Allotaxonometry

Last updated: 2024/09/26, 08:35:21 EDT

Principles of Complex Systems, Vols. 1, 2, & 3D CSYS/MATH 6701, 6713, & a pretend number, 2024–2025

Prof. Peter Sheridan Dodds

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



Outline

A plenitude of distances

Rank-turbulence divergence

Probability-turbulence divergence

Explorations

Nutshell

References

| Coal — Understand this: | O₁: Twitter on 2016/11/09 | Divergence contribution δD²_{1,12} × (10⁻³⁶C) | | Instrument, Rank-Turbulence Divergence | Instrument, Rank-Turbulence | I

The PoCSverse Allotaxonometry 1 of 70

A plenitude of distances

tank-turbulence livergence

divergence
Probability-

turbulence divergen Explorations

Allotaxonometry 2 of 70

Rank-turbulence

divergence Probability-

Foundational papers:

Site (papers, examples, code):



"Allotaxonometry and rank-turbulence divergence: A universal instrument for comparing complex systems"

Dodds et al.,

http://compstorylab.org/allotaxonometry/

EPJ Data Science, **12**, 1–42, 2023. ^[5]

EPJ Data Science version



"Probability-turbulence divergence: A tunable allotaxonometric instrument for comparing heavy-tailed categorical distributions"

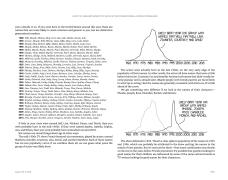
Dodds et al.,

Basic science = Describe + Explain:

- Dashboards of single scale instruments helps us understand, monitor, and control systems.
- Archetype: Cockpit dashboard for flying a plane
- Okay if comprehendible.
- Complex systems present two problems for dashboards:
 - Scale with internal diversity of components: We need meters for every species, every company, every word.
 - 2. Tracking change: We need to re-arrange meters on the fly.
- Goal—Create comprehendible, dynamically-adjusting, differential dashboards showing two pieces:¹
 - 1. 'Big picture' map-like overview,
 - 2. A tunable ranking of components.

¹See the lexicocalorimeter ✓

Baby names, much studied: [12]

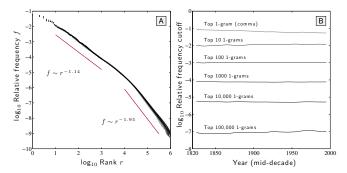


How to build a dynamical dashboard that helps sort through a massive number of interconnected time series?



"Is language evolution grinding to a halt? The scaling of lexical turbulence in English fiction suggests it is not" Pechenick, Danforth, Dodds, Alshaabi, Adams, Reagan, Danforth, Frank, Reagan, and Danforth.

Journal of Computational Science, 21, 24–37, 2017. [14]



The PoCSverse Allotaxonometry 5 of 70 A plenitude of

Rank-turbulence divergence

Probabilityturbulence divergence

Explorations Nutshell

Allotaxonometry

A plenitude of

Rank-turbulence divergence Probabilityturbulence divergence Explorations

С

Flux words)

Average $(\log_{10} \# v)$

- ■− φ_{down}

 \log_{10} Relative freq. threshold $f_{ ext{thr}}$

6 of 70

For language, Zipf's law has two scaling regimes: [19]

$$f \sim \left\{ \begin{array}{l} r^{-\alpha} \ {\rm for} \ r \ll r_{\rm b}, \\ r^{-\alpha'} \ {\rm for} \ r \gg r_{\rm b}, \end{array} \right. \label{eq:force}$$

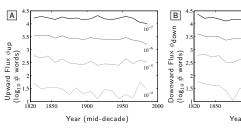
When comparing two texts, define Lexical turbulence as flux of words across a frequency threshold:

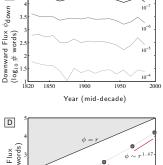
$$\phi \sim \left\{ egin{array}{l} f_{
m thr}^{-\mu} \ {
m for} \ f_{
m thr} \ll f_{
m b}, \ f_{
m thr}^{-\mu'} \ {
m for} \ f_{
m thr} \gg f_{
m b}, \end{array}
ight.$$

Estimates: $\mu \simeq 0.77$ and $\mu' \simeq 1.10$, and $f_{\rm b}$ is the scaling break point.

$$\phi \sim \left\{ \begin{array}{l} r^{\nu} = r^{\alpha \mu'} \text{ for } r \ll r_{\rm b}, \\ r^{\nu'} = r^{\alpha' \mu} \text{ for } r \gg r_{\rm b}. \end{array} \right.$$

Estimates: Lower and upper exponents $\nu \simeq 1.23$ and $\nu' \simeq 1.47$.





Allotaxonometry 8 of 70

A plenitude of

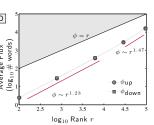
Rank-turbuleno

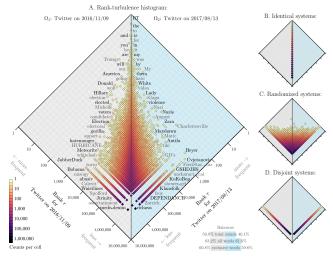
turbulence div

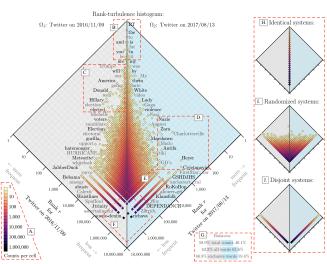
divergence

Probability

Exploration







Balances:

Top bar (optional)—Total size:

- Relative balance of system sizes.
- Examples: Total number of words in a book, total number of individuals in an ecology.

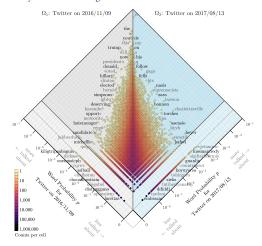
Middle bar—Types:

Fraction of types in each system as a percentage of the union of types from both systems.

Bottom bar—Exclusive types:

- Types that are present in one system only are 'exclusive types'.
- $\Omega^{(1)}$ -exclusive and $\Omega^{(2)}$ -exclusive indicate which system an exclusive type belongs to.
- Percentage of exclusive types in a system relative to that system's total number of types.

Probability-turbulence histogram:



So, so many ways to compare probability distributions:



"Families of Alpha- Beta- and Gamma- Divergences: Flexible and Robust Measures of Similarities" Cichocki and Amari, Entropy, **12**, 1532-1568, 2010. [2]

Methods in Applied Sciences, 1, 300–307, 2007. [1]

"Comprehensive survey on distance/similarity measures between probability density functions" Sung-Hyuk Cha, International Journal of Mathematical Models and

- & Comparisons are distances, divergences , similarities, inner products, fidelities ...
- 60ish kinds of comparisons grouped into 10 families
- A worry: Subsampled distributions with very heavy tails

A plenitude of distances

Shannon tried to slow things down in 1956: Rank-turbuleno

"The bandwagon" 🗗 Claude E Shannon,

IRE Transactions on Information Theory, 2, 3,

- 🚳 "Information theory has ... become something of a scientific
- While ... information theory is indeed a valuable tool ... [it] is certainly no panacea for the communication engineer or ... for anyone else.
- A few first rate research papers are preferable to a large number that are poorly conceived or half-finished.'

