Discussion on the properties of diversity functions

February 5, 2024

Morales et al. [2020], Stirling [2007], Smith and Wilson [1996], Hurley and Rickard [2009]

Objectives

1. Most diversity functions come from ecology, can we use them?
2. Among contradictory properties, which are desirable?

Example 1

"I just got of [1] the phone with Hai and he told me how to make [2a] an adjustement [2a] on a day to day basis in regards to incorrect liquidations but he also explained this is just to make the daily P&L #'s right, if nothing were done the month and P&L would still somehow work out [3] because adjustments [2b] would be made [2b].", email-enronsent44_01-0025 (typos from original text)

Example 1

"I just got of [1] the phone with Hai and he told me how to make [2a] an adjustement [2a] on a day to day basis in regards to incorrect liquidations but he also explained this is just to make the daily P&L #'s right, if nothing were done the month and P&L would still somehow work out [3] because adjustments [2b] would be made [2b].", email-enronsent44_01-0025 (typos from original text)

Example 2

"Does this mean that for June for a certain portion of July we should not do anything and just make adjustments [1] on a going forward [2] basis (and assume everything will work out [3] at month end)?", email-enronsent44_01-0026

Flavours of diversity

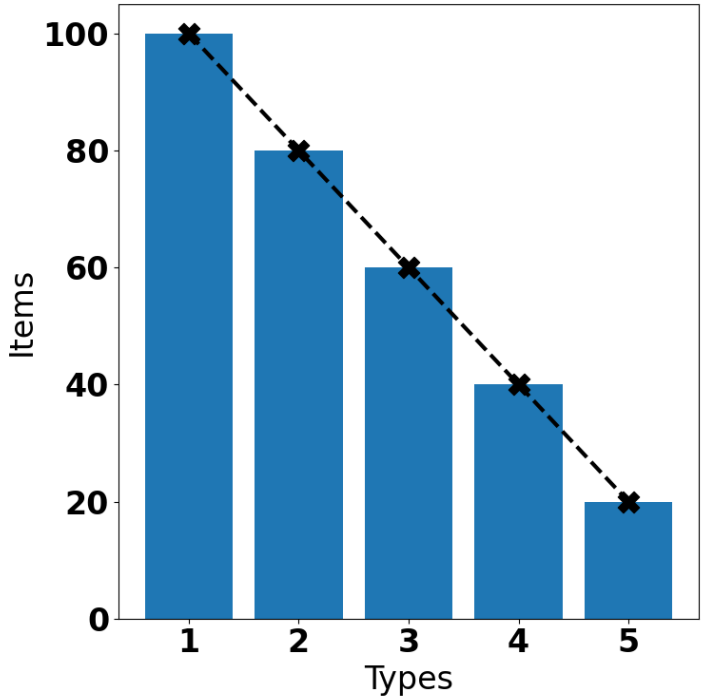
- Variety: how many types there are
- Balance: how even their distribution is
- Disparity: how fundamentally different they are

Lion-Bouton et al. [2022]

Types / items

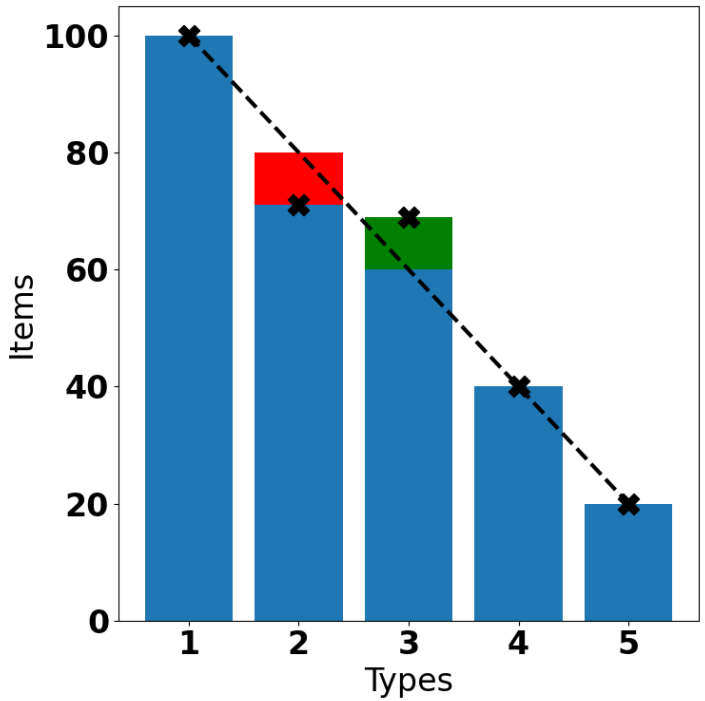
Concept		Possible types	Possible items
Segment annotation	Multi-Word Expressions	MWE classes (VID, LVC, VPC, ...)	Canonical forms
		Canonical forms	All observed forms
		All observed forms	Instances
	Named Entitites	Named entity classes (PER, ORG, LOC, ...)	Standardised forms
		Standardised entities	All observed forms
		All observed forms	Instances
Syntactic dependencies		Dependency type	Instances
		Dependency type + parent element	
		Dependency type + child element	

Base distribution



Convergence to equilibrium (1)

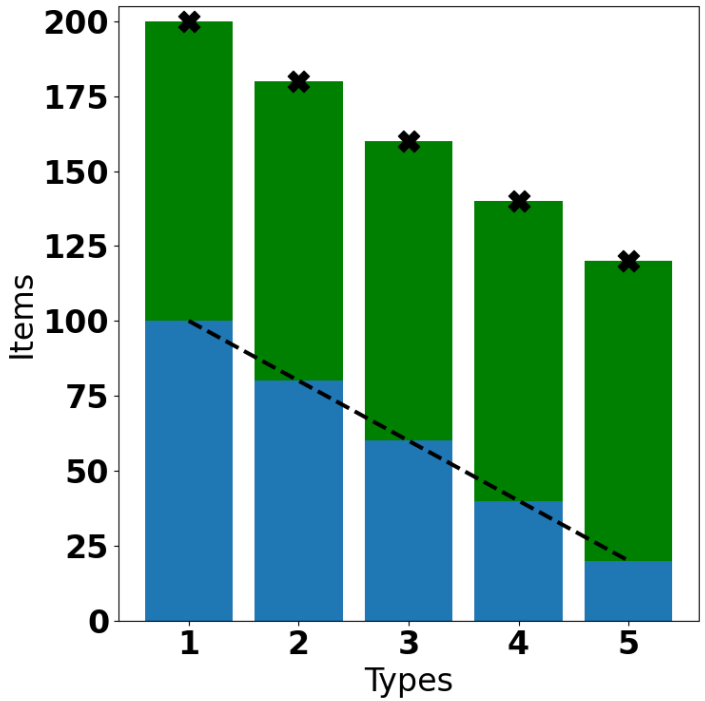
Pair convergence



Increase in diversity?

