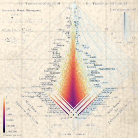
Allotaxonometry

Last updated: 2023/05/22, 06:32:04 CEST

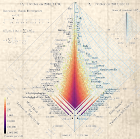
Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 300, 303, & 394, 2022–2023| @pocsvox

Prof. Peter Sheridan Dodds | @peterdodds

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



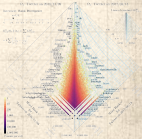
These slides are brought to you by:



These slides are also brought to you by:

Special Guest Executive Producer

On Instagram at pratchett_the_cat



Outline

A plenitude of distances

Rank-turbulence divergence

Probability-turbulence divergence

Explorations

Stories

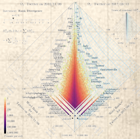
Mechanics of Fame

Superspreading

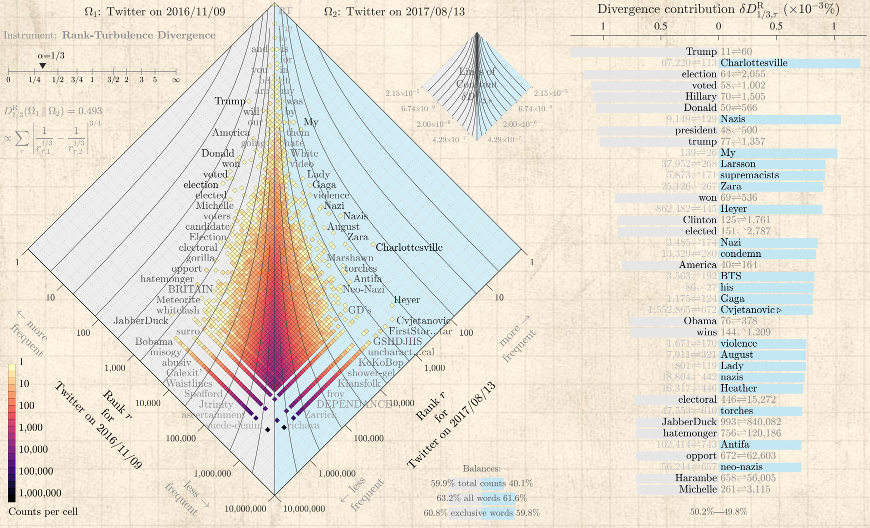
Lexical Ultrafame

Turbulent times

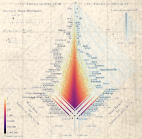
References



Goal—Understand this:

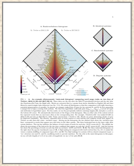


The Boggoracle Speaks:

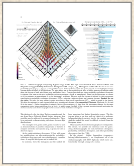


Site (papers, examples, code):
http://compstorylab.org/allotaxonometry/

Foundational papers:



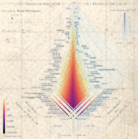
"Allotaxonometry and rank-turbulence divergence: A universal instrument for comparing complex systems"
Dodds et al.,
, 2020. [11]



"Probability-turbulence divergence: A tunable allotaxonometric instrument for comparing heavy-tailed categorical distributions"
Dodds et al.,
, 2020. [13]

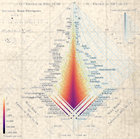
Basic science = Describe + Explain:

- Dashboards of single scale instruments helps us understand, monitor, and control systems.



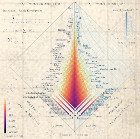
Basic science = Describe + Explain:

- Dashboards of single scale instruments helps us understand, monitor, and control systems.
- Archetype: Cockpit dashboard for flying a plane



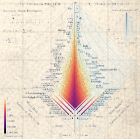
Basic science = Describe + Explain:

- Dashboards of single scale instruments helps us understand, monitor, and control systems.
- Archetype: Cockpit dashboard for flying a plane
- Okay if comprehendible.



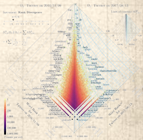
Basic science = Describe + Explain:

- 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:
  1. 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.



Basic science = Describe + Explain:

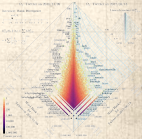
- Dashboards of single scale instruments helps us understand, monitor, and control systems.
- Archetype: Cockpit dashboard for flying a plane
- Okay if comprehendible.
- Complex systems present two problems for dashboards:
  1. Scale with internal diversity of components: We need meters for every species, 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.



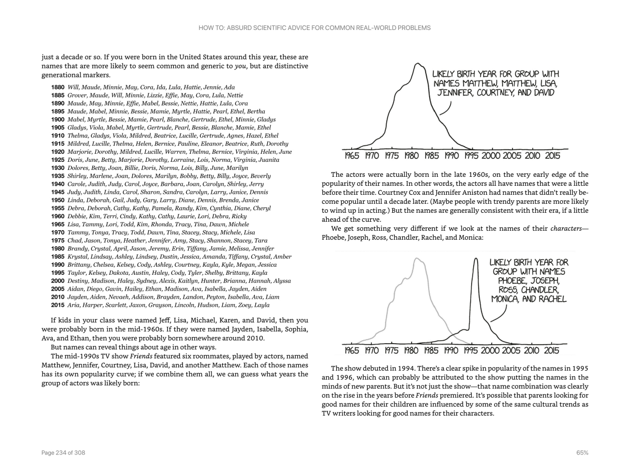
Basic science = Describe + Explain:

- 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:
  1. 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.

[1] See the lexicocalorimeter



Baby names, much studied: [26]



just a decade or so. If you were born in the United States around this year, these are names that are more likely to seem common and generic to *you*, but are distinctive generational markers.

1880 *Will, Maude, Minnie, May, Cora, Ida, Lula, Hattie, Jennie, Ada*
1885 *Grover, Maude, Will, Minnie, Lizzie, Effie, May, Cora, Lula, Nettie*
1890 *Maude, May, Minnie, Effie, Mabel, Bessie, Nettie, Hattie, Lula, Cora*
1895 *Maude, Mabel, Minnie, Bessie, Mamie, Myrtle, Hattie, Pearl, Ethel, Bertha*
1900 *Mabel, Myrtle, Bessie, Mamie, Pearl, Blanche, Gertrude, Ethel, Minnie, Gladys*
1905 *Gladys, Viola, Mabel, Myrtle, Gertrude, Pearl, Bessie, Blanche, Mamie, Ethel*
1910 *Thelma, Gladys, Viola, Mildred, Beatrice, Lucille, Gertrude, Agnes, Hazel, Ethel*
1915 *Mildred, Lucille, Thelma, Helen, Bernice, Pauline, Eleanor, Beatrice, Ruth, Dorothy*
1920 *Marjorie, Dorothy, Mildred, Lucille, Warren, Thelma, Bernice, Virginia, Helen, June*
1925 *Doris, June, Betty, Marjorie, Dorothy, Lorraine, Lois, Norma, Virginia, Juanita*
1930 *Dolores, Betty, Joan, Billie, Doris, Norma, Lois, Billy, June, Marilyn*
1935 *Shirley, Marlene, Joan, Dolores, Marilyn, Bobby, Betty, Billy, Joyce, Beverly*
1940 *Carole, Judith, Judy, Carol, Joyce, Barbara, Joan, Carolyn, Shirley, Jerry*
1945 *Judy, Judith, Linda, Carol, Sharon, Sandra, Carolyn, Larry, Janice, Dennis*
1950 *Linda, Deborah, Gail, Judy, Gary, Larry, Diane, Dennis, Brenda, Janice*
1955 *Debra, Deborah, Cathy, Kathy, Pamela, Randy, Kim, Cynthia, Diane, Cheryl*
1960 *Debbie, Kim, Terri, Cindy, Kathy, Cathy, Laurie, Lori, Debra, Ricky*
1965 *Lisa, Tammy, Lori, Todd, Kim, Rhonda, Tracy, Tina, Dawn, Michele*
1970 *Tammy, Tonya, Tracy, Todd, Dawn, Tina, Stacey, Stacy, Michele, Lisa*
1975 *Chad, Jason, Tonya, Heather, Jennifer, Amy, Stacy, Shannon, Stacey, Tara*
1980 *Brandy, Crystal, April, Jason, Jeremy, Erin, Tiffany, Jamie, Melissa, Jennifer*
1985 *Krystal, Lindsay, Ashley, Lindsey, Dustin, Jessica, Amanda, Tiffany, Crystal, Amber*
1990 *Brittany, Chelsea, Kelsey, Cody, Ashley, Courtney, Kayla, Kyle, Megan, Jessica*
1995 *Taylor, Kelsey, Dakota, Austin, Haley, Cody, Tyler, Shelby, Brittany, Kayla*
2000 *Destiny, Madison, Haley, Sydney, Alexis, Kaitlyn, Hunter, Brianna, Hannah, Alyssa*
2005 *Aidan, Diego, Gavin, Hailey, Ethan, Madison, Ava, Isabella, Jayden, Aiden*
2010 *Jayden, Aiden, Nevaeh, Addison, Brayden, Landon, Peyton, Isabella, Ava, Liam*
2015 *Aria, Harper, Scarlett, Jaxon, Grayson, Lincoln, Hudson, Liam, Zoey, Layla*

If kids in your class were named Jeff, Lisa, Michael, Karen, and David, then you were probably born in the mid-1960s. If they were named Jayden, Isabella, Sophia, Ava, and Ethan, then you were probably born somewhere around 2010.

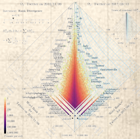
But names can reveal things about age in other ways.

The mid-1990s TV show *Friends* featured six roommates, played by actors, named Matthew, Jennifer, Courtney, Lisa, David, and another Matthew. Each of those names has its own popularity curve; if we combine them all, we can guess what years the group of actors was likely born:

The actors were actually born in the late 1960s, on the very early edge of the popularity of their names. In other words, the actors all have names that were a little before their time. Courtney Cox and Jennifer Aniston had names that didn't really become popular until a decade later. (Maybe people with trendy parents are more likely to wind up in acting.) But the names are generally consistent with their era, if a little ahead of the curve.

We get something very different if we look at the names of their *characters*—Phoebe, Joseph, Ross, Chandler, Rachel, and Monica:

The show debuted in 1994. There's a clear spike in popularity of the names in 1995 and 1996, which can probably be attributed to the show putting the names in the minds of new parents. But it's not just the show—that name combination was clearly on the rise in the years before *Friends* premiered. It's possible that parents looking for good names for their children are influenced by some of the same cultural trends as TV writers looking for good names for their characters.

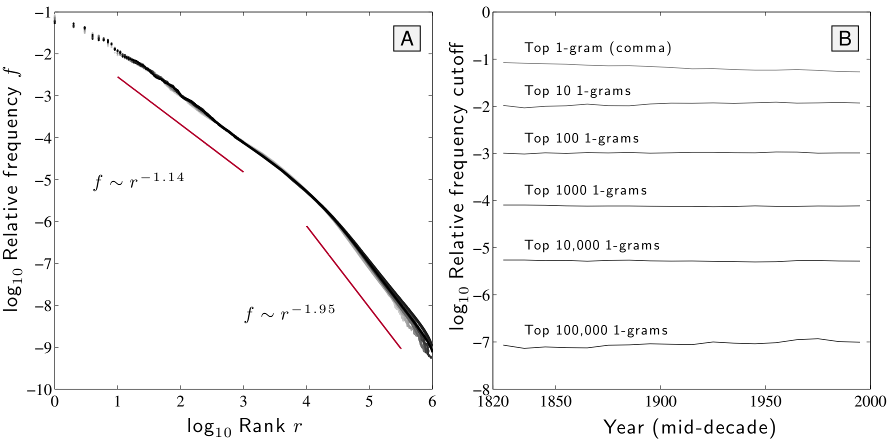


How to build a dynamical dashboard that helps sort through a massive number of interconnected time series?

"Is language evolution grinding to a halt? The scaling of lexical turbulence in English fiction suggests it is not"

Pechenick, Danforth, Dodds, Alshaabi, Adams, Dewhurst, Reagan, Danforth, Reagan, and Danforth.

Journal of Computational Science, **21**, 24–37, 2017. [29]



For language, Zipf's law has two scaling regimes: [38]

$$f \sim \begin{cases} r^{-\alpha} \text{ for } r \ll r_{\text{b}}, \\ r^{-\alpha'} \text{ for } r \gg r_{\text{b}}, \end{cases}$$

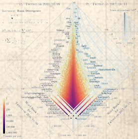
When comparing two texts, define Lexical turbulence as flux of words across a frequency threshold:

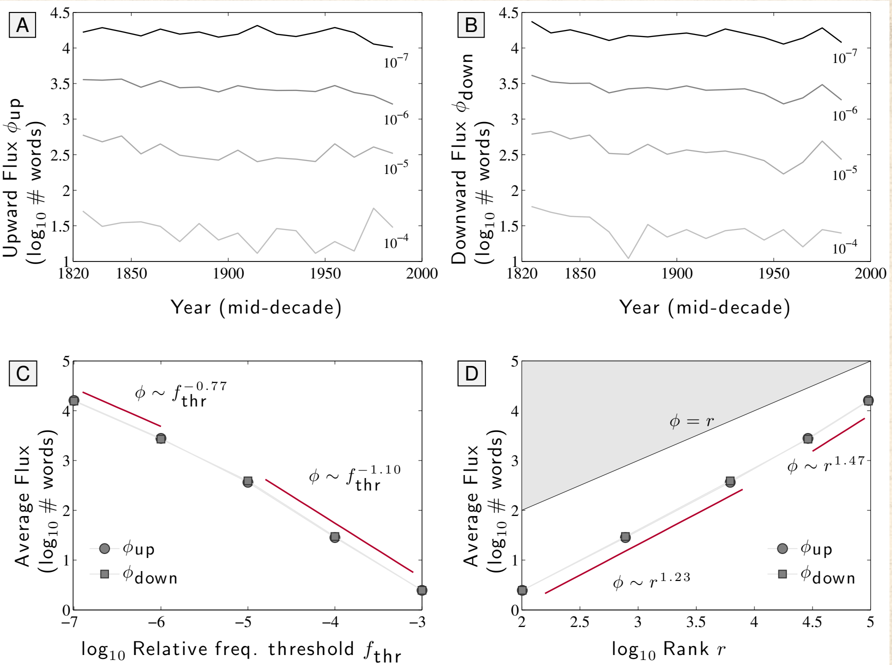
$$\phi \sim \begin{cases} f_{\text{thr}}^{-\mu} \text{ for } f_{\text{thr}} \ll f_{\text{b}}, \\ f_{\text{thr}}^{-\mu'} \text{ for } f_{\text{thr}} \gg f_{\text{b}}, \end{cases}$$

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

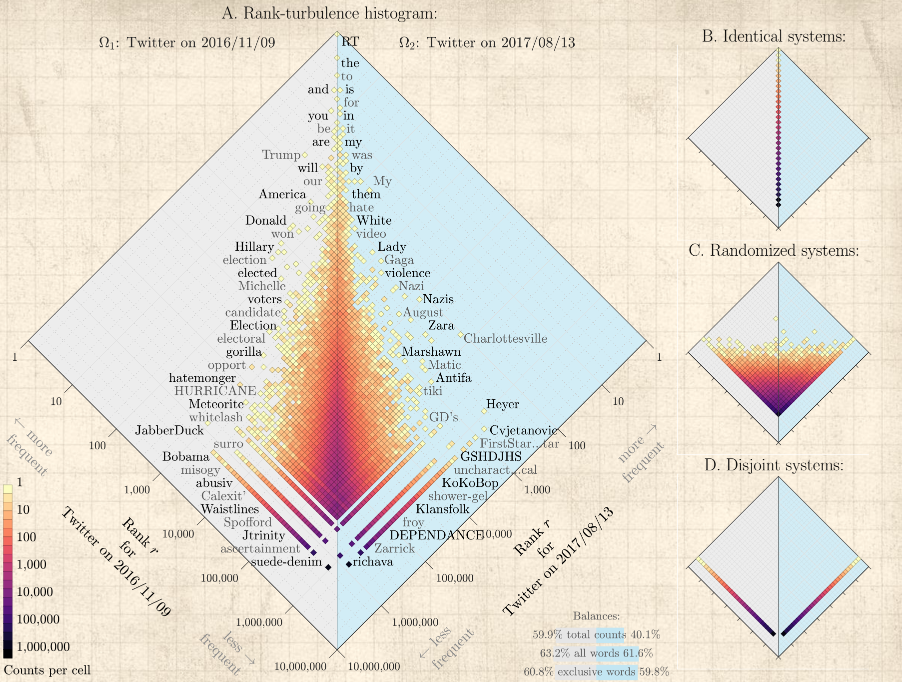
$$\phi \sim \begin{cases} r^{\nu} = r^{\alpha\mu'} \text{ for } r \ll r_{\text{b}}, \\ r^{\nu'} = r^{\alpha'\mu} \text{ for } r \gg r_{\text{b}}. \end{cases}$$

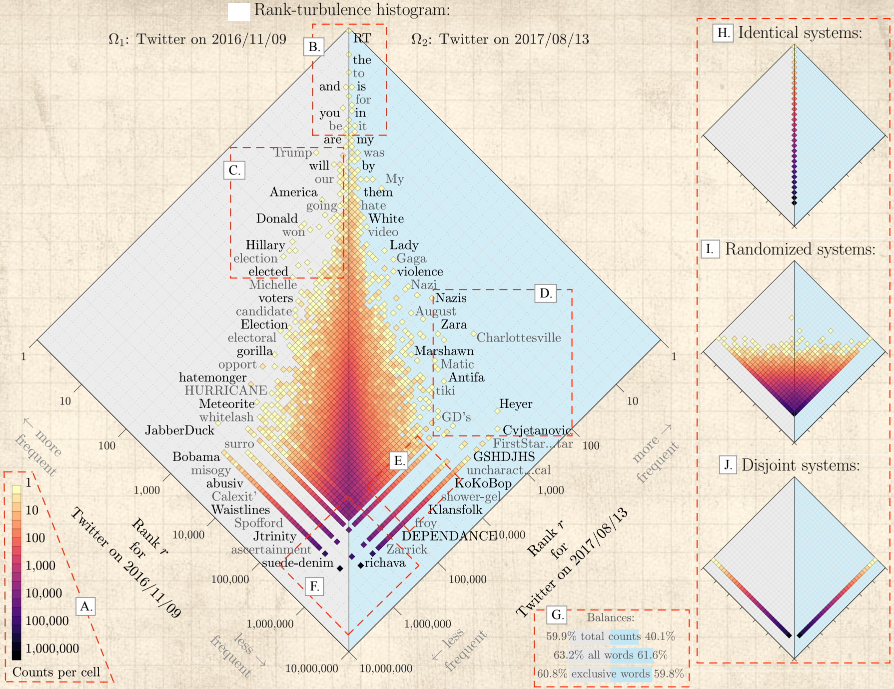
Estimates: Lower and upper exponents $\nu \simeq 1.23$ and $\nu' \simeq 1.47$.





A
Upward Flux ϕup
($\log_{10}$ # words)
10^{-7}
10^{-6}
10^{-5}
10^{-4}
Year (mid-decade)
B
Downward Flux ϕdown
($\log_{10}$ # words)
10^{-7}
10^{-6}
10^{-5}
10^{-4}
Year (mid-decade)
C
$\phi \sim f_{\text{thr}}^{-0.77}$
$\phi \sim f_{\text{thr}}^{-1.10}$
Average Flux
($\log_{10}$ # words)
ϕ_{up}
ϕ_{down}
$\log_{10}$ Relative freq. threshold f_{thr}
D
$\phi = r$
$\phi \sim r^{1.47}$
$\phi \sim r^{1.23}$
Average Flux
($\log_{10}$ # words)
ϕ_{up}
ϕ_{down}
$\log_{10}$ Rank r


A. Rank-turbulence histogram:
Ω_1: Twitter on 2016/11/09
Ω_2: Twitter on 2017/08/13
RT
the
to
and
is
for
you
in
be
it
are
my
Trump
was
will
by
our
My
America
them
going
hate
Donald
White
won
video
Hillary
Lady
election
Gaga
elected
violence
Michelle
Nazi
voters
Nazis
candidate
August
Election
Zara
electoral
Charlottesville
gorilla
Marshawn
opport
Matic
hatemonger
Antifa
HURRICANE
tiki
Meteorite
Heyer
whitelash
GD's
JabberDuck
Cvjetanovic
surro
FirstStar...tar
Bobama
GSHDJHS
misogy
uncharact...cal
abusiv
KoKoBop
Calexit'
shower-gel
Waistlines
Klansfolk
Spofford
froy
Jtrinity
DEPENDANCE
ascertainment
Zarrick
suede-denim
richava
1
10
100
1,000
10,000
100,000
1,000,000
10,000,000
← more frequent
less frequent →
more frequent →
← less frequent
Rank r for Twitter on 2016/11/09
Rank r for Twitter on 2017/08/13
Counts per cell
Balances:
59.9% total counts 40.1%
63.2% all words 61.6%
60.8% exclusive words 59.8%
B. Identical systems:
C. Randomized systems:
D. Disjoint systems:



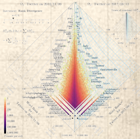
Rank-turbulence histogram:
Ω_1: Twitter on 2016/11/09
Ω_2: Twitter on 2017/08/13
B.
RT
the
to
and
is
for
you
in
be
it
are
my
Trump
was
C.
will
by
our
My
America
them
going
hate
Donald
White
won
video
Hillary
Lady
election
Gaga
elected
violence
Michelle
Nazi
voters
D.
Nazis
candidate
August
Election
Zara
electoral
Charlottesville
gorilla
Marshawn
opport
Matic
hatemonger
Antifa
HURRICANE
tiki
Meteorite
Heyer
whitelash
GD's
JabberDuck
Cvjetanovic
surro
FirstStar...tar
Bobama
E.
GSHDJHS
misogy
uncharact...cal
abusiv
KoKoBop
Calexit'
shower-gel
Waistlines
Klansfolk
Spofford
froy
Jtrinity
DEPENDANCE
ascertainment
Zarrick
suede-denim
richava
F.
1
10
100
1,000
10,000
100,000
1,000,000
10,000,000
← more frequent
less frequent →
← less frequent
more frequent →
Rank r for Twitter on 2016/11/09
Rank r for Twitter on 2017/08/13
A.
1
10
100
1,000
10,000
100,000
1,000,000
Counts per cell
G.
Balances:
59.9% total counts 40.1%
63.2% all words 61.6%
60.8% exclusive words 59.8%
H. Identical systems:
I. Randomized systems:
J. Disjoint systems:

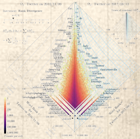
G. Balances:

59.9% total counts 40.1%

63.2% all words 61.6%

60.8% exclusive words 59.8%

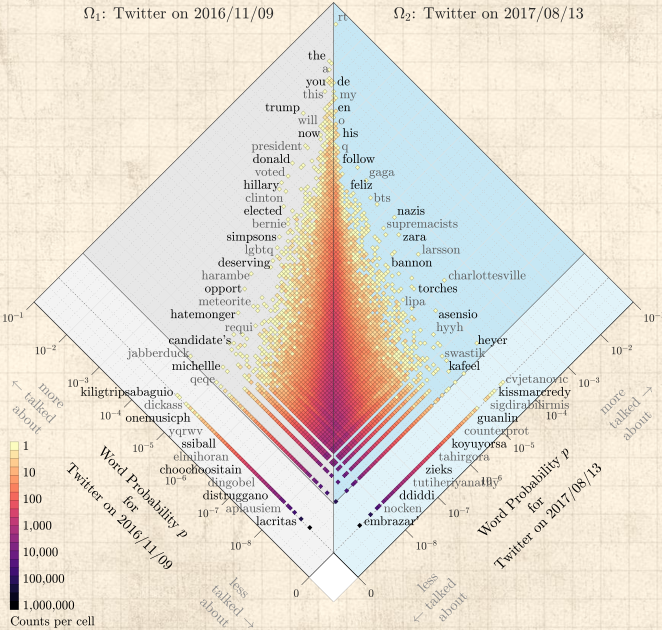


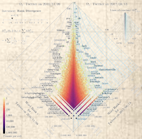


Exclusive types:

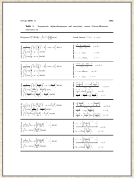
- We call types that are present in one system only 'exclusive types'.
- When warranted, we will use expressions of the form $\Omega^{(1)}$-exclusive and $\Omega^{(2)}$-exclusive to indicate to which system an exclusive type belongs.

Probability-turbulence histogram:

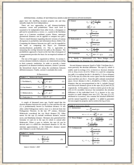




So, so many ways to compare probability distributions:

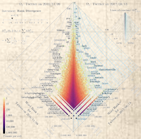


“Families of Alpha- Beta- and Gamma-Divergences: Flexible and Robust Measures of Similarities”
Cichocki and Amari,
Entropy, **12**, 1532-1568, 2010. [8]

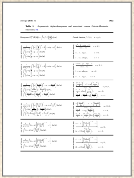


“Comprehensive survey on distance/similarity measures between probability density functions”
Sung-Hyuk Cha,
International Journal of Mathematical Models and Methods in Applied Sciences, **1**, 300–307, 2007. [4]

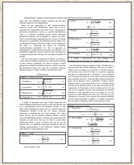
- Comparisons are distances, divergences, similarities, inner products, fidelities ...



So, so many ways to compare probability distributions:

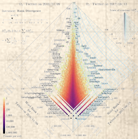


"Families of Alpha- Beta- and Gamma-Divergences: Flexible and Robust Measures of Similarities"
Cichocki and Amari,
Entropy, **12**, 1532-1568, 2010. [8]

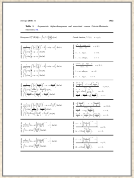


"Comprehensive survey on distance/similarity measures between probability density functions"
Sung-Hyuk Cha,
International Journal of Mathematical Models and Methods in Applied Sciences,
1, 300–307, 2007. [4]

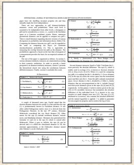
- Comparisons are distances, divergences, similarities, inner products, fidelities ...
- 60ish kinds of comparisons grouped into 10 families



So, so many ways to compare probability distributions:

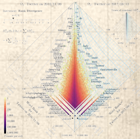


"Families of Alpha- Beta- and Gamma-Divergences: Flexible and Robust Measures of Similarities"
Cichocki and Amari,
Entropy, **12**, 1532-1568, 2010. [8]



"Comprehensive survey on distance/similarity measures between probability density functions"
Sung-Hyuk Cha,
International Journal of Mathematical Models and Methods in Applied Sciences, **1**, 300–307, 2007. [4]

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



Quite the festival:

Table 1. L_p Minkowski family		
1. Euclidean L_2	$d_{Euc} = \sqrt{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^2}$	(1)
2. City block L_1	$d_{CB} = \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(2)
3. Minkowski L_p	$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^p}$	(3)
4. Chebyshev L_∞	$d_{Cheb} = \max_i \lvert P_i - Q_i \rvert$	(4)

Table 2. L_1 family		
5. Sørensen	$d_{sor} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} (P_i + Q_i)}$	(5)
6. Gower	$d_{gow} = \frac{1}{d} \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{R_i}$	(6)
	$= \frac{1}{d} \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(7)
7. Soergel	$d_{sg} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)
8. Kulczynski d	$d_{kul} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)
9. Canberra	$d_{Can} = \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{P_i + Q_i}$	(10)
10. Lorentzian	$d_{Lor} = \sum_{i=1}^{d} \ln(1 + \lvert P_i - Q_i \rvert)$	(11)
* L_1 family ⊃ {Intersectoin (13), Wave Hedges (15), Czekanowski (16), Ruzicka (21), Tanimoto (23), etc}.		