The PoCSverse Allotaxonometry

Allotaxonometry 13 of 70

divergence

Probability-

Explorations

Nutshell

Rank-turbuleno

Probabilityturbulence divergence Explorations

- We want two main things:
 - 1. A measure of difference between systems 2. A way of sorting which
 - types/species/words contribute to that difference
- For sorting, many comparisons give the same ordering.
- A few basic building blocks:
 - $|P_i Q_i|$ (dominant) $max(P_i, Q_i)$
 - $min(P_i, Q_i)$
 - P_iQ_i
 - $|P_i^{1/2} Q_i^{1/2}|$ (Hellinger)

Table 1. Lp Minkov	vski family		The PoCSverse
1. Euclidean L ₂	$d_{Eac} = \sqrt{\sum_{i=1}^{d} P_i - Q_i ^2}$	(1)	Allotaxonometry 17 of 70
2. City block L ₁	$d_{CB} = \sum_{i=1}^{d} P_i - Q_i $	(2)	A plenitude of distances
3. Minkowski L _p	$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} P_i - Q_i ^p}$	(3)	Rank-turbulence divergence
4 Chebyshey I	$d_{Cheb} = \max P_i - Q_i $	(4)	D. J. 1995

Table 2. L₁ family $\sum_{i} |P_{i} - Q_{i}|$ (5) $\sum_{i=1}^{d} (P_i + Q_i)$

Probability

The PoCSverse

Allotaxonometry 16 of 70

A plenitude of

Rank-turbulenc

turbulence diver

divergence

Probability

Exploration

Nutshell

6. Gower	$d_{gone} = \frac{1}{d} \sum_{i=1}^{d} \frac{ P_i - Q_i }{R_i}$	(6)
	$= \frac{1}{d} \sum_{i=1}^{d} P_i - Q_i $	(7)
7. Soergel	$d_{sg} = \frac{\sum_{i=1}^{d} P_i - Q_i }{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)
8. Kulczynski d	$d_{kal} = \frac{\sum_{i=1}^{d} P_i - Q_i }{\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)
9. Canberra	$d_{Com} = \sum_{i=1}^{d} \frac{ P_i - Q_i }{P_i + Q_i}$	(10)
10. Lorentzian	$d_{lor} = \sum_{i=1}^{d} \ln(1 + P_i - Q_i)$	(11)
	Intersectoin (13), Wave Hed Ruzicka (21), Tanimoto (23), e	

Table 1. L _v Minkowski fan	nily	1	The PoCSve
1. Euclidean L_2 $d_{E_{BR}}$	$= \sqrt{\sum_{i=1}^{d} P_i - Q_i ^2}$	(1)	Allotaxonon 18 of 70
2. City block L_1 d_{CI}	$_{i} = \sum_{i=1}^{d} P_{i} - Q_{i} $	(2)	A plenitude distances
3. Minkowski L _p d _{Mk}	$= p \sum_{i=1}^{d} P_i - Q_i ^p$	(3)	Rank-turbul divergence
4. Chebyshev L_{∞} d_{Ch}	$= \max_{i} P_{i} - Q_{i} $	(4)	Probability-

Table 2. L_1 family		
5. Sørensen	$d_{sor} = \frac{\sum_{i=1}^{d} P_i - Q_i }{\sum_{i=1}^{d} (P_i + Q_i)}$	(5)

	Explorations
(5)	Nutshell
	References

Allotaxonometry

Rank-turbuleno

6. Gower	$d_{gow} = \frac{1}{d} \sum_{i=1}^{d} \frac{ P_i - Q_i }{R_i}$	(6)			
	$= \frac{1}{d} \sum_{i=1}^{d} P_i - \underline{Q}_i $	(7)			
7. Soergel	$d_{ig} = \frac{\sum_{i=1}^{d} P_i - Q_i }{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)			
8. Kulczynski d	$d_{kat} = \frac{\displaystyle\sum_{i=1}^{d} P_i - Q_i }{\displaystyle\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)			
9. Canberra	$d_{Com} = \sum_{i=1}^{d} \frac{ P_i - Q_i }{P_i + Q_i}$	(10)			
10. Lorentzian	$d_{lor} = \sum_{i=1}^{d} \ln(1 + P_i - Q_i)$	(11)			
* L₁ family ⊃ {Intersectoin (13), Wave Hedges (15), Czekanowski (16), Ruzicka (21), Tanimoto (23), etc}.					

Allotaxonometry 15 of 70

Rank-turbuleno

Probabilityturbulence divergen

Exploration

🚵 Information theoretic sortings are more opaque

No tunability

Quite the festival:

