

Shannon's Entropy:

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

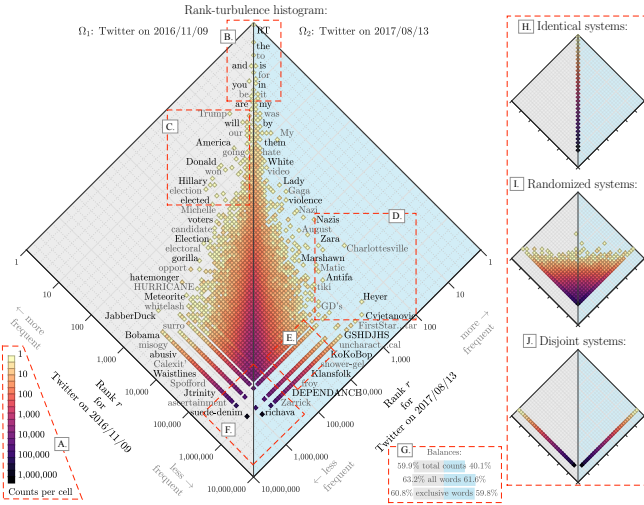
Kullback-Liebler (KL) divergence:

$$\begin{aligned} D^{\text{KL}}(P_2 \parallel P_1) &= \left\langle \log_2 \frac{1}{p_{2,\tau}} - \log_2 \frac{1}{p_{1,\tau}} \right\rangle_{P_2} \\ &= \sum_{\tau \in R_{1,2;\alpha}} p_{2,\tau} \left[\log_2 \frac{1}{p_{2,\tau}} - \log_2 \frac{1}{p_{1,\tau}} \right] \\ &= \sum_{\tau \in R_{1,2;\alpha}} p_{2,\tau} \log_2 \frac{p_{1,\tau}}{p_{2,\tau}}. \end{aligned} \quad (2)$$

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.



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:

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

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

The PoCSverse
Allotaxonomy
19 of 70
A plentitude of
distances
Rank-turbulence
divergence
Probability-
turbulence divergence
Explorations
Nutshell
References

Jensen-Shannon divergence (JSD): [9, 7, 13, 1]

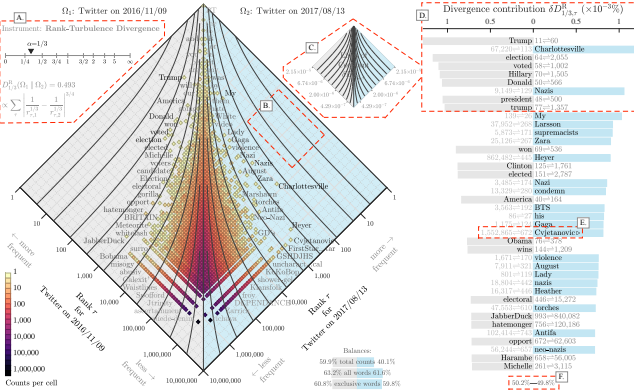
$$\begin{aligned} D^{\text{JS}}(P_1 \parallel P_2) &= \frac{1}{2} D^{\text{KL}}\left(P_1 \parallel \frac{1}{2}[P_1 + P_2]\right) + \frac{1}{2} D^{\text{KL}}\left(P_2 \parallel \frac{1}{2}[P_1 + P_2]\right) \\ &= \frac{1}{2} \sum_{\tau \in R_{1,2;\alpha}} \left(p_{1,\tau} \log_2 \frac{p_{1,\tau}}{\frac{1}{2}[p_{1,\tau} + p_{2,\tau}]} + p_{2,\tau} \log_2 \frac{p_{2,\tau}}{\frac{1}{2}[p_{1,\tau} + p_{2,\tau}]} \right). \end{aligned} \quad (3)$$

Involving a third intermediate averaged system means JSD is now finite:
 $0 \leq D^{\text{JS}}(P_1 \parallel P_2) \leq 1$.