Convergence to equilibrium (2)

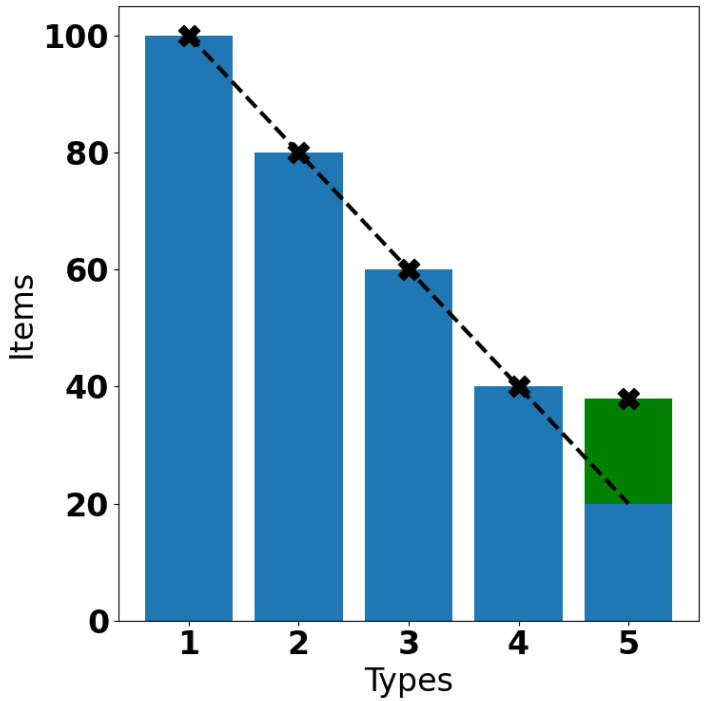
Global convergence



Increase in diversity?

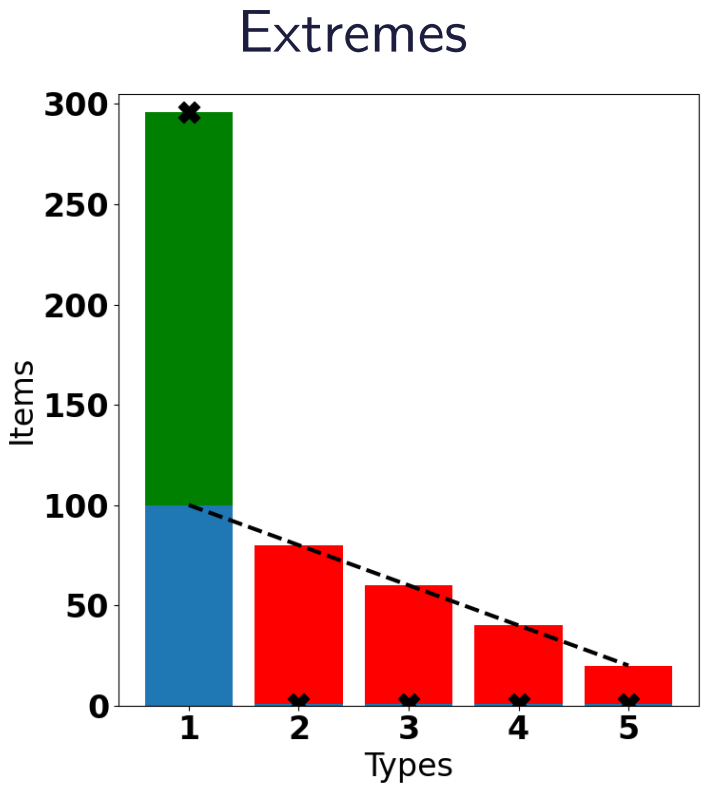
Convergence to equilibrium (3)

Increase in lowest type



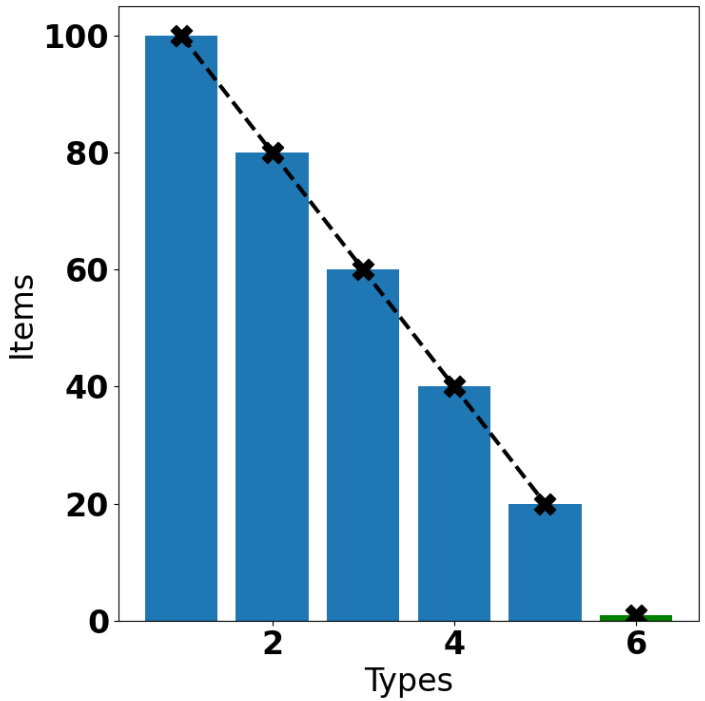
Increase in diversity?