Allotaxonometry

A plenitude of

12 of 70

				∑mm(7, (2)		
Table L. L. Minke	and family			d _m = 1 - s _m = $\frac{m}{\sum (r + \alpha)}$	(29)	
I. Euclidean L ₂	$d_{nu} = \sum_{i} (P_i - Q_i)^n$	(1)	15. Kulczynski s	Σ(F+0) , Σ==(F,0)	-	27 Manuto G. en
2. City block L:	$d_{ca} = \sum_{i=1}^{d} P_i - Q_i $	(2)		$A_{ab} = \frac{1}{A_{ab}} = \frac{\frac{2a}{2a}}{\sum_{i} P_i - Q_i }$	(20)	- P.W.
). Minkewski $L_{\rm p}$	$d_m = \sum_{i=1}^{m} P_i = Q_i P_i$	(3)	16: Rusicka	\$ min(F, g)	(21)	29. Squand-cloud
6. Chebyshev L.	$d_{ctot} = \max\{P_i - Q_i\}$	(4)		$\sum man(P,Q)$		z _w =1-d _w z _w =2∑ F (2)
		_	17. Tani-	\$1.00 miles		
Eable 2. L, family 5. Sauceura		_		\$1.50.5mm.01	(22)	Table 6. Sound J. Smily or v' Smil
	4_ = \frac{\sum_{i} \range i - \alpha \cdot }{\sum_{i} \range i - \alpha \cdot}	(5)		Emerge-metan		29. Squared $d_{sp} = \sum_{i=1}^{n} (P_i - Q_i)$
. Gover	7	_		Σman(7,(2)	(29)	30. Passon χ $d_{\chi}(P,Q) = \sum_{i=1}^{k} \frac{P}{2}$
	$d_{p,m} = \frac{1}{d} \sum_{i} \frac{(R-Q_i)}{R_i}$	(6)	Table 4, Issue Po	the first		31. Nayman χ' $d_{\gamma}(P,Q) = \sum_{i} \frac{Q^{i}}{2}$
. Soonad	$-\frac{1}{d}\sum_{i=1}^{d} P_i-Q_i $	O	18. Inner Product		(21)	32. Squared χ^{i} $d_{i_{0}m} = \sum_{i=1}^{m} \frac{d_{i}(P - q_{i})}{P + i}$
. seego	$d_{ij} = \frac{\sum_{j} P_j - Q_j}{\sum_{j \in \mathcal{M}} P_j Q_j}$	(9)	19 Harmonic mean	$s_{acc} = 2\sum_{i=1}^{n} \frac{P(Q)}{P_i + Q}$	(25)	33. Probabilistic Symmetric 2 ^t $d_{res} = 2\sum_{p} \frac{dp}{p}$
k Kulayesii d	\$15-01	-	20 Cosine	∑ eu.	(20)	34. Divergence $d_{m} = 2\sum_{i} \frac{d_{i}(P_{i}-1)}{d_{i}(P_{i}-1)}$
	Δ Σπάσ(P, Q)	(%)		Er Ec	(44)	15. Clark
A Camborra	$d_{nn} = \sum_{i} \frac{ P-Q_i }{ P+Q_i }$	(30)	21 Kansar- Hamshrook	Σng	$\overline{}$	St. Address Symmetric 2 ³ $d_{corn} = \sum_{i} \frac{(P_i - q_i)^2}{2}$
H. Lorentoian	4 \$100+18-00	(11)	(PCE)	"- " <u>Σπ' - Σα' - Σπα</u>	(27)	Symmetric y * Secured L. family > Occupal (29)
L family > 1	Interaction (17), Wave Holl	HK (15)	22: Jacourd	Σ40	_	
	Ruricka (21), Tanimoto (27), e	6).		*= *\frac{\bar{\Sigma} \chi - \bar{\Sigma} \chi a}{\Delta \chi - \bar{\Sigma} \chi a}	(28)	Table 7. Shannon's contrapy family 37. Kullback— $d_{ac} = \sum_{i} P_i \ln \frac{P_i}{P_i}$ Lubber $d_{ac} = \sum_{i} P_i \ln \frac{P_i}{P_i}$
Table 3. Intersection		_				No. 3 of Section 1
	$\kappa_m = \sum_{i=1}^n \min(P_i,Q_i)$	(12)		$\sum_{i=1-n_{i}} \frac{\sum_{j} (r_{j} - g_{j})^{i}}{\sum_{j} r_{j}^{i} + \sum_{j} g_{j}^{i} - \sum_{j} r_{i}g_{j}}$	(20)	$Z_{r} = \sum_{i=1}^{r} (P_{i}^{r} - Q_{i}^{r})^{2}$
	$\frac{1}{2} = 1 - \epsilon_n = \frac{1}{2} \sum_{i=1}^{n} P_i - Q_i$	(13)	23. Dice	15/10	-	19. K divergence $d_{ass} = \sum_{i} P_i \ln \frac{1}{P_i}$ 60. Tomore
12. Ware Hedges	$d_{mn} = \sum_{i=1}^{n} (1 - \frac{\min(P_i,Q_i)}{\max(P_i,Q_i)})$	(24)			(40)	$d_{i_0} = \sum_{i=1}^{i_0} \left(\frac{2P_i}{P_i + Q_i} \right) + Q_i$
	$-\sum_{i}\frac{(P_i-Q_i)}{\max(P_i,Q_i)}$	(15)		Y1701	_	41. Joneso-Shannon
13. Crekanowski	$2\sum_{t\in \mathcal{C}}\min(T_t(t))$	(16)	-	$\sum_{i=1}^{n} P_i^{i} = \sum_{i=1}^{n} Q_i^{i}$	(30)	$d_{in} = \frac{1}{2} \left(\sum_{j \in \mathcal{F}_i} \operatorname{tr} \left(\frac{2F_j}{F_j + Q_j} \right) \sum_{j \in \mathcal{F}_i} Q_j \operatorname{tr} \left(\frac{2}{F_j} \right) \right)$
	∑(₹+Q)		Table 5 States	Samily or Squared short Samily	=	42. Jones difference $d_{\infty} = \sum_{i} \left[\frac{P_i \ln P_i + Q_i \ln Q_i}{2} - \left(\frac{P_i + Q_i}{2} \right) \right]$
	$\sum_{i_{m}=1}^{m} x_{i_{m}} = \frac{\sum_{i}^{m} x_{i} - q_{i,1}}{\sum_{i}^{m} x_{i}^{m} + q_{i,1}}$	(17)	24 Fiddity	4 \$\sum_{100} \overline{P100}	(32)	2 2 2
	Dr.+01		25. Whattacharyy	$d_a = -\ln \sum_i \sqrt{2/2}$	(33)	

			∑mm(7, (2)				
_	_		$d_{max} = 1 - a_{max} = \frac{22}{\sum_{i=1}^{n} (P_i + Q_i)}$	(29)			
1.9-91	(1)	15. Kalczyteki r	$\sigma_{\rm min} = \frac{1}{d_{\rm min}} - \frac{\sum_{i} \min(\mathcal{T}_i(Q))}{\sum_{i} \min(\mathcal{T}_i(Q))}$	(28)	27. Matusita	4 EF-E1	(76)
P, -Q,	(2)	th Resida	214-61			- I1 <u>> Piz</u>	(37)
Pt-Qt !"	(2)	in name	$s_{min} = \frac{\sum_{ij} \min(P_i,Q_j)}{\sum_{ij} \max(P_i,Q_j)}$	(21)	29. Squand-chee	4	(75)
2-01	(4)	17. Tani-	∑nac, 25 ∑n + ∑n - 2∑mm, n	_	14-14-	$a_{ap} = 2\sum_{i} \sqrt{2(2i-1)}$	(29)
g i		moto	2 π + Σα - Σ mag α			L, family or g' family	Ξ
01	(5)		∑man(t,g) - min(t,g))		29. Squared Euclidean	$d_{\rm op} = \sum_{i=1}^{d} (P_i - Q_i)^2$	(40)
	=		Σmm(7,Q)	(29)	30. Postson g	$d_{r}(P,Q) = \sum_{i=1}^{r} \frac{(P_{i} - Q_{i})^{r}}{Q_{i}}$	(61
R-Q:	069	Table 4, Inner Po	sduct family	=	31. Nayssas x	$d_{n}(P,Q) = \sum_{i=1}^{n} \frac{(P-Q)^{i}}{P}$	(42
e.	(7)	18. Inner Product	$a_{\mu\nu} = P \bullet Q = \sum_{i} P_i Q_i$	(21)	32. Squared x*	$d_{nm} = \sum_{i} \frac{(P - Q)^{i}}{P + Q}$	(40
1.00	(%)	19 Harmonic moss	$s_{acc} = 2\sum_{i=1}^{n} \frac{P(Q)}{P_i + Q_i}$	(25)	33. Probabilistic Symmetric g ²	$d_{\text{cross}} = 2\sum_{i}\frac{(P_i - Q_i)^2}{P_i + Q_i}$	(64
-0.1	_	20. Cosine	Σ rsz.	(2n)	34 Divergence	$d_{m} = 2\sum_{i} \frac{(P - Q)^{i}}{(P + Q)^{i}}$	(45
5.01	(9)		Er Eu	,,	35. Clark	4 - 5(12-91)	100
0	(30)	21 Kamar- Hamebrook	Σna		76. Additio	4 - 5 (8-0) (8+0)	607
18-80	(11)	(PCE)	- <u>\$1.50.500</u>	(27)	* Squared L _i fam	By > (haccard (29), Dice (31))	(4.
, Wave State	gas (15);	22: Jacourd	Σng	(2f)	Table 7. Shannon	s entropy family	=
			$\sum_{i} p_i' + \sum_{i} q_i' - \sum_{i} p_i q_i$	(28)	37 Kullback- Labler	$d_{ac} = \sum_{i} P_i \ln \frac{P_i}{Q_i}$	(4
1.00	(12)		∑(r, -q,)'	- on	38. Jullinys	$d_{z} = \sum_{i} (P_i - Q_i) \ln \frac{P_i}{Q_i}$	(4
De-ai	(13)	23. Dice	$\sum P_i' + \sum Q_i' - \sum P_i$	e i	39 K divergence	$d_{Adv.} = \sum_{i=1}^{d} P_i \ln \frac{2P_i}{P_i + Q_i}$	(5
min(P,Q)	(14)			(40)	40. Topsau	$\sum_{i=1}^{n} \left[e_{i} \left(\frac{2\pi}{F+G} \right) - g_{i} \left(\frac{2g}{F+G} \right) \right]$	(5
20	(15)	-	\$1001		41 Answe-Shann	66	-
Ø.00	ne.	4	-1	(31)		$\frac{2F_{i}}{v + Q_{i}}$ $+ \sum_{i} Q_{i} \ln \left(\frac{2Q_{i}}{E + Q_{i}} \right)$	(5
+0.)	Comp	Table 5 Golden	Smily or Squared-short Smily	=	42 Januar differs	$\frac{e_{Q}}{2} = \frac{e_{Q}}{2} \cdot \left(\frac{P_{1} + Q_{1}}{2} \right) \cdot \left(\frac{P_{1} + Q_{2}}{2} \right)$	(S
-0.1	(17)	24. Fiddity	4 \$ P12	(32)	12.2	2 (2)(2)	- 14
+(2)		25. Shattacharvo		_			