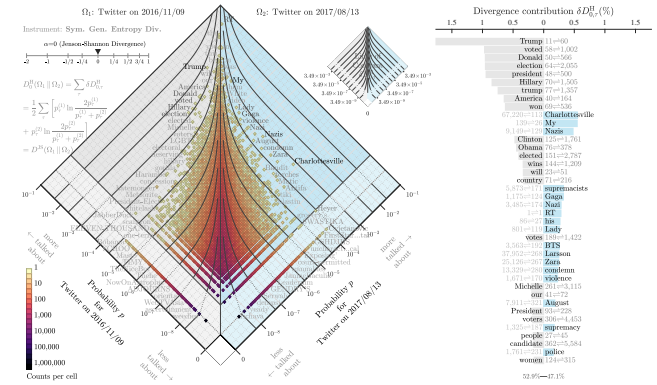
Generalized entropy divergence: [2]

$$\begin{aligned} D_{\alpha}^{\text{AS}}(P_1 \parallel P_2) &= \frac{1}{\alpha(\alpha-1)} \sum_{\tau \in R_{1,2;\alpha}} \left[\left(p_{1,\tau}^{1-\alpha} + p_{2,\tau}^{1-\alpha} \right) \left(\frac{p_{1,\tau} + p_{2,\tau}}{2} \right)^{\alpha} - \left(p_{1,\tau} + p_{2,\tau} \right) \right]. \end{aligned} \quad (4)$$

Produces JSD when $\alpha \rightarrow 0$.



The PoCSverse
Allotaxonomy
20 of 70
A plentitude of
distances
Rank-turbulence
divergence
Probability-
turbulence divergence
Explorations
Nutshell
References



Desirable rank-turbulence divergence features:

- Rank-based.
- Symmetric.
- Semi-positive: $D_{\alpha}^{\text{R}}(\Omega_1 \parallel \Omega_2) \geq 0$.
- Linearly separable, for interpretability.
- 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.).
- Turbulence-handling: Suited for systems with rank-ordered component size distribution that are heavy-tailed.
- Scalable: Allow for sensible comparisons across system sizes.
- Tunable.
- Story-finding: Features 1–8 combine to show which component types are most 'important'

Trouble:

The limit of $\alpha \rightarrow 0$ does not behave well for

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

The leading order term is:

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

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

Oops.

But the insides look nutritious:

$$\left| \ln \frac{r_{\tau,1}}{r_{\tau,2}} \right|$$

is a nicely interpretable log-ratio of ranks.

The PoCSverse
Allotaxonomy
24 of 70
A plentitude of
distances
Rank-turbulence
divergence
Probability-
turbulence divergence
Explorations
Nutshell
References

The PoCSverse
Allotaxonomy
27 of 70
A plentitude of
distances
Rank-turbulence
divergence
Probability-
turbulence divergence
Explorations
Nutshell
References

Some reworking:

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

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

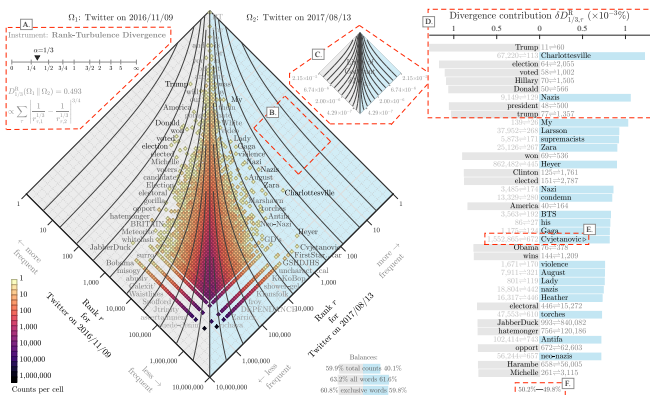
Rank-turbulence divergence:

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

General normalization:

- If 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 system.
- Similarly, $\Omega^{(2)}$'s merged ranking will have all of $\Omega^{(1)}$'s types in last place with rank $r = N_2 + \frac{1}{2}N_1$.
- The normalization is then:

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



Normalization:

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

Limit of $\alpha \rightarrow 0$:

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

where

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

- Largest rank ratios dominate.