Convergence to equilibrium (4)



Adding new types (1)

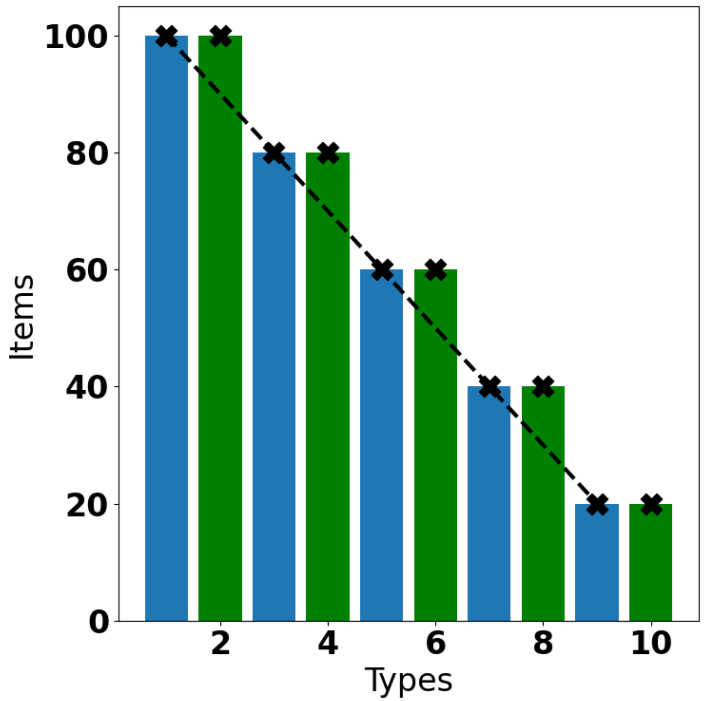
New lowest type



How should diversity react?

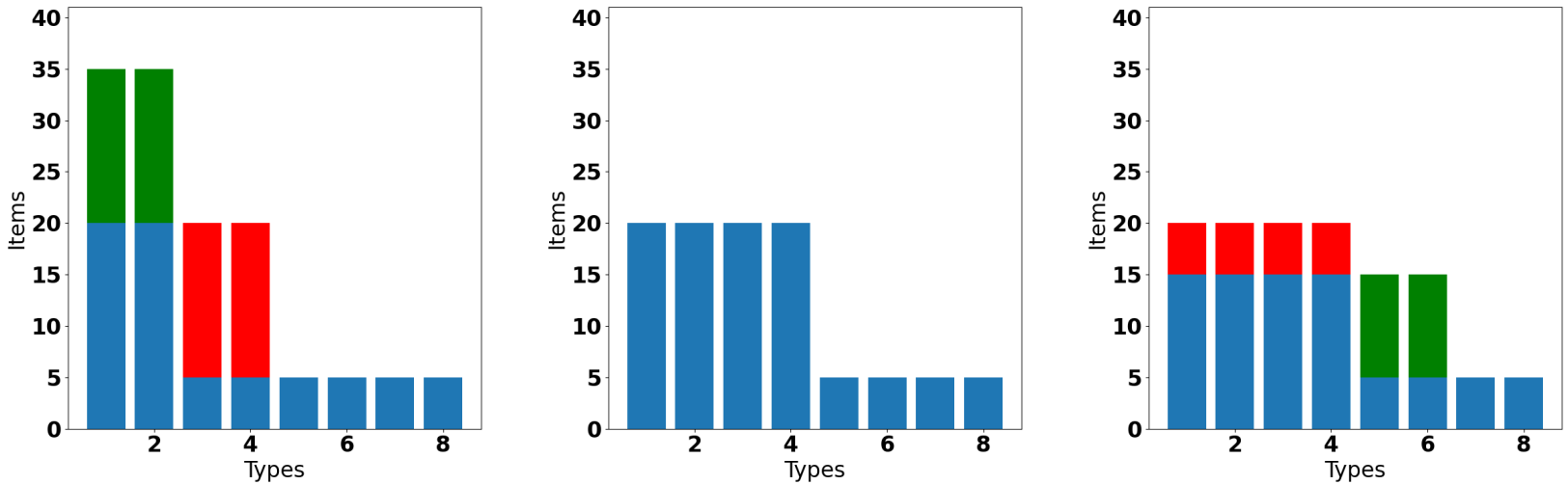
Adding new types (2)

Replication



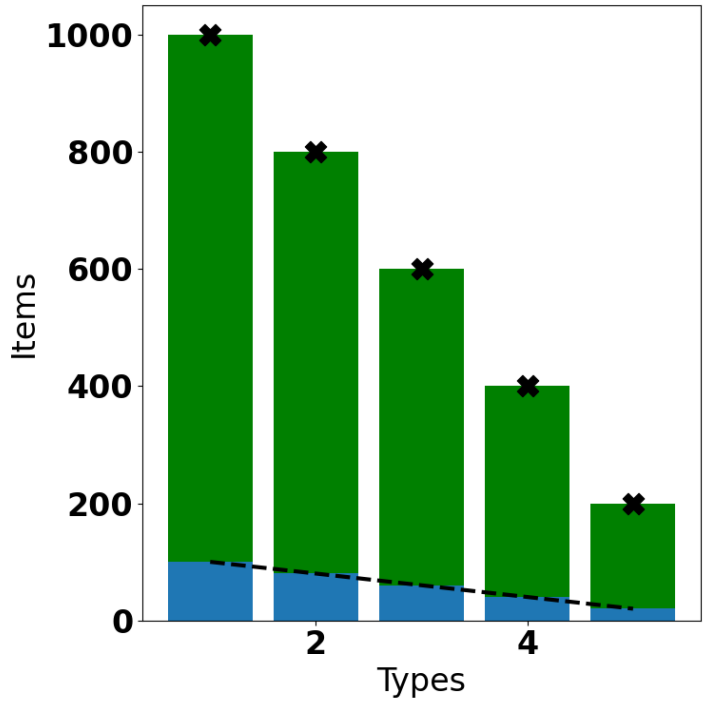
How should diversity react?

Behaviour in specific distributions



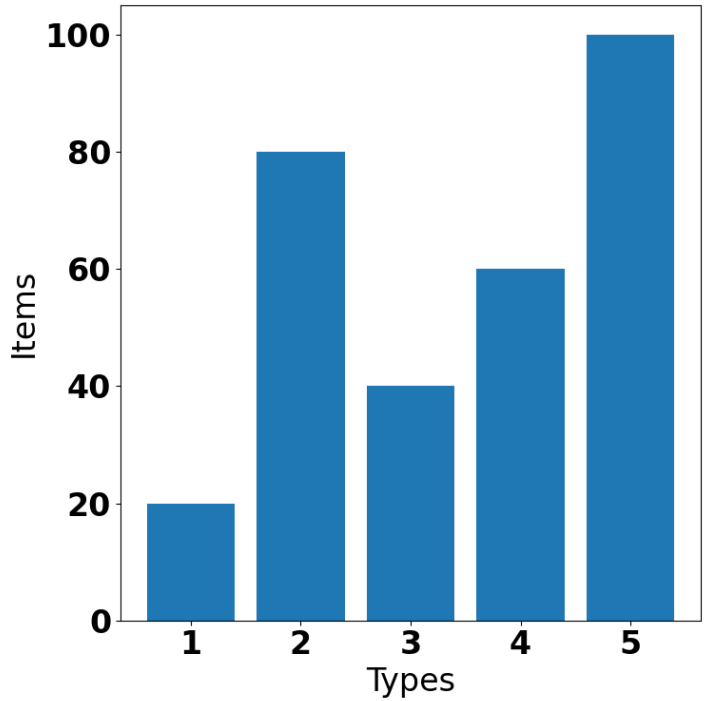
How should diversity react?

