d \geq 1$:

- C = network differentiation = # node types.
- N = network size = # nodes.
- d = combinatorial degree.
- Low d: strongly specialized parts.
- High d: strongly combinatorial in nature, parts are reused.
- Claim: Natural selection produces high d systems.
- Claim: Engineering/brains produces low d systems.

TABLE 1
*Summary of results**

Network	Node	No. data points	Range of log N	Log–log R^2	Semi-log R^2	p_{power}/p_{lin}	Relationship between C and N	Comb. degree	Exponent v for type-net scaling	Figure in text
Selected networks										
Electronic circuits	Component	373	2.12	0.747	0.602	0.05/4e−5	Power law	2.29	0.92	2
Legos[36]	Piece	391	2.65	0.903	0.732	0.09/1e−7	Power law	1.41	—	3
Businesses										
military vessels	Employee	13	1.88	0.971	0.832	0.05/3e−3	Power law	1.60	—	4
military offices	Employee	8	1.59	0.964	0.789	0.16/0.16	Increasing	1.13	—	4
universities	Employee	9	1.55	0.786	0.749	0.27/0.27	Increasing	1.37	—	4
insurance co.	Employee	52	2.30	0.748	0.685	0.11/0.10	Increasing	3.04	—	4
Universities										
across schools	Faculty	112	2.72	0.695	0.549	0.09/0.01	Power law	1.81	—	5
history of Duke	Faculty	46	0.94	0.921	0.892	0.09/0.05	Increasing	2.07	—	5
Ant colonies										
caste = type	Ant	46	6.00	0.481	0.454	0.11/0.04	Power law	8.16	—	6
size range = type	Ant	22	5.24	0.658	0.548	0.17/0.04	Power law	8.00	—	6
Organisms	Cell	134	12.40	0.249	0.165	0.08/0.02	Power law	17.73	—	7
Neocortex	Neuron	10	0.85	0.520	0.584	0.16/0.16	Increasing	4.56	—	9
Competitive networks										
Biotas	Organism	—	—	—	—	—	Power law	≈3	0.3 to 1.0	—
Cities	Business	82	2.44	0.985	0.832	0.08/8e−8	Power law	1.56	—	10

*(1) The kind of network, (2) what the nodes are within that kind of network, (3) the number of data points, (4) the logarithmic range of network sizes N (i.e. $\log(N_{max}/N_{min})$), (5) the log–log correlation, (6) the semi-log correlation, (7) the serial-dependence probabilities under, respectively, power-law and logarithmic models, (8) the empirically determined best-fit relationship between differentiation C and organization size N (if one of the two models can be refuted with $p < 0.05$; otherwise we just write "increasing" to denote that neither model can be rejected), (9) the combinatorial degree (i.e. the inverse of the best-fit slope of a log–log plot of C versus N), (10) the scaling exponent for how quickly the edge-degree δ scales with type-network size C (in those places for which data exist), (11) figure in this text where the plots are presented. Values for biotas represent the broad trend from the literature.

Language:

- The naturally-incorrectly-attributed[5] Heaps' Law
- For words: $N_{\text{types}} \sim N_{\text{tokens}}^{\beta}$ where $0\beta \leq 1$.
- Applies when reading through a sufficiently large coherent text (e.g., a book)
- Applies when sampling from a fixed power-law size distribution
- Misapplies when reading through sequential texts.
- Implied by this PoCS generated paper:

"Text mixing shapes the anatomy of rank-frequency distributions"
Williams et al.,
Physical Review E, **91**, 052811, 2015. [5]

[5] Plus one for Stigler's Law of Eponymy.

The type-thing scaling law:

- Most generally: $N_{\text{types}} \sim N_{\text{things}}^{\beta}$ where $0 < \beta \leq 1$.

$$t = \sum_{r=1}^{N_{t,\alpha}} S_{r,t,\alpha} \simeq N_{t,\alpha}^{\alpha} \int_{z=1}^{N_{t,\alpha}} z^{-\alpha} \mathrm{d}z \sim \frac{N_{t,\alpha}^{\alpha}}{1-\alpha} \left[N_{t,\alpha}^{1-\alpha} - 1 \right], \tag{1}$$

Type-token scaling for finite systems:

"Zipf's Law leads to Heaps' Law: Analyzing their relation in finite-size systems"
Lü, Zhang, and Zhou,
PLOS ONE, **5**, 1–11, 2010. [2]

- In a somewhat complicated way which we will fix up, Lü et al. determine that given a size-rank distribution:

$$S_r \sim r^{-\alpha} \tag{2}$$

for $r = 1, 2, \ldots, N_{\text{types}}$ then

$$N_{\text{things}} \simeq \frac{N_{t,\alpha}^{\alpha}}{1-\alpha} \left[N_{t,\alpha}^{1-\alpha} - 1 \right]. \tag{3}$$

Insert assignment question

Given:

$$N_{\text{things}} \simeq \frac{N_{t,\alpha}^{\alpha}}{1-\alpha} \left[N_{t,\alpha}^{1-\alpha} - 1 \right], \tag{4}$$

then

$$N_{\text{types}} \to \begin{cases} N_{\text{types}} & \text{for } \alpha = 0, \\ (1-\alpha) N_{\text{types}} & \text{for } 0 < \alpha \ll 1, \\ e^{W(N_{\text{types}})} \sim \dfrac{N_{\text{types}}}{\ln N_{\text{types}}} & \text{for } \alpha = 1, \\ (\alpha N_{\text{types}})^{1/\alpha} & \text{for } \alpha \gg 1, \\ 1 & \text{for } \alpha \to \infty, \end{cases} \tag{5}$$

where W is the Lambert[6] W function.

[6] There can be only one multivalued Lambert fuction.

References I

[1] M. A. Changizi, M. A. McDannald, and D. Widders.
Scaling of differentiation in networks: Nervous systems, organisms, ant colonies, ecosystems, businesses, universities, cities, electronic circuits, and Legos.
J. Theor. Biol, 218:215–237, 2002. pdf

[2] L. Lü, Z.-K. Zhang, and T. Zhou.
Zipf's Law leads to Heaps' Law: Analyzing their relation in finite-size systems.
PLOS ONE, 5(12):1–11, 12 2010. pdf

[3] T. A. McMahon and J. T. Bonner.
On Size and Life.
Scientific American Library, New York, 1983.

[4] C. S. S. Peirce.
Prolegomena to an apology for pragmaticism.
The Monist, 16(4):492–546, 1906. pdf

References II

[5] J. R. Williams, J. P. Bagrow, C. M. Danforth, and P. S. Dodds.
Text mixing shapes the anatomy of rank-frequency distributions.
Physical Review E, 91:052811, 2015. pdf