Probability-turbulence divergence:

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

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

Rank-turbulence divergence:

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

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

Limit of $\alpha \rightarrow \infty$:

$$D_{\infty}^R(R_1 \parallel R_2) = \sum_{\tau \in R_{1,2;\alpha}} \delta D_{\infty,\tau}^R = \frac{1}{\mathcal{N}_{1,2;\infty}} \sum_{\tau \in R_{1,2;\alpha}} (1 - \delta_{r_{\tau,1}r_{\tau,2}}) \max \left\{ \frac{1}{r_{\tau,1}}, \frac{1}{r_{\tau,2}} \right\}. \quad (14)$$

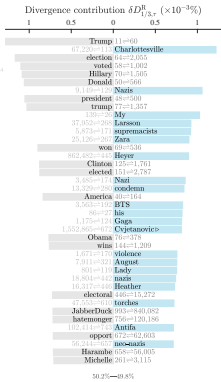
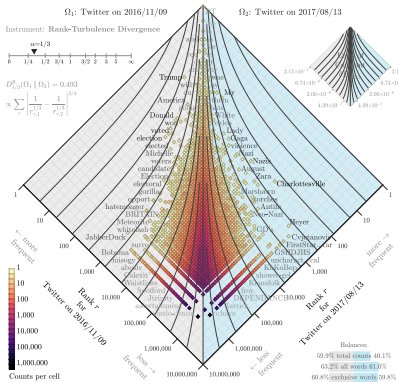
where

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

- Highest ranks dominate.

Type contribution ordering for the limit of $\alpha=0$

- 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 $\alpha \rightarrow 0$, 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 $D_0^p(P_1 \parallel P_2)$, we can still order them according to the log ratio of their probabilities.
- The overall ordering of types by divergence contribution for $\alpha=0$ is then: (1) exclusive types by descending probability and then (2) types appearing in both systems by descending log ratio.

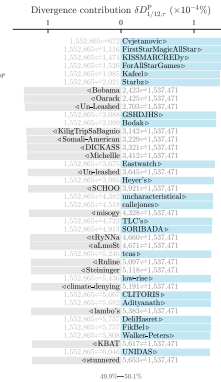
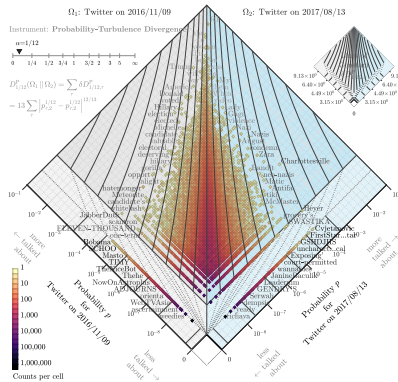


Limit of $\alpha=\infty$ for probability-turbulence divergence

$$D_\infty^p(P_1 \parallel P_2) = \frac{1}{2} \sum_{\tau \in R_{1,2;\infty}} (1 - \delta_{p_{\tau,1}, p_{\tau,2}}) \max(p_{\tau,1}, p_{\tau,2}) \quad (21)$$

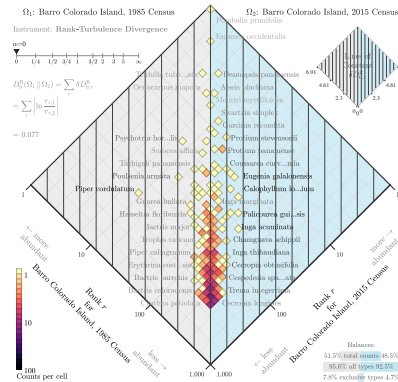
where

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



Connections for PTD:

- $\alpha = 0$: Similarity measure Sørensen-Dice coefficient [4, 17, 10], F_1 score of a test's accuracy [18, 15].
- $\alpha = 1/2$: Hellinger distance [8] and Mautusita distance [11].
- $\alpha = 1$: Many including all $L^{(p)}$ -norm type constructions.
- $\alpha = \infty$: Motyka distance [3].