Other (1): scale invariance



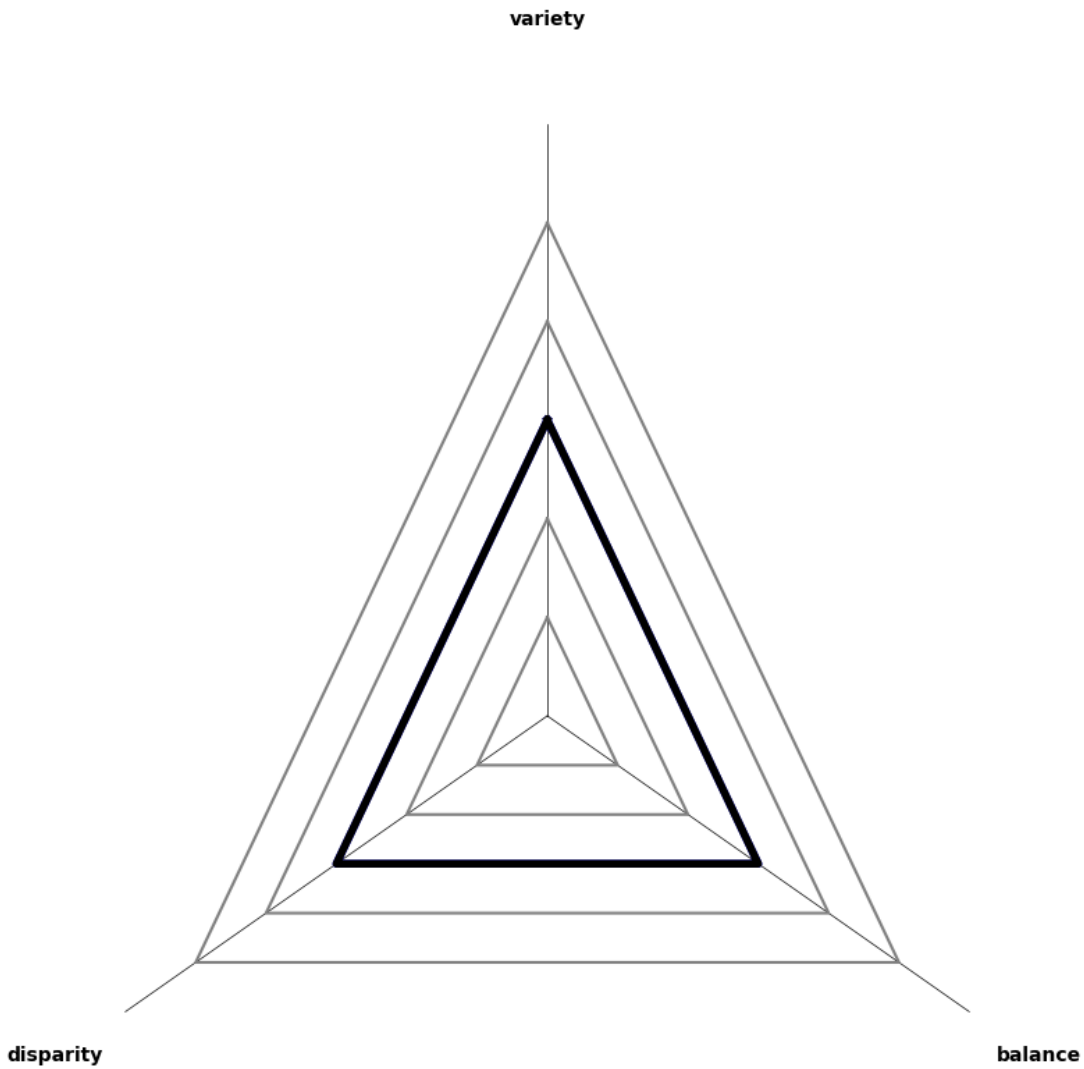
How should diversity react?