unananan n.,	AC - 42 A - 62	(1)	D. Kalczyloki i			27. Maturita	4-55-50
Sty block L	$d_{co} = \sum_{i=1}^{r} P_i - Q_i $	(2)		$r_{ac} = \frac{1}{d_{ac}} = \frac{2l}{\sum_{ij} p_i - q_i}$	(29)		- 1-1 <u>5 Fill</u>
diskowski $L_{\rm p}$	$d_{\rm col} = \sqrt{\sum_{ij} (P_i - Q_i)^2}$	(3)	16. Rusicka	*_ = \frac{\sum_{min}(P,Q)}{2}	(21)	29. Squared-cheed	4 \$(F-F2)
Subyshev L _n	$d_{chot} = \max\{P_i - Q_i\}$	(4)	17. Tani-	\(\sum_{max}(\tau_{i}(t)) \) \(\sum_{i} + \sum_{i}(t) - 2 \sum_{max}(\tau_{i}(t)) \)	-	14-14-	r_ = 2∑ - 1
de L. J., family		_	moto	$d_{max} = \frac{\sum_{i} x_i + \sum_{i} y_i - \sum_{i} \min(x_i, y_i)}{\sum_{i} x_i + \sum_{i} y_i - \sum_{i} \min(x_i, y_i)}$	(22)	TAL CO.	L family or y' family
orono	∑e-a:	(5)				29. Squared Enclided	$d_{ij} = \sum_{i} (P_i - Q_i)^i$
	$\sum_{i}(P_i + Q_i)$			$= \frac{\sum_{(max(P,Q)) - min(P,Q)}}{\sum_{(max(P,Q))}}$	(23)	30. Paston y	4,17,00-517-01
iower	$d_{p,m} = \frac{1}{d} \sum_{i} \frac{ P_i - Q_i }{R}$	00		2 man(2,42)	_	31. November V	
	- <u>4</u> ∑in-a:	a	Table 4, Inner P 18, Inner Produc				$d_{s}(P,Q) = \sum_{i=1}^{k} \frac{(P-Q)^{T}}{P_{i}}$
oergel		4.0		1,-P*Q-270	(21)	32. Squared x ²	$d_{npm} = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i + Q_i}$
	$d_{ij} = \frac{\sum_{i} P_i - Q_i}{\sum_{max} P_i Q_i}$	(9)	19 Harmonic mean	$s_{max}=2\sum_{i=1}^{n}\frac{P(Q)}{P_{i}+Q_{i}}$	(25)	33. Probabilistic Symmetric g ²	$d_{\text{cross}} = 2\sum_{i} \frac{(P_i - Q_i)^2}{P_i + Q_i}$
akayeda d	Σ(e-a)	_	20. Cosine	Σna.	(26)	34. Divergence	$d_m = 2\sum_{i=0}^{d} \frac{(P_i - Q_i)^i}{(P_i + Q_i)^i}$
	$d_{i,j} = \frac{2\pi}{\sum \min(P_i,Q_i)}$	(9)		Er Eu	100	35. Clark	((a-a))
anborra	$d_{ma} = \sum_{i=1}^{n} \frac{ P - Q_i }{P_i + Q_i}$	(20)	21 Kamar- Hambrook	Σπα	\neg	76. Addriso	4
Lorentrian	$d_{ab} = \sum_{i} \ln(1 + P_i - Q_i)$	(11)	(PCE)	*~ *\frac{\sum_{e'} - \sum_{e'}}{\sum_{e'} - \sum_{e'}} = \frac{\sum_{e'}}{\sum_{e'}} = \frac_{e'} = \frac{\sum_{e'}}{\sum_{e'}} = \frac_{e'} = \frac_{e'} =	(27)	* Squared L. Sam	N ⊃ theoret (24) Dice (31)
	ntersectoin (17), Wave Hel		22. Jacourd	Σπα	_	Table 7, Shannon	
	Ruricka (21), Tanimoto (23), e	61.		"= " <u>Σ</u> ε' + <u>Σ</u> α' - <u>Σ</u> εα	(28)	37. Kullback- Lebler	$d_{m} = \sum_{i} P \ln \frac{P_{i}}{Q_{i}}$
Me 3. Intersection Intersection		_		Yur-a/	_	18. Jeffleys	$d_1 = \sum_{i} (P_i - Q_i) \ln \frac{P_i}{Q_i}$
	$a_m = \sum_i \min(P_i,Q_i)$	(12)		1. 1-1. 2. 2. 2. 2. 2. 2. 2. 2. 2. 2. 2. 2. 2.	(20)	29 K diversessor	
	$-a = 1 - s_n = \frac{1}{2} \sum_{i=1}^{n} P_i - Q_i$	(13)	23. Dice	1 <u>\$</u> 70	-	40. Teasur	$d_{mn} = \sum_{i} P_i \ln \frac{1}{P_i + Q_i}$
Ware Hedges	$d_{mn} = \sum_{i=1}^n (1 - \frac{\min(P,Q)}{\max(P,Q)})$	(14)			(40)		$\sum_{i} \left(e^{i\phi} \left(\frac{2E_i}{E+Q_i} \right) \cdot Q_i \right) \left(\frac{2Q_i}{E+Q_i} \right)$
	- \(\sum_{\text{max}(P,Q)} \)	(15)		Σ(r,-g)'	_	41 Joneso Shann	as .
Crekanowski	2∑mm(₹,@)	OB.		$t_{m-1} = t_{m-1} = \frac{1}{\sum_{i} p_i^* + \sum_{i} p_i^*}$	(31)	$d_{in} = \frac{1}{2} \left(\sum_{j=1}^{n} P_j \ln \left(\frac{1}{p} \right) \right)$	$\frac{2P_{i}}{1+Q_{i}}$ $+ \sum_{i} Q_{i} \ln \left(\frac{2Q_{i}}{P_{i}+Q_{i}} \right)$
	∑(P+Q)	(10)			_	42 Janear difficu	0 40 (E+0), (E+0
	∑ r-q.	(17)	24 Fiddity	family or Squared chord family $s_{xx} = \sum_{i} \overline{P(i)}$	(32)	d = 2	2)(2)
	$\sum_{i}(x+g_i)$	0.0	25. Whattacharys		(33)		
			26 Hallinear				
				$d_{\alpha} = \sum_{i} (\sqrt{p_i} - \sqrt{p_i})^{\alpha}$	(34)		
				=2 11-∑ 47(£)	(35)		

$$H(P) = \langle \log_2 \frac{1}{p_\tau} \rangle = \sum_{\tau \in R_{1,2;\alpha}} p_\tau \log_2 \frac{1}{p_\tau} \tag{1} \label{eq:1}$$