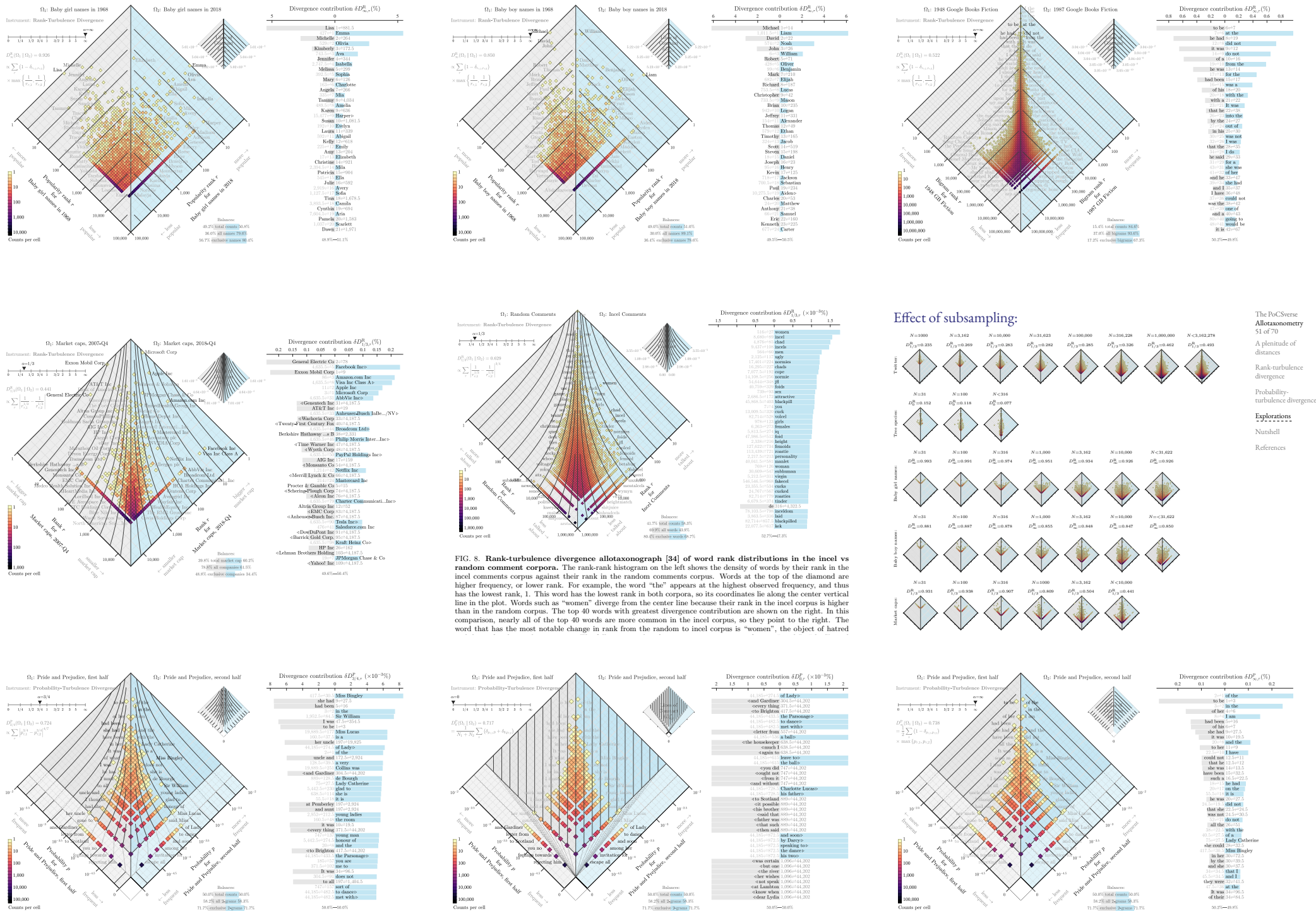
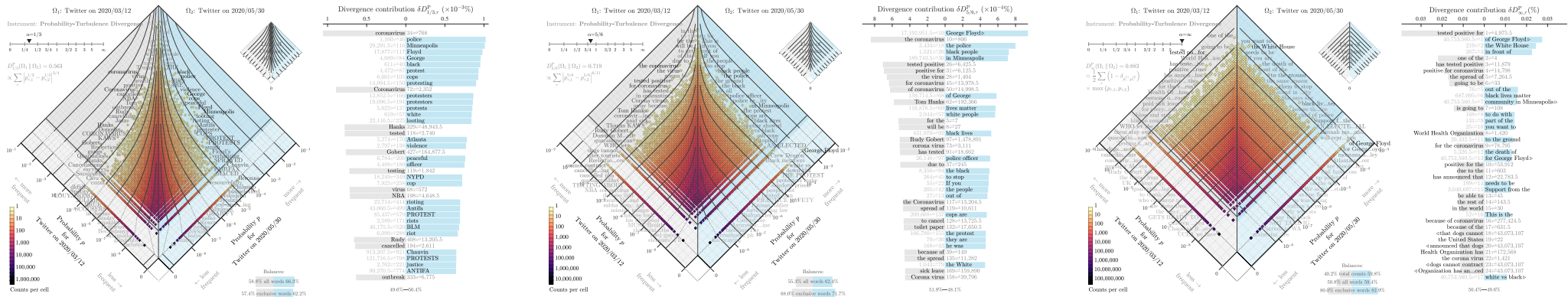