Other (2): shuffle invariance

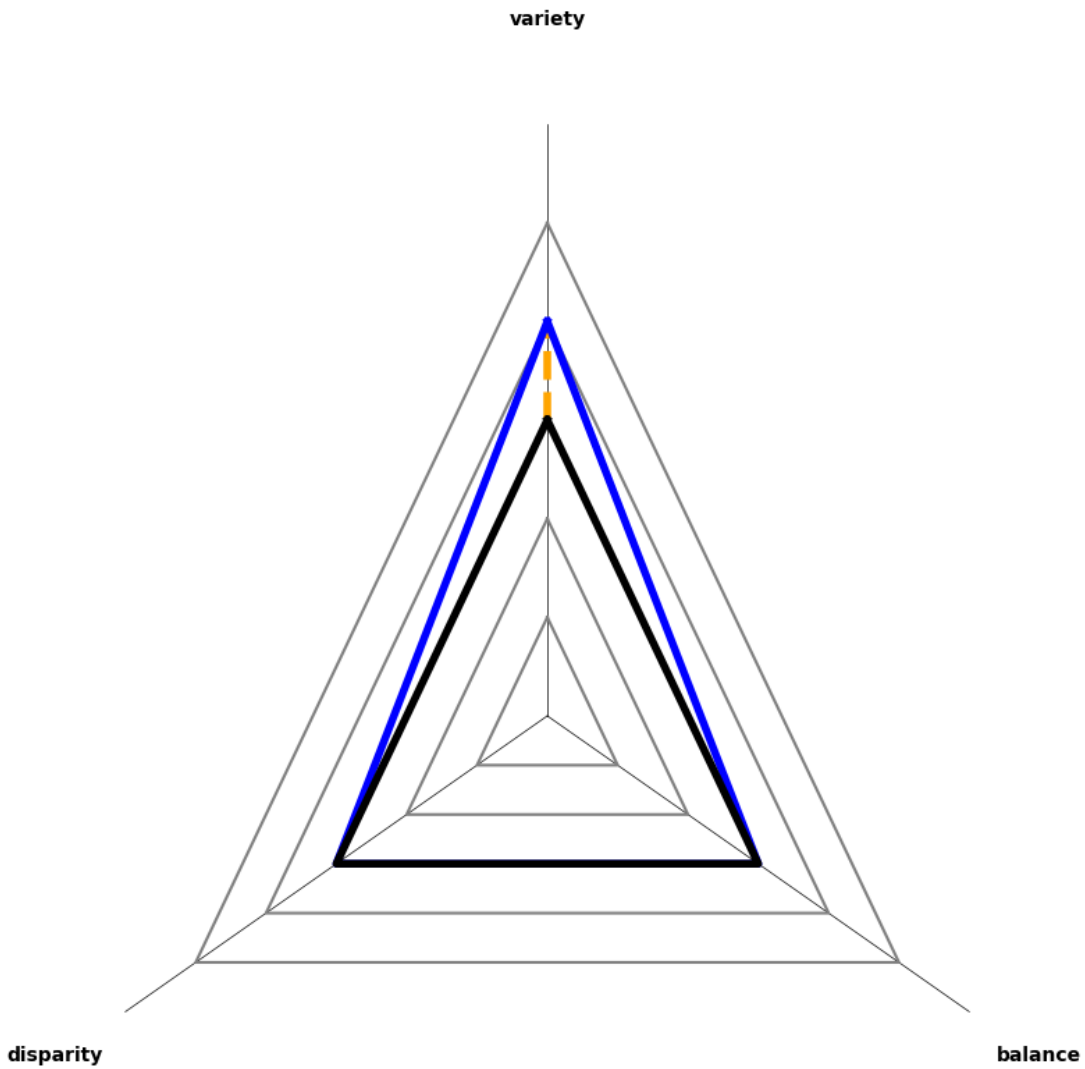


How should diversity react?

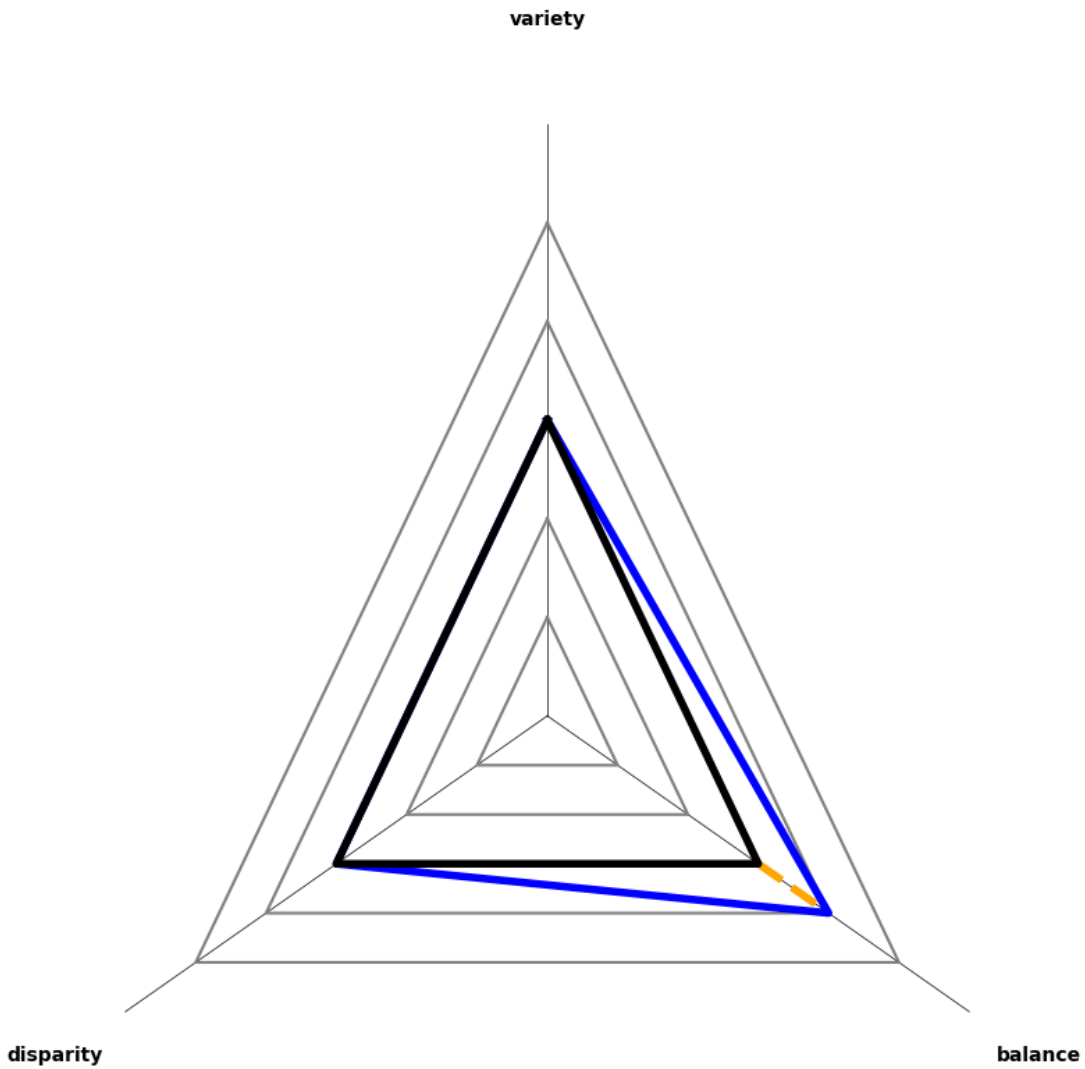
Aggregate diversity indices (1)