& Kullback-Liebler (KL) divergence:

$$\begin{split} &D^{\mathrm{KL}}\left(P_{2}\mid\mid P_{1}\right)=\left\langle\log_{2}\frac{1}{p_{2,\tau}}-\log_{2}\frac{1}{p_{1,\tau}}\right\rangle_{P_{2}}\\ &=\sum_{\tau\in R_{1,2;\alpha}}p_{2,\tau}\left[\log_{2}\frac{1}{p_{2,\tau}}-\log_{2}\frac{1}{p_{1,\tau}}\right]\\ &=\sum_{\tau\in R_{1,2;\alpha}}p_{2,\tau}\log_{2}\frac{p_{1,\tau}}{p_{2,\tau}}. \end{split} \tag{2}$$

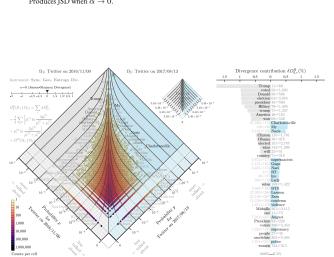
- & Problem: If just one component type in system 2 is not present in system 1, KL divergence = ∞ .
- Solution: If we can't compare a spork and a platypus directly, we create a fictional spork-platypus hybrid.
- New problem: Re-read solution.
- Sensen-Shannon divergence (JSD): [9, 7, 13, 1]

$$\begin{split} & D^{\text{IS}}\left(P_{1} \parallel P_{2}\right) \\ & = \frac{1}{2} D^{\text{KL}}\left(P_{1} \parallel \frac{1}{2}\left[P_{1} + P_{2}\right]\right) + \frac{1}{2} D^{\text{KL}}\left(P_{2} \parallel \frac{1}{2}\left[P_{1} + P_{2}\right]\right) \\ & = \frac{1}{2} \sum_{\tau \in R_{1,2;\alpha}} \left(p_{1,\tau} \log_{2} \frac{p_{1,\tau}}{\frac{1}{2}\left[p_{1,\tau} + p_{2,\tau}\right]} + p_{2,\tau} \log_{2} \frac{p_{2,\tau}}{\frac{1}{2}\left[p_{1,\tau} + p_{2,\tau}\right]}\right). \end{split} \tag{3}$$

- Involving a third intermediate averaged system means JSD is now finite: $0 \le D^{JS}(P_1 || P_2) \le 1.$
- & Generalized entropy divergence: [2]

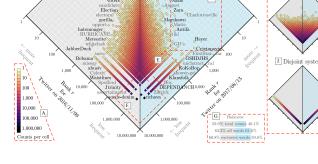
$$\begin{split} D_{\alpha}^{\text{AS2}}\left(P_{1} \parallel P_{2}\right) &= \\ \frac{1}{\alpha(\alpha-1)} \sum_{\tau \in R_{1,2;\alpha}} \left[\left(p_{\tau,1}^{1-\alpha} + p_{\tau,2}^{1-\alpha}\right) \left(\frac{p_{\tau,1} + p_{\tau,2}}{2}\right)^{\alpha} - \left(p_{\tau,1} + p_{\tau,2}\right) \right] \end{split} \tag{4}$$

Produces ISD when $\alpha \to 0$.



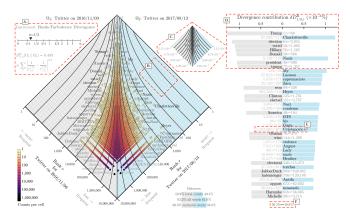
Probability-

Allotaxonometry 20 of 70



Rank-turbulence histogram

Ω₁: Twitter on 2016/11/09 B.



Desirable rank-turbulence divergence features:

- 1. Rank-based.
- 2. Symmetric.
- 3. Semi-positive: $D_{\alpha}^{\mathbb{R}}(\Omega_1 \mid\mid \Omega_2) \geq 0$.
- 4. Linearly separable, for interpretability.
- 5. Subsystem applicable: Ranked lists of any principled subset may be equally well compared (e.g., hashtags on Twitter, stock prices of a certain sector, etc.).
- 6. Turbulence-handling: Suited for systems with rank-ordered component size distribution that are heavy-tailed.
- 7. Scalable: Allow for sensible comparisons across system sizes.
- 8. Tunable.
- 9. Story-finding: Features 1-8 combine to show which component types are most 'important'

Some good things about ranks:

- Working with ranks is intuitive
- Affords some powerful statistics (e.g., Spearman's rank correlation coefficient)
- & Can be used to generalize beyond systems with probabilities

A start:

H. Identical systems:

I. Randomized systems:

$$\left| \frac{1}{r_{\tau,1}} - \frac{1}{r_{\tau,2}} \right|. \tag{5}$$

- Inverse of rank gives an increasing measure of 'importance'
- A High rank means closer to rank 1
- We assign tied ranks for components of equal 'size'
- A Issue: Biases toward high rank components

We introduce a tuning parameter:

$$\left| \frac{1}{\left[r_{\tau,1} \right]^{\alpha}} - \frac{1}{\left[r_{\tau,2} \right]^{\alpha}} \right|^{1/\alpha}. \tag{6}$$

- $As \alpha \to 0$, high ranked components are increasingly dampened
- For words in texts, for example, the weight of common words and rare words move increasingly closer together.
- $As \alpha \to \infty$, high rank components will dominate.
- For texts, the contributions of rare words will vanish.

Allotaxonometry

A plenitude of

Rank-turbulence

divergence

Exploration

24 of 70

 $\left| \frac{1}{\left[r_{\sigma,1} \right]^{\alpha}} - \frac{1}{\left[r_{\sigma,2} \right]^{\alpha}} \right|^{1/\alpha}.$

The leading order term is:

 $\left(1-\delta_{r_{\tau,1}r_{\tau,2}}\right)\alpha^{1/\alpha}\left|\ln\frac{r_{\tau,1}}{r_{-\beta}}\right|^{1/\alpha},$ (7)

which heads toward ∞ as $\alpha \to 0$.

- Oops.
- But the insides look nutritious:

$$\ln \frac{r_{\tau,1}}{r}$$

is a nicely interpretable log-ratio of ranks.

