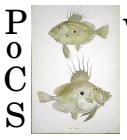


Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 6701, 6713, & a pretend number
University of Vermont, Fall 2023
Assignment 15
"Fake exclusive commentary by Christian Bale"

Due: Friday, March 1, by 11:59 pm
https://pdodds.w3.uvm.edu/teaching/courses/2023-2024pocsverse/assignments/15/
Some useful reminders:
Deliverator: Prof. Peter Sheridan Dodds (contact through Teams)
Assistant Deliverator: Chris O'Neil (contact through Teams)
Office: The Ether
Office hours: See Teams calendar
Course website: https://pdodds.w3.uvm.edu/teaching/courses/2023-2024pocsverse
Overleaf: LaTeX templates and settings for all assignments are available at https://www.overleaf.com/read/tsxfwwmwdgxj. If this link doesn't work, try https://www.overleaf.com/read/tsxfwwmwdgxj#456832

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

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

Via Brightspace or other preferred death vortex.

Please submit your project's current draft in pdf format via Brightspace by the same time specified for this assignment. For teams, please list all team member names clearly at the start.

- Please use Overleaf for writing up your project.
- Build your paper using:
  https://github.com/petersheridandodds/universal-paper-template
- Please use Github and Gitlab to share the code and data things you make.
- For this first assignment, just getting the paper template up is enough.

1. Continue to explore the emotional content of Pride and Prejudice, Frankenstein, and Moby Dick.

   Task: Compute emotional arcs by sliding a window of width T 1-grams through each book as described below.

   Reminder: The labMT word list was published with Ref. [1] in 2011, and has been occasionally upgraded when necessary to accommodate changes in language use.

   See https://hedonometer.org and https://storywrangling.org for the current version.

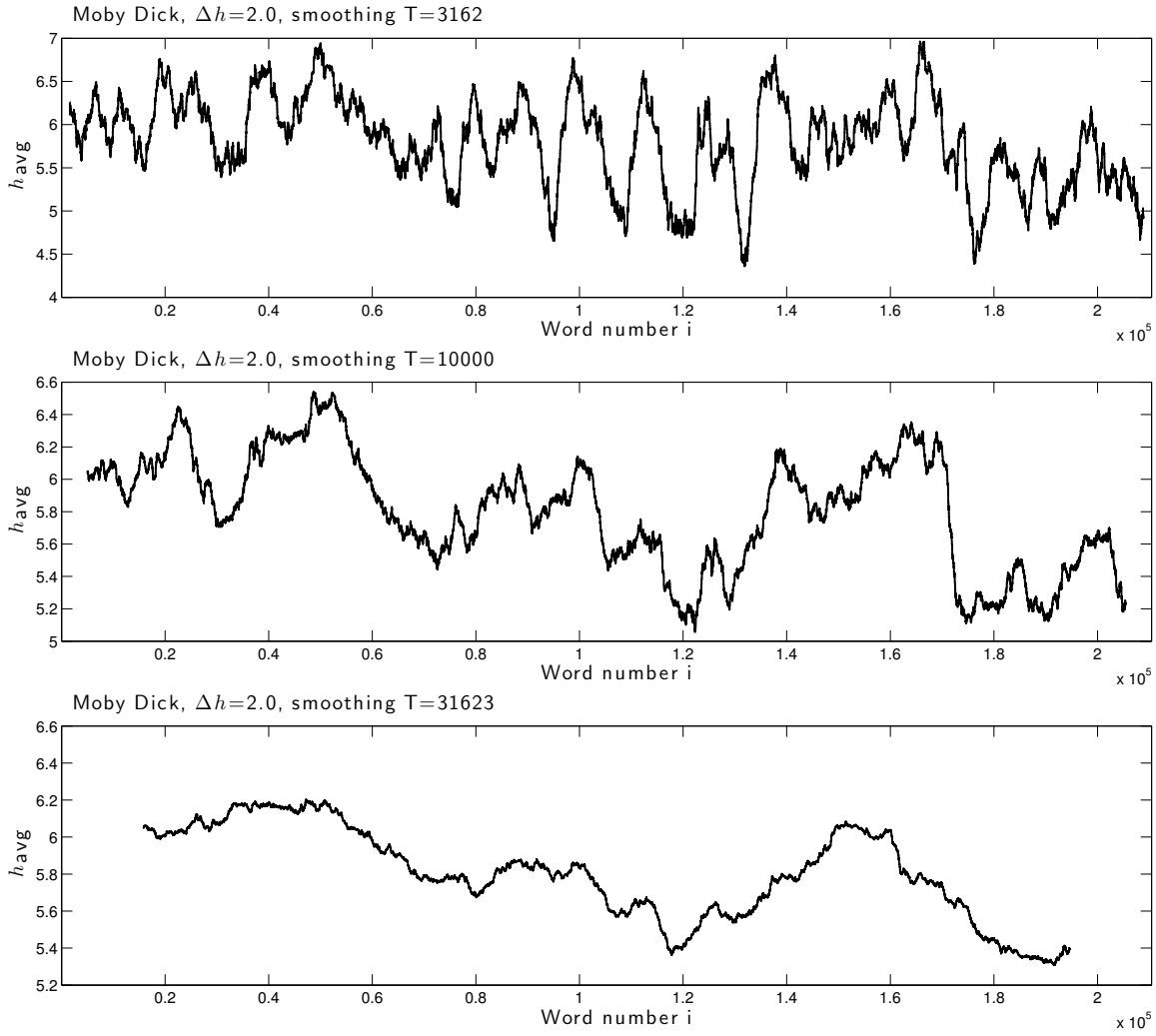
   (a) (3×3 = 9) First use the full lexical lens provided by labMT ($\delta h_{\rm avg} = 0$).
   Make a single figure containing a stacked set of 7 plots with text windows of size $T = [10^{\mu}]$ for $\mu = 1$, 1.5, 2, 2.5, 3, 3.5, and 4.0.
   Stacked here means separated and stacked vertically, as opposed to directly overlaid. See examples for Moby Dick at the end of this assignment.
   The notation $[\cdot]$ means round to the nearest integer.

   (b) (3×3 = 9) Choose a 'good' text window from above, and repeat the analysis with lenses which exclude the central words around the neutral point.
   The blocked words are $h_{\rm avg} \pm \delta h_{\rm avg}$ where $\delta h_{\rm avg} = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0,$ and 3.5.
   By 'good', we mean one that seems to you to produce a reasonable smoothing. Not too choppy, not too washed out.

   Notes:

   - The horizontal axis is "reading-experience time" corresponding to 1-grams in the text, running from 1 to N.
   - The windows should overlap, sliding one word ahead each time. This is a simple averaging filter.
   - Points should be located above the center of each window.
   - So the point for the window running from n to $n + T - 1$ (T words) will be located at $n + (T - 1)/2$.
   - Do not pre-filter the text for any given lens. Windows will contain variable numbers of words with and without happiness scores.
   - Only average 1-grams that are in the lexical lens—do not count all possible T words for the normalization.

   Three example averaging windows for Moby Dick with $\delta h_{\rm avg} = 2.0$:



1. Come up with some rich, text-based stories for analysis.

   For example: One (longish) book, or a book series, or a TV series.

   Data would be the original text (books), subtitles, screenplay, or scripts (TV series).

   - You must be able to obtain the full text.
   - You will want something with at least around 10^5 words. More than 10^6 would be great.
   - Transcripts of shows may be good for extracting temporal character interaction networks.

   Please talk about possibilities with others in the class.

   For this assignment, simply list at least one possibility, noting the approximate text size in number of words.

References

[1] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, 6:e26752, 2011. pdf