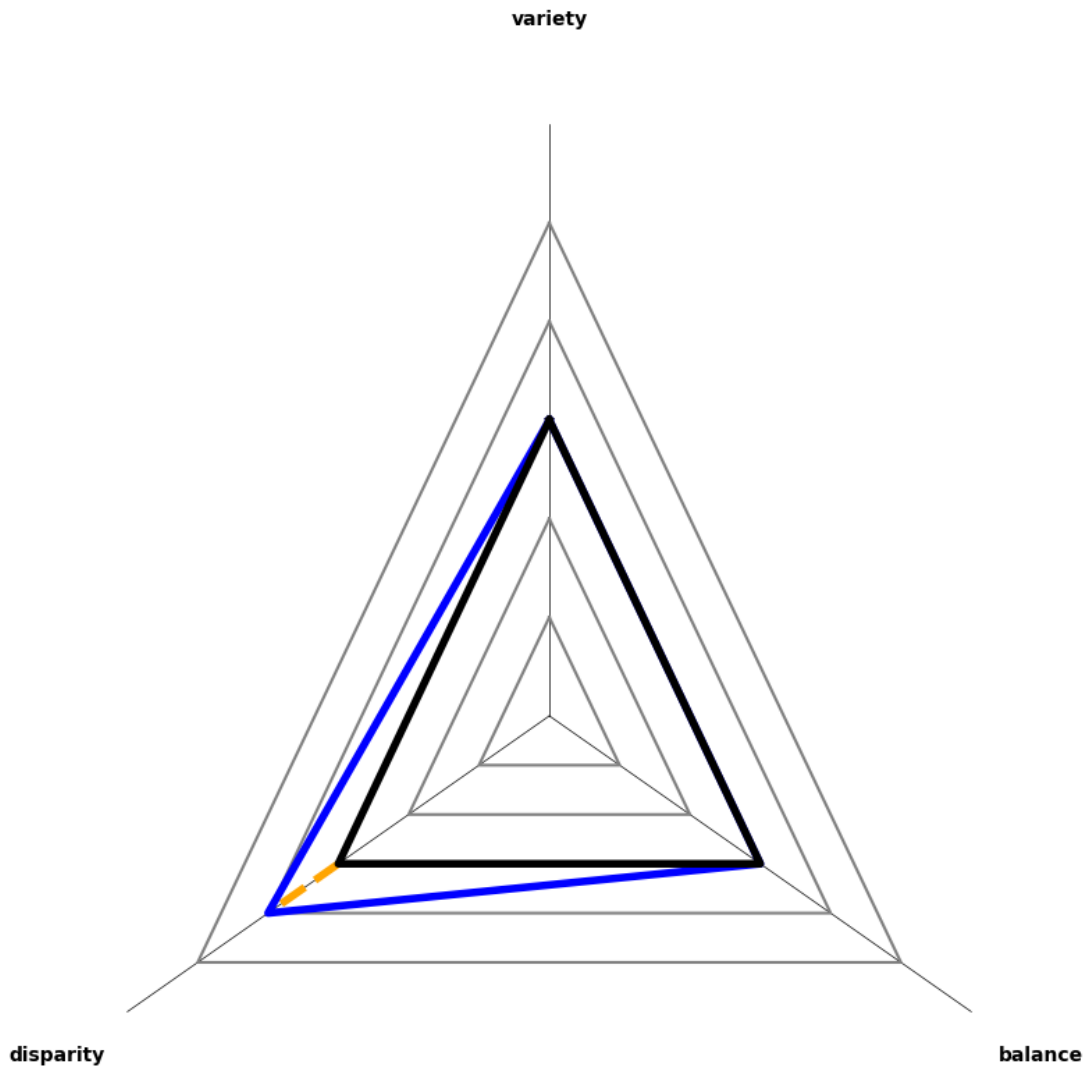
Aggregate diversity indices (2): Monotonicity of variety



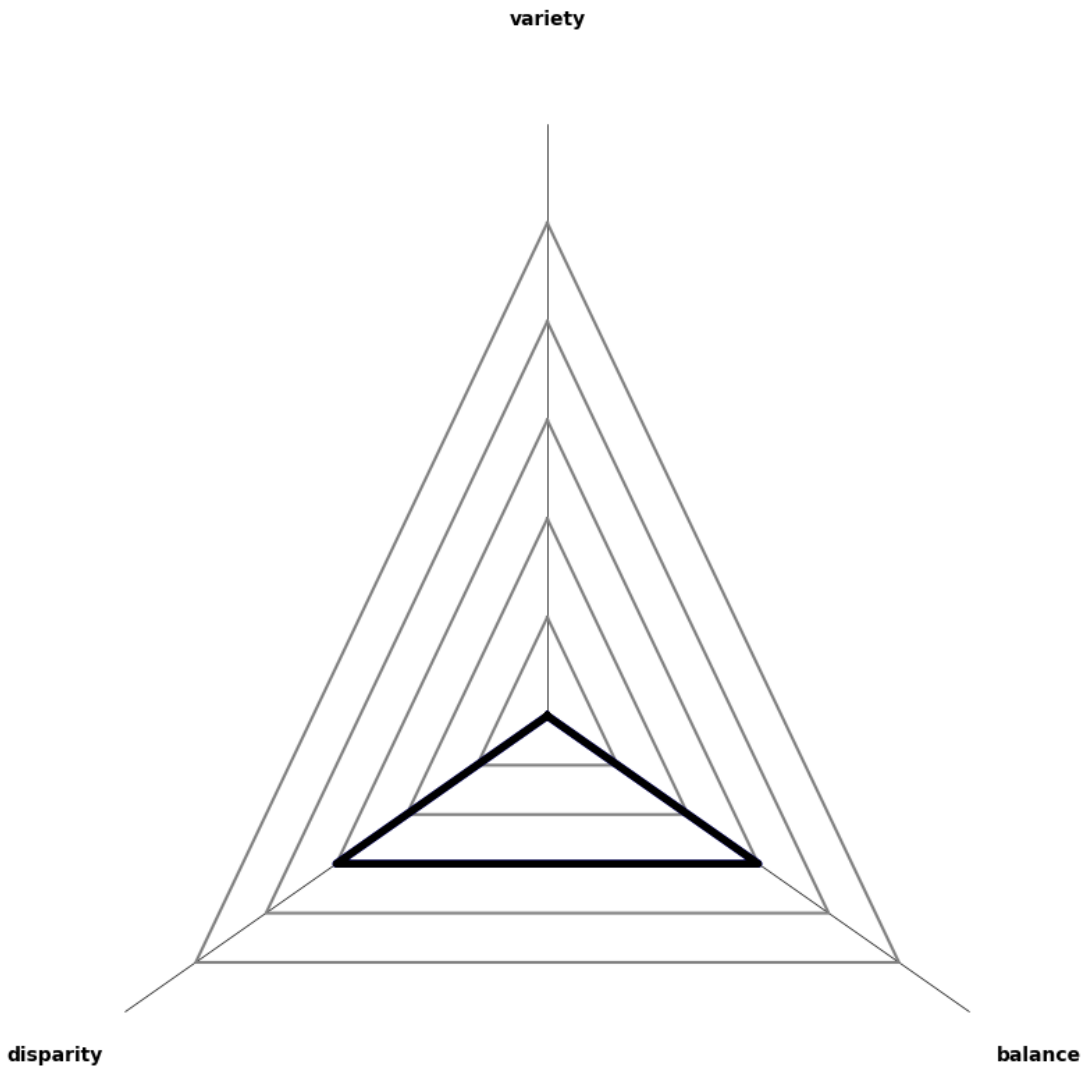
Aggregate diversity indices (3): Monotonicity of balance