The PoCSverse Allotaxonometry 25 of 70 A plenitude of

Rank-turbulence divergence

Probability turbulence diverge Exploration

Nutshell

Allotaxonometry 26 of 70

Rank-turbulence divergence

Nutshell

Allotaxonometry

Rank-turbulence

divergence

Some reworking:

$$\delta D_{\alpha,\tau}^{\mathbb{R}}(R_1 \parallel R_2) \propto \frac{\alpha+1}{\alpha} \left| \frac{1}{\left[r_{\tau,1}\right]^{\alpha}} - \frac{1}{\left[r_{\tau,2}\right]^{\alpha}} \right|^{1/(\alpha+1)}. \quad (8)$$

- & Keeps the core structure.
- & Large α limit remains the same.
- \rightarrow 0 limit now returns log-ratio of ranks.
- \aleph Next: Sum over τ to get divergence.
- Still have an option for normalization.

Rank-turbulence divergence:

$$D_{\alpha}^{\mathrm{R}}(R_1 \parallel R_2) = \frac{1}{\mathcal{N}_{1,2;\alpha}} \sum_{\tau \in R_{1,2;\alpha}} \delta D_{\alpha,\tau}^{\mathrm{R}}(R_1 \parallel R_2) \quad \ \ (9)$$

Normalization:

- Take a data-driven rather than analytic approach to determining $\mathcal{N}_{1,2:\alpha}$.
- $\ensuremath{\mathfrak{S}}$ Compute $\mathcal{N}_{1,2;\alpha}$ by taking the two systems to be disjoint while maintaining their underlying Zipf distributions.
- \Re Ensures: $0 \le D_{\alpha}^{\mathbb{R}}(R_1 \parallel R_2) \le 1$
- Limits of 0 and 1 correspond to the two systems having identical and disjoint Zipf distributions.

Rank-turbulence divergence:

Summing over all types, dividing by a normalization prefactor $\mathcal{N}_{1,2;\alpha}$ we have our prototype:

$$D_{\alpha}^{\mathbb{R}}(R_1 \parallel R_2) = \frac{1}{\mathcal{N}_{1,2;\alpha}} \frac{\alpha+1}{\alpha} \sum_{\tau \in R_{1,2;\alpha}} \left| \frac{1}{\left[r_{\tau,1}\right]^{\alpha}} - \frac{1}{\left[r_{\tau,2}\right]^{\alpha}} \right|^{1/(\alpha+1)}. \tag{10}$$

Allotaxonometry 29 of 70

Rank-turbulence divergence

Allotaxonometry

30 of 70

& Similarly, $\Omega^{(2)}$'s merged ranking will have all of $\Omega^{(1)}$'s types in last place with rank $r = N_2 + \frac{1}{2}N_1$.

 \mathbb{A} Iif the Zipf distributions are disjoint, then in $\Omega^{(1)}$'s merged

ranking, the rank of all $\Omega^{(2)}$ types will be $r = N_1 + \frac{1}{2}N_2$,

where N_1 and N_2 are the number of distinct types in each

The normalization is then:

General normalization:

$$\begin{split} \mathcal{N}_{1,2;\alpha} &= \frac{\alpha+1}{\alpha} \sum_{\tau \in R_1} \left| \frac{1}{\left[r_{\tau,1}\right]^{\alpha}} - \frac{1}{\left[N_1 + \frac{1}{2}N_2\right]^{\alpha}} \right|^{1/(\alpha+1)} \\ &+ \frac{\alpha+1}{\alpha} \sum_{\tau \in R_2} \left| \frac{1}{\left[N_2 + \frac{1}{2}N_1\right]^{\alpha}} - \frac{1}{\left[r_{\tau,2}\right]^{\alpha}} \right|^{1/(\alpha+1)} \end{split}$$
 (11

Limit of $\alpha \to 0$:

$$D_0^{\rm R}(R_1 \, \| \, R_2) = \sum_{\tau \in R_{1,2;\alpha}} \delta D_{0,\tau}^{\rm R} = \frac{1}{\mathcal{N}_{1,2;0}} \sum_{\tau \in R_{1,2;\alpha}} \left| \ln \frac{r_{\tau,1}}{r_{\tau,2}} \right|, \tag{12}$$

where

$$\mathcal{N}_{1,2;0} = \sum_{\tau \in R_1} \left| \ln \frac{r_{\tau,1}}{N_1 + \frac{1}{2}N_2} \right| + \sum_{\tau \in R_2} \left| \ln \frac{r_{\tau,2}}{\frac{1}{2}N_1 + N_2} \right|.$$
 (13)

Largest rank ratios dominate.

Limit of $\alpha \to \infty$:

$$\begin{split} &D_{\infty}^{\mathrm{R}}(R_1 \parallel R_2) = \sum_{\tau \in R_{1,2;\alpha}} \delta D_{\infty,\tau}^{\mathrm{R}} \\ &= \frac{1}{\mathcal{N}_{1,2;\infty}} \sum_{\tau \in R_{1,2;\alpha}} \left(1 - \delta_{r_{\tau,1} r_{\tau,2}}\right) \max_{\tau} \left\{\frac{1}{r_{\tau,1}}, \frac{1}{r_{\tau,2}}\right\}. \end{split} \tag{14}$$

where

$$\mathcal{N}_{1,2;\infty} = \sum_{\tau \in R_1} \frac{1}{r_{\tau,1}} + \sum_{\tau \in R_2} \frac{1}{r_{\tau,2}}.$$
 (15)

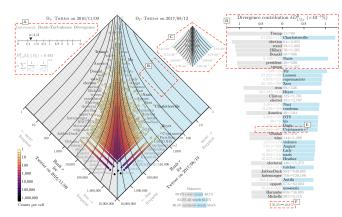
Highest ranks dominate.

Allotaxonometry 31 of 70

A plenitude of

Rank-turbuleno divergence

Probability-Exploration



Allotaxonometry 32 of 70

Rank-turbulence divergence

Allotaxonometry

33 of 70

Explorations

Probability-turbulence divergence:

$$D_{\alpha}^{\mathrm{P}}(P_1 \parallel P_2) = \frac{1}{\mathcal{N}_{1,2;\alpha}^{\mathrm{P}}} \frac{\alpha+1}{\alpha} \sum_{\tau \in R_{1,2;\alpha}} \left| \left[p_{\tau,1} \right]^{\alpha} - \left[p_{\tau,2} \right]^{\alpha} \right|^{1/(\alpha+1)}. \tag{16}$$

- & For the unnormalized version ($\mathcal{N}_{1,2;\alpha}^{P}$ =1), some troubles return with 0 probabilities and $\alpha \to 0$.
- Weep not: $\mathcal{N}_{1,2;\alpha}^{P}$ will save the day.

Normalization:

With no matching types, the probability of a type present in one system is zero in the other, and the sum can be split between the

$$\mathcal{N}_{1,2;\alpha}^{p} = \frac{\alpha+1}{\alpha} \sum_{\tau \in R_1} \left[p_{\tau,1} \right]^{\alpha/(\alpha+1)} + \frac{\alpha+1}{\alpha} \sum_{\tau \in R_2} \left[p_{\tau,2} \right]^{\alpha/(\alpha+1)} \tag{17}$$

Allotaxonometr

Rank-turbulence

Probability-turbulence

divergence

Limit of $\alpha = 0$ for probability-turbulence divergence

 \Leftrightarrow if both $p_{\tau,1} > 0$ and $p_{\tau,2} > 0$ then

$$\lim_{\alpha \rightarrow 0} \frac{\alpha+1}{\alpha} \; \Big| \; \left[\; p_{\tau,1} \right]^{\alpha} - \left[\; p_{\tau,2} \right]^{\alpha} \; \Big|^{1/(\alpha+1)} = \left| \ln \frac{p_{\tau,2}}{p_{\tau,1}} \right|. \; (18)$$

 \Re But if $p_{\tau,1} = 0$ or $p_{\tau,2} = 0$, limit diverges as $1/\alpha$.