Table 3. Intersection family		
11. Intersection	$s_{IS} = \sum_{i=1}^{d} \min(P_i, Q_i)$	(12)
	$d_{non-IS} = 1 - s_{IS} = \frac{1}{2} \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(13)
12. Wave Hedges	$d_{WH} = \sum_{i=1}^{d} (1 - \frac{\min(P_i, Q_i)}{\max(P_i, Q_i)})$	(14)
	$= \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{\max(P_i, Q_i)}$	(15)
13. Czekanowski	$s_{Cze} = \frac{2 \sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} (P_i + Q_i)}$	(16)
	$d_{Cze} = 1 - s_{Cze} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} (P_i + Q_i)}$	(17)

14. Motyka	$s_{Mot} = \frac{\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} (P_i + Q_i)}$	(18)
	$d_{Mot} = 1 - s_{Mot} = \frac{\sum_{i=1}^{d} \max(P_i, Q_i)}{\sum_{i=1}^{d} (P_i + Q_i)}$	(19)
15. Kulczynski s	$s_{Kul} = \frac{1}{d_{Kul}} = \frac{\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}$	(20)
16. Ruzicka	$s_{Ruz} = \frac{\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(21)
17. Tanimoto	$d_{Tani} = \frac{\sum_{i=1}^{d} P_i + \sum_{i=1}^{d} Q_i - 2\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} P_i + \sum_{i=1}^{d} Q_i - \sum_{i=1}^{d} \min(P_i, Q_i)}$	(22)
	$= \frac{\sum_{i=1}^{d} (\max(P_i, Q_i) - \min(P_i, Q_i))}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(23)

Table 4. Inner Product family		
18. Inner Product	$s_{IP} = P \bullet Q = \sum_{j=1}^{d} P_i Q_i$	(24)
19. Harmonic mean	$s_{HM} = 2 \sum_{i=1}^{d} \frac{P_i Q_i}{P_i + Q_i}$	(25)
20. Cosine	$s_{Cos} = \frac{\sum_{i=1}^{d} P_i Q_i}{\sqrt{\sum_{i=1}^{d} P_i^2} \sqrt{\sum_{i=1}^{d} Q_i^2}}$	(26)
21. Kumar-Hassebrook (PCE)	$s_{Jac} = \frac{\sum_{i=1}^{d} P_i Q_i}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2 - \sum_{i=1}^{d} P_i Q_i}$	(27)
22. Jaccard	$s_{Jac} = \frac{\sum_{i=1}^{d} P_i Q_i}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2 - \sum_{i=1}^{d} P_i Q_i}$	(28)
	$d_{Jac} = 1 - s_{Jac} = \frac{\sum_{i=1}^{d} (P_i - Q_i)^2}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2 - \sum_{i=1}^{d} P_i Q_i}$	(39)
23. Dice	$s_{Dice} = \frac{2 \sum_{i=1}^{d} P_i Q_i}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2}$	(40)
	$d_{Dice} = 1 - s_{Dice} = \frac{\sum_{i=1}^{d} (P_i - Q_i)^2}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2}$	(31)

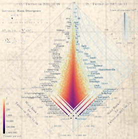
Table 5. Fidelity family or Squared-chord family		
24. Fidelity	$s_{Fid} = \sum_{i=1}^{d} \sqrt{P_i Q_i}$	(32)
25. Bhattacharyya	$d_B = -\ln \sum_{i=1}^{d} \sqrt{P_i Q_i}$	(33)
26. Hellinger	$d_H = \sqrt{2 \sum_{i=1}^{d} (\sqrt{P_i} - \sqrt{Q_i})^2}$	(34)
	$= 2 \sqrt{1 - \sum_{i=1}^{d} \sqrt{P_i Q_i}}$	(35)
27. Matusita	$d_M = \sqrt{\sum_{i=1}^{d} (\sqrt{P_i} - \sqrt{Q_i})^2}$	(36)
	$= \sqrt{2 - 2 \sum_{i=1}^{d} \sqrt{P_i Q_i}}$	(37)
28. Squared-chord	$d_{sqc} = \sum_{i=1}^{d} (\sqrt{P_i} - \sqrt{Q_i})^2$	(38)
s_{sqc} = 1-d_{sqc}	$s_{sqc} = 2 \sum_{i=1}^{d} \sqrt{P_i Q_i} - 1$	(39)

Table 6. Squared L_2 family or χ^2 family		
29. Squared Euclidean	$d_{sqe} = \sum_{i=1}^{d} (P_i - Q_i)^2$	(40)
30. Pearson χ^2	$d_P(P, Q) = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{Q_i}$	(41)
31. Neyman χ^2	$d_N(P, Q) = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i}$	(42)
32. Squared χ^2	$d_{SqChi} = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i + Q_i}$	(43)
33. Probabilistic Symmetric χ^2	$d_{PChi} = 2 \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i + Q_i}$	(44)
34. Divergence	$d_{Div} = 2 \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{(P_i + Q_i)^2}$	(45)
35. Clark	$d_{Clk} = \sqrt{\sum_{i=1}^{d} \left(\frac{\lvert P_i - Q_i \rvert}{P_i + Q_i} \right)^2}$	(46)
36. Additive Symmetric χ^2	$d_{AdChi} = \sum_{i=1}^{b} \frac{(P_i - Q_i)^2 (P_i + Q_i)}{P_i Q_i}$	(47)
* Squared L_2 family ⊃ {Jaccard (29), Dice (31)}		

Table 7. Shannon's entropy family		
37. Kullback–Leibler	$d_{KL} = \sum_{i=1}^{d} P_i \ln \frac{P_i}{Q_i}$	(48)
38. Jeffreys	$d_J = \sum_{i=1}^{d} (P_i - Q_i) \ln \frac{P_i}{Q_i}$	(49)
39. K divergence	$d_{Kdiv} = \sum_{i=1}^{d} P_i \ln \frac{2P_i}{P_i + Q_i}$	(50)
40. Topsøe	$d_{Top} = \sum_{i=1}^{d} \left(P_i \ln \left(\frac{2P_i}{P_i + Q_i} \right) + Q_i \ln \left(\frac{2Q_i}{P_i + Q_i} \right) \right)$	(51)
41. Jensen-Shannon	$d_{JS} = \frac{1}{2} \left[\sum_{i=1}^{d} P_i \ln \left(\frac{2P_i}{P_i + Q_i} \right) + \sum_{i=1}^{d} Q_i \ln \left(\frac{2Q_i}{P_i + Q_i} \right) \right]$	(52)
42. Jensen difference	$d_{JD} = \sum_{i=1}^{b} \left[\frac{P_i \ln P_i + Q_i \ln Q_i}{2} - \left(\frac{P_i + Q_i}{2} \right) \ln \left(\frac{P_i + Q_i}{2} \right) \right]$	(53)

Table 8. Combinations		
43. Taneja	$d_{TJ} = \sum_{i=1}^{d} \left(\frac{P_i + Q_i}{2} \right) \ln \left(\frac{P_i + Q_i}{2\sqrt{P_i Q_i}} \right)$	(54)
44. Kumar-Johnson	$d_{KJ} = \sum_{i=1}^{d} \left(\frac{(P_i^2 - Q_i^2)^2}{2(P_i Q_i)^{3/2}} \right)$	(55)
45. Avg(L_1, L_∞)	$d_{ACC} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert + \max_i \lvert P_i - Q_i \rvert}{2}$	(56)

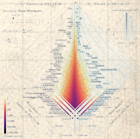
Table 10. Vicissitude		
Vicis-Wave Hedges	$d_{emanon1} = \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{\min(P_i, Q_i)}$	(60)
Vicis-Symmetric χ^2	$d_{emanon2} = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{\min(P_i, Q_i)^2}$	(61)
Vicis-Symmetric χ^2	$d_{emanon3} = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{\min(P_i, Q_i)}$	(62)
Vicis-Symmetric χ^2	$d_{emanon4} = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{\max(P_i, Q_i)}$	(63)
max-Symmetric χ^2	$d_{e5} = \max \left(\sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i}, \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{Q_i} \right)$	(64)
min-symmetric χ^2	$d_{e6} = \min \left(\sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i}, \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{Q_i} \right)$	(65)

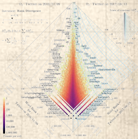


Shannon tried to slow things down in 1956:

"The bandwagon"
Claude E Shannon,
IRE Transactions on Information Theory, **2**, 3, 1956. [34]

- "Information theory has ... become something of a scientific bandwagon."





Shannon tried to slow things down in 1956:

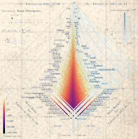
"The bandwagon"
Claude E Shannon,
IRE Transactions on Information Theory, **2**, 3, 1956. [34]

- "Information theory has ... become something of a scientific bandwagon."
- "While ... information theory is indeed a valuable tool ... [it] is certainly no panacea for the communication engineer or ... for anyone else.

Shannon tried to slow things down in 1956:

"The bandwagon"
Claude E Shannon,
IRE Transactions on Information Theory, **2**, 3, 1956. [34]

- "Information theory has ... become something of a scientific bandwagon."
- "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."



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

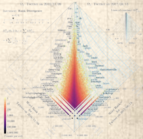
Table 1. L_p Minkowski family

1. Euclidean L_2	$d_{Euc} = \sqrt{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^2}$	(1)
2. City block L_1	$d_{CB} = \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(2)
3. Minkowski L_p	$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^p}$	(3)
4. Chebyshev L_∞	$d_{Cheb} = \max_i \lvert P_i - Q_i \rvert$	(4)

Table 2. L_1 family

5. Sørensen	$d_{sor} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} (P_i + Q_i)}$	(5)
6. Gower	$d_{gow} = \frac{1}{d} \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{R_i}$	(6)
	$= \frac{1}{d} \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(7)
7. Soergel	$d_{sg} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)
8. Kulczynski d	$d_{kul} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)
9. Canberra	$d_{Can} = \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{P_i + Q_i}$	(10)
10. Lorentzian	$d_{Lor} = \sum_{i=1}^{d} \ln(1 + \lvert P_i - Q_i \rvert)$	(11)

* L_1 family $\supset$ {Intersectoin (13), Wave Hedges (15), Czekanowski (16), Ruzicka (21), Tanimoto (23), etc}.



- 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.

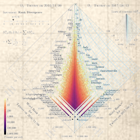
Table 1. L_p Minkowski family

1. Euclidean L_2	$d_{Euc} = \sqrt{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^2}$	(1)
2. City block L_1	$d_{CB} = \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(2)
3. Minkowski L_p	$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^p}$	(3)
4. Chebyshev L_∞	$d_{Cheb} = \max_i \lvert P_i - Q_i \rvert$	(4)

Table 2. L_1 family

5. Sørensen	$d_{sor} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} (P_i + Q_i)}$	(5)
6. Gower	$d_{gow} = \frac{1}{d} \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{R_i}$	(6)
	$= \frac{1}{d} \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(7)
7. Soergel	$d_{sg} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)
8. Kulczynski d	$d_{kul} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)
9. Canberra	$d_{Can} = \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{P_i + Q_i}$	(10)
10. Lorentzian	$d_{Lor} = \sum_{i=1}^{d} \ln(1 + \lvert P_i - Q_i \rvert)$	(11)

* L_1 family $\supset$ {Intersectoin (13), Wave Hedges (15), Czekanowski (16), Ruzicka (21), Tanimoto (23), etc}.



- 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)
  - $\max(P_i, Q_i)$
  - $\min(P_i, Q_i)$
  - $P_i Q_i$
  - $|P_i^{1/2} - Q_i^{1/2}|$ (Hellinger)

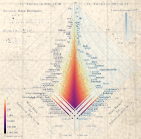
Table 1. L_p Minkowski family

1. Euclidean L_2	$d_{Euc} = \sqrt{\sum_{i=1}^{d} \|P_i - Q_i\|^2}$	(1)
2. City block L_1	$d_{CB} = \sum_{i=1}^{d} \|P_i - Q_i\|$	(2)
3. Minkowski L_p	$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} \|P_i - Q_i\|^p}$	(3)
4. Chebyshev L_∞	$d_{Cheb} = \max_i \|P_i - Q_i\|$	(4)

Table 2. L_1 family

5. Sørensen	$d_{sor} = \dfrac{\sum_{i=1}^{d} \|P_i - Q_i\|}{\sum_{i=1}^{d} (P_i + Q_i)}$	(5)
6. Gower	$d_{gow} = \dfrac{1}{d} \sum_{i=1}^{d} \dfrac{\|P_i - Q_i\|}{R_i}$	(6)
	$= \dfrac{1}{d} \sum_{i=1}^{d} \|P_i - Q_i\|$	(7)
7. Soergel	$d_{sg} = \dfrac{\sum_{i=1}^{d} \|P_i - Q_i\|}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)
8. Kulczynski d	$d_{kul} = \dfrac{\sum_{i=1}^{d} \|P_i - Q_i\|}{\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)
9. Canberra	$d_{Can} = \sum_{i=1}^{d} \dfrac{\|P_i - Q_i\|}{P_i + Q_i}$	(10)
10. Lorentzian	$d_{Lor} = \sum_{i=1}^{d} \ln(1 + \|P_i - Q_i\|)$	(11)

* L_1 family $\supset$ {Intersectoin (13), Wave Hedges (15), Czekanowski (16), Ruzicka (21), Tanimoto (23), etc}.

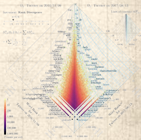


Information theoretic sortings are more opaque

Table 1. L_p Minkowski family		
1. Euclidean L_2	$d_{Euc} = \sqrt{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^2}$	(1)
2. City block L_1	$d_{CB} = \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(2)
3. Minkowski L_p	$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^p}$	(3)
4. Chebyshev L_∞	$d_{Cheb} = \max_i \lvert P_i - Q_i \rvert$	(4)

Table 2. L_1 family		
5. Sørensen	$d_{sor} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} (P_i + Q_i)}$	(5)
6. Gower	$d_{gow} = \frac{1}{d} \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{R_i}$	(6)
	$= \frac{1}{d} \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(7)
7. Soergel	$d_{sg} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)
8. Kulczynski d	$d_{kul} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)
9. Canberra	$d_{Can} = \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{P_i + Q_i}$	(10)
10. Lorentzian	$d_{Lor} = \sum_{i=1}^{d} \ln(1 + \lvert P_i - Q_i \rvert)$	(11)

* L_1 family $\supset$ {Intersectoin (13), Wave Hedges (15), Czekanowski (16), Ruzicka (21), Tanimoto (23), etc}.



- Information theoretic sortings are more opaque
- No tunability

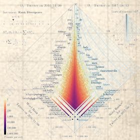
Table 1. L_p Minkowski family

1. Euclidean L_2	$d_{Euc} = \sqrt{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^2}$	(1)
2. City block L_1	$d_{CB} = \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(2)
3. Minkowski L_p	$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^p}$	(3)
4. Chebyshev L_∞	$d_{Cheb} = \max_i \lvert P_i - Q_i \rvert$	(4)

Table 2. L_1 family

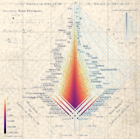
5. Sørensen	$d_{sor} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} (P_i + Q_i)}$	(5)
6. Gower	$d_{gow} = \frac{1}{d} \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{R_i}$	(6)
	$= \frac{1}{d} \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(7)
7. Soergel	$d_{sg} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)
8. Kulczynski d	$d_{kul} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)
9. Canberra	$d_{Can} = \sum_{i=1}^{d} \frac{\lvert P_i - Q_i \rvert}{P_i + Q_i}$	(10)
10. Lorentzian	$d_{Lor} = \sum_{i=1}^{d} \ln(1 + \lvert P_i - Q_i \rvert)$	(11)

* L_1 family $\supset$ {Intersectoin (13), Wave Hedges (15), Czekanowski (16), Ruzicka (21), Tanimoto (23), etc}.



Shannon's Entropy:

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

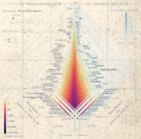


- Shannon's Entropy:

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

- Kullback-Liebler (KL) divergence:

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



- Shannon's Entropy:

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

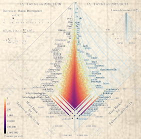
- Kullback-Liebler (KL) divergence:

$$D^{\mathrm{KL}}(P_2 \,||\, P_1) = \left\langle \log_2 \frac{1}{p_{2,\tau}} - \log_2 \frac{1}{p_{1,\tau}} \right\rangle_{P_2}$$

$$= \sum_{\tau \in R_{1,2;\alpha}} p_{2,\tau} \left[\log_2 \frac{1}{p_{2,\tau}} - \log_2 \frac{1}{p_{1,\tau}} \right]$$

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

- Problem: If just one component type in system 2 is not present in system 1, KL divergence = ∞.



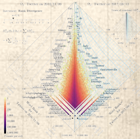
- Shannon's Entropy:

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

- Kullback-Liebler (KL) divergence:

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

- 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.



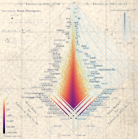
- Shannon's Entropy:

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

- Kullback-Liebler (KL) divergence:

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

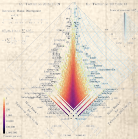
- Problem: If just one component type in system 2 is not present in system 1, KL divergence = ∞.
- Solution: If we can't compare a spork and a platypus directly, we create a fictional spork-platypus hybrid.
- New problem: Re-read solution.

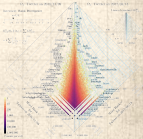


Jensen-Shannon divergence (JSD): [21, 15, 28, 4]

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

Involving a third intermediate averaged system means JSD is now finite: $0 \leq D^{\mathrm{JS}}\left(P_1 \,||\, P_2\right) \leq 1$.





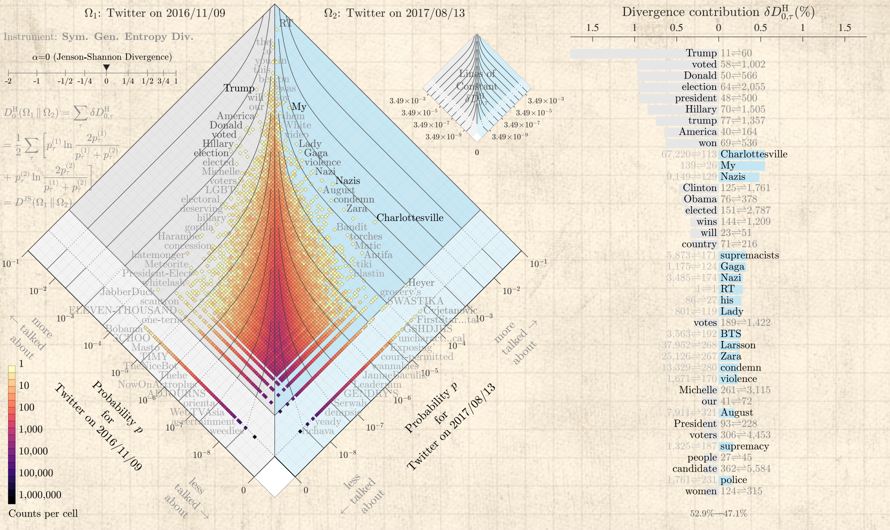
- Jensen-Shannon divergence (JSD): [21, 15, 28, 4]

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

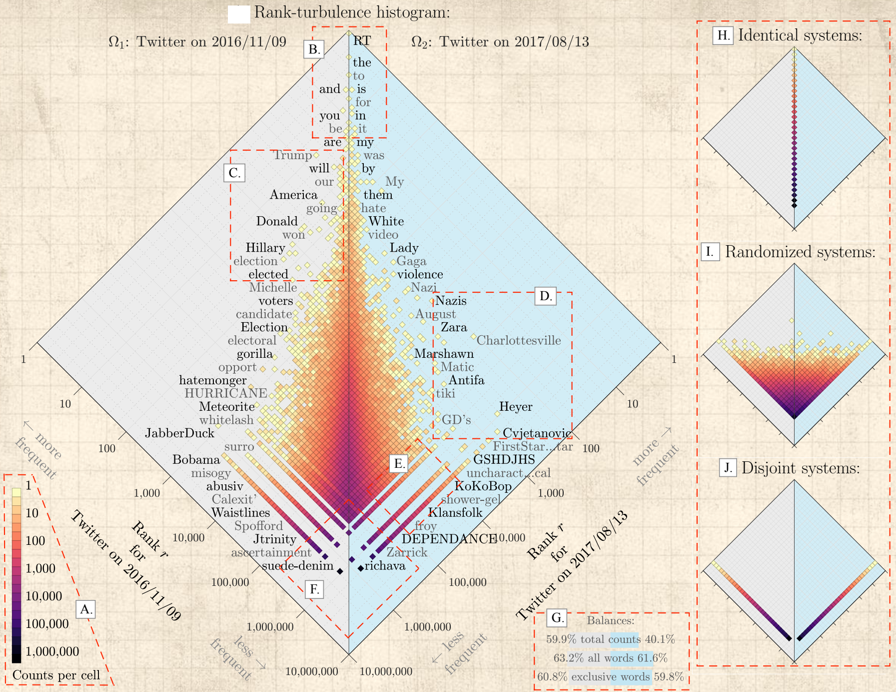
- Involving a third intermediate averaged system means JSD is now finite: $0 \le D^{\mathrm{JS}}\left(P_1 \,\|\, P_2\right) \le 1$.
- Generalized entropy divergence: [8]

$$
\begin{aligned}
&D^{\mathrm{AS2}}_{\alpha}\left(P_1 \,\|\, P_2\right) = \\
&\frac{1}{\alpha(\alpha-1)} \sum_{\tau \in R_{1,2;\alpha}} \left[\left(p_{\tau,1}^{1-\alpha} + p_{\tau,2}^{1-\alpha}\right) \left(\frac{p_{\tau,1} + p_{\tau,2}}{2}\right)^{\alpha} - \left(p_{\tau,1} + p_{\tau,2}\right) \right].
\end{aligned}
\tag{4}
$$

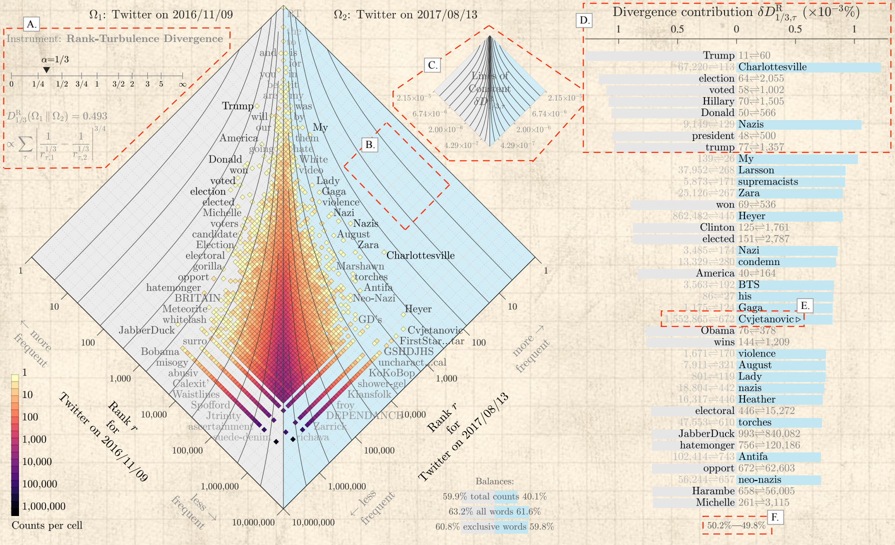
Produces JSD when $\alpha \to 0$.



Ω1: Twitter on 2016/11/09
Ω2: Twitter on 2017/08/13
Instrument: Sym. Gen. Entropy Div.
α=0 (Jenson-Shannon Divergence)
-2 -1 -1/2 -1/4 0 1/4 1/2 3/4 1
$D_0^{\mathrm{H}}(\Omega_1 \,\|\, \Omega_2) = \sum_\tau \delta D^{\mathrm{H}}_{0,\tau}$
$= \frac{1}{2}\sum_\tau \left[p_\tau^{(1)} \ln \frac{2p_\tau^{(1)}}{p_\tau^{(1)} + p_\tau^{(2)}} + p_\tau^{(2)} \ln \frac{2p_\tau^{(2)}}{p_\tau^{(1)} + p_\tau^{(2)}} \right]$
$= D^{\mathrm{JS}}(\Omega_1 \,\|\, \Omega_2)$
Lines of Constant $\delta D^{\mathrm{H}}_{0,\tau}$
3.49×10^{-3}
3.49×10^{-5}
3.49×10^{-7}
3.49×10^{-9}
RT
Trump
My
America
Donald
voted
Hillary
election
elected
Michelle
voters
LGBT
electoral
deserving
hillary
gorilla
Harambe
concession
hatemonger
Meteorite
President-Elect
whitelash
JabberDuck
ELEVEN-THOUSAND
one-term
Bobama
SCHOO
Masto
TIMY
TheNiceBot
Thehe
NowOnAstroplus
ADJOURNS
orienta
WebTVAsia
ascertainment
weedies
White
video
Lady
Gaga
violence
Nazi
Nazis
August
condemn
Zara
Charlottesville
Bandit
torches
Matic
Antifa
tiki
blastin
Heyer
grocery's
SWASTIKA
Cvjetanovic
FirstStar...ta
GSHDJHS
uncharact...cal
Exposing
court-permitted
wannabes
JanineBaculik
Leaderrim
GENDRYS
Serwah
dempsie
yeady
richava
10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 10^{-8} 0
← more talked about
less talked about →
more talked about →
← less talked about
Probability p for Twitter on 2016/11/09
Probability p for Twitter on 2017/08/13
1
10
100
1,000
10,000
100,000
1,000,000
Counts per cell
Divergence contribution $\delta D^{\mathrm{H}}_{0,\tau}$(%)
1.5 1 0.5 0 0.5 1 1.5
Trump 11⇌60
voted 58⇌1,002
Donald 50⇌566
election 64⇌2,055
president 48⇌500
Hillary 70⇌1,505
trump 77⇌1,357
America 40⇌164
won 69⇌536
67,220⇌113 Charlottesville
139⇌26 My
9,149⇌129 Nazis
Clinton 125⇌1,761
Obama 76⇌378
elected 151⇌2,787
wins 144⇌1,209
will 23⇌51
country 71⇌216
5,873⇌171 supremacists
1,175⇌124 Gaga
3,485⇌174 Nazi
1⇌1 RT
86⇌27 his
801⇌119 Lady
votes 189⇌1,422
3,563⇌192 BTS
37,952⇌268 Larsson
25,126⇌267 Zara
13,329⇌280 condemn
1,671⇌170 violence
Michelle 261⇌3,115
our 41⇌72
7,911⇌321 August
President 93⇌228
voters 306⇌4,453
1,325⇌187 supremacy
people 27⇌45
candidate 362⇌5,584
1,761⇌231 police
women 124⇌315
52.9%—47.1%



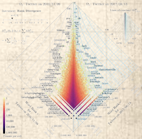
Rank-turbulence histogram:
Ω_1: Twitter on 2016/11/09
Ω_2: Twitter on 2017/08/13
B.
C.
D.
E.
F.
G.
A.
RT
the
to
and
is
for
you
in
be
it
are
my
Trump
was
will
by
our
My
America
them
going
hate
Donald
White
won
video
Hillary
Lady
election
Gaga
elected
violence
Michelle
Nazi
voters
Nazis
candidate
August
Election
Zara
electoral
Charlottesville
gorilla
Marshawn
opport
Matic
hatemonger
Antifa
HURRICANE
tiki
Meteorite
Heyer
whitelash
GD's
JabberDuck
Cvjetanovic
surro
FirstStar...tar
Bobama
GSHDJHS
misogy
uncharact...cal
abusiv
KoKoBop
Calexit'
shower-gel
Waistlines
Klansfolk
Spofford
froy
Jtrinity
DEPENDANCE
ascertainment
Zarrick
suede-denim
richava
1
10
100
1,000
10,000
100,000
1,000,000
10,000,000
← more frequent
less frequent →
more frequent →
← less frequent
Rank r for Twitter on 2016/11/09
Rank r for Twitter on 2017/08/13
Counts per cell
Balances:
59.9% total counts 40.1%
63.2% all words 61.6%
60.8% exclusive words 59.8%
H. Identical systems:
I. Randomized systems:
J. Disjoint systems:



Ω_1: Twitter on 2016/11/09
Ω_2: Twitter on 2017/08/13
A.
Instrument: Rank-Turbulence Divergence
α=1/3
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{\mathrm{R}}_{1/3}(\Omega_1 \| \Omega_2) = 0.493$
$\propto \sum_\tau \left| \frac{1}{r_{\tau,1}^{1/3}} - \frac{1}{r_{\tau,2}^{1/3}} \right|^{3/4}$
B.
C.
Lines of Constant $\delta D^{\mathrm{R}}_{1/3,\tau}$
2.15×10⁻⁵ 6.74×10⁻⁶ 2.00×10⁻⁶ 4.29×10⁻⁷
D.
Divergence contribution $\delta D^{\mathrm{R}}_{1/3,\tau}$ (×10⁻³%)
1 0.5 0 0.5 1
Trump 11⇌60
67,220⇌113 Charlottesville
election 64⇌2,055
voted 58⇌1,002
Hillary 70⇌1,505
Donald 50⇌566
9,149⇌129 Nazis
president 48⇌500
trump 77⇌1,357
139⇌26 My
37,952⇌268 Larsson
5,873⇌171 supremacists
25,126⇌267 Zara
won 69⇌536
862,482⇌445 Heyer
Clinton 125⇌1,761
elected 151⇌2,787
3,485⇌174 Nazi
13,329⇌280 condemn
America 40⇌164
3,563⇌192 BTS
86⇌27 his
1,175⇌124 Gaga
1,552,865⇌672 Cvjetanovic ▷
E.
Obama 76⇌378
wins 144⇌1,209
1,671⇌170 violence
7,911⇌321 August
801⇌119 Lady
18,804⇌442 nazis
16,317⇌446 Heather
electoral 446⇌15,272
47,553⇌610 torches
JabberDuck 993⇌840,082
hatemonger 756⇌120,186
102,414⇌743 Antifa
opport 672⇌62,603
56,244⇌657 neo-nazis
Harambe 658⇌56,005
Michelle 261⇌3,115
F.
50.2%—49.8%
RT the to is for in it and you be are my was by Trump will our My America going them hate White Donald won video voted election elected Lady Gaga violence Michelle voters Nazi candidate Nazis August Election electoral Zara Charlottesville gorilla opport Marshawn torches hatemonger Antifa BRITAIN Neo-Nazi Meteorite whitelash GD's Heyer JabberDuck Cvjetanovic surro FirstStar...tar Bobama GSHDJHS misogy uncharact...cal abusiv KoKoBop Calexit' shower-gel Waistlines Klansfolk Spofford froy Jtrinity DEPENDANCE ascertainment Zarrick suede-denim richava
1 10 100 1,000 10,000 100,000 1,000,000 10,000,000
← more frequent
more frequent →
less frequent →
← less frequent
Rank r for Twitter on 2016/11/09
Rank r for Twitter on 2017/08/13
1 10 100 1,000 10,000 100,000 1,000,000
Counts per cell
Balances:
59.9% total counts 40.1%
63.2% all words 61.6%
60.8% exclusive words 59.8%

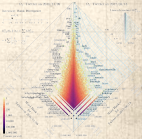
Desirable rank-turbulence divergence features:

1. Rank-based.



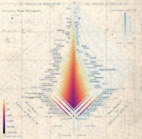
Desirable rank-turbulence divergence features:

1. Rank-based.
2. Symmetric.



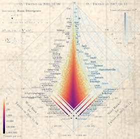
Desirable rank-turbulence divergence features:

1. Rank-based.
2. Symmetric.
3. Semi-positive: $D^{\mathsf{R}}_{\alpha}(\Omega_1 \,||\, \Omega_2) \geq 0$.



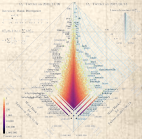
Desirable rank-turbulence divergence features:

1. Rank-based.
2. Symmetric.
3. Semi-positive: $D^{\mathsf{R}}_{\alpha}(\Omega_1 \,||\, \Omega_2) \geq 0$.
4. Linearly separable, for interpretability.



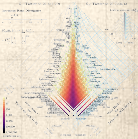
Desirable rank-turbulence divergence features:

1. Rank-based.
2. Symmetric.
3. Semi-positive: $D^{\mathsf{R}}_{\alpha}(\Omega_1 \,||\, \Omega_2) \geq 0$.
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.).



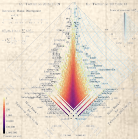
Desirable rank-turbulence divergence features:

1. Rank-based.
2. Symmetric.
3. Semi-positive: $D^{\mathsf{R}}_{\alpha}(\Omega_1 \,||\, \Omega_2) \geq 0$.
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.