Aggregate diversity indices (4): Monotonicity of disparity

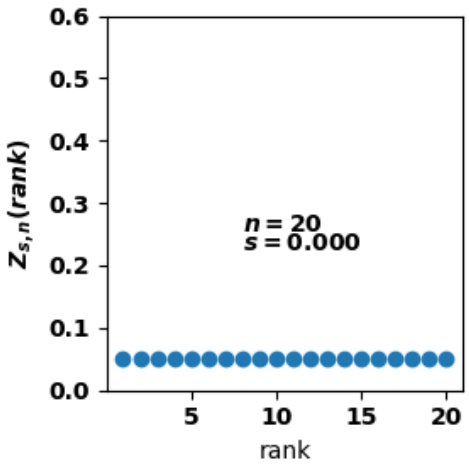


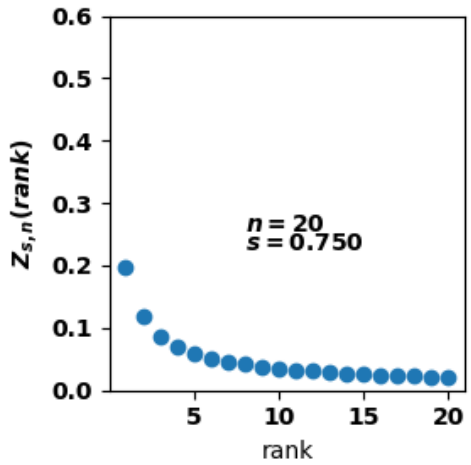
Aggregate diversity indices (5): Scaling of variety

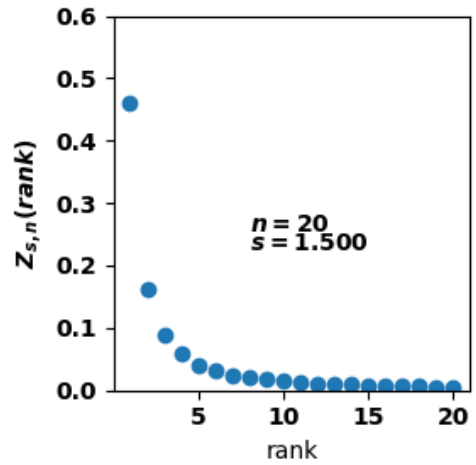


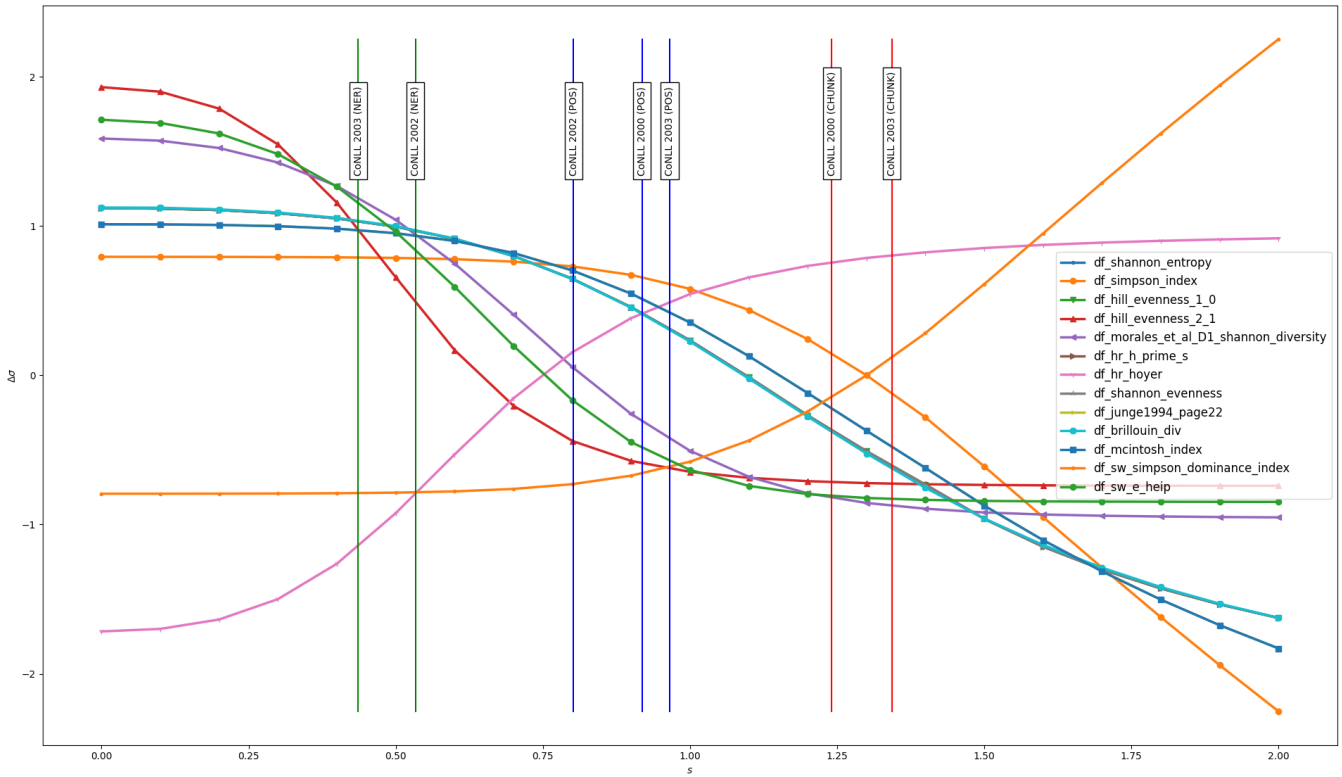
Aggregate diversity indices (6): Scaling of disparity

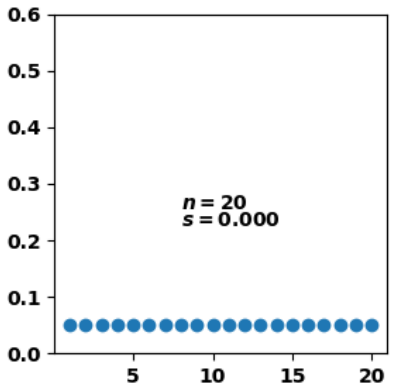
Behaviour depending on Zipfian parameters (1)

