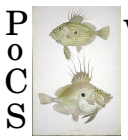

Principles of Complex Systems, Vols. 1 & 2, CSYS/MATH 300 and 303
University of Vermont, Fall 2021
Assignment 18
code name: Zari, Not Zari

Due: Wednesday, February 16, by 11:59 pm, 2021.
Relevant clips, episodes, and slides are listed on the assignment's page:
https://pdodds.w3.uvm.edu//teaching/courses/2021-2022principles-of-complex-systems//assignments/18/
Some useful reminders:
Deliverator: Prof. Peter Sheridan Dodds (contact through Teams)
Assistant Deliverator: Michael Arnold (contact through Teams)
Office: The Ether
Office hours: Tuesdays, 3:00 to 4:00 pm on Teams
Course website:
https://pdodds.w3.uvm.edu//teaching/courses/2021-2022principles-of-complex-systems

---

All parts are worth 3 points unless marked otherwise. Please show all your workingses clearly and list the names of others with whom you collaborated.

For coding, we recommend you improve your skills with Python, R, and/or Julia. The Deliverator uses Matlab.

Graduate students are requested to use LaTeX (or related TeX variant). If you are new to LaTeX, please endeavor to submit at least n questions per assignment in LaTeX, where n is the assignment number.

Assignment submission:

1. Please send to both the Deliverator and Assistant Deliverator via direct message on Teams.
2. PDF only! Please name your file as follows (where the number is to be padded by a 0 if less than 10 and names are all lowercase): `CSYS300assignment%02d$firstname-$lastname.pdf` as in `CSYS300assignment06michael-palin.pdf`

---

1. (3 points each)

   Using your text of choice, generate word shifts comparing two "interesting" regions of text.

   Use the Python package described in Ref. [1].

   (Various Matlab versions made by the Unreliable Deliverator do exist and need to shared.)

Links to paper versions (arXiv is always best), Github repository, and an exhilarating Twitter feed can be found here:
https://pdodds.w3.uvm.edu/research/papers/gallagher2021a/.

"Interesting" is anything you find interesting. Could be books 3 and 12 in a series, second half of a book compared to the first half, season 4 of a show versus all seasons, etc.

Aim to find two texts that are both reasonably large (more than 10^4 words) and fairly different in average happiness scores (though even the same scores can be meaningfully explored with word shifts).

Let's call the two texts $T^{(1)}$ and $T^{(2)}$. In your plots, you should label them meaningfully based on your choices).

Use a reasonable stop window of your choice, e.g., [4, 6] or [3, 7].

(a) Produce a word shift comparing text $T^{(2)}$ relative to text $T^{(1)}$. Use the average happiness of text $T^{(1)}$ as the baseline.

(b) Interpret the word shift. Does what you see make sense? Are there any surprises? Are some words being used in what the average person might not think is their primary meaning? For example, "crying" in Moby Dick means yelling, and "sick" can mean "awesome."

(c) Produce a word shift comparing text $T^{(1)}$ relative to text $T^{(2)}$. Use the average happiness of text $T^{(2)}$ as the baseline.

(d) Comment on any asymmetries you see (the basic word shifts we use are asymmetric).

(e) Produce a word shift comparing text $T^{(1)}$ relative to text $T^{(2)}$. Now use 5 as the baseline reference score (neutral on the happiness-sadness spectrum of 1–9).

(f) Compared to your first word shift, how interpretable is this one?

References

[1] R. J. Gallagher, M. R. Frank, L. Mitchell, A. J. Schwartz, A. J. Reagan, C. M. Danforth, and P. S. Dodds. Generalized word shift graphs: A method for visualizing and explaining pairwise comparisons between texts. *EPJ Data Science*, 10:4, 2021. Available online at https://arxiv.org/abs/2008.02250. pdf