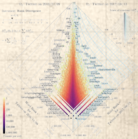
Desirable rank-turbulence divergence features:

1. Rank-based.
2. Symmetric.
3. Semi-positive: $D^{\mathsf{R}}_{\alpha}(\Omega_1 \,||\, \Omega_2) \geq 0$.
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.



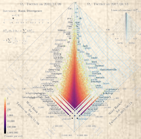
Desirable rank-turbulence divergence features:

1. Rank-based.
2. Symmetric.
3. Semi-positive: $D^{\mathsf{R}}_{\alpha}(\Omega_1 \,||\, \Omega_2) \geq 0$.
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.

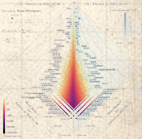


Desirable rank-turbulence divergence features:

1. Rank-based.
2. Symmetric.
3. Semi-positive: $D^{\mathsf{R}}_{\alpha}(\Omega_1 \,||\, \Omega_2) \geq 0$.
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’

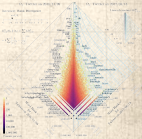


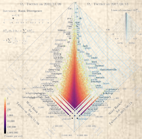
Some good things about ranks:



Some good things about ranks:

- Working with ranks is intuitive



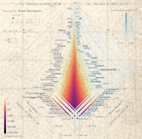


Some good things about ranks:

- Working with ranks is intuitive
- Affords some powerful statistics (e.g., Spearman's rank correlation coefficient)

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



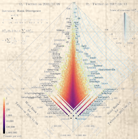
Some good things about ranks:

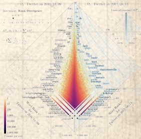
- Working with ranks is intuitive
- Affords some powerful statistics (e.g., Spearman's rank correlation coefficient)
- Can be used to generalize beyond systems with probabilities

A start:

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

- 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'





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

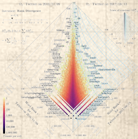
A start:

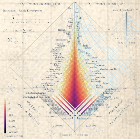
$$\left| \frac{1}{r_{\tau,1}} - \frac{1}{r_{\tau,2}} \right| . \tag{5}$$

- 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

We introduce a tuning parameter:

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

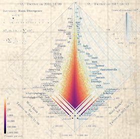




We introduce a tuning parameter:

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

- As $\alpha \rightarrow 0$, high ranked components are increasingly dampened



We introduce a tuning parameter:

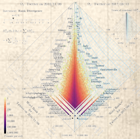
$$\left| \frac{1}{\left[r_{\tau,1}\right]^{\alpha}} - \frac{1}{\left[r_{\tau,2}\right]^{\alpha}} \right|^{1/\alpha} . \tag{6}$$

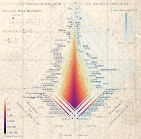
- As $\alpha \rightarrow 0$, high ranked components are increasingly dampened
- For words in texts, for example, the weight of common words and rare words move increasingly closer together.

We introduce a tuning parameter:

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

- As $\alpha \rightarrow 0$, 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 $\alpha \rightarrow \infty$, high rank components will dominate.





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.

Trouble:

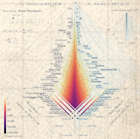
- The limit of $\alpha \to 0$ does not behave well for

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

- The leading order term is:

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

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



Trouble:

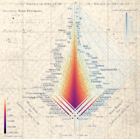
- The limit of $\alpha \to 0$ does not behave well for

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

- The leading order term is:

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

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



Trouble:

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

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

- The leading order term is:

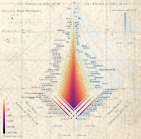
$$\left(1 - \delta_{r_{\tau,1} r_{\tau,2}}\right) \alpha^{1/\alpha} \left| \ln \frac{r_{\tau,1}}{r_{\tau,2}} \right|^{1/\alpha}, \qquad (7)$$

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

- Oops.
- But the insides look nutritious:

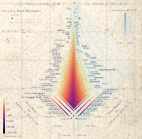
$$\left| \ln \frac{r_{\tau,1}}{r_{\tau,2}} \right|$$

is a nicely interpretable log-ratio of ranks.



Some reworking:

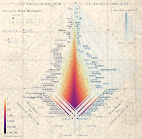
$$\delta D^{\mathsf{R}}_{\alpha,\tau}(R_1 \,\|\, R_2) \propto \frac{\alpha+1}{\alpha} \left| \frac{1}{\left[r_{\tau,1}\right]^{\alpha}} - \frac{1}{\left[r_{\tau,2}\right]^{\alpha}} \right|^{1/(\alpha+1)} . \tag{8}$$



Some reworking:

$$\delta D^{\mathsf{R}}_{\alpha,\tau}(R_1 \,\|\, R_2) \propto \frac{\alpha+1}{\alpha} \left| \frac{1}{\left[r_{\tau,1}\right]^{\alpha}} - \frac{1}{\left[r_{\tau,2}\right]^{\alpha}} \right|^{1/(\alpha+1)}. \tag{8}$$

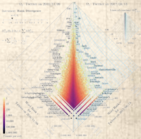
- Keeps the core structure.



Some reworking:

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

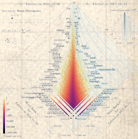
- Keeps the core structure.
- Large α limit remains the same.



Some reworking:

$$\delta D^{\mathrm{R}}_{\alpha,\tau}(R_1 \,\|\, R_2) \propto \frac{\alpha+1}{\alpha} \left| \frac{1}{\left[r_{\tau,1}\right]^{\alpha}} - \frac{1}{\left[r_{\tau,2}\right]^{\alpha}} \right|^{1/(\alpha+1)} . \tag{8}$$

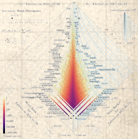
- Keeps the core structure.
- Large α limit remains the same.
- $\alpha \rightarrow 0$ limit now returns log-ratio of ranks.



Some reworking:

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

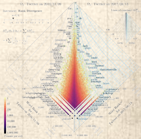
- Keeps the core structure.
- Large α limit remains the same.
- $\alpha \rightarrow 0$ limit now returns log-ratio of ranks.
- Next: Sum over τ to get divergence.



Some reworking:

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

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



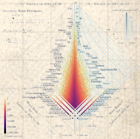
Some reworking:

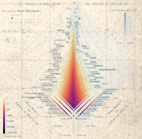
$$\delta D^{\mathrm{R}}_{\alpha,\tau}(R_1 \,||\, R_2) \propto \frac{\alpha+1}{\alpha} \left| \frac{1}{\left[r_{\tau,1}\right]^{\alpha}} - \frac{1}{\left[r_{\tau,2}\right]^{\alpha}} \right|^{1/(\alpha+1)} . \quad (8)$$

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

Rank-turbulence divergence:

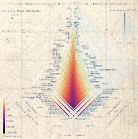
$$D^{\mathrm{R}}_{\alpha}(R_1 \,||\, R_2) = \frac{1}{\mathcal{N}_{1,2;\alpha}} \sum_{\tau \in R_{1,2;\alpha}} \delta D^{\mathrm{R}}_{\alpha,\tau}(R_1 \,||\, R_2) \quad (9)$$





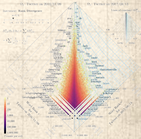
Normalization:

- Take a data-driven rather than analytic approach to determining $\mathcal{N}_{1,2;\alpha}$.



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.

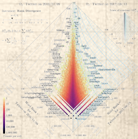


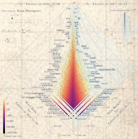
Normalization:

- 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: $0 \leq D^{\mathsf{R}}_{\alpha}(R_1 \,\|\, R_2) \leq 1$

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: $0 \leq D^{\mathsf{R}}_{\alpha}(R_1 \,\|\, R_2) \leq 1$
- Limits of 0 and 1 correspond to the two systems having identical and disjoint Zipf distributions.

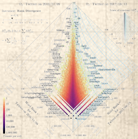




Rank-turbulence divergence:

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

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

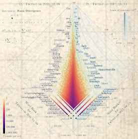


General normalization:

- 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 system.

General normalization:

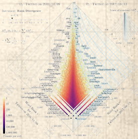
- Iif the Zipf distributions are disjoint, then in $\Omega^{(1)}$'s merged ranking, the rank of all $\Omega^{(2)}$ types will be $r = N_1 + \frac{1}{2}N_2$, where N_1 and N_2 are the number of distinct types in each system.
- Similarly, $\Omega^{(2)}$'s merged ranking will have all of $\Omega^{(1)}$'s types in last place with rank $r = N_2 + \frac{1}{2}N_1$.

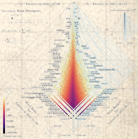


General normalization:

- 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 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:

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





Limit of $\alpha \rightarrow 0$:

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

where

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

- Largest rank ratios dominate.

Limit of $\alpha \to \infty$:

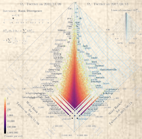
$$D^{\mathsf{R}}_{\infty}(R_1 \,\|\, R_2) = \sum_{\tau \in R_{1,2;\alpha}} \delta D^{\mathsf{R}}_{\infty,\tau}$$

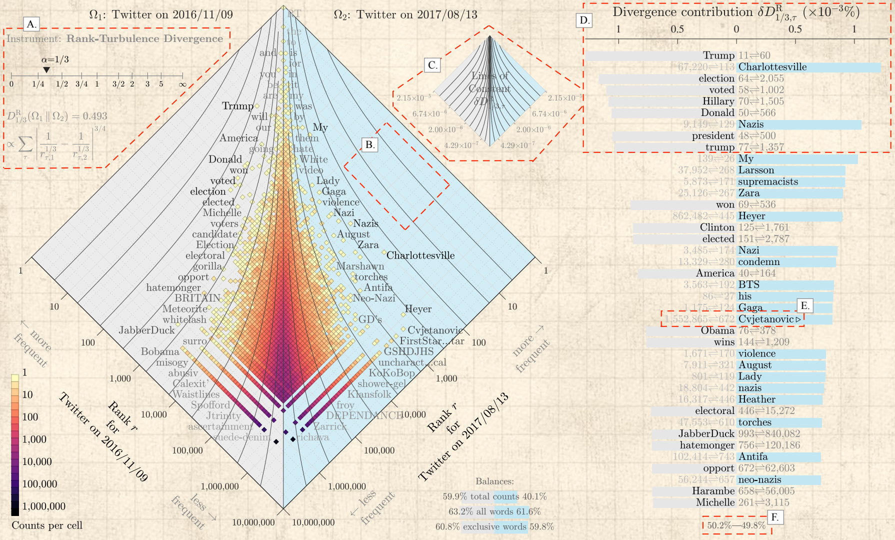
$$= \frac{1}{\mathcal{N}_{1,2;\infty}} \sum_{\tau \in R_{1,2;\alpha}} \left(1 - \delta_{r_{\tau,1} r_{\tau,2}}\right) \max_{\tau} \left\{ \frac{1}{r_{\tau,1}}, \frac{1}{r_{\tau,2}} \right\}. \tag{14}$$

where

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

- Highest ranks dominate.



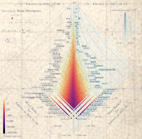


Ω_1: Twitter on 2016/11/09
Ω_2: Twitter on 2017/08/13
A.
Instrument: Rank-Turbulence Divergence
α=1/3
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{R}_{1/3}(\Omega_1 \| \Omega_2) = 0.493$
$\propto \sum_{\tau} \left| \frac{1}{r_{\tau,1}^{1/3}} - \frac{1}{r_{\tau,2}^{1/3}} \right|^{3/4}$
B.
C.
Lines of Constant $\delta D^{R}_{1/3,\tau}$
2.15×10^{-5}
6.74×10^{-6}
2.00×10^{-6}
4.29×10^{-7}
D.
Divergence contribution $\delta D^{R}_{1/3,\tau}$ (×10^{-3}%)
1 0.5 0 0.5 1
Trump 11⇌60
67,220⇌113 Charlottesville
election 64⇌2,055
voted 58⇌1,002
Hillary 70⇌1,505
Donald 50⇌566
9,149⇌129 Nazis
president 48⇌500
trump 77⇌1,357
139⇌26 My
37,952⇌268 Larsson
5,873⇌171 supremacists
25,126⇌267 Zara
won 69⇌536
862,482⇌445 Heyer
Clinton 125⇌1,761
elected 151⇌2,787
3,485⇌174 Nazi
13,329⇌280 condemn
America 40⇌164
3,563⇌192 BTS
86⇌27 his
1,175⇌124 Gaga
1,552,865⇌672 Cvjetanovic ▷
E.
Obama 76⇌378
wins 144⇌1,209
1,671⇌170 violence
7,911⇌321 August
801⇌119 Lady
18,804⇌442 nazis
16,317⇌446 Heather
electoral 446⇌15,272
47,553⇌610 torches
JabberDuck 993⇌840,082
hatemonger 756⇌120,186
102,414⇌743 Antifa
opport 672⇌62,603
56,244⇌657 neo-nazis
Harambe 658⇌56,005
Michelle 261⇌3,115
F.
50.2%—49.8%
RT the to is for in it my was by
and you be are will our going
Trump America Donald won voted election elected Michelle voters candidate Election electoral gorilla opport hatemonger BRITAIN Meteorite whitelash JabberDuck surro Bobama misogy abusiv Calexit' Waistlines Spofford Jtrinity ascertainment suede-denim
My them hate White video Lady Gaga violence Nazi Nazis August Zara Charlottesville Marshawn torches Antifa Neo-Nazi GD's Heyer Cvjetanovic FirstStar...tar GSHDJHS uncharact...cal KoKoBop shower-gel Klansfold froy DEPENDANCE Zarrick richava
1 10 100 1,000 10,000 100,000 1,000,000 10,000,000
← more frequent
more frequent →
less frequent →
← less frequent
Rank r for Twitter on 2016/11/09
Rank r for Twitter on 2017/08/13
1 10 100 1,000 10,000 100,000 1,000,000
Counts per cell
Balances:
59.9% total counts 40.1%
63.2% all words 61.6%
60.8% exclusive words 59.8%

Probability-turbulence divergence:

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

- For the unnormalized version ($\mathcal{N}^{\mathsf{P}}_{1,2;\alpha}$=1), some troubles return with 0 probabilities and $\alpha \to 0$.
- Weep not: $\mathcal{N}^{\mathsf{P}}_{1,2;\alpha}$ 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 two systems' types:

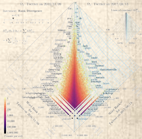
$$\mathcal{N}^{\mathsf{P}}_{1,2;\alpha} = \frac{\alpha+1}{\alpha}\sum_{\tau\in R_1}\left[p_{\tau,1}\right]^{\alpha/(\alpha+1)} + \frac{\alpha+1}{\alpha}\sum_{\tau\in R_2}\left[p_{\tau,2}\right]^{\alpha/(\alpha+1)} \tag{17}$$

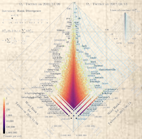
Limit of α=0 for probability-turbulence divergence

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

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

- 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:

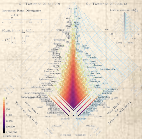
$$\mathcal{N}^{\mathsf{P}}_{1,2;\alpha} \rightarrow \frac{1}{\alpha}\left(N_1 + N_2\right). \tag{19}$$

- 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.

Combine these cases into a single expression:

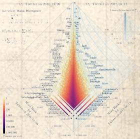
$$D_0^{\mathsf{P}}(P_1 \,\|\, P_2) = \frac{1}{(N_1 + N_2)} \sum_{\tau \in R_{1,2;0}} \left(\delta_{p_{\tau,1},0} + \delta_{0,p_{\tau,2}} \right). \tag{20}$$

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



Type contribution ordering for the limit of α=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 $D_0^{\mathsf{P}}(P_1 \,\|\, P_2)$, 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 ratio.

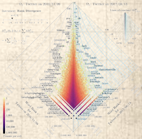


Limit of α=∞ for probability-turbulence divergence

$$D^{\mathsf{P}}_{\infty}(P_1 \,\|\, P_2) = \frac{1}{2} \sum_{\tau \in R_{1,2;\infty}} \left(1 - \delta_{p_{\tau,1},p_{\tau,2}}\right) \max\left(p_{\tau,1}, p_{\tau,2}\right) \tag{21}$$

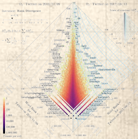
where

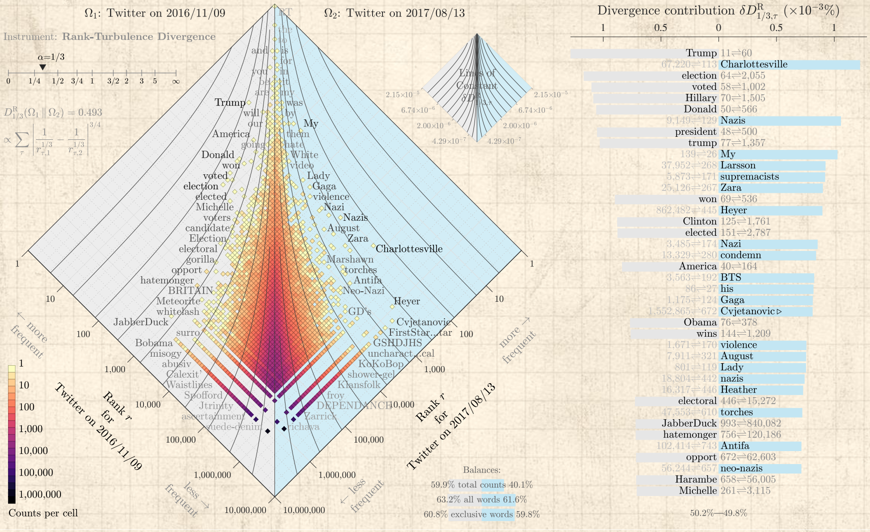
$$\mathcal{N}^{\mathsf{P}}_{1,2;\infty} = \sum_{\tau \in R_{1,2;\infty}} \left(p_{\tau,1} + p_{\tau,2} \right) = 1 + 1 = 2. \tag{22}$$



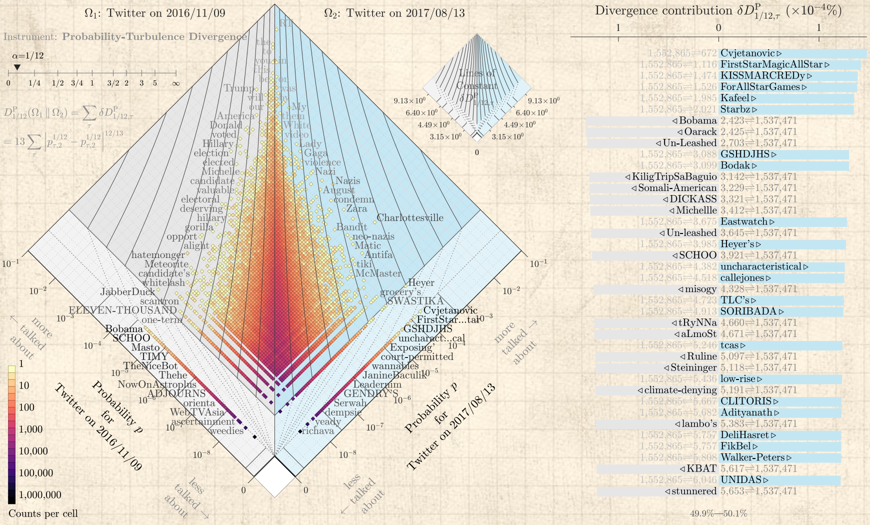
Connections for PTD:

- $\alpha = 0$: Similarity measure Sørensen-Dice coefficient [10, 35, 22], F_1 score of a test's accuracy [36, 33].
- $\alpha = 1/2$: Hellinger distance [18] and Mautusita distance [23].
- $\alpha = 1$: Many including all $L^{(p)}$-norm type constructions.
- $\alpha = \infty$: Motyka distance [9].

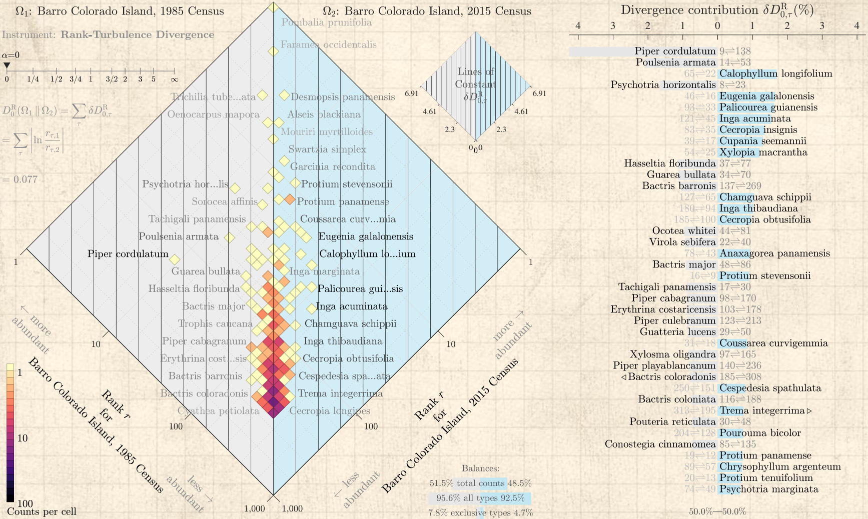


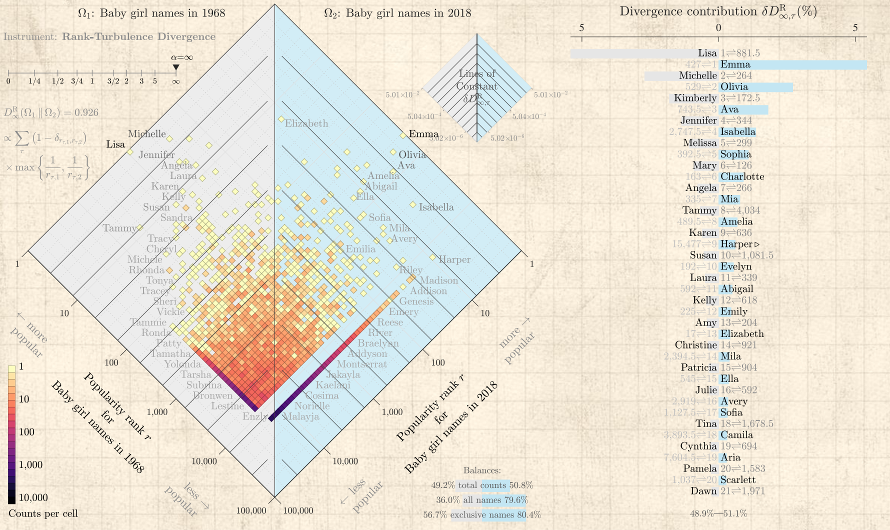


Ω1: Twitter on 2016/11/09
Ω2: Twitter on 2017/08/13
Instrument: Rank-Turbulence Divergence
α=1/3
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{R}_{1/3}(\Omega_1 \,\|\, \Omega_2) = 0.493$
$\propto \sum_{\tau} \left| \frac{1}{r_{\tau,1}^{1/3}} - \frac{1}{r_{\tau,2}^{1/3}} \right|^{3/4}$
Lines of Constant $\delta D^{R}_{1/3,\tau}$
2.15×10^-5 6.74×10^-6 2.00×10^-6 4.29×10^-7
Rank r for Twitter on 2016/11/09
Rank r for Twitter on 2017/08/13
← more frequent
more frequent →
less frequent →
← less frequent
1 10 100 1,000 10,000 100,000 1,000,000 10,000,000
Counts per cell
1 10 100 1,000 10,000 100,000 1,000,000
RT the to and is for you in be it are my was will by our My America them going hate White Donald won video voted Lady election Gaga elected violence Michelle Nazi voters Nazis candidate August Election Zara electoral Charlottesville gorilla Marshawn opport torches hatemonger Antifa BRITAIN Neo-Nazi Meteorite Heyer whitelash GD's JabberDuck Cvjetanovic surro FirstStar...tar Bobama GSHDJHS misogy uncharact...cal abusiv KoKoBop Calexit' shower-gel Waistlines Klansfolk Spofford froy Jtrinity DEPENDANCE ascertainment Zarrick suede-denim richava Trump
Balances:
59.9% total counts 40.1%
63.2% all words 61.6%
60.8% exclusive words 59.8%
Divergence contribution $\delta D^{R}_{1/3,\tau}$ (×10^-3%)
1 0.5 0 0.5 1
Trump 11⇌60
67,220⇌113 Charlottesville
election 64⇌2,055
voted 58⇌1,002
Hillary 70⇌1,505
Donald 50⇌566
9,149⇌129 Nazis
president 48⇌500
trump 77⇌1,357
139⇌26 My
37,952⇌268 Larsson
5,873⇌171 supremacists
25,126⇌267 Zara
won 69⇌536
862,482⇌445 Heyer
Clinton 125⇌1,761
elected 151⇌2,787
3,485⇌174 Nazi
13,329⇌280 condemn
America 40⇌164
3,563⇌192 BTS
86⇌27 his
1,175⇌124 Gaga
1,552,865⇌672 Cvjetanovic ▷
Obama 76⇌378
wins 144⇌1,209
1,671⇌170 violence
7,911⇌321 August
801⇌119 Lady
18,804⇌442 nazis
16,317⇌446 Heather
electoral 446⇌15,272
47,553⇌610 torches
JabberDuck 993⇌840,082
hatemonger 756⇌120,186
102,414⇌743 Antifa
opport 672⇌62,603
56,244⇌657 neo-nazis
Harambe 658⇌56,005
Michelle 261⇌3,115
50.2%—49.8%

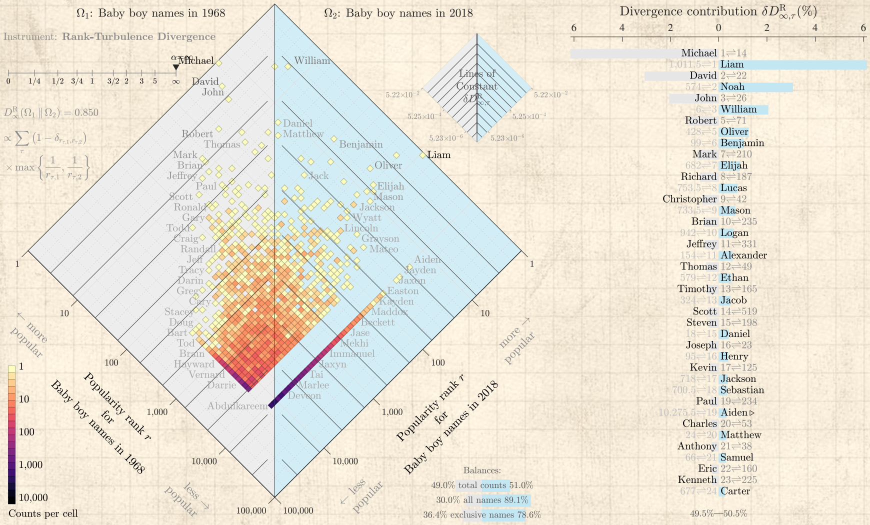


Ω1: Twitter on 2016/11/09
Ω2: Twitter on 2017/08/13
Instrument: Probability-Turbulence Divergence
α=1/12
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{\mathrm{P}}_{1/12}(\Omega_1 \,\|\, \Omega_2) = \sum_\tau \delta D^{\mathrm{P}}_{1/12,\tau}$
$= 13 \sum_\tau \left| p_{\tau,2}^{1/12} - p_{\tau,2}^{1/12} \right|^{12/13}$
Lines of Constant $\delta D^{\mathrm{P}}_{1/12,\tau}$
9.13×10^0 6.40×10^0 4.49×10^0 3.15×10^0
Probability p for Twitter on 2016/11/09
Probability p for Twitter on 2017/08/13
← more talked about
more talked about →
less talked about →
← less talked about
Counts per cell
1 10 100 1,000 10,000 100,000 1,000,000
RT the to you in this be on was Trump will our by My them America White Donald video voted Lady Hillary Gaga election violence elected Nazi Michelle Nazis candidate August valuable condemn electoral Zara deserving hillary Charlottesville gorilla Bandit opport neo-nazis alight Matic hatemonger Antifa Meteorite tiki candidate's McMaster whitelash Heyer JabberDuck grocery's scantron SWASTIKA ELEVEN-THOUSAND Cvjetanovic one-term FirstStar...tar Bobama GSHDJHS SCHOO uncharact...cal Masto Exposing' TIMY court-permitted TheNiceBot wannabies Thehe JanineBaculik NowOnAstroplus Leadernim ADJOURNS GENDRY'S orienta Serwah WebTVAsia dempsie ascertainment yeady weedies richava
Divergence contribution $\delta D^{\mathrm{P}}_{1/12,\tau}$ (×10^{-4}%)
1 0 1
1,552,865⇌672 Cvjetanovic ▷
1,552,865⇌1,116 FirstStarMagicAllStar ▷
1,552,865⇌1,474 KISSMARCREDy ▷
1,552,865⇌1,526 ForAllStarGames ▷
1,552,865⇌1,985 Kafeel ▷
1,552,865⇌2,021 Starbz ▷
◁ Bobama 2,423⇌1,537,471
◁ Oarack 2,425⇌1,537,471
◁ Un-Leashed 2,703⇌1,537,471
1,552,865⇌3,088 GSHDJHS ▷
1,552,865⇌3,099 Bodak ▷
◁ KiligTripSaBaguio 3,142⇌1,537,471
◁ Somali-American 3,229⇌1,537,471
◁ DICKASS 3,321⇌1,537,471
◁ Michellle 3,412⇌1,537,471
1,552,865⇌3,675 Eastwatch ▷
◁ Un-leashed 3,645⇌1,537,471
1,552,865⇌3,985 Heyer's ▷
◁ SCHOO 3,921⇌1,537,471
1,552,865⇌4,382 uncharacteristical ▷
1,552,865⇌4,518 callejones ▷
◁ misogy 4,328⇌1,537,471
1,552,865⇌4,723 TLC's ▷
1,552,865⇌4,913 SORIBADA ▷
◁ tRyNNa 4,660⇌1,537,471
◁ aLmoSt 4,671⇌1,537,471
1,552,865⇌5,246 tcas ▷
◁ Ruline 5,097⇌1,537,471
◁ Steininger 5,118⇌1,537,471
1,552,865⇌5,436 low-rise ▷
◁ climate-denying 5,191⇌1,537,471
1,552,865⇌5,667 CLITORIS ▷
1,552,865⇌5,682 Adityanath ▷
◁ lambo's 5,383⇌1,537,471
1,552,865⇌5,757 DeliHasret ▷
1,552,865⇌5,757 FikBel ▷
1,552,865⇌5,808 Walker-Peters ▷
◁ KBAT 5,617⇌1,537,471
1,552,865⇌6,046 UNIDAS ▷
◁ stunnered 5,653⇌1,537,471
49.9%—50.1%