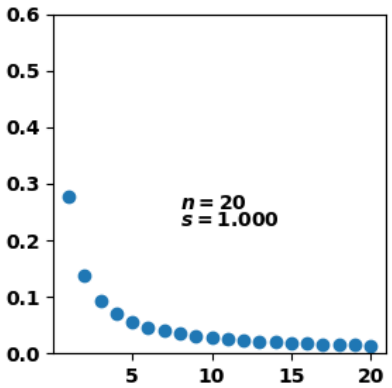


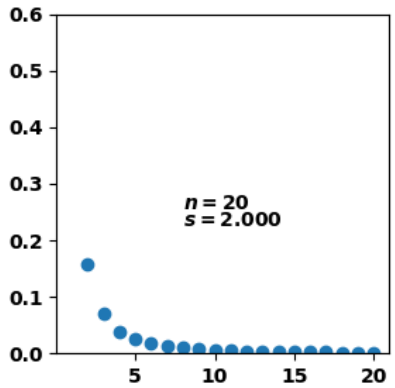




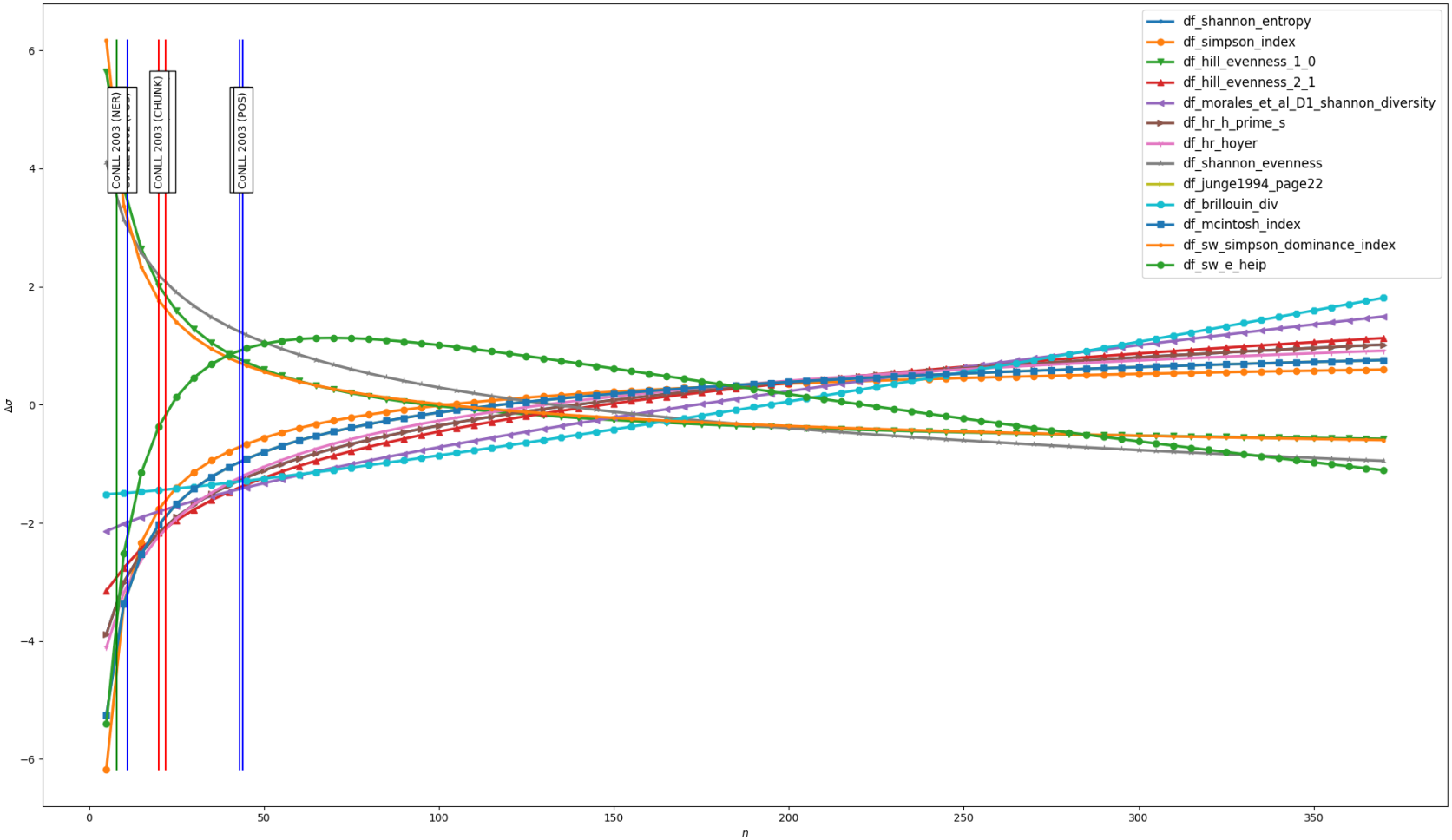








For the sake of readability, $n = 1000$ in the upper figure, and $n = 20$ in the lower figures.


df_shannon_entropy
df_simpson_index
df_hill_evenness_1_0
df_hill_evenness_2_1
df_morales_et_al_D1_shannon_diversity
df_hr_h_prime_s
df_hr_hoyer
df_shannon_evenness
df_junge1994_page22
df_brillouin_div
df_mcintosh_index
df_sw_simpson_dominance_index
df_sw_e_heip
CoNLL 2003 (NER)
CoNLL 2003 (CHUNK)
CoNLL 2003 (POS)
Δσ
n
6
4
2
0
−2
−4
−6
0
50
100
150
200
250
300
350

References

Niall P. Hurley and Scott T. Rickard. Comparing Measures of Sparsity, April 2009. URL http://arxiv.org/abs/0811.4706. arXiv:0811.4706 [cs, math].

Adam Lion-Bouton, Yagmur Ozturk, Agata Savary, and Jean-Yves Antoine. Evaluating Diversity of Multiword Expressions in Annotated Text. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3285–3295, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics. URL https://aclanthology.org/2022.coling-1.290.

Pedro Ramaciotti Morales, Robin Lamarche-Perrin, Raphael Fournier-S'niehotta, Remy Poulain, Lionel Tabourier, and Fabien Tarissan. Measuring Diversity in Heterogeneous Information Networks, December 2020. URL http://arxiv.org/abs/2001.01296. arXiv:2001.01296 [cs, math].

Benjamin Smith and J. Bastow Wilson. A Consumer's Guide to Evenness Indices. *Oikos*, 76(1):70–82, 1996. ISSN 0030-1299. doi: 10.2307/3545749. URL https://www.jstor.org/stable/3545749. Publisher: [Nordic Society Oikos, Wiley].

Andy Stirling. A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, 4(15):707–719, February 2007. doi: 10.1098/rsif.2007.0213. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2007.0213. Publisher: Royal Society.