Limit of α =0 for probability-turbulence divergence

Normalization:

$$\mathcal{N}_{1,2;\alpha}^{P} \to \frac{1}{\alpha} (N_1 + N_2).$$
 (19)

Because the normalization also diverges as $1/\alpha$, the divergence will be zero when there are no exclusive types and non-zero when there are exclusive types.

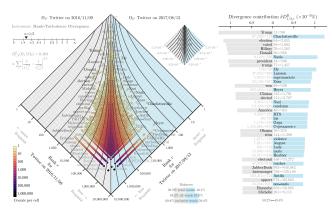
Type contribution ordering for the limit of α =0

- A In terms of contribution to the divergence score, all exclusive types supply a weight of $1/(N_1 + N_2)$. We can order them by preserving their ordering as $\alpha \to 0$, which amounts to ordering by descending probability in the system in which they appear.
- And while types that appear in both systems make no contribution to $D_0^{\rm p}(P_1 \, \| \, P_2),$ we can still order them according to the log ratio of their probabilities.
- The overall ordering of types by divergence contribution for α =0 is then: (1) exclusive types by descending probability and then (2) types appearing in both systems by descending log

Allotaxonometry 40 of 70 A plenitude of

Rank-turbuler Probability-turbulence

divergence Explorations



A plenitude of

divergence

divergence

Probability-turbulenc divergence

Limit of $\alpha = \infty$ for probability-turbulence divergence

$$D_{\infty}^{\mathrm{P}}(P_1 \, \| \, P_2) = \frac{1}{2} \sum_{\tau \in R_{1,2;\infty}} \left(1 - \delta_{p_{\tau,1},p_{\tau,2}}\right) \max \left(p_{\tau,1},p_{\tau,2}\right) \tag{2}$$

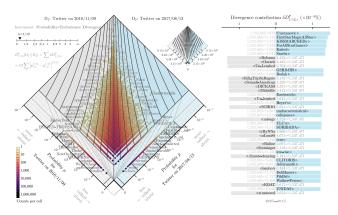
where

$$\mathcal{N}_{1,2;\infty}^p = \sum_{\tau \in R_{1,2;\infty}} \left(\ p_{\tau,1} + p_{\tau,2} \ \right) = 1 + 1 = 2. \tag{22}$$

Allotaxonometr

Rank-turbulenc

Probability-turbulenc divergence Exploration



Combine these cases into a single expression:

$$D_0^{\rm P}(P_1 \, \| \, P_2) = \frac{1}{(N_1 + N_2)} \sum_{\tau \in R_{1,2;0}} \left(\delta_{p_{\tau,1},0} + \delta_{0,p_{\tau,2}} \right). \tag{20}$$

- $\text{ The term } \left(\delta_{p_{\tau,1},0}+\delta_{0,p_{\tau,2}}\right) \text{ returns 1 if either } p_{\tau,1}=0 \text{ or } \\ p_{\tau,2}=0 \text{, and 0 otherwise when both } p_{\tau,1}>0 \text{ and } p_{\tau,2}>0.$
- Ratio of types that are exclusive to one system relative to the total possible such types,

Connections for PTD:



 $\alpha = 1/2$: Hellinger distance [8] and Mautusita distance [11].

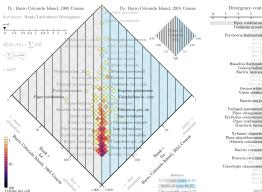
 $\alpha = 1$: Many including all $L^{(p)}$ -norm type constructions.

 $\alpha = \infty$: Motyka distance [3].

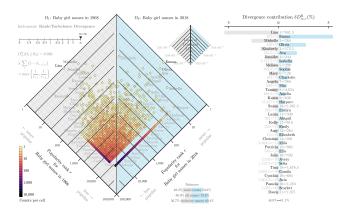
Allotaxonometry 42 of 70

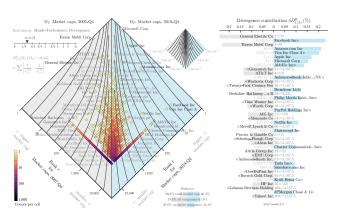
A plenitude of Rank-turbuleno

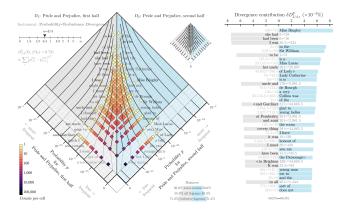
Probability-turbulenc divergence Exploration

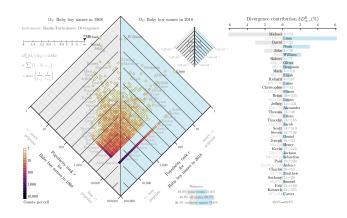












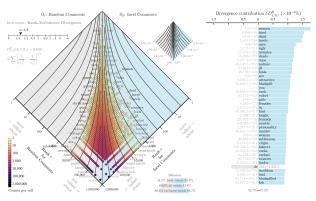


FIG. 8. Rank-turbulence divergence allotaxonograph [34] of word rank distributions in the incel vs random comment corpora. The rank-rank histogram on the left shows the density of words by their rank in the incel comments corpus. Words at the top of the diamond are higher frequency, or lower rank. For example, the word 'the' appears at the highest observed frequency, and thus has the lowest rank, I. This word has the lowest rank in both corpora, so its coordinates lie along the centre vertical line in the plot. Words such as "women" diverge from the center line because their rank in the incel corpus is higher than in the random corpus. The top 40 words with greatest divergence contribution are shown on the right. In this comparison, nearly all of the top 40 words are more common in the incel corpus, so they point to the right. The word that has the most notable change in rank from the random to incel corpus is "women", the object of hatred

Allotaxonometry 51 of 70

A plenitude of

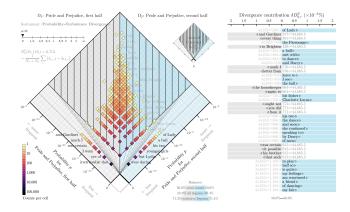
distances

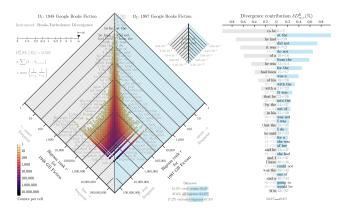
Rank-turbulence
divergence

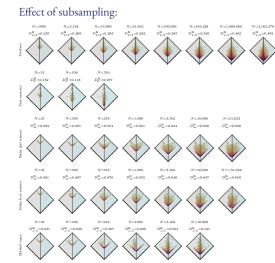
Probabilityturbulence divergence

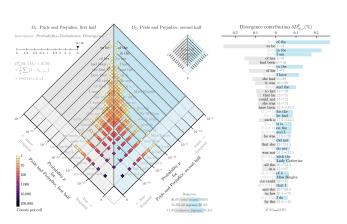
Explorations

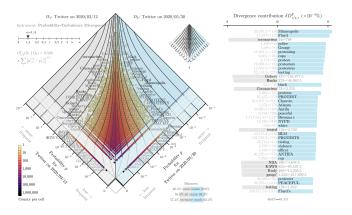
Nutshell

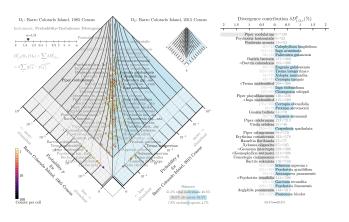


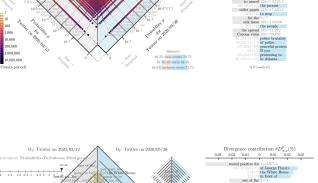












Flipbooks for RTD:

Market Twitter:

allotaxonometer-flipbook-1-rank-div.pdf 🖽 🗷 allotaxonometer-flipbook-2-probability-div.pdf allotaxonometer-flipbook-3-gen-entropy-div.pdf

Market caps:

allotaxonometer-flipbook-4-marketcaps-6years-rank-div.pdf

allotaxonometer-flipbook-5-babynames-girls-50years-rank-div.pdf allotaxonometer-flipbook-6-babynames-boys-50years-rank-div.pdf

Baby girl names over time relative to 1950 E Baby boy names over time relative to 1950

Google books:

allotaxonometer-flipbook-7-google-books-onegrams-rank-div.pdf allotaxonometer-flipbook-8-google-books-bigrams-rank-div.pdf allotaxonometer-flipbook-9-google-books-trigrams-rank-div.pdf

Flipbooks for PTD:

Jane Austen:

Pride and Prejudice, 1-grams Pride and Prejudice, 2-grams Pride and Prejudice, 3-grams

Social media:

Twitter, 1-grams ₩ 🗷 Twitter, 2-grams 🖽 🗷 Twitter, 3-grams

& Ecology:

Barro Colorado Island 🖽 🗗

Code:

Allotaxonometry

A plenitude of

Rank-turbulence

Probability-turbulence divergence

Explorations

60 of 70

https://gitlab.com/compstorylab/allotaxonometer

Allotaxonometry 62 of 70 A plenitude of

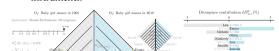
Rank-turbuleno divergence

Probability turbulence div

Explorations Nutshell

Claims, exaggerations, reminders:

- Needed for comparing large-scale complex systems: Comprehendible, dynamically-adjusting, differential dashboards.
- Many measures seem poorly motivated and largely unexamined (e.g., JSD).
- Of value: Combining big-picture maps with ranked lists.
- Online tunable versions of rank-turbulence divergence now
 - App version: https://allotaxp.vercel.app/
 - Observable version:
 - https://observablehq.com/@jstonge/allotaxonometer-4-all Github: https://github.com/jstonge/allotaxp
- 🚓 Future: Probability-turbulence divergence plus many other instruments.



The PoCSverse Allotaxonometry 63 of 70

> Rank-turbuleno divergence

A plenitude of

Explorations

References I

[1] S.-H. Cha.

Comprehensive survey on distance/similarity measures between probability density functions. International Journal of Mathematical Models and Methods

in Applied Sciences, 1:300–307, 2007. pdf ☑

[2] A. Cichocki and S.-i. Amari. Families of Alpha- Beta- and Gamma- divergences: Flexible and robust measures of similarities. Entropy, 12:1532-1568, 2010. pdf

[3] M.-M. Deza and E. Deza. Dictionary of Distances. Elsevier, 2006.

References II

[4] L. R. Dice.

Measures of the amount of ecologic association between species.

Ecology, 26:297-302, 1945.

[5] P. S. Dodds, J. R. Minot, M. V. Arnold, T. Alshaabi, J. L. Adams, D. R. Dewhurst, T. J. Gray, M. R. Frank, A. J. Reagan, and C. M. Danforth.

Allotaxonometry and rank-turbulence divergence: A universal instrument for comparing complex systems.

EPJ Data Science, 12(37):1-42, 2023.

Available online at https://arxiv.org/abs/2002.09770.pdf

References III

[6] P. S. Dodds, J. R. Minot, M. V. Arnold, T. Alshaabi, J. L. Adams, A. J. Reagan, and C. M. Danforth. Probability-turbulence divergence: A tunable allotaxonometric instrument for comparing heavy-tailed categorical distributions, 2020.

Available online at https://arxiv.org/abs/2008.13078. pdf

[7] D. M. Endres and J. E. Schindelin. A new metric for probability distributions. IEEE Transactions on Information theory, 2003. pdf

E. Hellinger.

Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen.

Journal für die reine und angewandte Mathematik (Crelles Journal), 1909(136):210-271, 1909. pdf

The PoCSverse Allotaxonometry 64 of 70

A plenitude of

divergence

Probabilityturbulence divergenc

References

The PoCSverse

Allotaxonometry 65 of 70

A plenitude of

Rank-turbulence

divergence

Probability-

References

Allotaxonometry 66 of 70 A plenitude of

Rank-turbulence

Probability-

References

Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory, 37(1):145-151,

1991. pdf 🗹

[10] J. Looman and J. B. Campbell.

Adaptation of Sørensen's k (1948) for estimating unit affinities in prairie vegetation. Ecology, 41(3):409-416, 1960. pdf

[11] K. Matusita et al.

References IV

[9] J. Lin.

Decision rules, based on the distance, for problems of fit, two samples, and estimation.

The Annals of Mathematical Statistics, 26(4):631-640, 1955. pdf 🖸

References V

[12] R. Munroe.

How To: Absurd Scientific Advice for Common Real-World Problems.

Penguin, 2019.

[13] F. Osterreicher and I. Vajda.

A new class of metric divergences on probability spaces and its applicability in statistics.

Annals of the Institute of Statistical Mathematics, 55(3):639-653, 2003.

[14] E. A. Pechenick, C. M. Danforth, and P. S. Dodds. Is language evolution grinding to a halt? The scaling of lexical turbulence in English fiction suggests it is not. Journal of Computational Science, 21:24–37, 2017. pdf ✓

[15] Y. Sasaki.

The truth of the f-measure, 2007.

References VII

Rank-turbulence

The PoCSverse

Allotaxonometr

68 of 70

Allotaxonometry 67 of 70

A plenitude of

Rank-turbuleno

turbulence divergence

divergence

Probability-

Explorations

Nutshell

References

Probability-

Exploration

References

Text mixing shapes the anatomy of rank-frequency distributions.

[19] J. R. Williams, J. P. Bagrow, C. M. Danforth, and P. S. Dodds.

Physical Review E, 91:052811, 2015. pdf

References VI

[16] C. E. Shannon.

[17] T. Sorensen.

commons.

[18] C. J. Van Rijsbergen.

Information retrieval.

The bandwagon.

IRE Transactions on Information Theory, 2(1):3, 1956.

sociology based on similarity of species content and its

application to analyses of the vegetation on Danish

Videnski Selskab Biologiske Skrifter, 5:1-34, 1948

Butterworth-Heinemann, 2nd edition, 1979.

A method of establishing groups of equal amplitude in plant

Allotaxonometry 70 of 70 A plenitude of

The PoCSverse

Allotaxonometry 69 of 70

A plenitude of

Rank-turbuleno

turbulence divergen

divergence

Probability

Explorations

Nurshell

References

Rank-turbulence divergence

Probability turbulence diverge

Exploration Nutshell

References