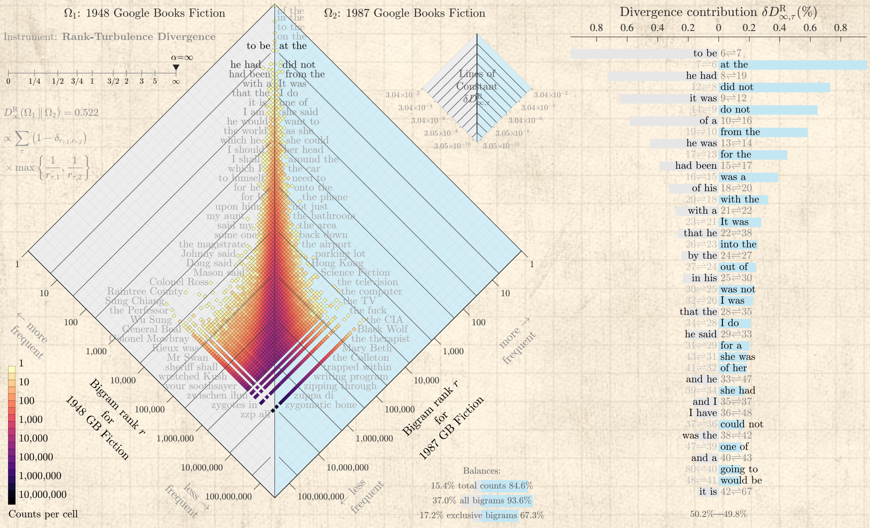

Ω_1: Barro Colorado Island, 1985 Census
Ω_2: Barro Colorado Island, 2015 Census
Instrument: Rank-Turbulence Divergence
α=0
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{R}_{0}(\Omega_1 \,\|\, \Omega_2) = \sum_\tau \delta D^{R}_{0,\tau}$
$= \sum_\tau \left| \ln \frac{r_{\tau,1}}{r_{\tau,2}} \right|$
= 0.077
Lines of Constant $\delta D^{R}_{0,\tau}$
6.91 4.61 2.3 0 0 2.3 4.61 6.91
Pombalia prunifolia
Faramea occidentalis
Trichilia tube...ata
Desmopsis panamensis
Oenocarpus mapora
Alseis blackiana
Mouriri myrtilloides
Swartzia simplex
Garcinia recondita
Psychotria hor...lis
Protium stevensonii
Sorocea affinis
Protium panamense
Tachigali panamensis
Coussarea curv...mia
Poulsenia armata
Eugenia galalonensis
Piper cordulatum
Calophyllum lo...ium
Guarea bullata
Inga marginata
Hasseltia floribunda
Palicourea gui...sis
Bactris major
Inga acuminata
Trophis caucana
Chamguava schippii
Piper cabagranum
Inga thibaudiana
Erythrina cost...sis
Cecropia obtusifolia
Bactris barronis
Cespedesia spa...ata
Bactris coloradonis
Trema integerrima
Cyathea petiolata
Cecropia longipes
1 10 100 1,000
← more abundant
less abundant →
more → abundant
← less abundant
Rank r for Barro Colorado Island, 1985 Census
Rank r for Barro Colorado Island, 2015 Census
1 10 100
Counts per cell
Balances:
51.5% total counts 48.5%
95.6% all types 92.5%
7.8% exclusive types 4.7%
Divergence contribution $\delta D^{R}_{0,\tau}$ (%)
4 3 2 1 0 1 2 3 4
Piper cordulatum 9⇌138
Poulsenia armata 14⇌53
65⇌22 Calophyllum longifolium
Psychotria horizontalis 8⇌23
46⇌16 Eugenia galalonensis
93⇌33 Palicourea guianensis
121⇌45 Inga acuminata
83⇌35 Cecropia insignis
39⇌17 Cupania seemannii
54⇌25 Xylopia macrantha
Hasseltia floribunda 37⇌77
Guarea bullata 34⇌70
Bactris barronis 137⇌269
127⇌65 Chamguava schippii
180⇌94 Inga thibaudiana
185⇌100 Cecropia obtusifolia
Ocotea whitei 44⇌81
Virola sebifera 22⇌40
78⇌43 Anaxagorea panamensis
Bactris major 48⇌86
16⇌9 Protium stevensonii
Tachigali panamensis 17⇌30
Piper cabagranum 98⇌170
Erythrina costaricensis 103⇌178
Piper culebranum 123⇌213
Guatteria lucens 29⇌50
31⇌18 Coussarea curvigemmia
Xylosma oligandra 97⇌165
Piper playablancanum 140⇌236
◃Bactris coloradonis 185⇌308
250⇌151 Cespedesia spathulata
Bactris coloniata 116⇌188
313⇌195 Trema integerrima▹
Pouteria reticulata 30⇌48
204⇌128 Pourouma bicolor
Conostegia cinnamomea 85⇌135
19⇌12 Protium panamense
89⇌57 Chrysophyllum argenteum
20⇌13 Protium tenuifolium
74⇌49 Psychotria marginata
50.0%—50.0%



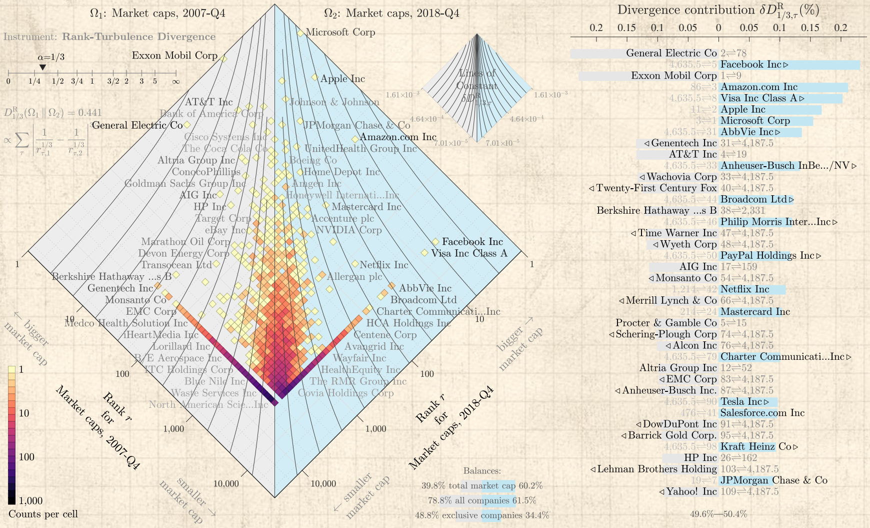
Ω_1: Baby girl names in 1968
Ω_2: Baby girl names in 2018
Instrument: Rank-Turbulence Divergence
α=∞
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{\mathrm{R}}_{\infty}(\Omega_1 \,\|\, \Omega_2) = 0.926$
$\propto \sum_{\tau} \left(1 - \delta_{r_{\tau,1}, r_{\tau,2}}\right)$
$\times \max\left\{\frac{1}{r_{\tau,1}}, \frac{1}{r_{\tau,2}}\right\}$
Lines of Constant $\delta D^{\mathrm{R}}_{\infty,\tau}$
5.01×10^{-2}
5.04×10^{-4}
5.02×10^{-6}
Michelle Lisa Jennifer Angela Laura Karen Kelly Susan Sandra Tammy Tracy Cheryl Michele Rhonda Tonya Tracey Sheri Vickie Tammie Ronda Patty Tamatha Yolonda Tarsha Subrina Bronwen Lestine Enzly
Elizabeth Emma Olivia Ava Amelia Abigail Ella Isabella Sofia Mila Avery Emilia Harper Riley Madison Addison Genesis Emery Reese River Braelynn Addyson Montserrat Jakayla Kaelani Cosima Norielle Malayja
1 10 100 1,000 10,000 100,000
← more popular
more → popular
less → popular
← less popular
Popularity rank r for Baby girl names in 1968
Popularity rank r for Baby girl names in 2018
1 10 100 1,000 10,000
Counts per cell
Balances:
49.2% total counts 50.8%
36.0% all names 79.6%
56.7% exclusive names 80.4%
Divergence contribution $\delta D^{\mathrm{R}}_{\infty,\tau}(\%)$
5 0 5
Lisa 1⇌881.5
427⇌1 Emma
Michelle 2⇌264
529⇌2 Olivia
Kimberly 3⇌172.5
743.5⇌3 Ava
Jennifer 4⇌344
2,747.5⇌4 Isabella
Melissa 5⇌299
392.5⇌5 Sophia
Mary 6⇌126
163⇌6 Charlotte
Angela 7⇌266
335⇌7 Mia
Tammy 8⇌4,034
489.5⇌8 Amelia
Karen 9⇌636
15,477⇌9 Harper ▷
Susan 10⇌1,081.5
192⇌10 Evelyn
Laura 11⇌339
592⇌11 Abigail
Kelly 12⇌618
225⇌12 Emily
Amy 13⇌204
17⇌13 Elizabeth
Christine 14⇌921
2,394.5⇌14 Mila
Patricia 15⇌904
545⇌15 Ella
Julie 16⇌592
2,919⇌16 Avery
1,127.5⇌17 Sofia
Tina 18⇌1,678.5
3,893.5⇌18 Camila
Cynthia 19⇌694
7,604.5⇌19 Aria
Pamela 20⇌1,583
1,037⇌20 Scarlett
Dawn 21⇌1,971
48.9%—51.1%



Ω_1: Baby boy names in 1968
Ω_2: Baby boy names in 2018
Instrument: **Rank-Turbulence Divergence**
$\alpha = \infty$
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{R}_{\infty}(\Omega_1 \| \Omega_2) = 0.850$
$\propto \sum_{\tau} (1 - \delta_{r_{\tau,1}, r_{\tau,2}})$
$\times \max \left\{ \frac{1}{r_{\tau,1}}, \frac{1}{r_{\tau,2}} \right\}$
Lines of Constant $\delta D^{R}_{\infty,\tau}$
5.22×10^{-2}
5.25×10^{-4}
5.23×10^{-6}
Michael
David
John
Robert
Thomas
Mark
Brian
Jeffrey
Paul
Scott
Ronald
Gary
Todd
Craig
Randall
Jeff
Tracy
Darin
Greg
Cary
Stacey
Doug
Bart
Tod
Brain
Hayward
Vernard
Darrie
Abdulkareem
William
Daniel
Matthew
Benjamin
Liam
Oliver
Jack
Elijah
Mason
Jackson
Wyatt
Lincoln
Grayson
Mateo
Aiden
Jayden
Jaxon
Easton
Kayden
Maddox
Beckett
Jase
Mekhi
Immanuel
Jaxyn
Tai
Marlee
Deveon
1
10
100
1,000
10,000
100,000
← more popular
more → popular
less → popular
← less popular
Popularity rank r for Baby boy names in 1968
Popularity rank r for Baby boy names in 2018
Counts per cell
1
10
100
1,000
10,000
Balances:
49.0% total counts 51.0%
30.0% all names 89.1%
36.4% exclusive names 78.6%
Divergence contribution $\delta D^{R}_{\infty,\tau}$(%)
6 4 2 0 2 4 6
Michael 1⇌14
1,011.5⇌1 Liam
David 2⇌22
574⇌2 Noah
John 3⇌26
6⇌3 William
Robert 5⇌71
428⇌5 Oliver
99⇌6 Benjamin
Mark 7⇌210
682⇌7 Elijah
Richard 8⇌187
753.5⇌8 Lucas
Christopher 9⇌42
733.5⇌9 Mason
Brian 10⇌235
942⇌10 Logan
Jeffrey 11⇌331
154⇌11 Alexander
Thomas 12⇌49
579⇌12 Ethan
Timothy 13⇌165
324⇌13 Jacob
Scott 14⇌519
Steven 15⇌198
18⇌15 Daniel
Joseph 16⇌23
95⇌16 Henry
Kevin 17⇌125
718⇌17 Jackson
700.5⇌18 Sebastian
Paul 19⇌234
10,275.5⇌19 Aiden ▷
Charles 20⇌53
24⇌20 Matthew
Anthony 21⇌38
66⇌21 Samuel
Eric 22⇌160
Kenneth 23⇌225
677⇌24 Carter
49.5%—50.5%



Ω1: 1948 Google Books Fiction
Ω2: 1987 Google Books Fiction
Instrument: Rank-Turbulence Divergence
α=∞
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{R}_{\infty}(\Omega_1 \,\|\, \Omega_2) = 0.522$
$\propto \sum_{\tau} (1 - \delta_{r_{\tau,1}, r_{\tau,2}})$
$\times \max\left\{\frac{1}{r_{\tau,1}}, \frac{1}{r_{\tau,2}}\right\}$
Lines of Constant $\delta D^{R}_{\infty,\tau}$
Counts per cell
1 10 100 1,000 10,000 100,000 1,000,000 10,000,000
Bigram rank r for 1948 GB Fiction
Bigram rank r for 1987 GB Fiction
← more frequent
more frequent →
less frequent →
← less frequent
Balances:
15.4% total counts 84.6%
37.0% all bigrams 93.6%
17.2% exclusive bigrams 67.3%
Divergence contribution $\delta D^{R}_{\infty,\tau}$ (%)
0.8 0.6 0.4 0.2 0 0.2 0.4 0.6 0.8
to be 6⇌7
7⇌6 at the
he had 8⇌19
12⇌8 did not
it was 9⇌12
14⇌9 do not
of a 10⇌16
19⇌10 from the
he was 13⇌14
17⇌13 for the
had been 15⇌17
16⇌15 was a
of his 18⇌20
20⇌18 with the
with a 21⇌22
23⇌21 It was
that he 22⇌38
26⇌23 into the
by the 24⇌27
27⇌24 out of
in his 25⇌30
30⇌25 was not
32⇌26 I was
that the 28⇌35
34⇌28 I do
he said 29⇌33
31⇌29 for a
43⇌31 she was
41⇌32 of her
and he 33⇌47
39⇌34 she had
and I 35⇌37
I have 36⇌48
37⇌36 could not
was the 38⇌42
47⇌39 one of
and a 40⇌43
80⇌40 going to
48⇌41 would be
it is 42⇌67
50.2%—49.8%



Ω1: Market caps, 2007-Q4
Ω2: Market caps, 2018-Q4
Instrument: Rank-Turbulence Divergence
α=1/3
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{R}_{1/3}(\Omega_1 \,\|\, \Omega_2) = 0.441$
$\propto \sum_{\tau} \left| \frac{1}{r_{\tau,1}^{1/3}} - \frac{1}{r_{\tau,2}^{1/3}} \right|$
Microsoft Corp
Exxon Mobil Corp
Apple Inc
AT&T Inc
Johnson & Johnson
Bank of America Corp
General Electric Co
JPMorgan Chase & Co
Amazon.com Inc
Cisco Systems Inc
The Coca Cola Co
UnitedHealth Group Inc
Altria Group Inc
Boeing Co
ConocoPhillips
Home Depot Inc
Goldman Sachs Group Inc
Amgen Inc
AIG Inc
Honeywell Internati...Inc
HP Inc
Mastercard Inc
Target Corp
Accenture plc
eBay Inc
NVIDIA Corp
Marathon Oil Corp
Facebook Inc
Devon Energy Corp
Visa Inc Class A
Transocean Ltd
Netflix Inc
Berkshire Hathaway ...s B
Allergan plc
Genentech Inc
AbbVie Inc
Monsanto Co
Broadcom Ltd
EMC Corp
Charter Communicati...Inc
Medco Health Solution Inc
HCA Holdings Inc
iHeartMedia Inc
Centene Corp
Lorillard Inc
Avangrid Inc
B/E Aerospace Inc
Wayfair Inc
ITC Holdings Corp
HealthEquity Inc
Blue Nile Inc
The RMR Group Inc
Waste Services Inc
Covia Holdings Corp
North American Scie...Inc
Lines of Constant $\delta D^{R}_{1/3,\tau}$
1.61×10⁻³
4.64×10⁻⁴
7.01×10⁻⁵
1
10
100
1,000
10,000
← bigger market cap
bigger market cap →
smaller market cap →
← smaller market cap
Rank r for Market caps, 2007-Q4
Rank r for Market caps, 2018-Q4
1
10
100
1,000
Counts per cell
Balances:
39.8% total market cap 60.2%
78.8% all companies 61.5%
48.8% exclusive companies 34.4%
Divergence contribution $\delta D^{R}_{1/3,\tau}$ (%)
0.2 0.15 0.1 0.05 0 0.05 0.1 0.15 0.2
General Electric Co 2⇌78
4,635.5⇌5 Facebook Inc ▷
Exxon Mobil Corp 1⇌9
86⇌3 Amazon.com Inc
4,635.5⇌8 Visa Inc Class A ▷
11⇌2 Apple Inc
3⇌1 Microsoft Corp
4,635.5⇌31 AbbVie Inc ▷
◁ Genentech Inc 31⇌4,187.5
AT&T Inc 4⇌19
4,635.5⇌33 Anheuser-Busch InBe.../NV ▷
◁ Wachovia Corp 33⇌4,187.5
◁ Twenty-First Century Fox 40⇌4,187.5
4,635.5⇌44 Broadcom Ltd ▷
Berkshire Hathaway ...s B 38⇌2,331
4,635.5⇌46 Philip Morris Inter...Inc ▷
◁ Time Warner Inc 47⇌4,187.5
◁ Wyeth Corp 48⇌4,187.5
4,635.5⇌50 PayPal Holdings Inc ▷
AIG Inc 17⇌159
◁ Monsanto Co 54⇌4,187.5
1,214⇌42 Netflix Inc
◁ Merrill Lynch & Co 66⇌4,187.5
214⇌24 Mastercard Inc
Procter & Gamble Co 5⇌15
◁ Schering-Plough Corp 74⇌4,187.5
◁ Alcon Inc 76⇌4,187.5
4,635.5⇌79 Charter Communicati...Inc ▷
Altria Group Inc 12⇌52
◁ EMC Corp 83⇌4,187.5
◁ Anheuser-Busch Inc. 87⇌4,187.5
4,635.5⇌90 Tesla Inc ▷
476⇌41 Salesforce.com Inc
◁ DowDuPont Inc 91⇌4,187.5
◁ Barrick Gold Corp. 95⇌4,187.5
4,635.5⇌98 Kraft Heinz Co ▷
HP Inc 26⇌162
◁ Lehman Brothers Holding 103⇌4,187.5
19⇌7 JPMorgan Chase & Co
◁ Yahoo! Inc 109⇌4,187.5
49.6%—50.4%

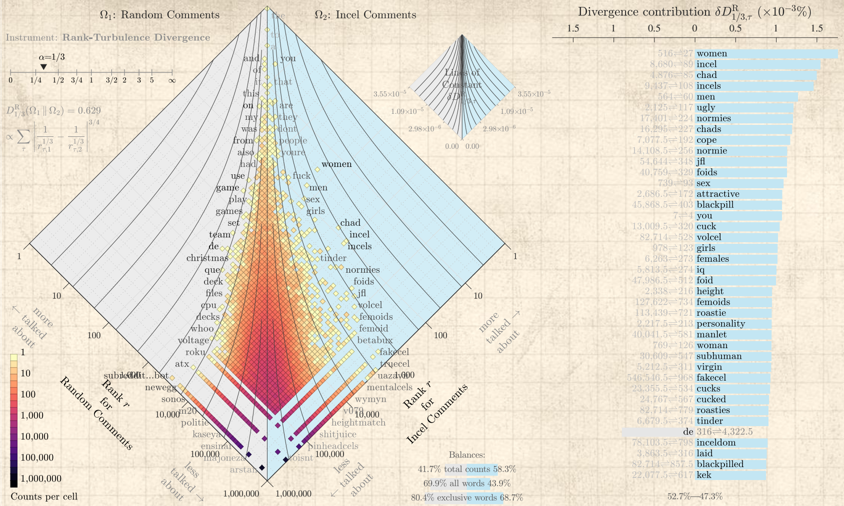
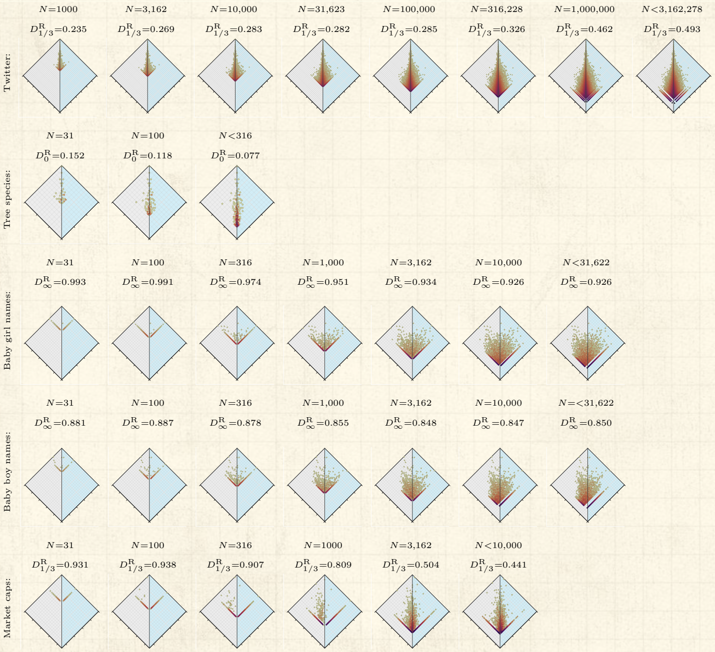
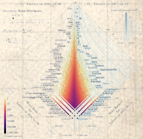
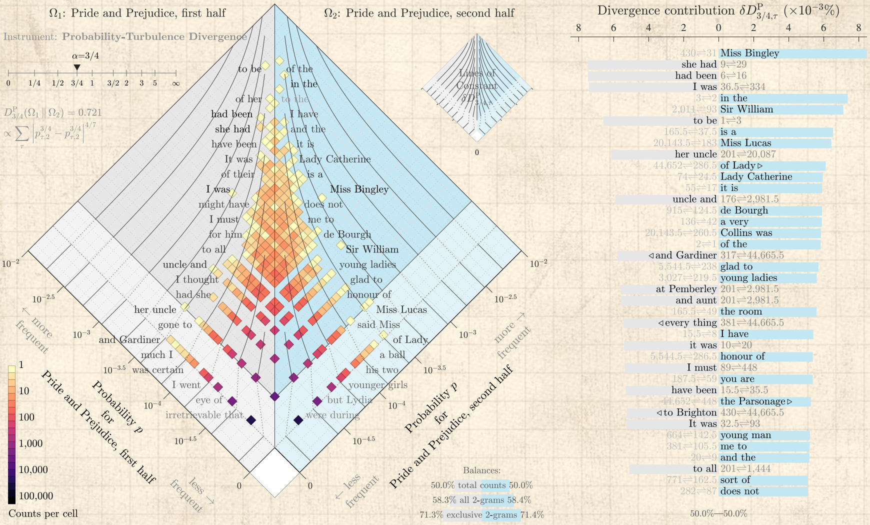


FIG. 8. **Rank-turbulence divergence allotaxonograph [34] of word rank distributions in the incel vs random comment corpora.** The rank-rank histogram on the left shows the density of words by their rank in the incel comments corpus against their rank in the random comments corpus. Words at the top of 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 incel corpus is "women", the object of hatred

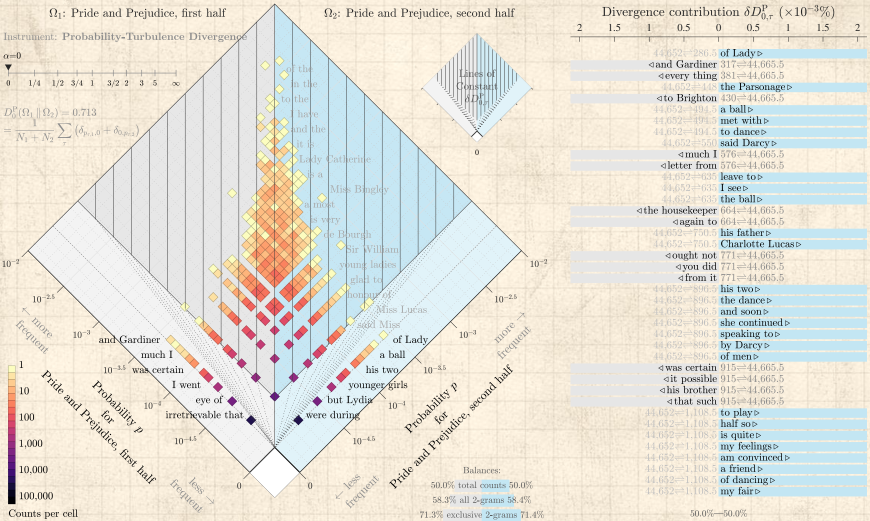
Effect of subsampling:



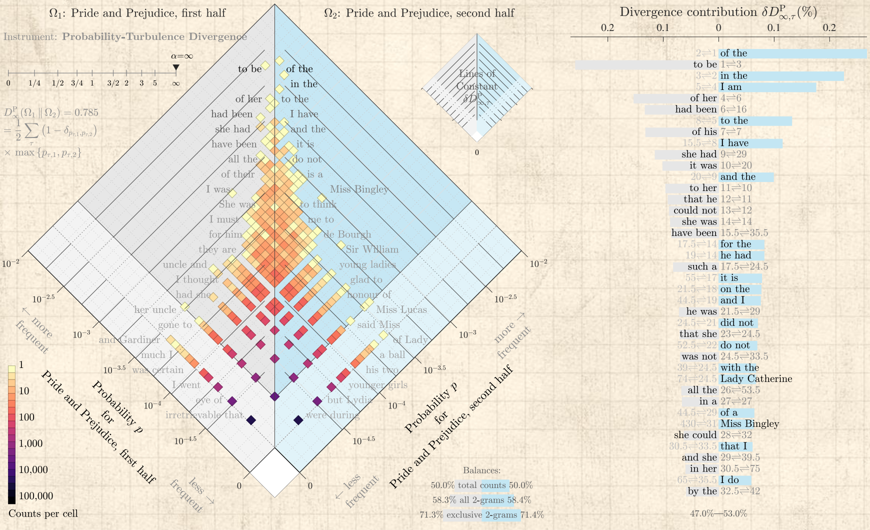




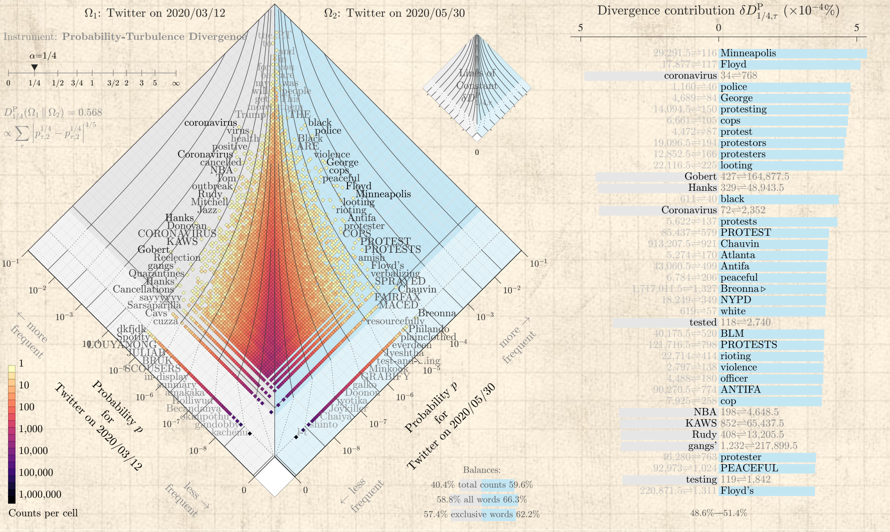
Ω1: Pride and Prejudice, first half
Ω2: Pride and Prejudice, second half
Instrument: Probability-Turbulence Divergence
α=3/4
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{P}_{3/4}(\Omega_1 \| \Omega_2) = 0.721$
$\propto \sum_{\tau} \left| p^{3/4}_{\tau,1} - p^{3/4}_{\tau,2} \right|^{4/7}$
Lines of Constant $\delta D^{P}_{3/4,\tau}$
to be
of the
in the
of her
to the
had been
I have
she had
and the
have been
it is
It was
Lady Catherine
of their
is a
I was
Miss Bingley
might have
does not
I must
me to
for him
de Bourgh
to all
Sir William
uncle and
young ladies
I thought
glad to
had she
honour of
her uncle
Miss Lucas
gone to
said Miss
and Gardiner
of Lady
much I
a ball
was certain
his two
I went
younger girls
eye of
but Lydia
irretrievable that
were during
← more frequent
more → frequent
less → frequent
← less frequent
Probability p for Pride and Prejudice, first half
Probability p for Pride and Prejudice, second half
Counts per cell
1
10
100
1,000
10,000
100,000
Balances:
50.0% total counts 50.0%
58.3% all 2-grams 58.4%
71.3% exclusive 2-grams 71.4%
Divergence contribution $\delta D^{P}_{3/4,\tau}$ ($\times 10^{-3}$%)
8 6 4 2 0 2 4 6 8
430⇌31 Miss Bingley
she had 9⇌29
had been 6⇌16
I was 36.5⇌334
3⇌2 in the
2,011⇌93 Sir William
to be 1⇌3
165.5⇌37.5 is a
20,143.5⇌183 Miss Lucas
her uncle 201⇌20,087
44,652⇌286.5 of Lady ▹
74⇌24.5 Lady Catherine
55⇌17 it is
uncle and 176⇌2,981.5
915⇌124.5 de Bourgh
136⇌42 a very
20,143.5⇌260.5 Collins was
2⇌1 of the
◃and Gardiner 317⇌44,665.5
5,544.5⇌238 glad to
3,027⇌219.5 young ladies
at Pemberley 201⇌2,981.5
and aunt 201⇌2,981.5
165.5⇌49 the room
◃every thing 381⇌44,665.5
15.5⇌8 I have
it was 10⇌20
5,544.5⇌286.5 honour of
I must 89⇌448
187.5⇌59 you are
have been 15.5⇌35.5
44,652⇌448 the Parsonage ▹
◃to Brighton 430⇌44,665.5
It was 32.5⇌93
664⇌142.5 young man
381⇌105.5 me to
20⇌9 and the
to all 201⇌1,444
771⇌162.5 sort of
282⇌87 does not
50.0%—50.0%