FIG. 8. Rank-turbulence divergence allotaxonomography [34] of word rank distributions in the incel vs random comment corpora. The rank-rank histogram on the left shows the density of words by their rank in the incel comments corpus against their rank in the random comments corpus. Words at the top of 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, 1. This word has the lowest rank in both corpora, so its coordinates lie along the center 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 the incel corpus is “women”, the object of hatred



Flipbooks for RTD:

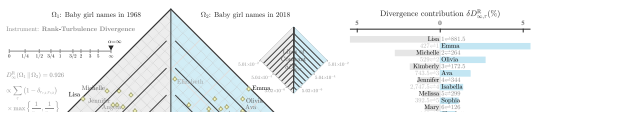
- Twitter:
 - [allotaxonometer-flipbook-1-rank-div.pdf](#)
 - [allotaxonometer-flipbook-2-probability-div.pdf](#)
 - [allotaxonometer-flipbook-3-gen-entropy-div.pdf](#)
 - Market caps:
 - [allotaxonometer-flipbook-4-marketcaps-6years-rank-div.pdf](#)
 - Baby names:
 - [allotaxonometer-flipbook-5-babynames-girls-50years-rank-div.pdf](#)
 - [allotaxonometer-flipbook-6-babynames-boys-50years-rank-div.pdf](#)
- Baby girl names over time relative to 1950
- Baby boy names over time relative to 1950

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](#)

Claims, exaggerations, reminders:

- Needed for comparing large-scale complex systems: Comprehensible, 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 exist:
 - App version: <https://allotax.vercel.app/>
 - Observable version: <https://observablehq.com/@jstonge/allotaxonometer-4-all>
 - Github: <https://github.com/jstonge/allotaxp>
- Future: Could do all probability based instruments



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 IV

[9] J. Lin.
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.
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 VII

[19] 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

The PoCSverse
Allotaxonomy
64 of 70
A plenitude of
distances

Rank-turbulence
divergence

Probability-
turbulence divergence

Explorations

Nutshell

References

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.
Allotaxonomy and rank-turbulence divergence: A universal instrument for comparing complex systems.
EPJ Data Science, 12(37):1–42, 2023.
arXiv version: <https://arxiv.org/abs/2002.09770>. 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.

The PoCSverse
Allotaxonomy
70 of 70
A plenitude of
distances

Rank-turbulence
divergence

Probability-
turbulence divergence

Explorations

Nutshell

References

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 allotaxonomic instrument for comparing heavy-tailed type-size 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, 49(7):1858–1860, 2003. pdf

[8] 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

References VI

[16] C. E. Shannon.
The bandwagon.
IRE Transactions on Information Theory, 2(1):3, 1956. pdf

[17] T. Sorensen.
A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons.
Videnski Selskab Biologiske Skrifter, 5:1–34, 1948.

[18] C. J. Van Rijsbergen.
Information retrieval.
Butterworth-Heinemann, 2nd edition, 1979.

The PoCSverse
Allotaxonomy
65 of 70
A plenitude of
distances

Rank-turbulence
divergence

Probability-
turbulence divergence

Explorations

Nutshell

References

The PoCSverse
Allotaxonomy
68 of 70
A plenitude of
distances

Rank-turbulence
divergence

Probability-
turbulence divergence

Explorations

Nutshell

References

The PoCSverse
Allotaxonomy
66 of 70
A plenitude of
distances

Rank-turbulence
divergence

Probability-
turbulence divergence

Explorations

Nutshell

References

The PoCSverse
Allotaxonomy
69 of 70
A plenitude of
distances

Rank-turbulence
divergence

Probability-
turbulence divergence

Explorations

Nutshell

References