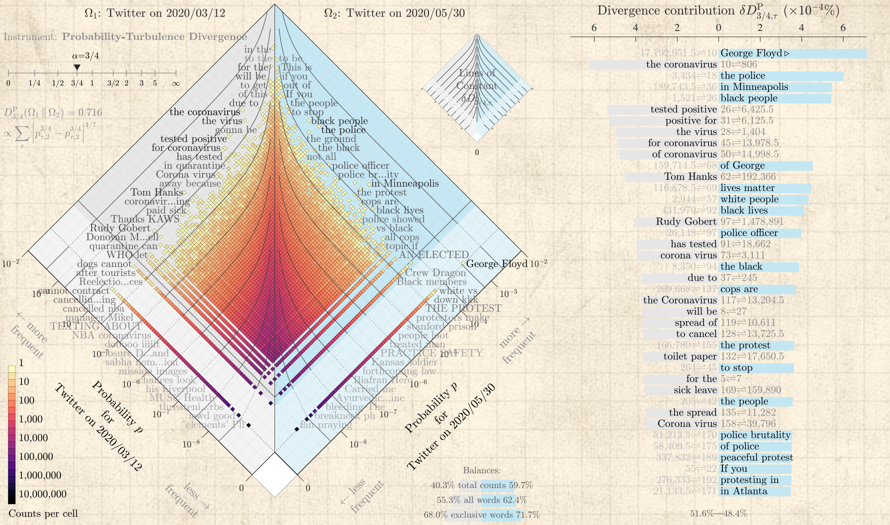
Ω1: Pride and Prejudice, first half
Ω2: Pride and Prejudice, second half
Instrument: Probability-Turbulence Divergence
α=0
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D_0^{\mathrm{P}}(\Omega_1 \| \Omega_2) = 0.713$
$= \frac{1}{N_1 + N_2} \sum_\tau (\delta_{p_{\tau,1},0} + \delta_{0,p_{\tau,2}})$
Lines of Constant $\delta D^{\mathrm{P}}_{0,\tau}$
of the
in the
to the
I have
and the
it is
Lady Catherine
is a
Miss Bingley
a most
is very
de Bourgh
Sir William
young ladies
glad to
honour of
Miss Lucas
said Miss
and Gardiner
much I
was certain
I went
eye of
irretrievable that
of Lady
a ball
his two
younger girls
but Lydia
were during
10^{-2}
$10^{-2.5}$
10^{-3}
$10^{-3.5}$
10^{-4}
$10^{-4.5}$
0
← more frequent
less frequent →
← less frequent
more frequent →
Probability p for Pride and Prejudice, first half
Probability p for Pride and Prejudice, second half
1
10
100
1,000
10,000
100,000
Counts per cell
Balances:
50.0% total counts 50.0%
58.3% all 2-grams 58.4%
71.3% exclusive 2-grams 71.4%
Divergence contribution $\delta D^{\mathrm{P}}_{0,\tau}$ ($\times 10^{-3}$%)
2 1.5 1 0.5 0 0.5 1 1.5 2
44,652⇌286.5 of Lady ▷
◁ and Gardiner 317⇌44,665.5
◁ every thing 381⇌44,665.5
44,652⇌448 the Parsonage ▷
◁ to Brighton 430⇌44,665.5
44,652⇌494.5 a ball ▷
44,652⇌494.5 met with ▷
44,652⇌494.5 to dance ▷
44,652⇌550 said Darcy ▷
◁ much I 576⇌44,665.5
◁ letter from 576⇌44,665.5
44,652⇌635 leave to ▷
44,652⇌635 I see ▷
44,652⇌635 the ball ▷
◁ the housekeeper 664⇌44,665.5
◁ again to 664⇌44,665.5
44,652⇌750.5 his father ▷
44,652⇌750.5 Charlotte Lucas ▷
◁ ought not 771⇌44,665.5
◁ you did 771⇌44,665.5
◁ from it 771⇌44,665.5
44,652⇌896.5 his two ▷
44,652⇌896.5 the dance ▷
44,652⇌896.5 and soon ▷
44,652⇌896.5 she continued ▷
44,652⇌896.5 speaking to ▷
44,652⇌896.5 by Darcy ▷
44,652⇌896.5 of men ▷
◁ was certain 915⇌44,665.5
◁ it possible 915⇌44,665.5
◁ his brother 915⇌44,665.5
◁ that such 915⇌44,665.5
44,652⇌1,108.5 to play ▷
44,652⇌1,108.5 half so ▷
44,652⇌1,108.5 is quite ▷
44,652⇌1,108.5 my feelings ▷
44,652⇌1,108.5 am convinced ▷
44,652⇌1,108.5 a friend ▷
44,652⇌1,108.5 of dancing ▷
44,652⇌1,108.5 my fair ▷
50.0%—50.0%



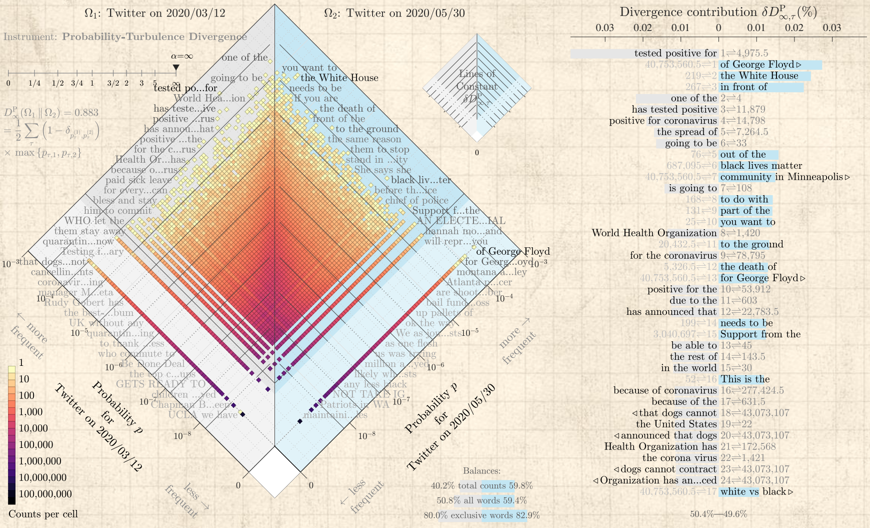
Ω_1: Pride and Prejudice, first half
Ω_2: Pride and Prejudice, second half
Instrument: Probability-Turbulence Divergence
α=∞
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{\mathrm{P}}_{\infty}(\Omega_1 \,\|\, \Omega_2) = 0.785$
$= \frac{1}{2} \sum_{\tau} \left(1 - \delta_{p_{\tau,1}, p_{\tau,2}}\right)$
$\times \max\{p_{\tau,1}, p_{\tau,2}\}$
Lines of Constant $\delta D^{\mathrm{P}}_{\infty,\tau}$
to be
of the
in the
of her
to the
had been
I have
she had
and the
have been
it is
all the
do not
of their
is a
I was
Miss Bingley
She was
to think
I must
me to
for him
de Bourgh
they are
Sir William
uncle and
young ladies
I thought
glad to
had she
honour of
her uncle
Miss Lucas
gone to
said Miss
and Gardiner
of Lady
much I
a ball
was certain
his two
I went
younger girls
eye of
but Lydia
irretrievable that
were during
10^{-2}
$10^{-2.5}$
10^{-3}
$10^{-3.5}$
10^{-4}
$10^{-4.5}$
0
← more frequent
more → frequent
less → frequent
← less frequent
Probability p for Pride and Prejudice, first half
Probability p for Pride and Prejudice, second half
1
10
100
1,000
10,000
100,000
Counts per cell
Balances:
50.0% total counts 50.0%
58.3% all 2-grams 58.4%
71.3% exclusive 2-grams 71.4%
Divergence contribution $\delta D^{\mathrm{P}}_{\infty,\tau}$ (%)
0.2 0.1 0 0.1 0.2
2⇌1 of the
to be 1⇌3
3⇌2 in the
5⇌4 I am
of her 4⇌6
had been 6⇌16
8⇌5 to the
of his 7⇌7
15.5⇌8 I have
she had 9⇌29
it was 10⇌20
20⇌9 and the
to her 11⇌10
that he 12⇌11
could not 13⇌12
she was 14⇌14
have been 15.5⇌35.5
17.5⇌14 for the
19⇌14 he had
such a 17.5⇌24.5
55⇌17 it is
21.5⇌18 on the
44.5⇌19 and I
he was 21.5⇌29
24.5⇌21 did not
that she 23⇌24.5
52.5⇌22 do not
was not 24.5⇌33.5
39⇌24.5 with the
74⇌24.5 Lady Catherine
all the 26⇌53.5
in a 27⇌27
44.5⇌29 of a
430⇌31 Miss Bingley
she could 28⇌32
30.5⇌33.5 that I
and she 29⇌39.5
in her 30.5⇌75
65⇌35.5 I do
by the 32.5⇌42
47.0%—53.0%



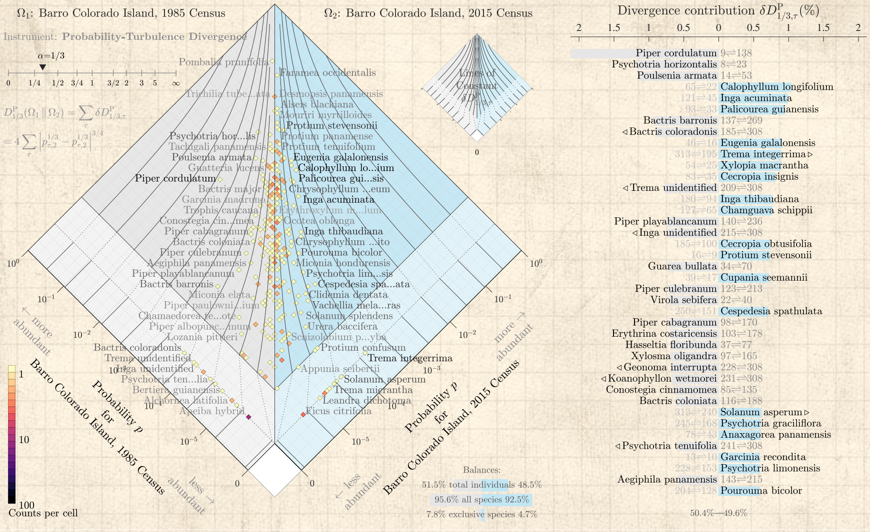
Ω1: Twitter on 2020/03/12
Ω2: Twitter on 2020/05/30
Instrument: Probability-Turbulence Divergence
α=1/4
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{P}_{1/4}(\Omega_1 \| \Omega_2) = 0.568$
$\propto \sum_{\tau} \left| p_{\tau,2}^{1/4} - p_{\tau,2}^{1/4} \right|^{4/5}$
Lines of Constant $\delta D^{P}_{1/4,\tau}$
Probability p for Twitter on 2020/03/12
Probability p for Twitter on 2020/05/30
← more frequent
more frequent →
less frequent →
← less frequent
Counts per cell
1
10
100
1,000
10,000
100,000
1,000,000
Balances:
40.4% total counts 59.6%
58.8% all words 66.3%
57.4% exclusive words 62.2%
Divergence contribution $\delta D^{P}_{1/4,\tau}$ (×10⁻⁴%)
29,291.5⇌116 Minneapolis
17,877⇌117 Floyd
coronavirus 34⇌768
1,160⇌46 police
4,689⇌84 George
14,094.5⇌150 protesting
6,661⇌105 cops
4,472⇌87 protest
19,096.5⇌194 protestors
12,852.5⇌166 protesters
22,116.5⇌225 looting
Gobert 427⇌164,877.5
Hanks 329⇌48,943.5
611⇌40 black
Coronavirus 72⇌2,352
5,622⇌137 protests
85,437⇌579 PROTEST
913,207.5⇌921 Chauvin
5,274⇌170 Atlanta
43,060.5⇌499 Antifa
6,784⇌206 peaceful
1,717,011.5⇌1,327 Breonna ▹
18,249⇌349 NYPD
619⇌57 white
tested 118⇌2,740
40,175.5⇌520 BLM
121,716.5⇌798 PROTESTS
22,714⇌414 rioting
2,797⇌138 violence
4,488⇌180 officer
90,270.5⇌774 ANTIFA
7,925⇌258 cop
NBA 198⇌4,648.5
KAWS 852⇌65,437.5
Rudy 408⇌13,205.5
gangs’ 1,232⇌217,899.5
46,280⇌763 protester
92,973⇌1,024 PEACEFUL
testing 119⇌1,842
220,871.5⇌1,311 Floyd’s
48.6%—51.4%



Ω1: Twitter on 2020/03/12
Ω2: Twitter on 2020/05/30
Instrument: Probability-Turbulence Divergence
α=3/4
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{P}_{3/4}(\Omega_1 \| \Omega_2) = 0.716$
$\propto \sum_{\tau} \left| p_{\tau,2}^{3/4} - p_{\tau,2}^{3/4} \right|^{4/7}$
Lines of Constant $\delta D^{P}_{3/4,\tau}$
in the
to the
for the
will be
to get
of this
due to
the coronavirus
the virus
gonna be
tested positive
for coronavirus
has tested
in quarantine
Corona virus
away because
Tom Hanks
coronavir...ing
paid sick
Thanks KAWS
Rudy Gobert
Donovan M...ell
quarantine can
WHO let
dogs cannot
after tourists
Reelectio...ces
cannot contract
cancellin...ing
cancelled nba
manager Mikel
TEXTING ABOUT
NBA coronavirus
doooo iiiiit
closure D...and
sabha hom...ion
mission images
chances look
his Liverpool
MUSC Health
threateni...rbs
lewd good
elements' I'll
to be
This is
if you
out of
If you
the people
to stop
black people
the police
the ground
the black
not all
police officer
police br...ity
in Minneapolis
the protest
cops are
black lives
police showed
vs black
all cops
topic if
AN ELECTED
George Floyd
Crew Dragon
Black members
white vs
down kkk
THE PROTEST
protestors make
stanford prison
people loot
treated man
PRACTICE SAFETY
Kansas soldier
forthcoming law
Biafran Hero
Carried me
Ayurvedic...ine
bleeding. The
breakneck ph
fan praying
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
0
← more frequent
less frequent →
more frequent →
← less frequent
Probability p for Twitter on 2020/03/12
Probability p for Twitter on 2020/05/30
1
10
100
1,000
10,000
100,000
1,000,000
10,000,000
Counts per cell
Balances:
40.3% total counts 59.7%
55.3% all words 62.4%
68.0% exclusive words 71.7%
Divergence contribution $\delta D^{P}_{3/4,\tau}$ ($\times 10^{-4}$%)
6 4 2 0 2 4 6
17,192,951.5⇌10 George Floyd ▷
the coronavirus 10⇌806
3,434⇌18 the police
189,743.5⇌36 in Minneapolis
1,521⇌26 black people
tested positive 26⇌6,425.5
positive for 31⇌6,125.5
the virus 28⇌1,404
for coronavirus 45⇌13,978.5
of coronavirus 50⇌14,998.5
159,714.5⇌68 of George
Tom Hanks 62⇌192,366
116,878.5⇌69 lives matter
2,944⇌57 white people
431,970⇌92 black lives
Rudy Gobert 97⇌1,478,891
26,148⇌97 police officer
has tested 91⇌18,662
corona virus 73⇌3,111
8,350⇌94 the black
due to 37⇌245
209,668⇌137 cops are
the Coronavirus 117⇌13,204.5
will be 8⇌27
spread of 119⇌10,611
to cancel 128⇌13,725.5
166,780⇌155 the protest
toilet paper 132⇌17,650.5
264⇌45 to stop
for the 5⇌7
sick leave 169⇌159,890
205⇌42 the people
the spread 135⇌11,282
Corona virus 158⇌39,796
51,213.5⇌170 police brutality
58,409.5⇌175 of police
337,832⇌189 peaceful protest
55⇌22 If you
276,333⇌192 protesting in
21,133.5⇌171 in Atlanta
51.6%—48.4%



Ω1: Twitter on 2020/03/12
Ω2: Twitter on 2020/05/30
Instrument: Probability-Turbulence Divergence
α=∞
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{P}_{\infty}(\Omega_1 \| \Omega_2) = 0.883$
$= \frac{1}{2}\sum_{\tau}\left(1 - \delta_{p^{(1)}_{\tau}, p^{(2)}_{\tau}}\right)$
$\times \max\{p_{\tau,1}, p_{\tau,2}\}$
Lines of Constant $\delta D^{P}_{\infty,\tau}$
Counts per cell
1
10
100
1,000
10,000
100,000
1,000,000
10,000,000
100,000,000
Probability p for Twitter on 2020/03/12
Probability p for Twitter on 2020/05/30
← more frequent
more frequent →
less frequent →
← less frequent
Balances:
40.2% total counts 59.8%
50.8% all words 59.4%
80.0% exclusive words 82.9%
Divergence contribution $\delta D^{P}_{\infty,\tau}$ (%)
0.03 0.02 0.01 0 0.01 0.02 0.03
tested positive for 1⇌4,975.5
40,753,560.5⇌1 of George Floyd ▷
219⇌2 the White House
267⇌3 in front of
one of the 2⇌4
has tested positive 3⇌11,879
positive for coronavirus 4⇌14,798
the spread of 5⇌7,264.5
going to be 6⇌33
76⇌5 out of the
687,095⇌6 black lives matter
40,753,560.5⇌7 community in Minneapolis ▷
is going to 7⇌108
168⇌8 to do with
131⇌9 part of the
25⇌10 you want to
World Health Organization 8⇌1,420
20,432.5⇌11 to the ground
for the coronavirus 9⇌78,795
5,326.5⇌12 the death of
40,753,560.5⇌13 for George Floyd ▷
positive for the 10⇌53,912
due to the 11⇌603
has announced that 12⇌22,783.5
199⇌14 needs to be
3,040,697⇌15 Support from the
be able to 13⇌45
the rest of 14⇌143.5
in the world 15⇌30
52⇌16 This is the
because of coronavirus 16⇌277,424.5
because of the 17⇌631.5
◁ that dogs cannot 18⇌43,073,107
the United States 19⇌22
◁ announced that dogs 20⇌43,073,107
Health Organization has 21⇌172,568
the corona virus 22⇌1,421
◁ dogs cannot contract 23⇌43,073,107
◁ Organization has an...ced 24⇌43,073,107
40,753,560.5⇌17 white vs black ▷
50.4%—49.6%



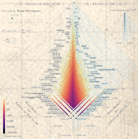
Ω1: Barro Colorado Island, 1985 Census
Ω2: Barro Colorado Island, 2015 Census
Instrument: Probability-Turbulence Divergence
α=1/3
0 1/4 1/2 3/4 1 3/2 2 3 5 ∞
$D^{P}_{1/3}(\Omega_1 \| \Omega_2) = \sum_\tau \delta D^{P}_{1/3,\tau}$
$= 4 \sum_\tau \left| p_{\tau,2}^{1/3} - p_{\tau,2}^{1/3} \right|^{3/4}$
Lines of Constant $\delta D^{P}_{1/3,\tau}$
Pombalia prunifolia
Faramea occidentalis
Trichilia tube...ata
Desmopsis panamensis
Alseis blackiana
Mouriri myrtilloides
Protium stevensonii
Psychotria hor...lis
Protium panamense
Tachigali panamensis
Protium tenuifolium
Poulsenia armata
Eugenia galalonensis
Guatteria lucens
Calophyllum lo...ium
Piper cordulatum
Palicourea gui...sis
Bactris major
Chrysophyllum ...eum
Garcinia madruno
Inga acuminata
Trophis caucana
Erythroxylum m...lum
Conostegia cin...mea
Ocotea oblonga
Piper cabagranum
Inga thibaudiana
Bactris coloniata
Chrysophyllum ...ito
Piper culebranum
Pourouma bicolor
Aegiphila panamensis
Miconia hondurensis
Piper playablancanum
Psychotria lim...sis
Bactris barronis
Cespedesia spa...ata
Miconia elata
Clidemia dentata
Piper paulowni...ium
Vachellia mela...ras
Chamaedorea te...ote
Solanum splendens
Piper albopunc...mum
Urera baccifera
Lozania pittieri
Schizolobium p...yba
Bactris coloradonis
Protium confusum
Trema unidentified
Trema integerrima
Inga unidentified
Appunia seibertii
Psychotria ten...lia
Solanum asperum
Bertiera guianensis
Trema micrantha
Alchornea latifolia
Leandra dichotoma
Apeiba hybrid
Ficus citrifolia
10^0 10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 0
← more abundant
less abundant →
more → abundant
← less abundant
Probability p for Barro Colorado Island, 1985 Census
Probability p for Barro Colorado Island, 2015 Census
1
10
100
Counts per cell
Balances:
51.5% total individuals 48.5%
95.6% all species 92.5%
7.8% exclusive species 4.7%
Divergence contribution $\delta D^{P}_{1/3,\tau}$ (%)
2 1.5 1 0.5 0 0.5 1 1.5 2
Piper cordulatum 9⇌138
Psychotria horizontalis 8⇌23
Poulsenia armata 14⇌53
65⇌22 Calophyllum longifolium
121⇌45 Inga acuminata
93⇌33 Palicourea guianensis
Bactris barronis 137⇌269
◁Bactris coloradonis 185⇌308
46⇌16 Eugenia galalonensis
313⇌195 Trema integerrima ▷
54⇌25 Xylopia macrantha
83⇌35 Cecropia insignis
◁Trema unidentified 209⇌308
180⇌94 Inga thibaudiana
127⇌65 Chamguava schippii
Piper playablancanum 140⇌236
◁Inga unidentified 215⇌308
185⇌100 Cecropia obtusifolia
16⇌9 Protium stevensonii
Guarea bullata 34⇌70
39⇌17 Cupania seemannii
Piper culebranum 123⇌213
Virola sebifera 22⇌40
250⇌151 Cespedesia spathulata
Piper cabagranum 98⇌170
Erythrina costaricensis 103⇌178
Hasseltia floribunda 37⇌77
Xylosma oligandra 97⇌165
◁Geonoma interrupta 228⇌308
◁Koanophyllon wetmorei 231⇌308
Conostegia cinnamomea 85⇌135
Bactris coloniata 116⇌188
313⇌240 Solanum asperum ▷
245⇌168 Psychotria graciliflora
78⇌43 Anaxagorea panamensis
◁Psychotria tenuifolia 241⇌308
13⇌10 Garcinia recondita
228⇌153 Psychotria limonensis
Aegiphila panamensis 143⇌215
204⇌128 Pourouma bicolor
50.4%—49.6%

Flipbooks for RTD:

- Twitter:
  instrument-flipbook-1-rank-div.pdf
  instrument-flipbook-2-probability-div.pdf
  instrument-flipbook-3-gen-entropy-div.pdf
- Market caps:
  instrument-flipbook-4-marketcaps-6years-rank-div.pdf
- Baby names:
  instrument-flipbook-5-babynames-girls-50years-rank-div.pdf
  instrument-flipbook-6-babynames-boys-50years-rank-div.pdf
- Google books:
  instrument-flipbook-7-google-books-onegrams-rank-div.pdf
  instrument-flipbook-8-google-books-bigrams-rank-div.pdf
  instrument-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

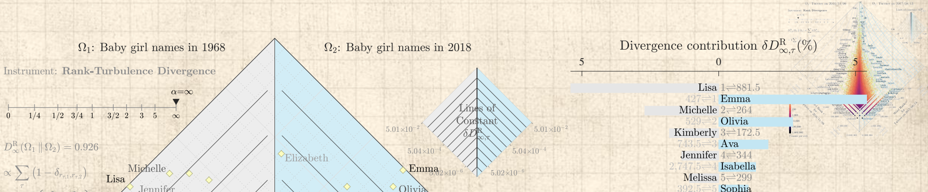


Code:

https://gitlab.com/compstorylab/allotaxonometer

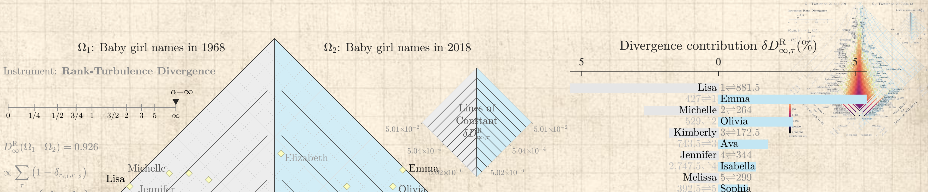
Claims, exaggerations, reminders:

- Needed for comparing large-scale complex systems:
  Comprehendible, dynamically-adjusting, differential dashboards



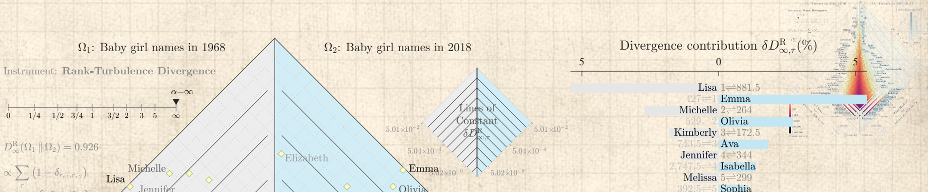
Claims, exaggerations, reminders:

- Needed for comparing large-scale complex systems:
  Comprehendible, dynamically-adjusting, differential dashboards
- Many measures seem poorly motivated and largely unexamined (e.g., JSD)



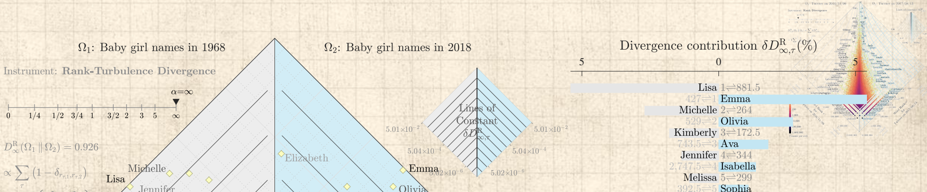
Claims, exaggerations, reminders:

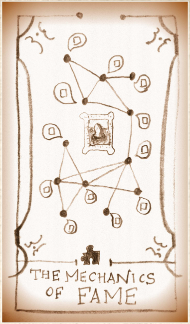
- Needed for comparing large-scale complex systems:
  Comprehendible, dynamically-adjusting, differential dashboards
- Many measures seem poorly motivated and largely unexamined (e.g., JSD)
- Of value: Combining big-picture maps with ranked lists

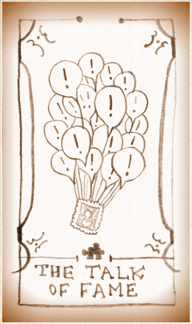


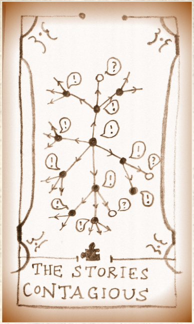
Claims, exaggerations, reminders:

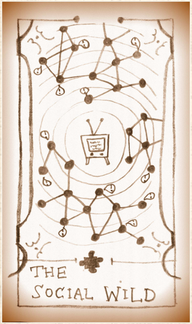
- Needed for comparing large-scale complex systems:
  Comprehendible, dynamically-adjusting, differential dashboards
- Many measures seem poorly motivated and largely unexamined (e.g., JSD)
- Of value: Combining big-picture maps with ranked lists
- Maybe one day: Online tunable version of rank-turbulence divergence (plus many other instruments)

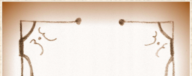


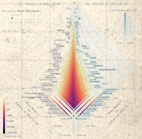

THE MECHANICS OF FAME


THE TALK OF FAME

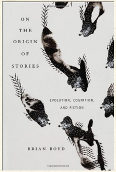

THE STORIES CONTAGIOUS


THE SOCIAL WILD

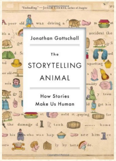




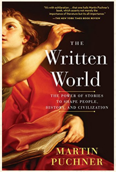
The everywhereness of algorithms and stories:



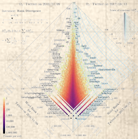
"On the Origin of Stories: Evolution, Cognition, and Fiction"
by Brian Boyd (2010). [3]



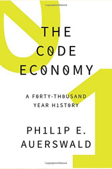
"The Storytelling Animal: How Stories Make Us Human"
by Jonathan Gottschall (2013). [17]



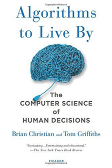
"The Written World: How Literature Shaped Civilization"
by Martin Puchner (2017). [31]



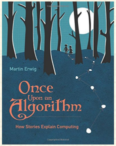
Algorithms, recipes, stories, ...



"The Code Economy: A Forty-Thousand Year History"
by Philip E Auerswald (2017). [1]

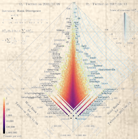


"Algorithms to Live By"
by Christian and Griffiths (2016). [7]

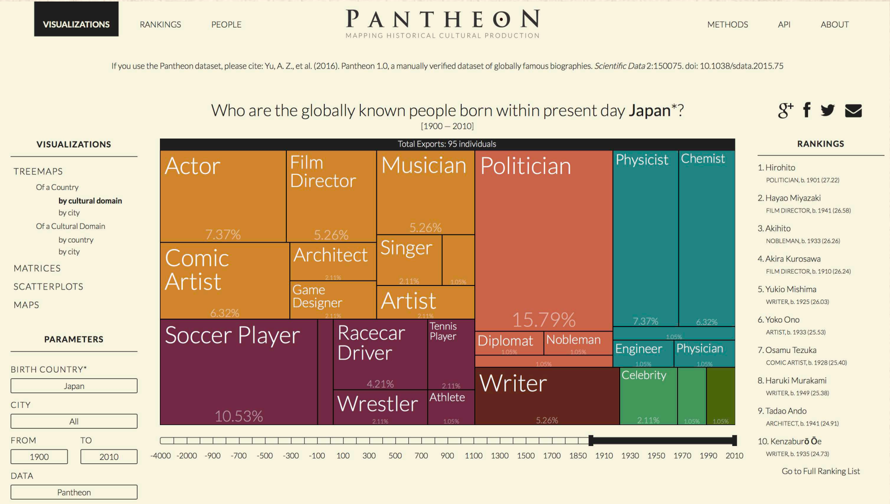


"Once Upon an Algorithm"
by Martin Erwig (2017). [16]

Also: Numerical Recipes in C [30] and How to Bake π [5]

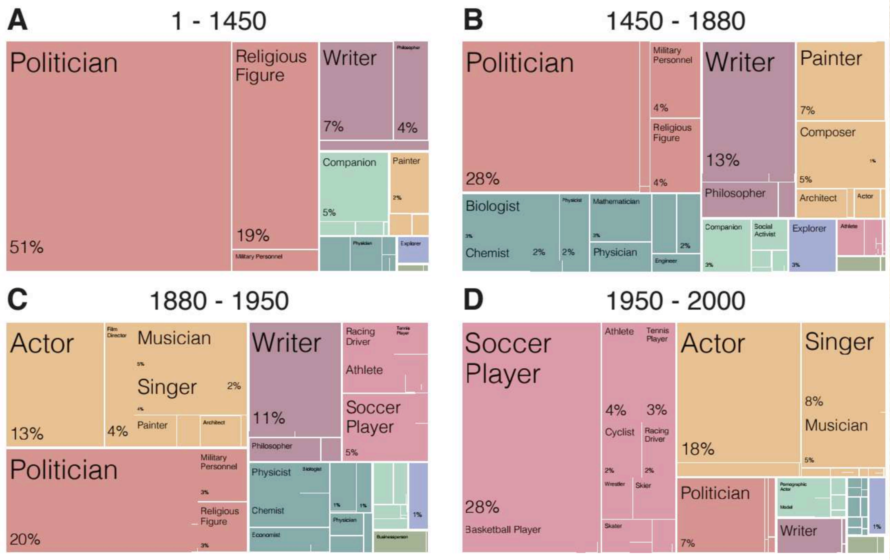


The famous are storytellers—Japan:



For people born 1950–

http://pantheon.media.mit.edu/treemap/country_exports/JP/all/1900/2010/H15/pantheon

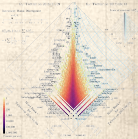


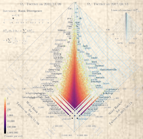
https://www.media.mit.edu/projects/pantheon-new/overview/

Super Survival of the Stories:

The Desirability of Storytellers, The Atlantic, Ed Yong, 2017-12-05.

- Study of Agta, Filipino hunter-gatherers.
- Storytelling valued well above all other skills including hunting.
- Stories encode prosocial norms such as cooperation.



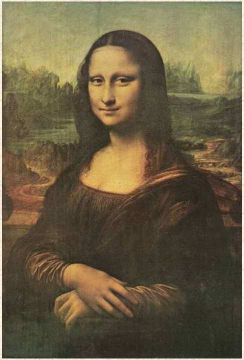


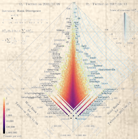
Super Survival of the Stories:

The Desirability of Storytellers, The Atlantic, Ed Yong, 2017-12-05.

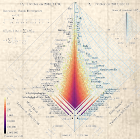
- Study of Agta, Filipino hunter-gatherers.
- Storytelling valued well above all other skills including hunting.
- Stories encode prosocial norms such as cooperation.
- Like the best stories, the best storytellers reproduce more successfully.

The most famous painting in the world:





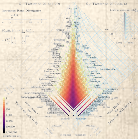
The dismal predictive powers of editors ......



Twelve ...

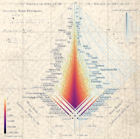
The completely unpredicted fall of Eastern Europe:

Timur Kuran: [20] “Now Out of Never: The Element of Surprise in the East European Revolution of 1989”



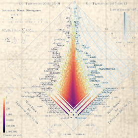
We understand bushfire stories:

1. Sparks start fires.



We understand bushfire stories:

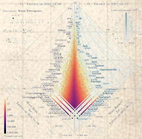
1. Sparks start fires.
2. System properties control a fire's spread.



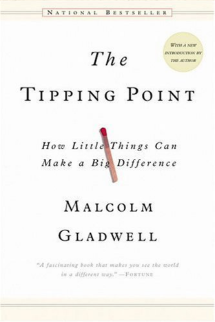
We understand bushfire stories:

1. Sparks start fires.
2. System properties control a fire's spread.

3. But for three reasons, we make two mistakes about Social Fires ...

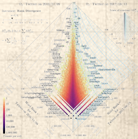


We understand bushfire stories:

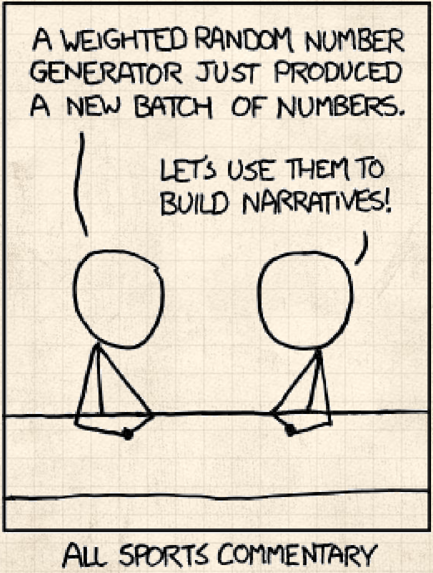


1. Sparks start fires.
2. System properties control a fire's spread.

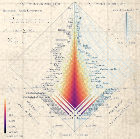
3. But for three reasons, we make two mistakes about Social Fires ...



Reason 1—We are Homo Narrativus.



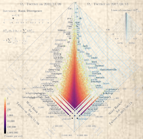
http://xkcd.com/904/



Reason 2—“We are all individuals.”

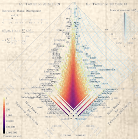
Archival footage:

Individual narratives are not enough to understand distributed, networked minds.

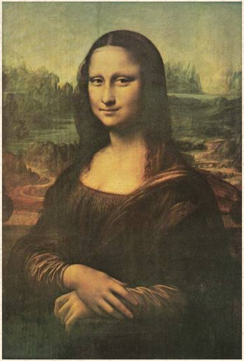


Reason 3—We are spectacular imitators.

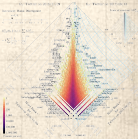
BBC/David Attenborough.



Mistake 1:
Success is due to intrinsic properties



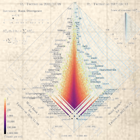
See “Becoming Mona Lisa” by David Sassoon



Mistake 1:
Success is due to intrinsic properties

it’s just so disappointingly small

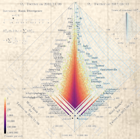
See “Becoming Mona Lisa” by David Sassoon



Mistake 1:
Success is due to intrinsic properties

Stolen in 1913, recovered in 1915.

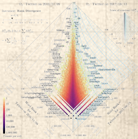
See “Becoming Mona Lisa” by David Sassoon



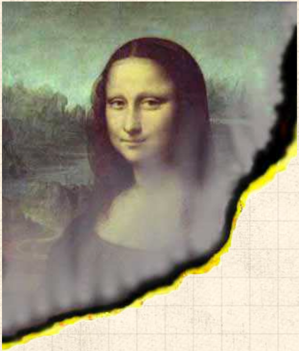
Mistake 1:
Success is due to intrinsic properties

Hidden during WWII.

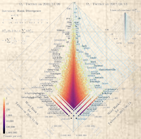
See “Becoming Mona Lisa” by David Sassoon



Mistake 1:
Success is due to intrinsic properties



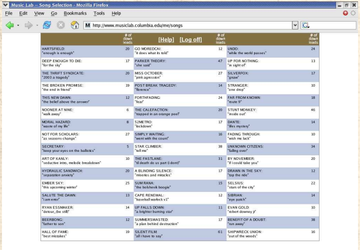
Repeatedly vandalised and attacked.



See "Becoming Mona Lisa" by David Sassoon

48 songs
30k participants

Exp 1— weak social

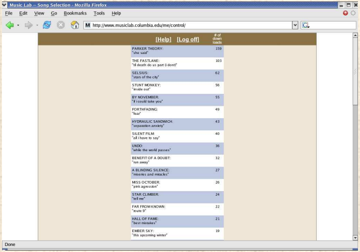


Music Lab – Song Selection - Mozilla Firefox

http://www.musiclab.columbia.edu/me/songs

	# of down loads	[Help] [Log off]	# of down loads		# of down loads
HARTSFIELD: "enough is enough"	20	GO MOREDCAI: "it does what its told"	12	UNDO: "while the world passes"	24
DEEP ENOUGH TO DIE: "for the sky"	17	PARKER THEORY: "she said"	47	UP FOR NOTHING: "in sight of"	13
THE THRIFT SYNDICATE: "2003 a tragedy"	20	MISS OCTOBER: "pink agression"	27	SILVERFOX: "gnaw"	17
THE BROKEN PROMISE: "the end in friend"	19	POST BREAK TRAGEDY: "florence"	14	STRANGER: "one drop"	10
THIS NEW DAWN: "the belief above the answer"	12	FORTHFADING: "fear"	24	FAR FROM KNOWN: "route 9"	18
NOONER AT NINE: "walk away"	6	THE CALEFACTION: "trapped in an orange peel"	20	STUNT MONKEY: "inside out"	46
MORAL HAZARD: "waste of my life"	8	52METRO: "lockdown"	17	DANTE: "lifes mystery"	14
NOT FOR SCHOLARS: "as seasons change"	27	SIMPLY WAITING: "went with the count"	16	FADING THROUGH: "wish me luck"	10
SECRETARY: "keep your eyes on the ballistics"	5	STAR CLIMBER: "tell me"	38	UNKNOWN CITIZENS: "falling over"	34
ART OF KANLY: "seductive intro, melodic breakdown"	10	THE FASTLANE: "til death do us part (i dont)"	31	BY NOVEMBER: "if i could take you"	20
HYDRAULIC SANDWICH: "separation anxiety"	20	A BLINDING SILENCE: "miseries and miracles"	17	DRAWN IN THE SKY: "tap the ride"	12
EMBER SKY: "this upcoming winter"	25	SUM RANA: "the bolshevik boogie"	15	SELSIUS: "stars of the city"	22
SALUTE THE DAWN: "i am error"	13	CAPE RENEWAL: "baseball warlock v1"	12	SIBRIAN: "eye patch"	14
RYAN ESSMAKER: "detour_(be still)"	14	UP FALLS DOWN: "a brighter burning star"	11	EVAN GOLD: "robert downey jr"	10
BEERBONG: "father to son"	12	SUMMERSWASTED: "a plan behind destruction"	17	BENEFIT OF A DOUBT: "run away"	38
HALL OF FAME: "best mistakes"	19	SILENT FILM: "all i have to say"	61	SHIPWRECK UNION: "out of the woods"	16

Exp. 2—strong social

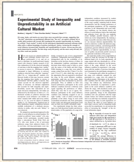


Music Lab – Song Selection - Mozilla Firefox

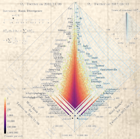
http://www.musiclab.columbia.edu/me/control/

[Help] [Log off]	# of down loads
PARKER THEORY: "she said"	159
THE FASTLANE: "til death do us part (i dont)"	103
SELSIUS: "stars of the city"	62
STUNT MONKEY: "inside out"	56
BY NOVEMBER: "if i could take you"	55
FORTHFADING: "fear"	49
HYDRAULIC SANDWICH: "separation anxiety"	43
SILENT FILM: "all i have to say"	40
UNDO: "while the world passes"	36
BENEFIT OF A DOUBT: "run away"	32
A BLINDING SILENCE: "miseries and miracles"	27
MISS OCTOBER: "pink agression"	26
STAR CLIMBER: "tell me"	24
FAR FROM KNOWN: "route 9"	22
HALL OF FAME: "best mistakes"	21
EMBER SKY: "this upcoming winter"	19

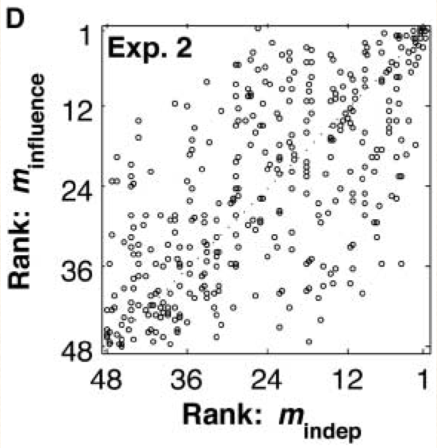
Done



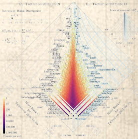
"An experimental study of inequality and unpredictability in an artificial cultural market"
Salganik, Dodds, and Watts, Science, **311**, 854–856, 2006. [32]



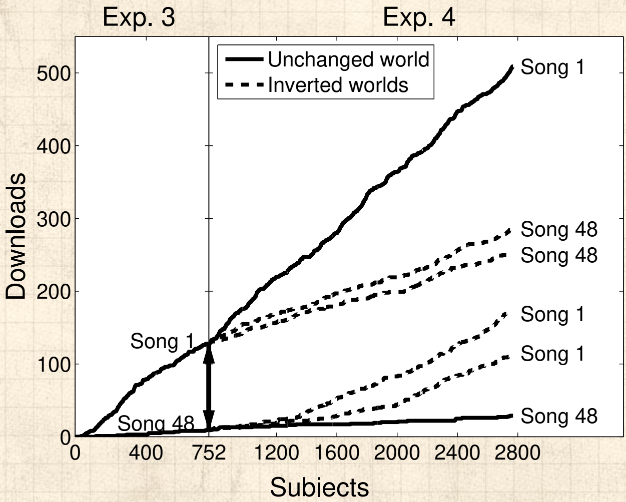
Resolving the paradox:

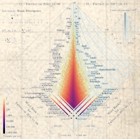


Increased social awareness leads to
Stronger inequality + Less predictability.

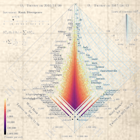


Payola/Deceptive advertising hurts us all:





“Mistake” 2: Seeing success is ‘due to social’ and wanting to say ‘all your interactions are belong to us’



“This is truly the last time, believe me”

The Washington Post

Business • Analysis

14 years of Mark Zuckerberg saying sorry, not sorry

By Geoffrey A. Fowler and Chiqui Esteban April 9, 2018

Do you trust Mark Zuckerberg?

From the moment the Facebook founder entered the public eye in 2003 for creating a Harvard student hot-or-not rating site, he's been apologizing. So we collected this abbreviated history of his public mea culpas.

It reads like a record on repeat. Zuckerberg, who made “move fast and break things” his slogan, says sorry for being naive, and then promises solutions such as privacy “controls,” “transparency” and better policy “enforcement.” And then he promises it again the next time. You can track his sorries in orange and promises in blue in the timeline below.

All the while, Facebook's access to our personal data increases and little changes about the way Zuckerberg handles it. So as Zuckerberg prepares to apologize for the first time in front of Congress, the question that lingers is: What will be different this time?

December 2007

After launching Beacon, which opted-in everyone to sharing with advertisers what they were doing in outside websites and apps.

“ We simply did a bad job with this release, and I apologize for it. ... People need to be able to explicitly choose what they share. ”

February 2009

After unveiling new terms of service that angered users.

“ Over the past couple of days, we received a lot of questions and comments. ... Based on this feedback, we have decided to return to our previous terms of use while we resolve the issues. ”

Commission for deceiving consumers about privacy.

“ I'm the first to admit that we've made a bunch of mistakes. ... Facebook has always been committed to being transparent about the information you have stored with us — and we have led the internet in building tools to give people the ability to see and control what they share. ”

July 2014

After an academic paper exposed that Facebook conducted psychological tests on nearly 700,000 users without their knowledge.

(Apology by Facebook COO Sheryl Sandberg)

Robert Godwin Sr.

“ Our hearts go out to the family and friends of Robert Godwin Sr., and we have a lot of work — and we will keep doing all we can to prevent tragedies like this from happening. ”

September 2017

While revealing a nine-step plan to stop nations from using Facebook to interfere in one another's elections, noting that the amount of “problematic content” found so far is “relatively small.”

“ I care deeply about the democratic process and protecting its integrity. ... It is a new challenge for internet communities to deal with

“ We won't prevent all mistakes or abuse, but we currently make too many errors enforcing our policies and preventing misuse of our tools. ... This will be a serious year of self-improvement and I'm looking forward to learning from working to fix our issues together. ”

March 2018

After details emerged about Cambridge Analytica taking user data.

“ We have a responsibility to protect your data, and if we can't then we don't deserve to serve you. ... We will learn from this experience to secure our platform further and make our community safer for everyone going forward. ”

“ It was my mistake, and I'm sorry. ... There's more we can do here to limit the information developers can access and put more safeguards in place to prevent abuse. ”

Related stories

Facebook: Most users may have had publica data 'scraped'

Facebook COO Sheryl Sandberg on data leak: 'I am really sorry, we are late'

As Facebook confronts data misuse, foreign governments might force real change

What if we paid for Facebook — instead of letting it spy on us for free?

About this story

Photoillustrations based on photos by Tony Avelar/Bloomberg News, Drew Angerer/Getty Images, Jeff Roberson/AP, Jim Watson/Getty Images, Craig Ruttle/AP, Paul Sakuma/AP, Stephen Lam/Reuters, Jose Gomez/Reuters, Richard Drew/AP.

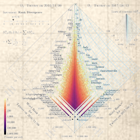
332 Comments

More stories

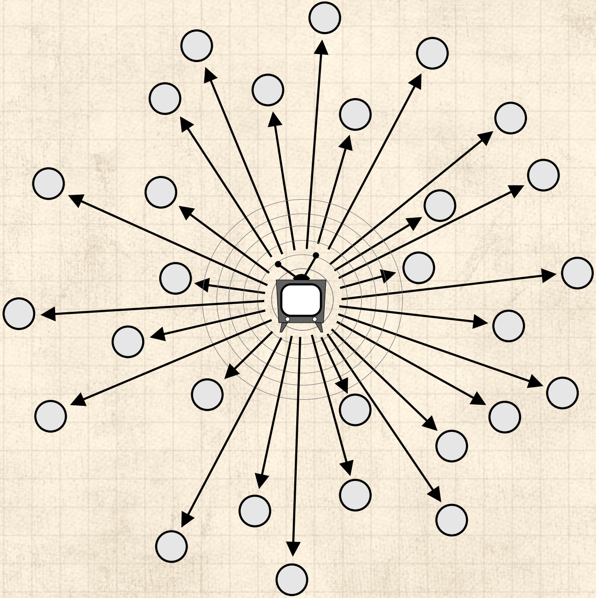
The Facebook ads Russians showed to different groups

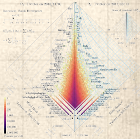
Facebook has said these ads were created by the Internet

WaPo article

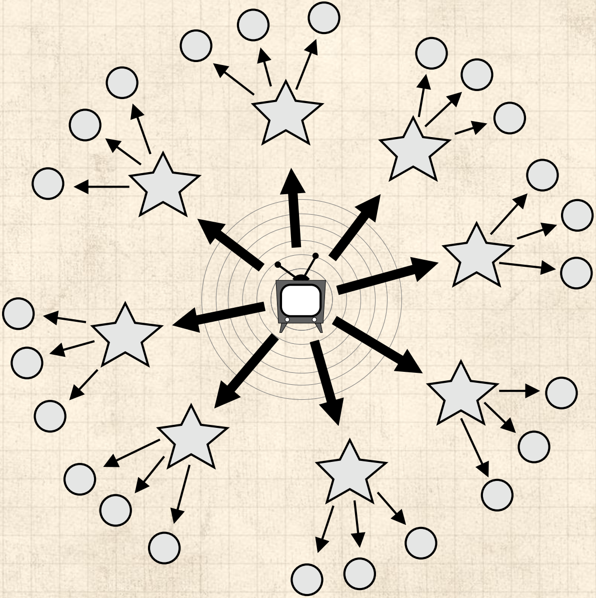


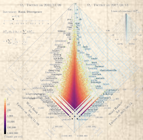
The hypodermic model of influence:



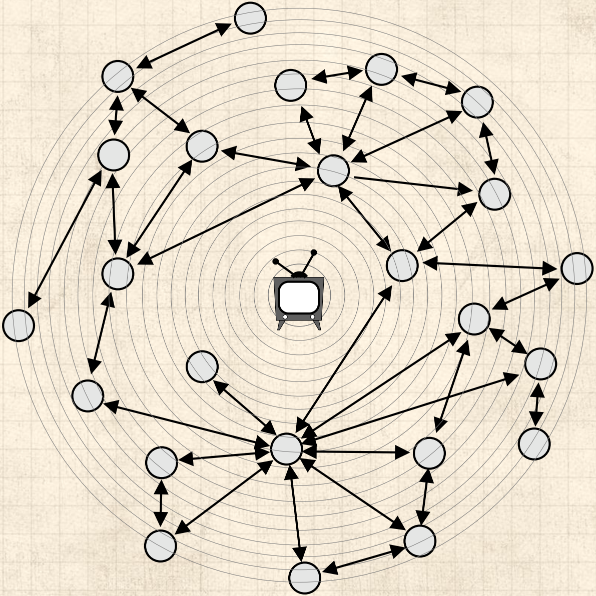


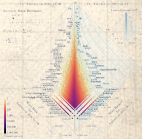
The two step model of influence: [19]



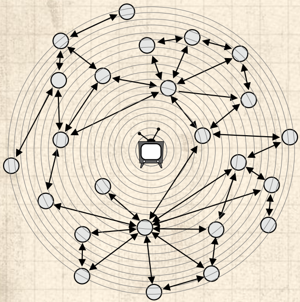


The network model of influence:



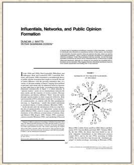


The network model of influence:

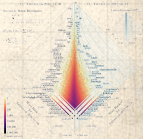


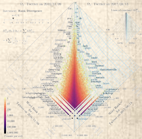
How superspreading works:

Many interconnected, average, trusting people
must benefit from both
receiving and sharing a message
far from its source.


Influentials, Networks, and Public Opinion Formation
DUNCAN J. WATTS
PETER SHERIDAN DODDS

"Influentials, Networks, and Public Opinion Formation"
Watts and Dodds,
J. Consum. Res., **34**, 441–458, 2007. [37]



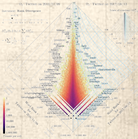


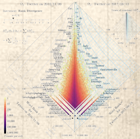
Etymological clarity:

- Fate—from the Latin *fatus*: meaning “spoken”.

Etymological clarity:

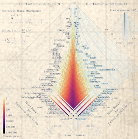
- Fate—from the Latin *fatus*: meaning “spoken”.
- Fate is talk that has been done.
  “It is written”, fore-tell, pre-dict.





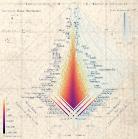
Etymological clarity:

- Fate—from the Latin *fatus*: meaning "spoken".
- Fate is talk that has been done.
  "It is written", fore-tell, pre-dict.
- "There is no such thing as fate, only the story of fate."



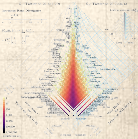
Etymological clarity:

- Fate—from the Latin *fatus*: meaning "spoken".
- Fate is talk that has been done.
  "It is written", fore-tell, pre-dict.
- "There is no such thing as fate, only the story of fate."
- Destiny is probablistic.



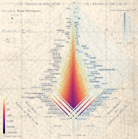
Etymological clarity:

- Fate—from the Latin *fatus*: meaning "spoken".
- Fate is talk that has been done.
  "It is written", fore-tell, pre-dict.
- "There is no such thing as fate, only the story of fate."
- Destiny is probablistic.
- Fame—from the Latin *fāma*: meaning "to talk."



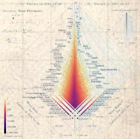
Etymological clarity:

- Fate—from the Latin *fatus*: meaning "spoken".
- Fate is talk that has been done.
  "It is written", fore-tell, pre-dict.
- "There is no such thing as fate, only the story of fate."
- Destiny is probablistic.
- Fame—from the Latin *fāma*: meaning "to talk."
- Fame is inherently the social discussion about the thing, not the thing itself.



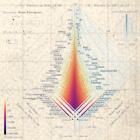
Etymological clarity:

- Fate—from the Latin *fatus*: meaning “spoken”.
- Fate is talk that has been done.
  “It is written”, fore-tell, pre-dict.
- “There is no such thing as fate, only the story of fate.”
- Destiny is probablistic.
- Fame—from the Latin *fāma*: meaning “to talk.”
- Fame is inherently the social discussion about the thing, not the thing itself.
- Renown: Repeatedly named, talked about. Old French renon, from re- + non (“name”).



Etymological clarity:

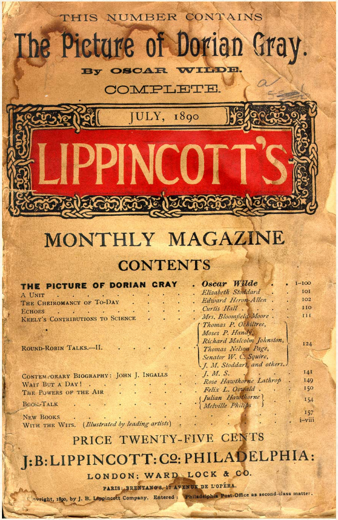
- Fate—from the Latin *fatus*: meaning “spoken”.
- Fate is talk that has been done.
  “It is written”, fore-tell, pre-dict.
- “There is no such thing as fate, only the story of fate.”
- Destiny is probablistic.
- Fame—from the Latin *fāma*: meaning “to talk.”
- Fame is inherently the social discussion about the thing, not the thing itself.
- Renown: Repeatedly named, talked about. Old French renon, from re- + non (“name”).
- Réclame. “Clamo”—Proto-Indo-European: “to shout” (again).



Etymological clarity:

- Fate—from the Latin *fatus*: meaning “spoken”.
- Fate is talk that has been done.
  “It is written”, fore-tell, pre-dict.
- “There is no such thing as fate, only the story of fate.”
- Destiny is probablistic.
- Fame—from the Latin *fāma*: meaning “to talk.”
- Fame is inherently the social discussion about the thing, not the thing itself.
- Renown: Repeatedly named, talked about. Old French renon, from re- + non (“name”).
- Réclame. “Clamo”—Proto-Indo-European: “to shout” (again). Connected to “lowing”.

Oscar Wilde, The Picture of Dorian Gray: Raw Fame



THIS NUMBER CONTAINS

The Picture of Dorian Gray.

By OSCAR WILDE.

COMPLETE.

JULY, 1890

LIPPINCOTT'S

MONTHLY MAGAZINE

CONTENTS

THE PICTURE OF DORIAN GRAY	*Oscar Wilde*	1–100
A UNIT	*Elizabeth Stoddard*	101
THE CHEIROMANCY OF TO-DAY	*Edward Heron-Allen*	102
ECHOES	*Curtis Hall*	110
KEELY'S CONTRIBUTIONS TO SCIENCE	*Mrs. Bloomfield-Moore*	111
ROUND-ROBIN TALKS.—II.	*Thomas P. Ochiltree, Moses P. Handy, Richard Malcolm Johnston, Thomas Nelson Page, Senator W. C. Squire, J. M. Stoddart, and others.*	124
CONTEMPORARY BIOGRAPHY: JOHN J. INGALLS	*J. M. S.*	141
WAIT BUT A DAY!	*Rose Hawthorne Lathrop*	149
THE POWERS OF THE AIR	*Felix L. Oswald*	150
BOOK-TALK	*Julian Hawthorne, Melville Philips*	154
NEW BOOKS		157
WITH THE WITS. (*Illustrated by leading artists*)		i–viii

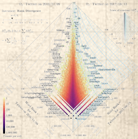
PRICE TWENTY-FIVE CENTS

J:B:LIPPINCOTT:Co: PHILADELPHIA:

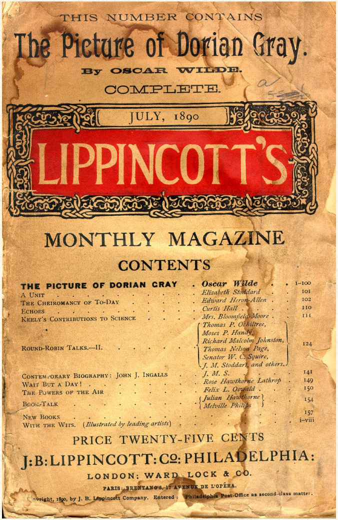
LONDON: WARD LOCK & CO.

PARIS: BRENTANO'S, 17 AVENUE DE L'OPÉRA.

Copyright, 1890, by J. B. Lippincott Company. Entered at Philadelphia Post-Office as second-class matter.



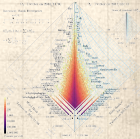
Oscar Wilde, The Picture of Dorian Gray: Raw Fame


THIS NUMBER CONTAINS
The Picture of Dorian Gray.
By OSCAR WILDE.
COMPLETE.
JULY, 1890
LIPPINCOTT'S
MONTHLY MAGAZINE

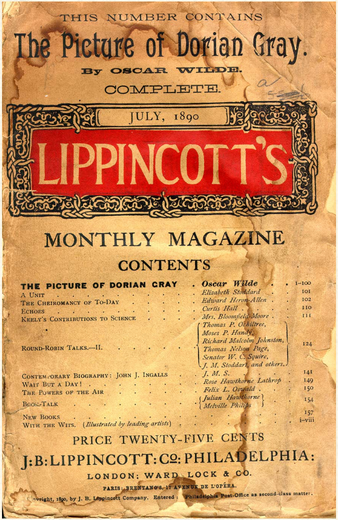
CONTENTS

PRICE TWENTY-FIVE CENTS
J:B:LIPPINCOTT:Co: PHILADELPHIA:
LONDON: WARD LOCK & CO.
PARIS: BRENTANO'S, 17 AVENUE DE L'OPÉRA.

"There is only one thing in the world



Oscar Wilde, The Picture of Dorian Gray: Raw Fame



THIS NUMBER CONTAINS

The Picture of Dorian Gray.

By OSCAR WILDE.

COMPLETE.

JULY, 1890

LIPPINCOTT'S

MONTHLY MAGAZINE

CONTENTS

THE PICTURE OF DORIAN GRAY	*Oscar Wilde*	1–100
A UNIT	*Elizabeth Stoddard*	101
THE CHEIROMANCY OF TO-DAY	*Edward Heron-Allen*	102
ECHOES	*Curtis Hall*	110
KEELY'S CONTRIBUTIONS TO SCIENCE	*Mrs. Bloomfield-Moore*	111
ROUND-ROBIN TALKS.—II.	*Thomas P. Ochiltree, Moses P. Handy, Richard Malcolm Johnston, Thomas Nelson Page, Senator W. C. Squire, J. M. Stoddart, and others.*	124
CONTEMPORARY BIOGRAPHY: JOHN J. INGALLS	*J. M. S.*	141
WAIT BUT A DAY!	*Rose Hawthorne Lathrop*	149
THE POWERS OF THE AIR	*Felix L. Oswald*	150
BOOK-TALK	*Julian Hawthorne, Melville Philips*	154
NEW BOOKS		157
WITH THE WITS. (*Illustrated by leading artists*)		i–viii

PRICE TWENTY-FIVE CENTS

J:B:LIPPINCOTT:CO: PHILADELPHIA:

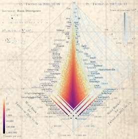
LONDON: WARD LOCK & CO.

PARIS: BRENTANO'S, 17 AVENUE DE L'OPÉRA.

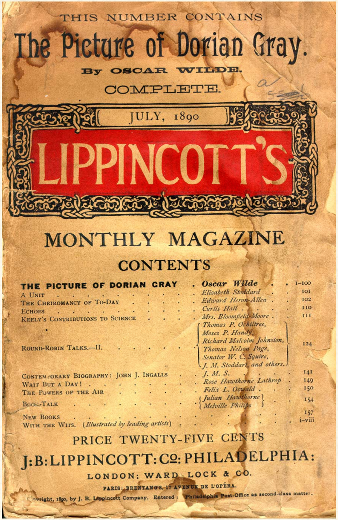
Copyright, 1890, by J. B. Lippincott Company. Entered at Philadelphia Post-Office as second-class matter.

"There is only one thing in the world

worse than being talked about,



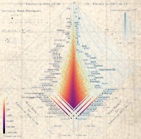
Oscar Wilde, The Picture of Dorian Gray: Raw Fame



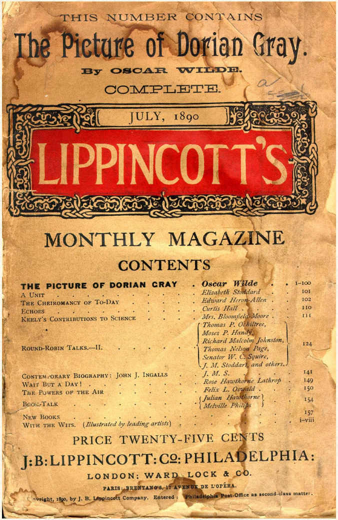
"There is only one thing in the world

worse than being talked about,

and that is



Oscar Wilde, The Picture of Dorian Gray: Raw Fame


THIS NUMBER CONTAINS

The Picture of Dorian Gray.

By OSCAR WILDE.

COMPLETE.

JULY, 1890

LIPPINCOTT'S

MONTHLY MAGAZINE

CONTENTS

PRICE TWENTY-FIVE CENTS

J. B. LIPPINCOTT CO. PHILADELPHIA.

LONDON: WARD, LOCK & CO.

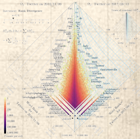
PARIS: BRENTANO'S, 17 AVENUE DE L'OPÉRA.

"There is only one thing in the world

worse than being talked about,

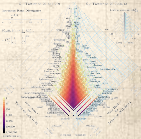
and that is

not being talked about."




THE
DATA POOR

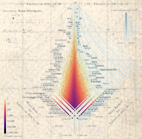

THE POURING
DATA

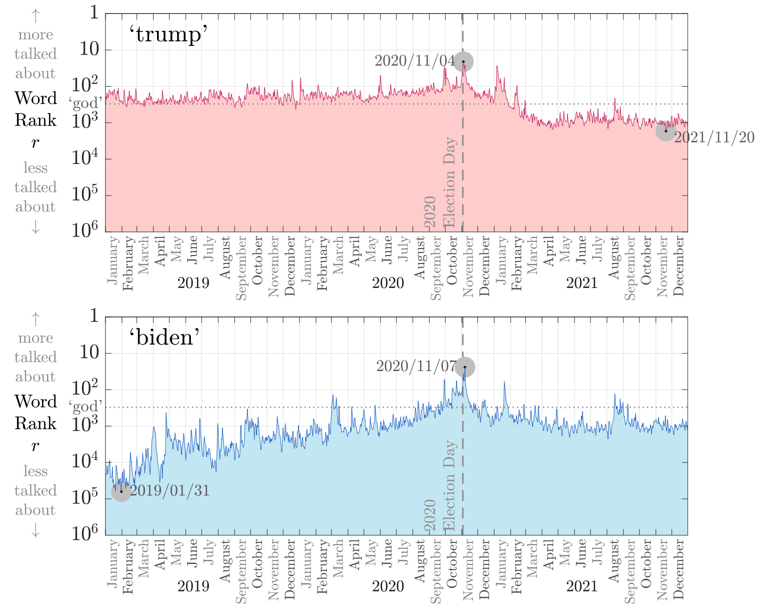


"Fame and Ultrafame: Measuring and comparing daily levels of 'being talked about' for United States' presidents, their rivals, God, countries, and K-pop"
Dodds et al.,
Available online at
https://arxiv.org/abs/1910.00149, 2019. [12]

"Computational timeline reconstruction of the stories surrounding Trump: Story turbulence, narrative control, and collective chronopathy"
Dodds et al.,
, 2020. [14]

- POTUSometer with the Smorgasdashbord:
  http://compstorylab.org/potusometer/
- Stories surrounding Trump:
  http://compstorylab.org/trumpstoryturbulence/




'trump'
2020/11/04
2021/11/20
'biden'
2020/11/07
2019/01/31
↑ more talked about
Word Rank r
less talked about ↓
'god'
2020 Election Day
1
10
10^2
10^3
10^4
10^5
10^6
January
February
March
April
May
June
July
August
September
October
November
December
2019
2020
2021

Ultrafame:
Nobody expects ~~the Spanish Inquisition~~ K-pop:

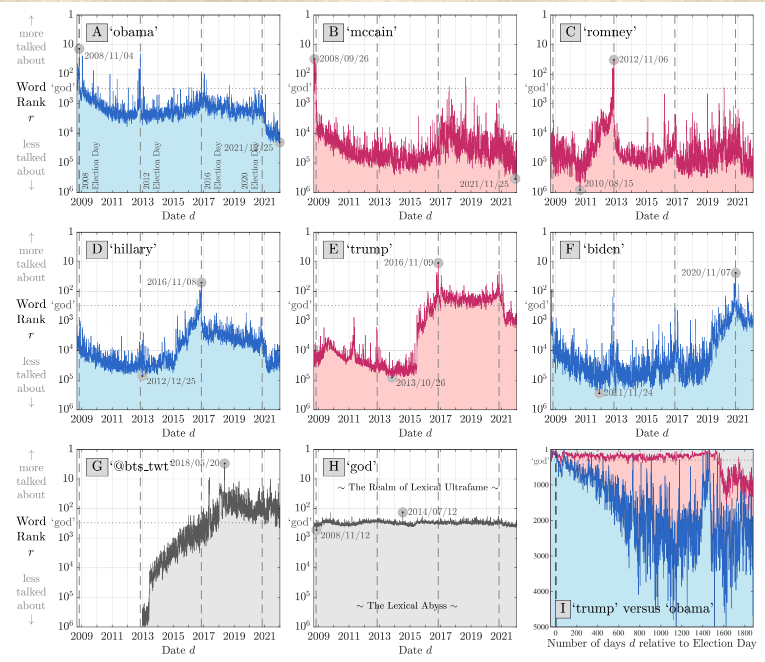
Vox (2019-04-17):
BTS, the band that changed K-pop, explained

Telegnomics

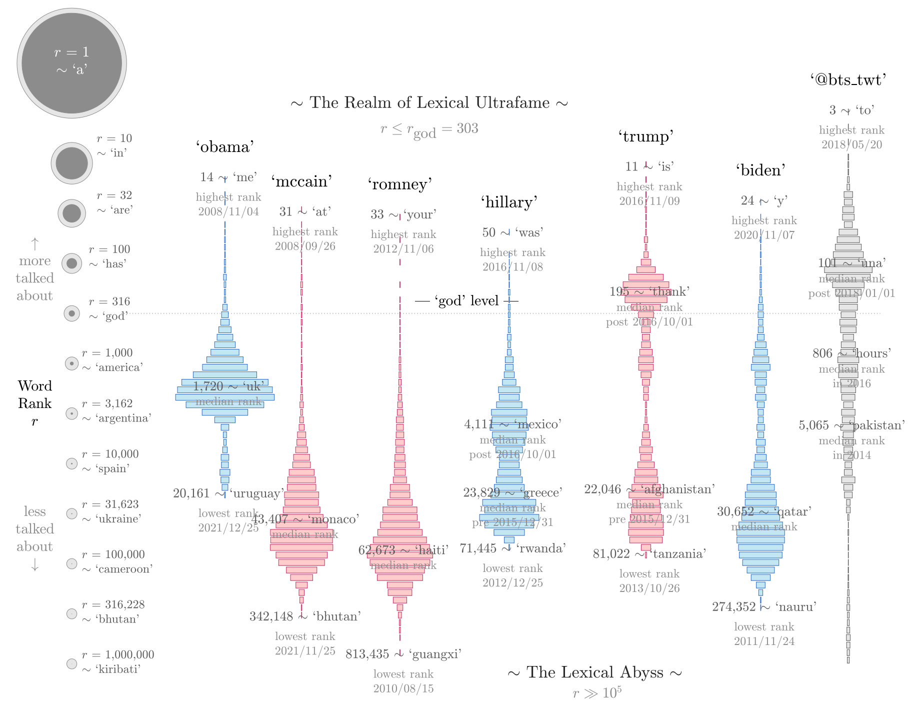
Distant reading by smashing texts into storyons:

```
cd ~/work/stories/2019-10story-turbulence-trump/
261G
more updateall.sh
file names:
compute_rank_turbulence_divergence_sweep_the_leg

Zip files:
zless 2018-01-06/1grams/en_*.tar.tsv
zless 2021-01-05/1grams/en_*.tar.tsv
zless 2021-01-06/1grams/en_*.tar.tsv
zless 2021-01-07/1grams/en_*.tar.tsv
```



2011 Whitehouse Correspondents' Dinner



r = 1 ∼ 'a'
r = 10 ∼ 'in'
r = 32 ∼ 'are'
r = 100 ∼ 'has'
r = 316 ∼ 'god'
r = 1,000 ∼ 'america'
r = 3,162 ∼ 'argentina'
r = 10,000 ∼ 'spain'
r = 31,623 ∼ 'ukraine'
r = 100,000 ∼ 'cameroon'
r = 316,228 ∼ 'bhutan'
r = 1,000,000 ∼ 'kiribati'
↑ more talked about
Word Rank r
less talked about ↓
∼ The Realm of Lexical Ultrafame ∼
r ≤ r_god = 303
— 'god' level —
∼ The Lexical Abyss ∼
r ≫ 10^5
'obama'
14 ∼ 'me'
highest rank 2008/11/04
1,720 ∼ 'uk'
median rank
20,161 ∼ 'uruguay'
lowest rank 2021/12/25
'mccain'
31 ∼ 'at'
highest rank 2008/09/26
43,407 ∼ 'monaco'
median rank
342,148 ∼ 'bhutan'
lowest rank 2021/11/25
'romney'
33 ∼ 'your'
highest rank 2012/11/06
62,673 ∼ 'haiti'
median rank
813,435 ∼ 'guangxi'
lowest rank 2010/08/15
'hillary'
50 ∼ 'was'
highest rank 2016/11/08
4,111 ∼ 'mexico'
median rank post 2016/10/01
23,829 ∼ 'greece'
median rank pre 2015/12/31
71,445 ∼ 'rwanda'
lowest rank 2012/12/25
'trump'
11 ∼ 'is'
highest rank 2016/11/09
195 ∼ 'thank'
median rank post 2016/10/01
22,046 ∼ 'afghanistan'
median rank pre 2015/12/31
81,022 ∼ 'tanzania'
lowest rank 2013/10/26
'biden'
24 ∼ 'y'
highest rank 2020/11/07
30,652 ∼ 'qatar'
median rank
274,352 ∼ 'nauru'
lowest rank 2011/11/24
'@bts_twt'
3 ∼ 'to'
highest rank 2018/05/20
101 ∼ 'una'
median rank post 2018/01/01
806 ∼ 'hours'
median rank in 2016
5,065 ∼ 'pakistan'
median rank in 2014

Ultrafame—Percentage of days per year ranked above ‘god’

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
‘barack’	1.8%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
‘obama’	54.4%	6.9%	0.5%	0.5%	2.2%	0.3%	0.0%	0.3%	2.2%	2.2%	0.5%	0.0%	0.3%	0.0%
‘@barackobama’	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
‘john’	3.5%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.8%	0.3%	0.5%	0.0%
‘mccain’	39.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	1.1%	0.0%	0.0%	0.0%
‘@senjohnmccain’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
‘mitt’	0.0%	0.0%	0.0%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
‘romney’	0.0%	0.0%	0.0%	0.0%	1.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%
‘@mittromney’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
‘hillary’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.4%	0.0%	0.0%	0.0%	0.0%	0.0%
‘clinton’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	7.7%	0.0%	0.0%	0.0%	0.0%	0.0%
‘@hillaryclinton’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.1%	0.0%	0.0%	0.0%	0.0%	0.0%
‘donald’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.7%	0.5%	0.0%	0.0%	1.6%	0.6%
‘trump’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	47.8%	98.6%	93.7%	92.3%	100.0%	10.2%
‘@realdonaldtrump’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.7%	26.8%	41.4%	62.7%	90.2%	2.2%
‘joe’	3.5%	2.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.2%	0.6%
‘biden’	1.8%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	23.8%	6.1%
‘@joebiden’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.1%	0.3%
‘@bts_twt’	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	8.5%	50.7%	100.0%	100.0%	98.9%	93.1%

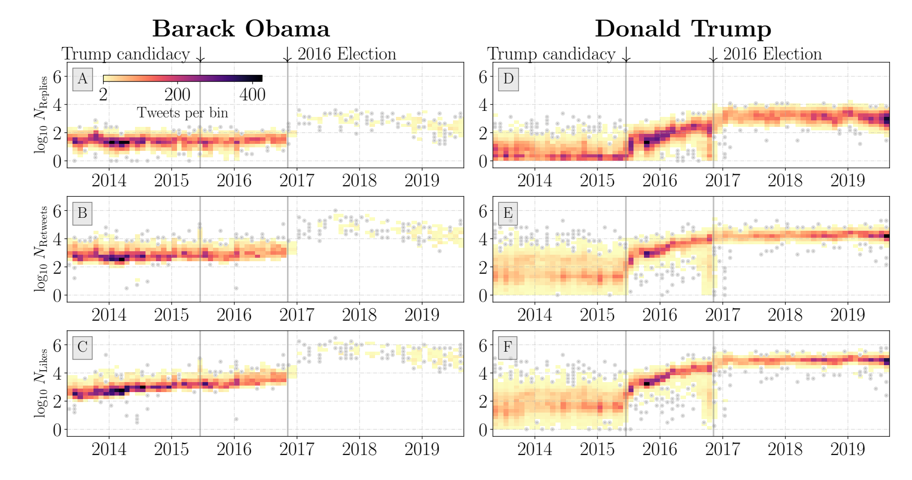
Relative median rates of ‘being talked about’
in the 8 weeks (56 days) pre-election day:

	2008	2012	2016	2020
‘barack’	128	11	11	13
‘obama’	1000	132	135	71
‘@barackobama’	9	24	10	17
‘john’	307	66	72	65
‘mccain’	757	1	2	3
‘@senjohnmccain’	0	0	1	0
‘mitt’	3	50	2	2
‘romney’	2	120	3	3
‘@mittromney’	0	14	1	1
‘hillary’	20	3	357	30
‘clinton’	42	8	326	23
‘@hillaryclinton’	0	0	130	19
‘donald’	7	5	178	135
‘trump’	4	3	656	1001
‘@realdonaldtrump’	0	4	219	656
‘joe’	128	39	32	287
‘biden’	67	4	5	504
‘@joebiden’	0	1	1	212
‘@bts_twt’	0	0	166	1037
‘god’	400	484	362	380

Relative median rates of ‘being talked about’ per year:

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
‘barack’	150	38	17	9	10	7	8	11	14	15	14	14	19	3
‘obama’	897	379	152	87	97	79	91	103	156	160	129	106	104	17
‘@barackobama’	10	8	11	10	17	15	16	13	13	17	17	13	24	5
‘john’	405	274	188	126	117	104	113	121	118	129	128	114	108	82
‘mccain’	579	11	4	2	2	2	1	1	3	15	7	5	3	2
‘@senjohnmccain’	0	2	1	0	0	1	1	1	1	9	2	0	0	0
‘mitt’	5	8	5	6	25	6	5	4	4	2	2	3	3	2
‘romney’	3	1	1	4	42	2	1	1	4	1	1	3	4	1
‘@mittromney’	0	0	0	0	5	0	0	0	1	0	0	1	1	0
‘hillary’	28	10	5	3	3	4	6	30	169	72	61	43	33	6
‘clinton’	62	25	16	10	8	6	8	27	140	65	62	45	32	8
‘@hillaryclinton’	0	0	0	0	0	0	1	11	71	22	19	21	23	3
‘donald’	11	17	11	11	8	6	7	44	166	145	114	104	143	43
‘trump’	7	20	10	7	4	3	3	77	583	1000	865	808	1134	229
‘@realdonaldtrump’	0	0	0	1	2	3	2	32	219	468	555	652	888	1
‘joe’	157	187	138	87	66	58	44	46	50	48	44	78	197	117
‘biden’	72	7	3	1	2	2	2	3	5	3	4	52	284	221
‘@joebiden’	0	0	0	0	0	0	0	0	1	1	2	18	162	28
‘@bts_twt’	0	0	0	0	0	5	36	123	242	595	2487	1802	1440	1437
‘god’	666	851	687	694	791	719	607	616	601	590	612	611	612	510

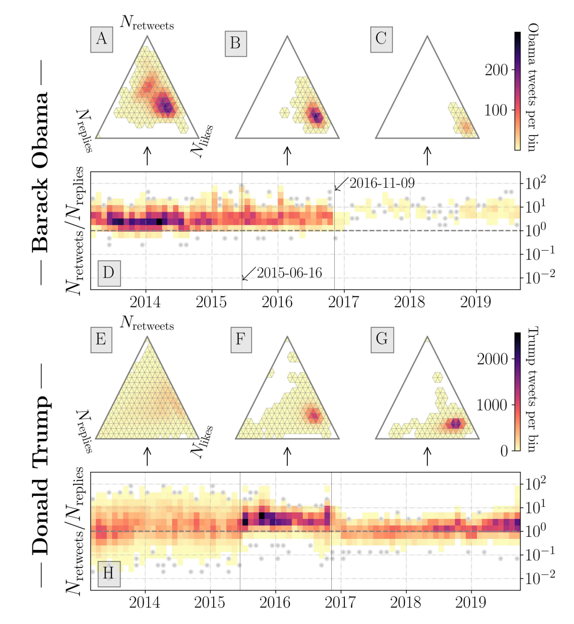
Ratiometrics:



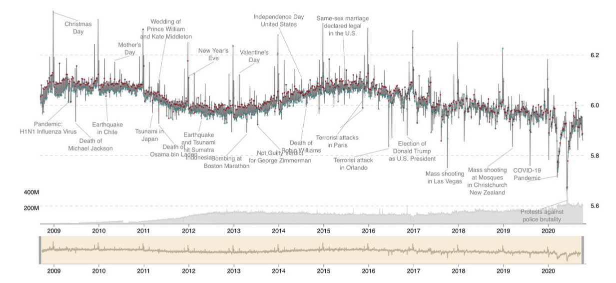
"Ratioing the President: An exploration of public engagement with Obama and Trump on Twitter,"

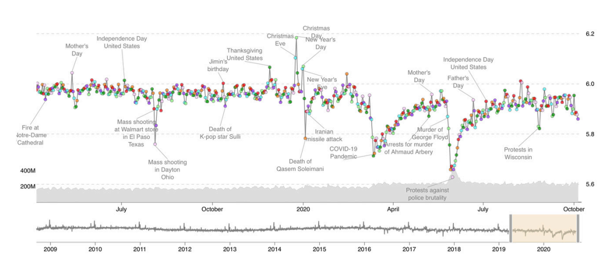
Minot et al., 2020 [24]

Ratiometrics:



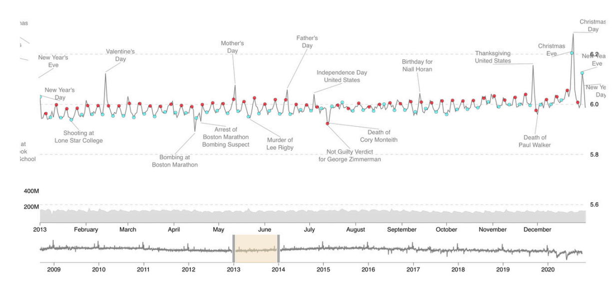
Emotional turbulence:

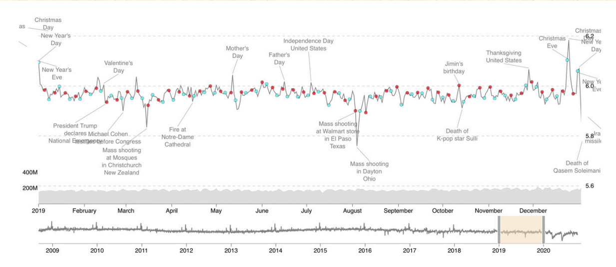




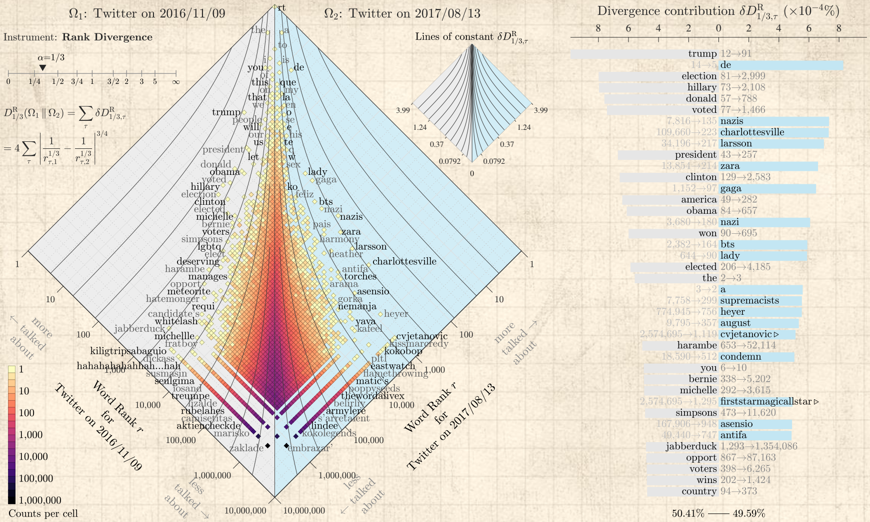
http://hedonometer.org/

Emotional turbulence:



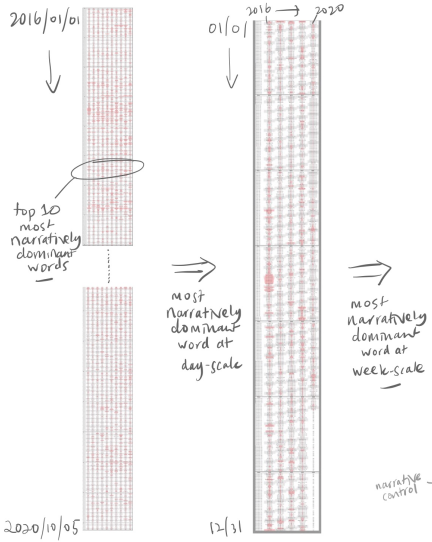


http://hedonometer.org/



Allotaxonometry—
the comparison of complex systems:

http://compstorylab.org/allotaxonometry/



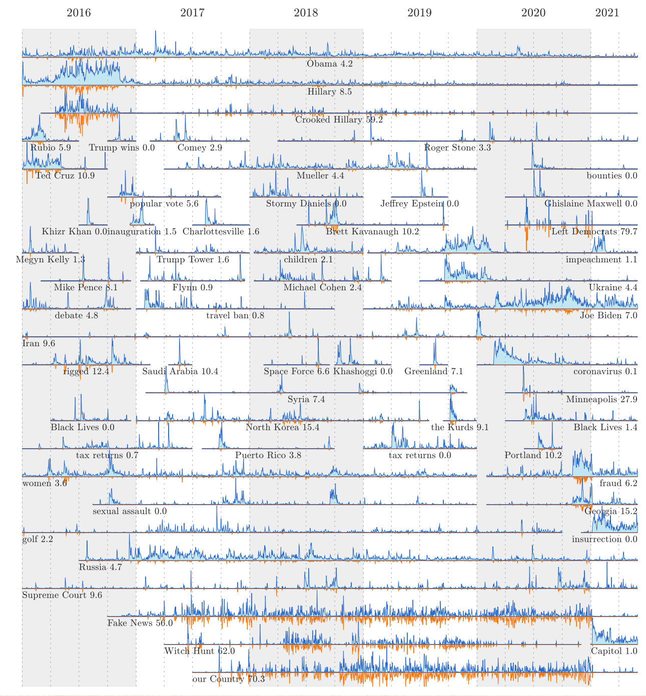
Week	2016	2017	2018	2019	2020
1. 01/01–01/07	Hillary 34.7	hacking 28.6	Bannon 2.2	shutdown 0.0	Iran 9.6
2. 01/08–01/14	Cruz 1.0	Meryl 5.0	Mueller 0.0	shutdown 0.0	Soleimani 5.9
3. 01/15–01/21	Cruz 10.7	inauguration 0.6	DACA 6.7	Pelosi 6.8	Parnas 0.0
4. 01/22–01/28	Cruz 10.6	inauguration 3.1	Mueller 0.0	Pelosi 2.6	Ukraine 5.5
5. 01/29–02/04	Cruz 11.2	ban 2.1	Mueller 0.0	border 0.0	impeachment 0.0
6. 02/05–02/11	Cruz 5.1	Bannon 0.0	memo 2.3	Whitaker 0.0	Vindman 2.5
7. 02/12–02/18	Cruz 6.9	Flynn 0.0	Mueller 0.0	emergency 0.0	Barr 2.2
8. 02/19–02/25	Rubio 3.8	Sweden 4.9	Parkland 0.3	Jussie 0.0	Bloomberg 6.3
9. 02/26–03/04	Rubio 9.2	Russia 6.4	Mueller 0.0	Cohen 3.7	coronavirus 0.0
10. 03/05–03/11	Cruz 1.0	Russian 4.8	Mueller 0.0	Nadler 13.7	coronavirus 0.0
11. 03/12–03/18	Cruz 5.7	tax 1.8	Mueller 2.2	emergency 1.6	coronavirus 0.0
12. 03/19–03/25	Arizona 16.8	Nunes 0.0	Mueller 2.2	Barr 0.0	coronavirus 0.0
13. 03/26–04/01	women 8.3	Russia 9.9	Stormy 0.0	Schiff 5.2	coronavirus 0.5
14. 04/02–04/08	Cruz 1.5	Russia 2.8	Mueller 0.0	returns 0.0	coronavirus 0.0
15. 04/09–04/15	Cruz 1.7	Syria 0.4	Mueller 2.0	Barr 2.4	coronavirus 0.0
16. 04/16–04/22	Cruz 10.5	Russia 0.5	Mueller 0.1	Barr 0.1	coronavirus 0.0
17. 04/23–04/29	Cruz 3.0	days 0.1	Kanye 8.0	Biden 6.0	coronavirus 0.0
18. 04/30–05/06	Indiana 11.5	Trumpcare 0.0	Mueller 0.0	Barr 0.0	coronavirus 0.0
19. 05/07–05/13	Ryan 2.5	Comey 2.8	Iran 6.6	Barr 0.0	coronavirus 0.0
20. 05/14–05/20	Bernie 25.3	Comey 1.0	ZTE 4.5	Barr 0.0	coronavirus 0.0
21. 05/21–05/27	Clinton 9.5	budget 0.0	Korea 18.2	Barr 0.0	pandemic 0.0
22. 05/28–06/03	Hillary 11.9	Kathy 4.4	Roseanne 4.0	USS 3.0	Minneapolis 32.1
23. 06/04–06/10	Clinton 11.1	Comey 0.8	pardon 0.0	Mexico 27.6	police 4.2
24. 06/11–06/17	Orlando 12.4	Mueller 0.0	Kim 4.1	foreign 2.0	Tulsa 4.5
25. 06/18–06/24	Hillary 23.9	Trumpcare 0.0	children 1.0	Iran 12.9	Tulsa 2.1
26. 06/25–07/01	Clinton 13.0	Russia 5.8	Justice 8.3	Moon 29.9	bounties 0.0
27. 07/02–07/08	Crooked 80.6	CNN 0.7	toddlers 0.0	parade 0.0	Rushmore 2.3
28. 07/09–07/15	Crooked 71.5	Russian 1.2	NATO 13.0	Epstein 0.0	coronavirus 0.0
29. 07/16–07/22	Pence 2.9	Mueller 0.0	Helsinki 3.1	racist 0.8	coronavirus 0.0
30. 07/23–07/29	DNC 6.1	Scouts 0.0	Cohen 0.0	Baltimore 13.6	Portland 11.8
31. 07/30–08/05	Khan 6.5	Mueller 0.0	LeBron 0.7	Baltimore 9.4	pandemic 0.0
32. 08/06–08/12	Crooked 55.2	Korea 5.8	Omarosa 0.4	Paso 7.6	USPS 0.0
33. 08/13–08/19	Manafort 0.0	Charlottesville 1.5	Omarosa 9.5	Greenland 6.9	USPS 0.0
34. 08/20–08/26	Clinton 7.6	Charlottesville 3.8	Cohen 2.7	Greenland 8.0	Biden 6.6
35. 08/27–09/02	Crooked 57.4	Harvey 0.0	Ohr 14.0	Dorian 12.2	Kenosha 9.5
36. 09/03–09/09	Bondi 0.0	DACA 2.4	Kavanaugh 2.1	Dorian 12.6	Atlantic 4.8
37. 09/10–09/16	deplorable 0.0	ESPN 2.7	Puerto 7.5	flavored 0.0	Woodward 2.6
38. 09/17–09/23	Clinton 6.5	Kim 4.9	Kavanaugh 1.7	Ukraine 4.5	coronavirus 0.0
39. 09/24–09/30	debate 4.9	Puerto 4.7	Kavanaugh 9.5	Ukraine 6.8	ballots 0.7
40. 10/01–10/07	Pence 4.9	Puerto 2.1	Kavanaugh 6.8	Ukraine 5.1	Covid 0.0
41. 10/08–10/14	sexual 0.3	Puerto 1.8	Kavanaugh 4.3	Kurds 8.2	
42. 10/15–10/21	rigged 10.1	Puerto 0.2	Saudi 5.3	Kurds 3.7	
43. 10/22–10/28	star 0.0	Mueller 0.0	caravan 0.0	impeachment 0.0	
44. 10/29–11/04	FBI 5.9	Mueller 0.0	caravan 0.0	impeachment 0.0	
45. 11/05–11/11	Clinton 0.9	Gillespie 12.0	Whitaker 6.2	Ukraine 6.2	
46. 11/12–11/18	Bannon 0.0	sexual 1.7	caravan 0.0	Ukraine 5.2	
47. 11/19–11/25	Hamilton 12.4	LaVar 21.3	Saudi 1.6	Ukraine 3.5	
48. 11/26–12/02	recount 0.0	Moore 0.0	Moscow 0.1	impeachment 3.1	
49. 12/03–12/09	Taiwan 7.8	Mueller 0.0	Cohen 2.1	impeachment 0.0	
50. 12/10–12/16	Russia 2.9	Mueller 0.0	Cohen 6.9	impeachment 0.0	
51. 12/17–12/23	inauguration 11.8	Mueller 0.0	wall 9.8	impeachment 1.4	
52. 12/24–12/31	inauguration 3.2	Mueller 0.0	wall 20.4	impeachment 7.6	

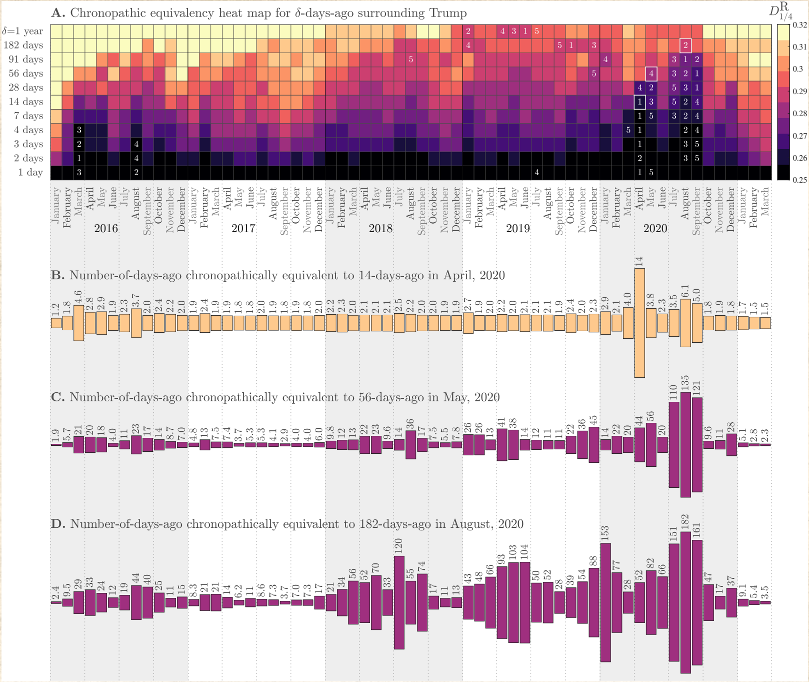
narrative control

http://compstorylab.org/trumpstoryturbulence/

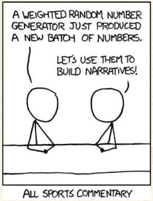
Week	2016	2017	2018	2019	2020	2021
1. 01/01–01/07	Hillary 34.7	hacking 28.6	Bannon 2.2	shutdown 0.0	Iran 9.6	Georgia 14.7
2. 01/08–01/14	Cruz 1.0	Meryl 5.0	Mueller 0.0	shutdown 0.0	Soleimani 5.9	Capitol 0.1
3. 01/15–01/21	Cruz 10.7	inauguration 0.6	DACA 6.7	Pelosi 6.8	Parnas 0.0	Capitol 0.0
4. 01/22–01/28	Cruz 10.6	inauguration 3.1	Mueller 0.0	Pelosi 2.6	Ukraine 5.5	insurrection 0.0
5. 01/29–02/04	Cruz 11.2	ban 2.1	Mueller 0.0	border 0.0	impeachment 0.0	Greene 0.0
6. 02/05–02/11	Cruz 5.1	Bannon 0.0	memo 2.3	Whitaker 0.0	Vindman 2.5	insurrection 0.0
7. 02/12–02/18	Cruz 6.9	Flynn 0.0	Mueller 0.0	emergency 0.0	Barr 2.2	Capitol 0.0
8. 02/19–02/25	Rubio 3.8	Sweden 4.9	Parkland 0.3	Jussie 0.0	Bloomberg 6.3	Capitol 0.0
9. 02/26–03/04	Rubio 9.2	Russia 6.4	Mueller 0.0	Cohen 3.7	coronavirus 0.0	Capitol 0.0
10. 03/05–03/11	Cruz 1.0	Russian 4.8	Mueller 0.0	Nadler 13.7	coronavirus 0.0	insurrection 0.0
11. 03/12–03/18	Cruz 5.7	tax 1.8	Mueller 2.2	emergency 1.6	coronavirus 0.0	Biden 0.0
12. 03/19–03/25	Arizona 16.8	Nunes 0.0	Mueller 2.2	Barr 0.0	coronavirus 0.0	Biden 0.0
13. 03/26–04/01	women 8.3	Russia 9.9	Stormy 0.0	Schiff 5.2	coronavirus 0.5	Capitol 0.0
14. 04/02–04/08	Cruz 1.5	Russia 2.8	Mueller 0.0	returns 0.0	coronavirus 0.0	Matt 0.0
15. 04/09–04/15	Cruz 1.7	Syria 0.4	Mueller 2.0	Barr 2.4	coronavirus 0.0	Capitol 0.0
16. 04/16–04/22	Cruz 10.5	Russia 0.5	Mueller 0.1	Barr 0.1	coronavirus 0.0	Capitol 0.0
17. 04/23–04/29	Cruz 3.0	days 0.1	Kanye 8.0	Biden 6.0	coronavirus 0.0	audit 0.0
18. 04/30–05/06	Indiana 11.5	Trumpcare 0.0	Mueller 0.0	Barr 0.0	coronavirus 0.0	Cheney 0.0
19. 05/07–05/13	Ryan 2.5	Comey 2.8	Iran 6.6	Barr 0.0	coronavirus 0.0	Cheney 0.0
20. 05/14–05/20	Bernie 25.3	Comey 1.0	ZTE 4.5	Barr 0.0	coronavirus 0.0	Cheney 0.0
21. 05/21–05/27	Clinton 9.5	budget 0.0	Korea 18.2	Barr 0.0	pandemic 0.0	Weisselberg 0.0
22. 05/28–06/03	Hillary 11.9	Kathy 4.4	Roseanne 4.0	USS 3.0	Minneapolis 32.1	reinstated 0.0
23. 06/04–06/10	Clinton 11.1	Comey 0.8	pardon 0.0	Mexico 27.6	police 4.2	McGahn 0.0
24. 06/11–06/17	Orlando 12.4	Mueller 0.0	Kim 4.1	foreign 2.0	Tulsa 4.5	DOJ 0.0
25. 06/18–06/24	Hillary 23.9	Trumpcare 0.0	children 1.0	Iran 12.9	Tulsa 2.1	Capitol 0.0
26. 06/25–07/01	Clinton 13.0	Russia 5.8	Justice 8.3	Moon 29.9	bounties 0.0	Organization 0.0
27. 07/02–07/08	Crooked 80.6	CNN 0.7	toddlers 0.0	parade 0.0	Rushmore 2.3	Weisselberg 0.0
28. 07/09–07/15	Crooked 71.5	Russian 1.2	NATO 13.0	Epstein 0.0	coronavirus 0.0	CPAC 0.0
29. 07/16–07/22	Pence 2.9	Mueller 0.0	Helsinki 3.1	racist 0.8	coronavirus 0.0	vaccinated 0.0
30. 07/23–07/29	DNC 6.1	Scouts 0.0	Cohen 0.0	Baltimore 13.6	Portland 11.8	Jan 0.0
31. 07/30–08/05	Khan 6.5	Mueller 0.0	LeBron 0.7	Baltimore 9.4	pandemic 0.0	Capitol 0.0
32. 08/06–08/12	Crooked 55.2	Korea 5.8	Omarosa 0.4	Paso 7.6	USPS 0.0	Rosen 0.0
33. 08/13–08/19	Manafort 0.0	Charlottesville 1.5	Omarosa 9.5	Greenland 6.9	USPS 0.0	Taliban 0.0
34. 08/20–08/26	Clinton 7.6	Charlottesville 3.8	Cohen 2.7	Greenland 8.0	Biden 6.6	Taliban 0.0
35. 08/27–09/02	Crooked 57.4	Harvey 0.0	Ohr 14.0	Dorian 12.2	Kenosha 9.5	Taliban 0.0
36. 09/03–09/09	Bondi 0.0	DACA 2.4	Kavanaugh 2.1	Dorian 12.6	Atlantic 4.8	Afghanistan 0.0
37. 09/10–09/16	deplorable 0.0	ESPN 2.7	Puerto 7.5	flavored 0.0	Woodward 2.6	Milley 0.0
38. 09/17–09/23	Clinton 6.5	Kim 4.9	Kavanaugh 1.7	Ukraine 4.5	coronavirus 0.0	Eastman 0.0
39. 09/24–09/30	debate 4.9	Puerto 4.7	Kavanaugh 9.5	Ukraine 6.8	ballots 0.7	audit 0.0
40. 10/01–10/07	Pence 4.9	Puerto 2.1	Kavanaugh 6.8	Ukraine 5.1	Covid 1.4	Bannon 0.0
41. 10/08–10/14	sexual 0.3	Puerto 1.8	Kavanaugh 4.3	Kurds 8.2	COVID 1.4	Jan 0.0
42. 10/15–10/21	rigged 10.1	Puerto 0.2	Saudi 5.3	Kurds 3.7	Biden 8.2	Powell 0.0
43. 10/22–10/28	star 0.0	Mueller 0.0	caravan 0.0	impeachment 0.0	Biden 9.2	Jan 0.0
44. 10/29–11/04	FBI 5.9	Mueller 0.0	caravan 0.0	impeachment 0.0	Biden 10.0	Youngkin 0.0
45. 11/05–11/11	Clinton 0.9	Gillespie 12.0	Whitaker 6.2	Ukraine 6.2	votes 3.4	infrastructure 0.0
46. 11/12–11/18	Bannon 0.0	sexual 1.7	caravan 0.0	Ukraine 5.2	Dominion 23.2	Christie 0.0
47. 11/19–11/25	Hamilton 12.4	LaVar 21.3	Saudi 1.6	Ukraine 3.5	Sidney 0.1	Rittenhouse 0.0
48. 11/26–12/02	recount 0.0	Moore 0.0	Moscow 0.1	impeachment 3.1	votes 24.1	Waukesha 0.0
49. 12/03–12/09	Taiwan 7.8	Mueller 0.0	Cohen 2.1	impeachment 0.0	Georgia 20.2	Meadows 0.0
50. 12/10–12/16	Russia 2.9	Mueller 0.0	Cohen 6.9	impeachment 0.0	vaccine 11.1	Meadows 0.0
51. 12/17–12/23	inauguration 11.8	Mueller 0.0	wall 9.8	impeachment 1.4	vaccine 15.4	Manchin 0.0
52. 12/24–12/31	inauguration 3.2	Mueller 0.0	wall 20.4	impeachment 7.6	Election 60.2	Brandon 0.0

Week	2016	2017	2018	2019	2020	2021
1. 01/01–01/07	Hillary Clinton 32.7	plant in 85.1	Steve Bannon 5.7	the government 0.0	a war 6.6	in Georgia 20.2
2. 01/08–01/14	Trump rally 0.0	Meryl Streep 6.6	shithole countries 0.0	the border 1.0	impeachment trial 0.0	the Capitol 0.0
3. 01/15–01/21	Ted Cruz 26.0	Trump's inauguration 0.0	the government 1.4	Cohen to 0.0	impeachment trial 0.0	the Capitol 0.0
4. 01/22–01/28	Megyn Kelly 4.9	executive order 0.0	the FBI 5.6	the government 0.0	impeachment trial 0.0	the Capitol 0.0
5. 01/29–02/04	Ted Cruz 19.7	travel ban 1.6	the FBI 9.4	Ralph Northam 26.0	impeachment trial 0.0	the Capitol 0.0
6. 02/05–02/11	New Hampshire 19.5	travel ban 1.1	military parade 0.0	El Paso 4.7	Alexander Vindman 0.0	the Capitol 0.0
7. 02/12–02/18	Ted Cruz 15.7	Michael Flynn 0.0	school shooting 3.1	national emergency 0.0	Roger Stone 4.0	the Capitol 0.0
8. 02/19–02/25	Ted Cruz 30.1	Trump administration 0.0	the NRA 0.0	Jussie Smollett 0.0	Bernie Sanders 13.6	the Capitol 0.0
9. 02/26–03/04	vote for 4.4	to Russia 22.0	Hope Hicks 0.0	Michael Cohen 5.3	the coronavirus 0.0	the Capitol 0.0
10. 03/05–03/11	Ted Cruz 2.4	travel ban 0.0	Stormy Daniels 0.0	Tim Apple 0.0	the coronavirus 0.0	voted for 0.0
11. 03/12–03/18	Trump is 0.1	Meals on 0.0	Stormy Daniels 0.0	New Zealand 17.9	the coronavirus 0.0	Lara Trump 0.0
12. 03/19–03/25	Lyin' Ted 66.2	health care 0.0	Cambridge Analytica 0.0	Mueller report 0.0	the coronavirus 0.0	the border 0.0
13. 03/26–04/01	Trump is 0.0	Freedom Caucus 20.8	Stormy Daniels 0.0	Mueller report 0.0	the coronavirus 0.0	Matt Gaetz 0.0
14. 04/02–04/08	Ted Cruz 3.9	Susan Rice 0.3	National Guard 0.0	tax returns 0.0	the coronavirus 0.0	Matt Gaetz 0.0
15. 04/09–04/15	New York 19.3	in Syria 0.2	Michael Cohen 0.0	sanctuary cities 5.3	the coronavirus 0.0	Matt Gaetz 0.0
16. 04/16–04/22	Ted Cruz 28.1	turnout for 0.0	Michael Cohen 2.4	Mueller report 0.0	the coronavirus 0.0	Maxine Waters 0.0
17. 04/23–04/29	Trump rally 0.0	tax plan 0.0	the Korean 0.0	Mueller report 0.0	the coronavirus 0.0	Liz Cheney 0.0
18. 04/30–05/06	Ted Cruz 5.5	health care 0.0	Stormy Daniels 0.0	Mueller report 0.0	treated worse 0.0	Liz Cheney 0.0
19. 05/07–05/13	Paul Ryan 2.0	James Comey 6.7	the Iran 9.0	tax returns 0.0	tested positive 0.0	Liz Cheney 0.0
20. 05/14–05/20	Hillary Clinton 26.5	Saudi Arabia 12.5	are animals 0.0	Lindsey Graham 0.0	the pandemic 0.0	Kevin McCarthy 0.0
21. 05/21–05/27	Hillary Clinton 24.8	Saudi Arabia 8.2	the FBI 23.3	Nancy Pelosi 12.5	a mask 6.3	the January 0.0
22. 05/28–06/03	Trump University 3.4	Kathy Griffin 5.7	Samantha Bee 4.4	John McCain 0.0	photo op 0.0	Memorial Day 0.0
23. 06/04–06/10	Hillary Clinton 18.6	James Comey 0.2	Justin Trudeau 8.5	with Mexico 39.2	Left Democrats 75.1	Jean Carroll 0.0
24. 06/11–06/17	Trump is 0.0	obstruction of 12.6	their parents 0.0	the FBI 8.5	in Tulsa 7.4	Trump DOJ 0.0
25. 06/18–06/24	Hillary Clinton 20.6	Karen Handel 16.6	their parents 3.4	need soap 0.0	in Tulsa 2.2	the Capitol 0.0
26. 06/25–07/01	Hillary Clinton 20.5	Fake News 37.6	Supreme Court 3.7	Jean Carroll 0.0	American soldiers 0.0	Trump Organization 0.0
27. 07/02–07/08	Crooked Hillary 82.8	North Korea 28.6	Trump administration 0.0	Jeffrey Epstein 0.0	Mount Rushmore 3.9	Ashli Babbitt 0.0
28. 07/09–07/15	Crooked Hillary 73.3	Trump Jr 0.0	Supreme Court 7.9	Jeffrey Epstein 0.0	Roger Stone 0.0	the Capitol 0.0
29. 07/16–07/22	Mike Pence 6.8	Secret Service 0.0	in Helsinki 1.7	a racist 0.0	in Portland 0.0	Tom Barrack 0.0
30. 07/23–07/29	Crooked Hillary 79.6	Boy Scouts 0.0	Walk of 0.0	Elijah Cummings 27.2	in Portland 8.9	the Capitol 0.0
31. 07/30–08/05	Khizr Khan 0.0	Maxine Waters 0.0	enemy of 22.2	El Paso 11.1	the election 3.4	the Capitol 0.0
32. 08/06–08/12	Hillary Clinton 10.5	North Korea 5.7	Space Force 11.1	El Paso 7.7	Social Security 0.0	overturn the 0.0
33. 08/13–08/19	Trump campaign 0.0	white supremacists 0.0	security clearance 0.0	New Hampshire 26.5	the USPS 0.0	the Taliban 0.0
34. 08/20–08/26	Hillary Clinton 19.1	Joe Arpaio 3.5	Michael Cohen 4.3	Prime Minister 28.7	Joe Biden 5.9	the Taliban 0.0
35. 08/27–09/02	Crooked Hillary 61.8	Hurricane Harvey 0.1	John McCain 0.2	Hurricane Dorian 9.6	Joe Biden 2.7	the Taliban 0.0
36. 09/03–09/09	in Detroit 0.0	to end 0.0	Brett Kavanaugh 7.6	the Taliban 3.0	Joe Biden 3.4	Robert E 0.0
37. 09/10–09/16	tax returns 0.0	white supremacist 0.0	Puerto Rico 8.4	Dan Bishop 37.7	Joe Biden 13.3	the Taliban 0.0
38. 09/17–09/23	Trump Jr 0.0	North Korea 12.8	Blasey Ford 0.0	a foreign 6.4	Supreme Court 7.3	to overturn 0.0
39. 09/24–09/30	Hillary Clinton 7.5	Puerto Rico 5.2	Brett Kavanaugh 15.7	impeachment inquiry 0.0	Supreme Court 5.7	debt ceiling 0.0
40. 10/01–10/07	Mike Pence 8.9	Puerto Rico 2.6	Supreme Court 6.9	Adam Schiff 13.3	Walter Reed 5.7	the debt 0.0
41. 10/08–10/14	sexual assault 0.0	Puerto Rico 2.2	Kanye West 0.0	the Kurds 11.3	Biden is 26.5	the January 0.0
42. 10/15–10/21	Hillary Clinton 19.9	families of 0.0	Saudi Arabia 6.6	the Kurds 3.8	Joe Biden 12.1	the January 0.0
43. 10/22–10/28	Hillary Clinton 11.7	Myeshia Johnson 0.0	the bombs 0.0	World Series 0.0	Joe Biden 10.1	Alec Baldwin 0.0
44. 10/29–11/04	Hillary Clinton 6.5	Twitter employee 0.0	birthright citizenship 0.0	the impeachment 0.0	Joe Biden 12.6	in Virginia 0.0
45. 11/05–11/11	Trump wins 0.0	mental health 0.0	Jim Acosta 0.0	pro quo 8.1	the election 2.2	infrastructure bill 0.0
46. 11/12–11/18	Steve Bannon 0.0	ban on 0.0	president who 0.0	impeachment inquiry 0.0	the election 7.5	Chris Christie 0.0
47. 11/19–11/25	Mike Pence 24.3	Roy Moore 0.0	Saudi Arabia 2.5	quid pro 1.3	the election 6.7	Kyle Rittenhouse 0.0
48. 11/26–12/02	popular vote 17.4	Native American 0.1	Trump Tower 2.5	Hong Kong 0.0	voter fraud 32.2	Donald Trump 0.0
49. 12/03–12/09	Air Force 18.2	Roy Moore 3.5	campaign finance 0.0	to impeach 7.7	in Georgia 12.9	Donald Trump 0.0
50. 12/10–12/16	of State 7.6	of sexual 0.0	Michael Cohen 7.8	articles of 0.0	the election 9.0	Mark Meadows 0.0
51. 12/17–12/23	Electoral College 5.8	tax bill 0.0	the wall 13.7	Christianity Today 8.1	election fraud 13.9	the Capitol 0.0


2016
2017
2018
2019
2020
2021
Obama 4.2
Hillary 8.5
Crooked Hillary 59.2
Rubio 5.9
Trump wins 0.0
Comey 2.9
Roger Stone 3.3
Ted Cruz 10.9
Mueller 4.4
bounties 0.0
popular vote 5.6
Stormy Daniels 0.0
Jeffrey Epstein 0.0
Ghislaine Maxwell 0.0
Khizr Khan 0.0
inauguration 1.5
Charlottesville 1.6
Brett Kavanaugh 10.2
Left Democrats 79.7
Megyn Kelly 1.3
Trump Tower 1.6
children 2.1
impeachment 1.1
Mike Pence 8.1
Flynn 0.9
Michael Cohen 2.4
Ukraine 4.4
debate 4.8
travel ban 0.8
Joe Biden 7.0
Iran 9.6
rigged 12.4
Saudi Arabia 10.4
Space Force 6.6
Khashoggi 0.0
Greenland 7.1
coronavirus 0.1
Syria 7.4
Minneapolis 27.9
Black Lives 0.0
North Korea 15.4
the Kurds 9.1
Black Lives 1.4
tax returns 0.7
Puerto Rico 3.8
tax returns 0.0
Portland 10.2
women 3.6
fraud 6.2
sexual assault 0.0
Georgia 15.2
golf 2.2
insurrection 0.0
Russia 4.7
Supreme Court 9.6
Fake News 56.0
Witch Hunt 62.0
Capitol 1.0
our Country 70.3


A. Chronopathic equivalency heat map for δ-days-ago surrounding Trump
$D^{R}_{1/4}$
0.32
0.31
0.3
0.29
0.28
0.27
0.26
0.25
δ=1 year
182 days
91 days
56 days
28 days
14 days
7 days
4 days
3 days
2 days
1 day
2016
2017
2018
2019
2020
B. Number-of-days-ago chronopathically equivalent to 14-days-ago in April, 2020
C. Number-of-days-ago chronopathically equivalent to 56-days-ago in May, 2020
D. Number-of-days-ago chronopathically equivalent to 182-days-ago in August, 2020

Understanding the Sociotechnocene—Stories:

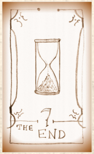


xkcd.com/904/

- Toward a Science of Stories.
- Claim: Homo narrativus—we run on stories.
- “What’s the John Dory?”
- “They’ve lost the plot/thread”
- Narrative hierarchies and scalability of stories.
- Research: Real-time and offline extraction of metaphors, frames, plots, narratives, conspiracy theories, and stories from large-scale text.
- Research: The taxonomy of human stories.
- To be built: Storyscopes—improvable, online, interactive instruments.

ding!

On Instagram at pratchett_the_cat

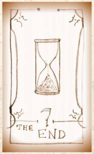


References I

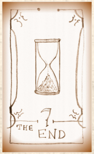
[1] P. E. Auerswald.
The Code Economy: A Forty-Thousand Year History.
Oxford University Press, 2017.

[2] P. Bak, K. Christensen, L. Danon, and T. Scanlon.
Unified scaling law for earthquakes.
Phys. Rev. Lett., 88:178501, 2002. pdf

[3] B. Boyd.
On the Origin of Stories: Evolution, Cognition, and Fiction.
Belknap Press, 2010.



References II



[4] 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

[5] E. Cheng.
How to bake pi: An edible exploration of the mathematics of mathematics.
Basic Books, 2015.

[6] K. Christensen, L. Danon, T. Scanlon, and P. Bak.
Unified scaling law for earthquakes.
Proc. Natl. Acad. Sci., 99:2509–2513, 2002. pdf

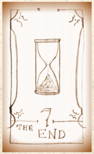
References III

[7] B. Christian and T. Griffiths.
Algorithms to Live By.
Macmillan, 2016.

[8] 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

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

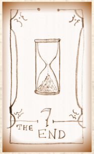
[10] L. R. Dice.
Measures of the amount of ecologic association between species.
Ecology, 26:297–302, 1945.



References IV

[11] 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, 2020.
Available online at https://arxiv.org/abs/2002.09770. pdf

[12] P. S. Dodds, J. R. Minot, M. V. Arnold, T. Alshaabi, J. L. Adams, D. R. Dewhurst, A. J. Reagan, and C. M. Danforth.
Fame and Ultrafame: Measuring and comparing daily levels of 'being talked about' for United States' presidents, their rivals, God, countries, and K-pop, 2019.

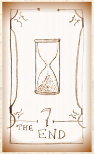


References V

Available online at https://arxiv.org/abs/1910.00149. pdf

[13] P. S. Dodds, J. R. Minot, M. V. Arnold, T. Alshaabi, J. L. Adams, D. R. Dewhurst, 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

[14] P. S. Dodds, J. R. Minot, M. V. Arnold, T. Alshaabi, J. L. Adams, A. J. Reagan, and C. M. Danforth.
Computational timeline reconstruction of the stories surrounding Trump: Story turbulence, narrative control, and collective chronopathy, 2020.



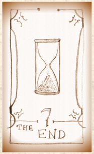
References VI

https://arxiv.org/abs/2008.07301. pdf

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

[16] M. Erwig.
Once Upon an Algorithm.
MIT Press, 2017.

[17] J. Gottschall.
The Storytelling Animal: How Stories Make Us Human.
Mariner Books, 2013.

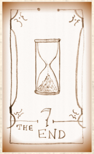


References VII

[18] 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

[19] E. Katz and P. F. Lazarsfeld.
Personal Influence.
The Free Press, New York, 1955.

[20] T. Kuran.
Now out of never: The element of surprise in the east european revolution of 1989.
World Politics, 44:7–48, 1991. pdf

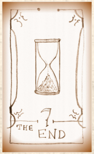


References VIII

[21] J. Lin.
Divergence measures based on the Shannon entropy.
IEEE Transactions on Information theory, 37(1):145–151, 1991. pdf

[22] 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

[23] 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

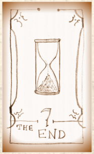


References IX

[24] J. R. Minot, M. V. Arnold, T. Alshaabi, C. M. Danforth, and P. S. Dodds.
Ratioing the President: An exploration of public engagement with Obama and Trump on Twitter, 2020.
Available online at https://arxiv.org/abs/2006.03526. pdf

[25] R. Munroe.
Thing Explainer: Complicated Stuff in Simple Words.
Houghton Mifflin Harcourt, 2015.

[26] R. Munroe.
How To: Absurd Scientific Advice for Common Real-World Problems.
Penguin, 2019.

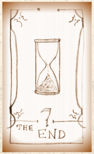


References X

[27] M. I. Norton and D. Ariely.
Building a better America—One wealth quintile at a time.
Perspectives on Psychological Science, 6:9–12, 2011. pdf

[28] 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.

[29] 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



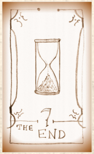
References XI

[30] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery.
Numerical Recipes in C.
Cambridge University Press, second edition, 1992.

[31] M. Puchner.
The Written World: How Literature Shaped Civilization.
Random, 2017.

[32] M. J. Salganik, P. S. Dodds, and D. J. Watts.
An experimental study of inequality and unpredictability in an artificial cultural market.
Science, 311:854–856, 2006. pdf

[33] Y. Sasaki.
The truth of the f-measure, 2007.

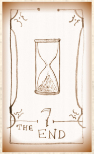


References XII

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

[35] 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.

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



References XIII

[37] D. J. Watts and P. S. Dodds.
Influentials, networks, and public opinion formation.
Journal of Consumer Research, 34:441–458, 2007. pdf

[38] 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

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